DOCUMENT RESUME

ED 278 066 CS 505 474

AUTHOR	O'Brien, Nancy, Ed.
TITLE	Status Report on Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications, April 1-September 30, 1986.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Institutes of Health (DHHS), Bethesda, Md.; National Science Foundation, Washington, D.C.; Office of Naval Research, Washington, D.C.
REPORT NO	SR-86/87(1986)
PUB DATE	86
CONTRACT	NICHHD-NO1-HD-5-2910; ONR-N00014-83-K-0083
GRANT	NICHHD-HD-01994; NIH-BRS-RR-05596; NINCDS-NS-13617; NINCDS-NS-13870; NINCDS-NS-18010; NSF-BNS-8111470; NSF-BNS-8520709
NOTE	319p.; For the previous report, see ED 274 022.
AVAILABLE FROM	U.S. Department of Commerce, National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22151.
PUB TYPE	Reports - Research/Technical (143) -- Information Analyses (070) -- Collected Works - General (020)
EDRS PRICE	MF01/PC13 Plus Postage.
DESCRIPTORS	*Communication Research; *Morphology (Languages); *Research Methodology; *Research Utilization; *Speech Communication

ABSTRACT

Focusing on the status, progress, instrumentation, and applications of studies on the nature of speech, this report contains the following research studies: "The Role of Psychophysics in Understanding Speech Perception" (B. H. Repp); "Specialized Perceiving Systems for Speech and Other Biologically Significant Sounds" (I. G. Mattingly; A. M. Liberman); "'Voicing' in English: A Catalog of Acoustic Features Signaling /b/ versus /p/ in Trochees" (L. Lisker); "Categorical Tendencies in Imitating Self-Produced Isolated Vowels" (B. H. Repp; D. R. Williams); "An Acoustic Analysis of V-to-C and V-to-V: Coarticulatory Effects in Catalan and Spanish VCV Sequences" (D. Recasens); "The Sound of Two Hands Clapping: An Exploratory Study" (B. H. Repp); "An Aeroacoustics Approach to Phonation; Some Experimental and Theoretical Observations" (R. S. McGowan); "Pattern Formation in Speech and Limb Movements Involving Many Degrees of Freedom" (J. A. S. Kelso); "The Space-Time Behavior of Single and Bimanual Rhythmical Movements: Data and Model" (B. A. Kay and others); "Language Mechanisms and Reading Disorder: A Modular Approach" (D. Shankweiler; S. Crain); "Syntactic Complexity and Reading Acquisition" (S. Crain; D. Shankweiler); "Phonological Coding in Word Reading: Evidence from Hearing and Deaf Readers" (V. L. Hanson; C. A. Fowler); "Strategies for Visual Word Recognition and Orthographic Depth: A Multi-Lingual Comparison (R. Frost and others); "The Inflected Noun System in Serbo-Croatian: Lexical Representation of Morphological Structure" (L. B. Feldman; C. A. Fowler); and "Repetition Priming Is Not Purely Episodic in Origin" (L. B. Feldman; J. Moskovljevic). Also included is a list of publications and an appendix listing these Status Reports by report number and providing DTIC and ERIC numbers. (JD)

SR-86/87 (1986)

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 April - 30 September 1986

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06511

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

(The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order number of previous Status Reports.)

Ignatius G. Mattingly, Acting Editor-in-Chief
Nancy O'Brien, Editor

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT

Grant HD-01994

NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT

Contract NO1-HD-5-2910

NATIONAL INSTITUTES OF HEALTH

Biomedical Research Support Grant RR-05596

NATIONAL SCIENCE FOUNDATION

Grant BNS-8111470
Grant BNS-8520709

NATIONAL INSTITUTE OF NEUROLOGICAL AND COMMUNICATIVE DISORDERS AND STROKE

Grant NS 13870
Grant NS 13617
Grant NS 18010

OFFICE OF NAVAL RESEARCH

Contract N00014-83-K-0083

HASKINS LABORATORIES PERSONNEL IN SPEECH RESEARCH

Investigators

Arthur S. Abramson*
Peter J. Alfonso*
Cinzia Avesani[1]
Thomas Baer
Fredericka Bell-Berti*
Catherine Best*
Geoffrey Bingham†
Gloria J. Borden*
Susan Brady*
Catherine P. Browman
Etienne Colomb††
Franklin S. Cooper*
Stephen Crain*
Robert Crowder*
Laurie B. Feldman*
Janet Fodor*

Anne Fowler†
Carol A. Fowler*
Ram Frost†††
David Garrett††††
Louis Goldstein*
Vicki L. Hanson
Katherine S. Harris*
Amelia I. Hudson[2]
Leonard Katz*
J. A. Scott Kelso*
Andrea G. Levitt*
Alvin M. Liberman*
Isabelle Y. Liberman*
Leigh Lisker*
Virginia Mann*
Ignatius G. Mattingly*

Nancy S. McGarr*
Richard S. McGowan
Kevin G. Munhall
Hiroshi Muta[3]
Susan Nittrouer††††
Patrick W. Nye
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Michael Studdert-Kennedy*
Betty Tuller*
Michael T. Turvey*
Mario Vayra[1]
Douglas H. Whalen

Technical/Support

Philip J. Chagnon
Alice Dadourian
Michael D'Angelo
Betty J. DeLise
Vincent Gulisano

Donald Hailey
Raymond C. Huey*
Sabina D. Koroluk
Yvonne Manning
Bruce Martin

Nancy O'Brien
William P. Scully
Richard S. Sharkany
Edward R. Wiley

Students*

Joy Armson
Dragana Barac
Eric Bateson
Suzanne Boyce
André Cooper
Margaret Dunn
Carole E. Gelfer
Bruce Kay

Noriko Kobayashi
Rena A. Krakow
Deborah Kuglitsch
Hwei-Bing Lin
Katrina Lukatela
Harriet Magen
Sharon Manuel
Jerry McRoberts

Lawrence D. Rosenblum
Arlyne Russo
Richard C. Schmidt
John Scholz
Suzanne Smith
Robin Story
Katyanee Svastikula
David Williams

---

*Part-time
[1]Visiting from Scuola Normale Superiore, Pisa, Italy
[2]Visiting from Louisiana State University, Baton Rouge, LA
[3]Visiting from University of Tokyo, Japan
†NIH Research Fellow
††Fogarty International Fellow, Lausanne, Switzerland
†††Postdoctoral Fellow, Hebrew University, Israel
††††NRSA Training Fellow

Status Report on Speech Research

Haskins Laboratories

CONTENTS

THE ROLE OF PSYCHOPHYSICS IN UNDERSTANDING SPEECH PERCEPTION*

Bruno H. Repp

Introduction

The purpose of this workshop is to discuss the psychophysics of speech perception. The program includes a variety of topics that presumably fall under this heading and that demonstrate that the psychophysics of speech perception is alive and well. Yet it is not really obvious what the psychophysics of speech perception is, what its goals and limitations are, and whether it is indeed a circumscribed area of investigation. It seems useful, therefore, to pose these basic questions explicitly and to include them in our discussions along with the many specific issues addressed by our research. The purpose of my paper is to stimulate such discussion by presenting a particular, possibly controversial, view of speech perception, psychophysics, and the relation between the two.

My presentation has five parts. First, I will attempt to define the psychophysics of speech perception and to discuss some of its assumptions and limitations. Then, turning to the second half of my title, I will consider briefly what it might mean to "understand" speech perception. Next, I will sketch a general view of phonetic perception and follow this with a discussion of what I believe to be the major research questions from that perspective. Finally, I will suggest a relatively novel application of psychophysics in the research enterprise I have envisioned.

1. What Is the Psychophysics of Speech Perception?

I am starting with the assumption that there is indeed a psychophysics of speech perception--a particular area of scientific inquiry that the title of this workshop is intended to refer to. If so, what distinguishes the psychophysics of speech perception from the investigation of speech perception in general?

Psychophysics, as traditionally defined, is the science of describing the relationships between objective (physical) and subjective (psychological) dimensions. In a typical experiment, physical characteristics of a series of stimuli are measured or manipulated, and the subjects' judgments are obtained

*In M. E. H. Schouten (Ed.), The psychophysics of speech perception. The Hague: Martinus Nijhoff Publishers, in press. Invited paper presented at the NATO Advanced Research Workshop on the Psychophysics of Speech Perception, Utrecht, The Netherlands, June 30 - July 4, 1986.
Acknowledgment. Preparation of this manuscript was supported by NICHD Grant HD-01994 to Haskins Laboratories. I am grateful to Al Bregman, Bob Crowder, Jim Flege, Ignatius Mattingly, Robert Remez, and Michael Studdert-Kennedy for helpful comments on an earlier draft.

on an explicit or derived numerical scale. The resulting stimulus-response relationship is often described in the form of a function, such as Weber's law or Stevens' exponential curves. However, there are many other ways of describing stimulus-response relationships, and it would be unwise to exclude any particular descriptions from the domain of psychophysics. Since virtually all speech perception research involves eliciting subjects' responses to stimuli that have been manipulated in some way, it seems to me that, at first blush, the psychophysics of speech perception is the only kind of research on speech perception that exists, especially if we exclude psycholinguistic topics such as word recognition and sentence comprehension, which concern the perception of meaning.

Is the title of this workshop then a tautology? Perhaps not. In fact, the term "psychophysics" is not commonly applied to all of the research on speech perception. Therefore, it has certain connotations that derive from the kinds of experiments it is explicitly associated with. That is, even though the boundaries of psychophysics are not clearly defined and may include a large variety of topics and methods, those researchers who consider themselves psychophysicists represent certain typical theoretical attitudes and preferences. Thus, psychophysics may be considered a particular approach to the study of speech perception that, without necessarily being programmatic, characterizes a fair amount of work in the field. I presume that, in choosing the title for this workshop, the organizers wished to highlight this approach, which I will now attempt to characterize.

1.1. Focus on the Auditory Modality

One attitude I associate with a psychophysical approach to speech perception is a preoccupation with psychoacoustics. Indeed, all presentations at this workshop are concerned with aspects of auditory speech perception. This is not to say that research on speech perception via the visual and tactile senses is not often psychophysical in character; in fact, much of it is, and several participants in this workshop have made important contributions to it. Nevertheless, this research has often been the province of specialists outside the mainstream of speech perception research. One consequence of this is that many speech perception researchers place special emphasis on auditory processes and thereby miss the more general insights to be gained from a multimodal approach.

Tactile speech perception, to be sure, is uncommon and requires special transduction devices; moreover, it is not clear whether tactile information feeds directly into the speech perception system the way auditory and visual information does (except for the Tadoma method, where articulation is felt directly). Visual speech perception, by contrast, is extremely common, especially in conjunction with listening. The extent to which auditory and visual information is integrated was strikingly demonstrated by McGurk and MacDonald (1976), who presented conflicting information in the two modalities and found that visual information may override the auditory information without the perceiver's awareness. In such instances, subjects believe they heard what in fact they saw. More often, the conscious percept represents a compromise between the inputs from the two modalities (Massaro & Cohen, 1983a; Summerfield, in press). It appears, therefore, that speech information from the two sensory modalities converges upon a common mental representation. As Summerfield (1979) and others have argued, the information seems to be represented internally in a common metric that is amodal in nature.

If this kind of argument is accepted, it follows that not too much weight should be attached to descriptions of speech information that are tied to one modality. Rather, the basis for speech perception must be sought in information that is modality-independent and can be described in a common vocabulary. Such a vocabulary is provided by articulatory kinematics and/or by the dynamic parameters that underlie articulatory processes. To be sure, articulations taking place in the back of the vocal tract are transmitted exclusively by acoustic means, whereas movements of lips and jaw are prominent in the optic signal. This partial dissociation should not detract us from the fact, however, that in each case the information is about articulatory position and motion or, more abstractly, about the changing area function of the speaker's oral cavity.

Alternatively, it might be assumed that cues from different modalities are integrated in the process of categorical decision making, without recourse to a common metric (Massaro & Cohen, 1983a; Summerfield, in press). However, the question then arises: What motivates the integration in the first place? If the internal representations of stimuli are modality-specific, they can be related only through some form of association, either innate or acquired. In Massaro and Cohen's model, the associations reside in attribute lists that constitute phonetic category prototypes. Although this model seems to account well for audiovisual syllable perception, it seems less able to handle the intersensory integration of continuous dimensions such as speaking rate (Green & Miller, 1985) or prelinguistic infants' ability to recognize auditory-visual or visual-proprioceptive correspondences (Kuhl & Meltzoff, 1982; Meltzoff & Moore, 1985). A description of the stimulus information in articulatory terms eliminates the need to hypothesize independent mental representations of modality-specific correlates of articulation (see Yates, 1985), and it emphasizes the fact that the relation between visual and auditory manifestations of speech is nonarbitrary and possibly innately specified.

While it is generally taken for granted that we see the moving articulators when we look at them, not abstract optic patterns, there has been some reluctance in the field to accept the analogous proposition (Gibson, 1966; Neisser, 1976; Studdert-Kennedy, 1985) that, when we listen to speech, we hear the moving articulators and not the auditory patterns that constitute the proximal stimulus. Instead, researchers have been intensely preoccupied with acoustic variables such as formant transitions, delayed voicing onset, rise time, and so forth, as if the corresponding auditory percepts were the primary objects of speech perception. Whether they are is open to question, however (see, e.g., Liberman & Mattingly, 1985). Their prominent role in speech research may in large part be due to traditional techniques of acoustic analysis and synthesis, rather than to any compelling theoretical considerations. Many issues in the psychoacoustics of speech perception might never have been considered, had methods of articulatory analysis and synthesis preceded spectrographic and formant-based methods. As it is, we need to ponder whether these psychoacoustic issues are really pertinent to speech perception, or whether they merely have been forced upon us by the instruments we have had available. In other words, if we had only articulatory synthesizers as well as devices that extract area functions from the acoustic (and/or optic) signal, what would be the theoretical status of phenomena such as backward masking, adaptation, contrast, spectral integration, etc., in speech perception research? How much would we lose if we talked only about articulation and not about acoustics at all?

1.2. Focus on Methodology

A second tendency that may reasonably be associated with a psychophysical approach is a focus on methodology. Certainly in classical psychophysics the methods by which stimulus-response mappings are obtained have been of overriding concern. There are many examples of a similar concern in speech perception research. Many experiments have compared performance in different discrimination paradigms, such as AX, 4IAX, ABX, fixed versus roving standard, etc. (e.g., MacKain, Best, & Strange, 1981; Macmillan et al., in press; Pisoni & Lazarus, 1974; Rosner, 1984). and even in the many studies using only a single method its choice has usually been a matter of concern. Other studies have compared different identification tasks, such as binary classification, numerical rating scales, absolute identification, and perceptual distance scaling (e.g., Ganong & Zatorre, 1980; Massaro & Cohen, 1983b; Vinegrad, 1972). In fact, it may be argued that most of categorical perception research, as well as much research on selective adaptation, contrast, auditory memory, etc., has been exercises in methodology. To be sure, the variations in methods have usually served to test some reasonable models or hypotheses, and I do not mean to imply that this research has been worthless. Nevertheless, the questions asked in such experiments often are somewhat removed from the original phenomena that stimulated the research; in other words, they have become methodological variations on a common theme, and sometimes variations themselves have become the themes for further variations.

Take categorical perception. The category boundary effect (Wood, 1976)--the well-known finding that discrimination performance is higher across a phonetic category boundary than within categories--is important because it tells us that the acoustic structure of speech is not very transparent to the typical listener, who habitually focuses only on linguistically significant information. Numerous studies have shown that the strength of the effect varies with methodological factors such as discrimination paradigm, interstimulus interval, training, instructions, language experience, types of stimuli, etc. (see review by Repp, 1984). The large majority of these studies has been concerned with subjects' ability to discriminate small acoustic differences among speech stimuli. This ability, not surprisingly, can be enhanced by training, reduction of stimulus uncertainty, short interstimulus intervals, etc. The studies that have shown this are prime examples of the psychophysics of speech perception, and they include many an elegant piece of experimentation. However, the important aspect of categorical perception that seems directly relevant to speech communication is not subjects' apparent inability to discriminate linguistically irrelevant differences along certain stimulus continua but rather their attention to linguistically distinctive information in the speech signal. To be sure, statements have been made in the literature (Liberman & Mattingly, 1985; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970) to the effect that human listeners simply _cannot_ perceive certain auditory properties of speech sounds, and this has, of course, been grist for the psychophysical mill. Apart from dismissing such extreme claims, however, little has been learned from all these studies about speech perception beyond the truism that perception within categories is not categorical. Rather, they have revealed some things about auditory discrimination and the methodological variables affecting it. Equivalent information could have been obtained by using nonspeech stimuli, and indeed one of the aims of psychophysical methodology (though this is rarely acknowledged) is to enable listeners to perceive speech as if it were a collection of arbitrary sounds. This leads me to another, related bias I associate with the psychophysics of speech perception.

1.3. Focus on the Sounds of Speech

One possible definition of the psychophysics of speech perception is that it is the study of the perception of the sounds of speech. Unfortunately, the term "speech sounds" has often been used indiscriminately to denote both linguistically significant categories and acoustic components of the speech signal (and/or the auditory impressions associated with them). A clear distinction needs to be made between the auditory/acoustic and linguistic/articulatory domains, however (cf. Repp, 1981); the term "speech sounds" is appropriate for the former, whereas "phonetic categories" (or "phonemes") is appropriate for the latter. With this distinction in mind, my claim is that psychophysics is concerned, for the most part, with speech sound perception rather than with phoneme perception. It seems likely, however, that, except in very special circumstances, the sounds of speech as such do not play an important role in speech communication (see also Liberman & Mattingly, 1985; Linell, 1982; Traunmüller, in press). Rather, I presume it is the more abstract, articulatory information that is used by listeners to decode the linguistic message. In fact, the only context in which the auditory qualities of speech segments may have a communicative function is in poetry, where an (unconscious) apprehension of the segmental sound pattern may enhance connotative and aesthetic qualities of the text (Fónagy, 1961; Hrushovski, 1980). Paradoxically, it seems that, so far, poetry has not attracted the attention of psychophysicists. (See, however, Marks, 1978.)

Why should one be interested in perceptual qualities that do not serve any important function in speech communication? There could be many valid reasons, such as questions about the auditory processing of complex sounds, the consequences of hearing impairment, skills of analytic perception, etc.--all topics worthy of scientific investigation. Nevertheless, these topics may be largely irrelevant to the perception of phonetic structure, and their study may therefore not contribute to our understanding of speech perception. To the hard-core psychophysicist, speech is primarily an acoustic signal of unusual complexity, which presents a challenge to the auditory system and to the experimenter's ingenuity. However, since this acoustic complexity is precisely what the speech perception system is equipped to handle, the speech signal actually has a very simple structure when viewed from the inside, as it were. For the speech perceiver, and for the speech researcher, perceptual complexity is defined by different criteria, such as the relative familiarity of a language, dialect, or foreign accent, the rate of speech, or the fidelity of the acoustic signal. In other words, perceptual complexity is defined not absolutely but in terms of deviations from expectancies. In the case of synthetic or degraded speech, an acoustically simpler signal may pose a perceptual problem.

1.4. Focus on the Naive Listener

The bias that I have just portrayed--that psychophysics tends to be concerned with linguistically irrelevant aspects of speech--may seem to apply only to a small portion of speech research. After all, most speech perception experiments do require subjects to respond with phonemic categories (strictly speaking, with alphabetic symbols) to the speech sounds they hear, and not with numerical ratings or other kinds of nonphonetic responses. However, it is often assumed, if only implicitly, that the phonemic or orthographic symbols employed by listeners are simply convenient labels for auditory experiences. Hand in hand with that assumption goes the much-discussed hypothesis that phonetic categories, and particularly the boundaries between them, reflect constraints imposed by the mammalian auditory system (see, e.g., Kuhl, 1981; Liberman & Mattingly, 1985). This hypothesis dovetails with another bias of psychophysical research.

Classical psychophysics is rarely concerned with subjects' experience prior to an experimental session, except for task-specific training received under controlled conditions. Essentially, psychophysics is about basic processes of perceptual translation, most often from a continuous physical dimension to a continuous psychological dimension. If categories are to be employed as responses in a psychophysical task, they are usually defined within the limited context of the experimental situation, often exemplified by the extremes of a stimulus dimension. The boundaries between such categories are either arbitrary--e.g., they may just bisect a stimulus continuum and hence depend on its range--or, if they are not (as is more often the case with speech) they are assumed to coincide with a psychoacoustic discontinuity that gave rise to the categories in the first place. Although subjects obviously have much experience with the categories of speech outside the laboratory, this experience is often considered irrelevant because the psychoacoustic basis for the category division is assumed to be present in the stimuli. (If the stimuli are synthetic and unfamiliar-sounding, so much the better.) At best, language experience may have taught subjects to attend to one particular discontinuity and to ignore another; hence certain cross-language differences in boundary location.

These assumptions are perfectly appropriate within the framework of psychophysics. Indeed, in the quest for an elegant description of the perceptual translation from the objective to the subjective realm, any intrusion of pre-experimental knowledge is undesirable. Imagine an experiment involving the perceived similarity of various round shapes, in which subjects judge two shapes as more similar than the others because both happen to look like the same familiar object (e.g., an apple). This would be an undesirable artifact (Titchener, 1909, called it the "object error") that might distort the <u>true</u> psychophysical function underlying the similarity judgments. This function is assumed to be universal and independent of prior experience.

There is considerable evidence, however, that many, perhaps all, phonetic distinctions rest on linguistic, not psychoacoustic criteria (see Repp & Liberman, in press; Rosen & Howell, in press). These criteria are acquired--or, if innate, are modified--through experience with spoken language. Rather than referring to particular auditory experiences, phonetic category labels--once certain orthographic and linguistic conventions are stripped off--denote specific <u>articulatory</u> maneuvers whose auditory correlates, though systematic, are largely irrelevant. This is most strikingly demonstrated by the finding that phonetic structure can be perceived in auditorily anomalous stimuli composed of time-varying sinusoids that imitate formant movements and thus retain information about the changing shape of the vocal tract (Remez, in press; Remez, Rubin, Pisoni, & Carrell, 1981). The articulatory patterns characteristic of a language presumably have evolved according to articulatory and linguistic constraints (Lindblom, 1983; Ohala, 1983), and it seems unlikely that auditory limitations have played a significant role, except in the very general sense that phonetic contrasts that are difficult to discriminate tend to be avoided or, if they occur, may lead to language change (Bladon, in press; Ohala, 1981). I will argue below that listeners refer to their knowledge of language-specific articulatory norms when listening to speech. This reference is <u>external</u> to the experimental situation and inside the listener. Rather than emerging from acoustic properties of the stimulus or the stimulus ensemble, the phonetic structure imposed by the talker and recovered by the listener represents a learned conventional pattern constrained by universal articulatory possibilities.

Since it is the linguistic structure that is important in speech communication, and not the auditory properties of speech components, it is natural that human listeners focus their attention on the former and not on the latter. This attention to a discrete representation of speech influences subjects' judgments in a variety of psychophysical tasks designed to assess the psychological transformation of acoustic stimulus dimensions. For example, it is probably responsible for the category boundary effect in categorical perception experiments, as hypothesized long ago by proponents of the so-called dual-process model (e.g., Fujisaki & Kawashima, 1969, 1970; Pisoni, 1973; Samuel, 1977). However, some researchers committed to psychophysical approaches (e.g., Macmillan, Kaplan, & Creelman, 1977) have taken these perceptual nonlinearities to be inherent in the auditory stimulus representation. Although auditory nonlinearities do seem to occur along certain acoustic dimensions of speech, they may be unrelated to the discontinuities imposed by the mental organization of the listener (see, e.g., Howell & Rosen, 1983; Rosen & Howell, in press; Schouten, in press; Watson, Kewley-Port, & Foyle, 1985). The same may be said about so-called phonetic trading relations and context effects (see review by Repp, 1982) which, for the most part, reflect not psychoacoustic interactions among signal components but the listener's imposition of multidimensional criteria in the process of phonetic categorization (Derr & Massaro, 1980; Massaro, in press-b; Repp, 1983; however, see also Diehl, in press).

By these arguments, speech is a particularly unwieldy object for psychophysical and psychoacoustic experimentation. If questions of auditory perception are to be addressed, why not use simpler stimuli? If questions of speech communication are to be addressed, why use a psychophysical approach? As Massaro (in press-b) aptly points out, a large part of modern speech perception research consists of either (a) applying reductionistic models to laboratory phenomena in a search for the auditory mechanisms that accomplish phonetic categorization, or (b) appealing to "special" mechanisms that do the job. Both enterprises have been sterile--the first in that it has not revealed any relevant mechanisms, and the second in that it has postponed or even relinquished the search for them. One problem with both approaches is that they represent models of speech perception according to which linguistically distinctive information somehow must emerge from the stimulus alone, without recourse to long-term mental representations of linguistic knowledge. One notable exception has been the work of Massaro and his collaborators, who have consistently pursued the idea that speech perception proceeds by reference to internal category "prototypes" (see Massaro, in press-b; Massaro & Oden, 1980a). Their model, and similar ideas in the literature, lead the way toward a <u>relational</u> (or systemic) theory of speech perception, to be sketched further below.

2. <u>Understanding Speech Perception</u>

The goal of speech perception researchers is to understand (or explain) speech perception--that much is obvious. However, what does this really mean? What <u>is</u> speech perception, and what does understanding (explaining) it entail? Probing these questions too deeply leads to profound espistemological issues. I offer only a few comments for discussion.

2.1. Two Definitions of Perception

The term "perception" is being used in different ways by different researchers, as has been pointed out by Chistovich (1971) and Shepard (1984), among others. An example of one usage is provided by Massaro's recent writings on categorical perception (Hary & Massaro, 1982; Massaro, in press-a, in press-b; Massaro & Cohen, 1983b). He argues that "categorical results do not imply categorical perception": The perception of speech continua is revealed to be continuous if only the right methods are employed. According to Hary and Massaro (1982), "a central issue in auditory information processing is whether certain auditory continua are perceived categorically rather than continuously" (p. 409). That is, it must be one or the other: Perception is entirely a function of the input. Perception is thus equated with sensory transduction--an immutable process that is insensitive to attention and experience. Of course, this is exactly what psychophysics is concerned with. The goal of speech perception research, in this view, is to find out what speech perception really is like, once all constraints imposed by attentional and experiential factors have been removed. The classification by reference to prototypes, which plays such a prominent part in Massaro's model, apparently is a post-perceptual process in his definition.

This view needs to be contrasted with a definition of perception that includes categorization and attentional filtering. According to this (my preferred) view, perception is what occurs when the transduced stimulus meets the mental structures (the "model of the world") laid down by past experience and possibly by genetic transmission (Hayek, 1952; Shepard, 1980, 1984; Yates, 1985). The result of perception is the outcome of that encounter, not the input to it. According to Fodor (1983, p. 40), "what perception must do is so to represent the world as to make it accessible to thought" through processes of transduction and inference. Categorical perception, and the apparent invariance of the categorical percept, represent the outcome of the inferential process. To find behavioral evidence of the (largely) continuous, transduced information that feeds into this process, a listener's perceptual strategy must be altered through instructions and training, or some measure of decision uncertainty (e.g., reaction time) must be obtained. Since there are a variety of mental structures a stimulus may relate to, there are often alternative ways of perceiving the same input, depending on the perceiver's experience (i.e., form of the mental representations) and attention (i.e., selection from among them). Thus, in this view, categorical results do imply categorical perception, and noncategorical results imply noncategorical perception.

Speech perception thus can mean different things depending on the situation and the subject's strategies. In addition, it has a double meaning from another perspective, depending on whether "speech" is taken to refer to the stimulus or the percept. Psychophysical research can be snugly accommodated under the stimulus-based definition that speech perception is whatever occurs when speech signals are presented to a listener. I favor a percept-based definition--that speech perception occurs when a stimulus is perceived as speech, that is, when the listener interprets the stimulus in relation to the linguistic system. By that definition, many psychophysical experiments deal not with speech perception but with the perception of speechlike auditory stimuli. This distinction is not intended as a value judgment (indeed, psychophysical research generally surpasses speech perception research in rigor and methodological sophistication), but as a separation of largely independent domains of inquiry.

2.2. Two Definitions of Understanding

What does it mean to understand (or explain) speech perception? According to one view, it involves building or programming a machine to recognize speech. For example, Chistovich (1980) presented this approach as the one taken by the Leningrad group. This pragmatic goal of "teachability" deserves our respect (for a critique, see Studdert-Kennedy, 1985). Even though the operations of the machine may not resemble those of the human brain, a speech recognition algorithm approximating human capabilities would represent a useful model of speech perception and thus increase our understanding of the process. Unfortunately, it seems that psychophysics has little to contribute to this enterprise. Psychoacoustic and physiological research has uncovered transformations in the auditory system that could be simulated by a speech processing system. However, incorporating auditory transforms into the machine representation of speech apparently does not improve speech recognition scores (Blomberg, Carlson, Elenius, & Granström, 1986). This is perhaps not surprising. Machine representations need to capture the relationships between stimulus properties and precompiled knowledge structures (Shepard, 1980), and relational properties are likely to be largely invariant under transformations. Moreover, transformations of the input cannot result in an information gain, let alone in the magical emergence of properties that cannot also be computed by a central algorithm, so the most detailed coding of the speech signal is likely to be the most useful one for machines. Unless the goal is to build an analog of a complex biological system (and we are far from that stage), insights derived from psychophysical and psychophysiological research are likely to be of little use to computers. The essential problem to be solved in speech recognition research, I presume, is not that of stimulus coding but that of phonetic knowledge representation and utilization.

The alternative approach to scientific explanation is a purely theoretical one. Scientists and other human beings, of course, can perceive speech and need not (cannot) be taught explicitly, so the teachability criterion does not apply. This approach to explanation, therefore, is fundamentally different from that provided by the automatic speech recognition research. Theory construction, in psychology at least, is a cognitive act subject to individual preferences, sociological factors, and philosophical considerations (see Toulmin, 1972). One person's explanation may be another's tautology.

A variety of scientific philosophies are evident in the speech perception field, and their coexistence for a number of years suggests that they represent, in large part, individual preferences and not theories subject to empirical disconfirmation. What is worse, they do not agree on what really needs to be explained about speech perception. Rather than discussing the current theories or endorsing any of them, I am going to present a personal view below, at the danger of adding to the general confusion. My own ideas are neither fully worked out nor entirely original. (See, for example, Bregman, 1977; Elman & McClelland, 1984, 1986; Hayek, 1952; Liberman & Mattingly, 1985; Massaro & Oden, 1980a; Shepard, 1980, 1984; Yates, 1985.) Whatever their merit, however, they may serve as a useful basis for discussion at this workshop. After presenting my view, I will discuss what seem to be the major research questions from this perspective and what role psychophysics might play in this enterprise.

3. Speech Perception as a Relational Process

Phonetic perception--that is, the perception of the phonological structure of speech without regard to its semantic content--has often been considered a purely input-driven process, to be contrasted with the largely knowledge-driven processes of language understanding (e.g., Marslen-Wilson and Welsh, 1978; Studdert-Kennedy, 1982). That is, it is often assumed that phonological structure is in the speech signal (e.g., Fowler, 1984; Gibson, 1966; Stevens & Blumstein, 1981) or emerges from it via specialized neural processes (Liberman & Mattingly, 1985). The present proposal contrasts with these views in that it assumes that speech perception requires two complementary ingredients: the input signal and the perceiver's internal representation of the speech domain. In other words, I am assuming that phonological structure emerges, especially in its language-specific details, from the relation between a stimulus and a "phonetic lexicon" in the perceiver's head that (in mature individuals) provides an exhaustive knowledge base representing all the characteristics associated with the structural units of a language.

In this view, it is not the stimulus as such (or its auditory transform) that is perceived, but rather its relationship to the phonetic knowledge base; perception thus is a relational process, a two-valued function. Its output is also two-valued: The relation of the input to the pre-existing internal structures yields (potential) awareness of the structure that provides the best fit, plus some measure of goodness of fit, which may be experienced as degree of confidence or uncertainty.

How is the phonetic knowledge represented in the brain? One possible conceptualization is in terms of "prototypes" (schemata, norms, ideals, logogens, basic categories) abstracted from language experience (cf. Flege, in press; Massaro & Oden, 1980a, 1980b; Yates, 1985). The mechanisms enabling this abstraction during language acquisition are unknown and may either reside in a specialized "module" (Fodor, 1983; Liberman & Mattingly, 1985) or represent general neural design principles (e.g., Grossberg, in press). Language-specific phonetic categories are assumed to "crystallize" around central tendencies extracted from the variable input under the guidance of linguistic distinctiveness criteria. How this occurs is one of the great unsolved questions in speech research.

Just like the stimulus itself, the contents of the listener's knowledge base can be described in acoustic (optic), auditory (visual), or articulatory terms; that is, the lexicon is assumed to contain information about typical articulatory motions and their acoustic and optic concomitants, as well as possibly about their underlying dynamic parameters. The articulatory information is primary in so far as it also serves to control speech production and silent (imagined) speech, because it relates more directly to linguistic and orthographic symbols, and because it unites the different sensory modalities (as pointed out earlier). Whatever metaphor is used to describe the knowledge base--and we cannot expect to capture in words the state of a complex neural network--the important consequence of having it is that a perceiver is able, at each moment in time, to evaluate the information in the speech signal as to whether it fits the language norms. Deviations from these expectations may be perceived as unnaturalness, foreign accent, or individual speaker characteristics; or they may pass unnoticed.

Speech that is pronounced clearly, free of noise, and typical of the language is perceived "directly": The appropriate prototypes "resonate" to the input (Shepard, 1984). Ambiguous or degraded speech is represented in terms of its relative similarities to the most relevant prototypes. Whenever a decision is required, one prototype is selected that provides the best fit to the input (cf. Massaro & Oden, 1980a, 1980b). *Explicit* linguistic category decisions, however, are basically a response phenomenon governed by (laboratory) task requirements. Whether or not overt categorical decisions are made, the structural linguistic information is always present, being *implicit* in the prototypes and their relations to each other (cf. Lindblom, MacNeilage, & Studdert-Kennedy, 1983). The size of the "perceptual units," and with it the size of the prototypes, is variable, being a joint function of cognitive accessibility and real-time task requirements (cf. Warren's, 1981, LAME model). Thus, even though explicit recognition of individual phonemes is likely to be a function of literacy and linguistic awareness (cf. Mattingly, 1972; Morais, Cary, Alegria, & Bertelson, 1979), phonemic structure is nevertheless implicit in the prototype inventory: For example, a /b/ is perceived when all prototypes transcribable as /b.../ are "active," i.e., resemble the input (cf. Elman & McClelland, 1984, 1986).

Properties of the speech signal become linguistic information only by virtue of their relation to the listener's knowledge base. One could imagine that the stimulus is represented in terms of a "similarity vector" (Chistovich, 1985) containing relative deviations from prototypes in some perceptual metric. This form of coding may be viewed as an effective way of information reduction, though it is by no means clear that the brain *needs* such a reduction the way we need it when thinking about the system's operation. That is, a similarity vector is better thought of as a set of potentials or relationships, not of physically instantiated quantities.

In my view, the "special" nature of speech, which has received so much emphasis in the past (e.g., Liberman, 1982), resides primarily in the fact that speech is a unique system of articulatory and acoustic events. In contrast to adherents of the modularity hypothesis (Fodor, 1983; Liberman & Mattingly, 1985) I suspect that the *mechanisms* of speech perception are general--i.e., that they can be conceptualized in terms of domain-independent models, such as adaptive systems theory (Grossberg, in press), interactive activation theory (Elman & McClelland, 1984), or information integration theory (Massaro & Oden, 1980a). In other words, I believe that the specialness of speech lies in those properties that define it as a unique phenomenon (i.e., its production mechanism, its peculiar acoustic properties, its linguistic structure and function) but not in the way the input makes contact with mental representations in the course of perception. That is, as long as we can only rely on *models* of the perceptual mechanism, it is likely that significant similarities *will* obtain across different domains, even though the physiological substrates may be quite different. This is a consequence of the relatively limited options we have for constructing models of perception and decision making.

To go one step further: If speech is special but speech perception is not, it follows that there is a lot to be learned about speech, but relatively little about speech perception. This conclusion, for what it is worth, suggests a "vertical" research strategy (giving a twist to Fodor's, 1983, arguments): The way to learn more about the speech *system* is to investigate its many special characteristics. This is a multidisciplinary venture, a task for the specialist called "speech researcher." By contrast, study of speech perception as such is open to a "horizontal" approach by psychologists interested in perception in general. However, there is comparatively little to be learned about that process. While there are lots of interesting facts to be uncovered about speech, the "mechanisms" of perception are a figment of the scientist's imagination (as is the mechanistic analogy itself). It is quite likely that, once we know enough about speech and have characterized the perceiver's knowledge in a suitably economic form, we also will have explained speech perception in its essential aspects.

4. A Program for Speech Perception Research

From the perspective I have adopted, there are four major questions for research on speech perception: What is the phonetic knowledge? How is it used? How is it acquired? How can it be modified?

4.1. Description of the Knowledge Base

Before we can ask any questions about speech perception, we need to know what speech is, so we can account for the perceiver's expectations. This seemingly obvious requirement is often neglected by psychologists who plunge into speech perception experiments without considering the relevance of acoustic, articulatory, and linguistic phonetics. Even so useful a tool as Massaro's "fuzzy logical model" of information integration (Massaro & Oden, 1980a, 1980b) yields parameters characterizing phonetic prototypes whose relation to the normative properties of English utterances often remain unclear. It is often assumed that these properties will emerge from studies involving the classification of acoustically impoverished stimuli (see also Samuel, 1982). This is unlikely, however, because perceivers have detailed expectations about the full complement of acoustic properties, including those held constant in a given experiment, and they will often shift their criteria for stimulus classification along some critical dimension to compensate for the constancy or absence of others. While demonstration of this fact may be a worthwhile goal of some experiments, a more important point is that the perceivers' expectations can be assessed directly and independently (at least to a first approximation) by collecting facts about the acoustic and articulatory norms of their language, which constitute their knowledge base. Ever since Chomsky's (1965, 1968) seminal publications, the study of syntax, semantics, and phonology has been considered part of cognitive science, leading to a description of the language user's knowledge. I would like to add (normative) phonetics: The study of articulatory and acoustic norms, too, yields a description of the average listener-speaker's "competence" (cf. Tatham, 1980).

I am thus proposing that the study of acoustic and articulatory phonetics be part and parcel of speech perception research. Incidentally, psychologists, with their thorough understanding of measurement and sampling problems, are especially well equipped to conduct phonetic and articulatory research, which too often has taken a case study approach in the past. Representative measurements are also important for automatic speech recognition research (Klatt, 1986). They would not make experimental determinations of prototypical perceptual parameters superfluous but rather provide a basis for their interpretation: The normative characteristics of a language are what a perceiver ought to have internalized. If deviations from the norm and/or individual differences emerge from such a comparison, the search for their causes should be an interesting and important undertaking.

In what form phonetic knowledge is represented in the brain is a question that cannot be answered conclusively by psychologists, who may choose from a number of alternative conceptualizations. As Shepard (1980, p. 181) has aptly stated, "there are many possible levels of description, and although they may appear very different in character, the various levels all pertain to the same underlying system. In this respect, the internal representation is no different from the external object." Choosing one particular level of description is basically a matter of preference and, perhaps, parsimony.

4.2. Perceptual Weights and Distances

One empirical question that psychologists may usefully address, however, is how phonetic knowledge is applied. Since a clear, unambiguous stimulus poses no challenge to the perceptual system and therefore cannot reveal its workings (cf. Shepard, 1984), the principal question is how phonetic ambiguities created by realistic signal degradation or by deliberate signal manipulation are resolved (explicitly) by the perceiver in the absence of lexical, syntactic, or other higher-order constraints. In such a situation, the perceiver must make a decision based on the perceptual distances of the input from the possible phonetic alternatives (prototypes) stored in his or her permanent knowledge base. The decision rule may be assumed to be straightforward: Select the prototype that matches the input most closely. However, what determines the degree of the match? What makes an ambiguous utterance more similar to one prototype than another? In other words, what is the phonetic distance metric, what are the dimensions of the perceptual space in which it operates, and what are the perceptual weights of these dimensions?

There are opportunities for the useful application of psychophysical methods here, since the distance metric may be, in part, a function of auditory parameters (see, e.g., Bladon & Lindblom, 1981). However, the relative importance of different acoustic dimensions for a given phonetic contrast cannot be predicted from psychophysical data alone, since it depends heavily on the nature and magnitude of the differences among the relevant prototypes, in combination with their auditory salience. Traditional psychophysics is concerned with perceptual similarities and differences between stimuli, whereas the present application requires a multidimensional psychophysics dealing with the similarity of stimuli to mental representations. The many confusion studies in the literature (beginning with Miller & Nicely, 1955) would seem to be about this issue, but the data have always been analyzed in terms of stimulus-stimulus, not stimulus-prototype similarities (which they indeed represent), and it is possible that important information has been missed. Research such as Massaro's modeling of information integration in phoneme identification (e.g., Derr & Massaro, 1980; Massaro & Oden, 1980a, 1980b) is an exemplary effort from the present viewpoint, despite certain limitations. Massaro has found again and again that stimulus attributes are evaluated in an independent and multiplicative (or log-additive) fashion in phonetic classification, and this has obvious implications for the nature of a phonetic distance metric. Many experiments on the perceptual integration and relative power of acoustic cues (e.g., Abramson & Lisker, 1985; Bailey & Summerfield, 1980; Lisker, Liberman, Erickson, Dechovitz, & Mandler, 1977; Repp, 1982) also contribute relevant information. Experiments that avoid the fractionation of acoustic signals into "cues" and search for a phonetic distance metric based on more global spectral properties (Klatt, 1982, 1986) are promising but still at a very early stage.

Even though perceptual distances may reflect certain facts about auditory processing, these influences on phonetic perception are probably limited. The principal reason is that the mental structures that determine speech categorization have been built up from past experience with speech that underwent essentially the same auditory transformations as the current input is undergoing. That is, all transformations occurring during stimulus transduction are necessarily represented in the central knowledge base. Therefore, it makes relatively little difference whether we think of the input as sequences of raw spectra and of the mental categories as prototypical spectral sequences (e.g., Klatt, 1979), or whether we consider both in terms of some auditory transform or collection of discrete cues. It is the relation between the two that matters, and that relation is likely to remain topologically invariant under transformations. Only nonlinear transformations will have some influence on phonetic distances (Klatt, 1986).

4.3. Perceptual Development

In addition to asking how phonetic knowledge is utilized, we must ask how and when it is acquired. Much developmental and comparative research in the past has focused on auditory discrimination abilities, and the approach has been quite psychophysical in character. The "categorical" effects that have been observed in infants and animals may not reflect phonetic perception but certain psychoacoustic discontinuities on speech continua (Jusczyk, 1985, 1986), although this suggestion becomes doubtful in view of findings (Sachs & Grant, 1976; Soli, 1983; Watson et al., 1985) that the category boundary effect can be trained away in adults. Alternatively, category boundary effects in infants may reflect an innate predisposition for perceiving a universal articulatory inventory (Werker, Gilbert, Humphrey, & Tees, 1981). The interpretation of these data is uncertain at present. Speech perception research in older children (e.g., Elliott, Longinotti, Clifton, & Meyer, 1981; Tallal & Stark, 1981) also has often focused on their auditory abilities, not specifically on their criteria for phonetic identification and on the nature of their phonetic knowledge. Only more recently, following the lead of researchers such as Kuhl (1979) and Werker et al. (1981), has phonetic categorization in infancy been studied more carefully. A finding of special significance is the discovery (Werker & Tees, 1984) that infants' ability to perceive phonetic contrasts foreign to their parents' language declines precipitously before 1 year of age. This stage seems to mark the beginnings of a language-specific phonetic lexicon. It is an important research endeavor to trace the accumulation and refinement of phonetic knowledge through different stages of development, and much work remains to be done (see Jusczyk, in press).

4.4. Perceptual Learning

Another question of great theoretical and practical importance is how the phonetic knowledge, once it is established in the mature adult, can be augmented and modified. This concerns the process of second language learning and also, to some extent, the skills acquired by professional phoneticians (and even by subjects in a laboratory task, although their skills may be rather temporary). Furthermore, there is the very interesting question of bilingualism--the separation and interaction of two different, fully established phonetic knowledge bases. Until recently, little rigorous research has been carried out in this predominantly education-oriented area. Research is burgeoning, however, and is yielding interesting results (see Flege, in press).

Another, related question is to what extent reduced or distorted auditory input over longer time periods affects the internal representation of phonetic knowledge. For example, it has been reported recently that otitis media in childhood (Welsh, Welsh, & Healy, 1983) or monaural hearing deprivation in adulthood (Silman, Gelfand, & Silverman, 1984) may result in reduced speech perception capabilities. Certainly, the congenitally hearing-impaired must have a very different representation of their limited phonetic experiences, and hearing impairments acquired later in life may distort the knowledge base as well. It has often been observed that the speech perception of the hearing-impaired is not completely predictable from assessments of auditory capacity (e.g., Tyler, Summerfield, Wood, & Fernandes, 1982). One reason for this may be that there are distortions, not only in the auditory processing of speech (to which they are commonly attributed), but also in the mental representations that hearing-impaired listeners refer to in phonetic classification. Such distortions are especially likely to result when hearing deteriorates progressively at a rate that exceeds the rate at which mental prototypes can be modified: A listener then expects to hear things that the auditory system cannot deliver. On the other hand, if the prototypes are degraded from many years of impoverished auditory experience, then there is little hope of improving speech perception by "improving" the acoustic signal, at least not without extensive training to rebuild the prototypes (cf. Sidwell & Summerfield, 1985).

5. Making Psychophysics More Relevant to Speech Research

One characteristic of the psychophysical approach is that it is domain-independent. The psychophysical methods applied in the study of speech perception are essentially the same as those applied in research on auditory, visual, or tactile perception of nonspeech stimuli. Indeed, the generality across different stimulus domains and modalities of Weber's law or the law of temporal summation has been an important discovery. Such laws are in accord with behaviorist and information-processing orientations in psychology, which assume that perception and cognition are governed by general-purpose, domain-independent processes. The description of such processes is an important part of psychological research.

By focusing on domain-independent laws of perception, however, psychophysics essentially ignores those features that are specific to speech and whose investigation is critical to an understanding of speech perception as distinct from perception in general. Of course, there are many aspects that speech shares with nonspeech sounds and even with stimuli in other modalities. Research on the perception of those, however, leads only to an understanding of sound perception, temporal change perception, timbre perception, even categorization--in short, of all the things that speech perception has in common with nonspeech perception. What is missing is the main ingredient: the content. To understand speech perception fully, research needs to focus on the unique properties of speech, which include the facts that it is articulated (and hence peculiarly structured), capable of being imitated by a perceiver, and perceived as segmentally structured for purposes of linguistic communication. I see at least one way in which the sophisticated methods of psychophysics could be adapted to these special features and thus be made more relevant to speech research.

Psychoacoustic approaches to speech perception deal with both stimulus and response at some remove from the mechanism that is directly responsible for most (if not all) special properties of speech: the vocal tract. A more speech-relevant psychophysics might examine the articulatory source of the acoustic signal in relation to what is probably the most direct evidence that perception has occurred--the perceiver's vocal reproduction of what has been heard (or seen). I am thus proposing an articulatory psychophysics based on the realization that speech is constituted of motor events (cf. Liberman & Mattingly, 1985). Its goal would be to describe the lawful relationships between a talker's articulations and a listener's perception or imitation of them.

A first step in this enterprise would be to look at the speech signal not in terms of its acoustic properties, but in terms of the articulatory information that it conveys. This is done most easily by generating the stimuli using an articulatory synthesizer or an actual human talker, perhaps in conjunction with analytic methods for extracting the vocal tract area function from the acoustic signal (e.g., Atal, Chang, Mathews, & Tukey, 1978; Ladefoged, Harshman, Goldstein, & Rice, 1978; Schroeder & Strube, 1979). Articulatory synthesis studies in the literature (e.g., Abramson, Nye, Henderson, & Marshall, 1981; Kasuya, Takeuchi, Sato, & Kido, 1982; Lindblom & Sundberg, 1971; Rubin, Baer, & Mermelstein, 1981) illustrate this approach. A second step would be to examine subjects' articulatory (rather than just written) response to speech stimuli. Studies of vocal imitation (e.g., Chistovich, Fant, de Serpa-Leitao, & Tjernlund, 1966; Kent, 1973; Repp & Williams, 1985) commonly have analyzed stimulus-response relationships in terms of acoustic parameters and thus fall somewhat short of the stated goal. In the wide field of speech production research, there are few studies that have required subjects to listen to speech stimuli and reproduce them; almost always the task has been to read words or nonsense materials, and measurements have focused on normative productions characteristic of a language, not on talkers' imitative or articulatory skills. The final step towards a true articulatory psychophysics would be to measure subjects' articulatory response to articulatorily defined stimuli, generated either by an articulatory synthesizer or by a human model whose articulators are likewise monitored. An important (though necessarily crude) example of this still rare approach is the work of Meltzoff and Moore (see 1985) on facial imitation in infancy. More detailed studies of adult subjects should benefit from the development of more economic descriptions of articulation and its underlying control parameters (Browman & Goldstein, 1985; Kelso, Vatikiotis-Bateson, Saltzman, & Kay, 1985).

Such studies would assess how articulatory dimensions such as jaw height, lip rounding, mouth opening, or velar elevation--or perhaps more global articulatory parameters such as the vocal tract area function--are apprehended by a listener/speaker, and how they are translated and rescaled to fit his or her own articulatory dimensions. Rather than relating physical stimulus parameters to some subjective auditory scale that is irrelevant to speech communication, the psychophysical function would relate equivalent articulatory measures in the model speaker and the imitator. Such functions would relate more directly to questions of speech acquisition and phonetic language learning than any measure of auditory perception. Even though articulatory psychophysics is likely to encounter various influences of linguistic categories on the subject's articulatory response, reflecting aspects of motor control that have become established through habit and practice, at least it would bypass the stage of overt categorical decisions (cf. Chistovich et al., 1966) that characterizes so many laboratory tasks. It may be possible to overcome these articulatory habits through training, and such training may not only yield better estimates of articulatory information transfer but also potential practical benefits for second-language learners and speech pathologists (more so than training in auditory discrimination). An ancillary, hitherto little-investigated topic is that of articulatory awareness--a talker's ability to consciously observe and manipulate his or her articulators.

6. Summary

Before summing up, one qualification is in order concerning the role of psychophysics in understanding speech perception. I have argued that this role is limited, and undoubtedly many will disagree with this opinion. In addition, however, I have followed the custom of the mainstream speech perception literature (and my own proclivities) by considering speech perception to be synonymous with the perception of phonetic structure. There are many other aspects of speech, however, such as intonation, stress, speaking rate, effort, rhythm, emotion, voice quality, speaker characteristics, room reverberation, and separation from other environmental sounds. All these aspects are worthy of detailed investigation, and although speech-specific knowledge also plays a role in their perception (e.g., Ainsworth & Lindsay, 1986; Darwin, 1984; Tuller & Fowler, 1980), auditory psychophysics probably has a more important contribution to make to research on these topics. The perception of subtle gradations becomes especially important in the registration of paralinguistic information. Thus, in yet another sense, the relevance of psychophysics to speech perception depends on how broadly or narrowly the field of speech perception research is defined.

In this paper I have tried to do five things. First, I have attempted to characterize the psychophysics of speech perception in terms of certain biases: heavy emphasis on the auditory modality; preoccupation with methodology; treatment of speech as a collection of sounds; neglect of the perceiver's knowledge and expectations. This characterization may well seem a caricature to those who espouse a broad definition of psychophysics. However, even though only a small part of speech perception research may fit my description, it represents an extreme (a prototype of psychophysical orthodoxy, as it were) that, though only rarely instantiated in its pure form, nevertheless exerts a certain "pull" on research in the field.

Second, I have tried to ask what it means to understand speech perception. Far from giving a satisfactory answer to this difficult question, I have made two points: Perception can be defined narrowly as a rigid process of transduction, or more broadly as a flexible process of relating the input to a knowledge base; I favor the second definition. As to understanding, it can mean producing some tangible evidence, such as a good recognition algorithm, or it can remain largely a matter of personal indulgence. My sympathies are with the former approach, but my own research has been very much within the latter.

Third, I have characterized speech perception as the application of detailed phonetic knowledge. I have argued that the mechanisms of speech perception may be quite general, but that the system as a whole is unique, thus stating a modified (possibly trivial) version of the modularity hypothesis (Fodor, 1983; Liberman & Mattingly, 1985). This has led me further to suggest that speech perception, when considered divorced from the whole system, is a relatively shallow topic for investigation, and that a better understanding of speech perception will result indirectly from studying the whole "speech chain" (Denes & Pinson, 1963).

Fourth, I have discussed four major research questions that follow from the view taken here: Description of the phonetic knowledge; rules of its application; time course of its acquisition; and its modifiability in adulthood. The first and third of these topics are considered central to speech research. Many traditional core questions of speech perception, together with opportunities for the application for psychophysical methods, are contained in the second topic and thus are assigned a secondary role. Special emphasis is placed on articulatory and acoustic phonetics as a means for gaining insight into the language user's perceptual knowledge.

Finally, I have proposed the possibility of an articulatory psychophysics as a way of increasing the relevance of psychophysical methods to speech research.

In sum, I have painted a somewhat pessimistic picture of speech perception research, and in particular of the contribution of psychophysical approaches. This should not be taken as an assault on auditory psychophysics as such; on the contrary, the investigation of auditory function is an important area in which much excellent work is being done, as illustrated by many contributions to this workshop. What is at issue is the relevance of this general approach to the study of speech perception. If my paper stimulates discussion of this fundamental question, it will have served its purpose.

References

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F_0 versus voice timing. In V. A. Fromkin (Ed.), Phonetic linguistics. Essays in honor of Peter Ladefoged (pp. 25-33). New York: Academic.

Abramson, A. S., Nye, P. W., Henderson, J. B., & Marshall, C. W. (1981). Vowel height and the perception of consonantal nasality. Journal of the Acoustical Society of America, 70, 329-393.

Ainsworth, W. A., & Lindsay, D. (1986). Perception of pitch movement on tonic syllables in British English. Journal of the Acoustical Society of America, 79, 472-480.

Atal, B. S., Chang, J. J., Mathews, M. V., & Tukey, J. W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. Journal of the Acoustical Society of America, 63, 1535-1555.

Bailey, P. J., & Summerfield, Q. (1980). Information in speech: Observations on the perception of [s]-stop clusters. Journal of Experimental Psychology: Human Perception and Performance, 6, 536-563.

Bladon, R. A. W., & Lindblom, B. E. F. (1981). Modeling the judgment of vowel quality. Journal of the Acoustical Society of America, 69, 1414-1422.

Blomberg, M., Carlson, R., Elenius, K., & Granström, B. (1986). Auditory models as front ends in speech recognition systems. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability in speech processes (pp. 108-114). Hillsdale, NJ: Erlbaum.

Browman, C. P., & Goldstein, L. M. (1985). Dynamic modeling of phonetic structure. In V. A. Fromkin (Ed.), Phonetic linguistics. Essays in honor of Peter Ladefoged (pp. 35-53). New York: Academic.

Bregman, A. S. (1977). Perception and behavior as compositions of ideals. Cognitive Psychology, 9, 250-292.

Chistovich, L. A. (1971). Problems of speech perception. In L. L. Hammerich, R. Jakobson, & E. Zwirner (Eds.), Form and substance (pp. 83-93). Copenhagen: Akademisk Forlag.

Chistovich, L. A. (1980). Auditory processing of speech. Language and Speech, 23, 67-75.
Chistovich, L. A. (1985). Central auditory processing of peripheral vowel spectra. Journal of the Acoustical Society of America, 77, 789-805.
Chistovich, L. A., Fant, G., de Serpa-Leitao, A., & Tjernlund, P. (1966). Mimicking and perception of synthetic vowels. Quarterly Progress and Status Report (Royal Technical University, Speech Transmission Laboratory, Stockholm), 2, 1-18.
Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge, MA: MIT Press.
Chomsky, N. (1968). Language and mind. New York: Harcourt, Brace & World.
Darwin, C. J. (1984). Perceiving vowels in the presence of another sound: Constraints on formant perception. Journal of the Acoustical Society of America, 76, 1636-1647.
Denes, P. B., & Pinson, E. N. (1963). The speech chain. Murray Hill, NJ: Bell Telephone Laboratories.
Derr, M. A., & Massaro, D. W. (1980). The contribution of vowel duration, F_0 contour, and frication duration as cues to the /juz/-/jus/ distinction. Perception & Psychophysics, 27, 51-59.
Diehl, R. (in press). Auditory constraints on the perception of speech. In M. E. H. Schouten (Ed.), The psychophysics of speech perception. The Hague: Martinus Nijhoff Publishers.
Elliott, L. L., Longinotti, C., Clifton, L.-A., & Meyer, D. (1981). Detection and identification thresholds for consonant-vowel syllables. Perception & Psychophysics, 30, 411-416.
Elman, J. L., & McClelland, J. L. (1984). Speech perception as a cognitive process: The interactive activation model. In N. J. Lass (Ed.), Speech and language: Advances in research and practice (Vol. 10, pp. 337-374). New York: Academic.
Elman, J. L., & McClelland, J. L. (1986). Exploiting lawful variability in the speech wave. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability in speech processes (pp. 360-380). Hillsdale, NJ: Erlbaum.
Flege, J. E. (in press). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), Human communication and its disorders, Vol. 1. Norwood, NJ: Ablex.
Fodor, J. (1983). The modularity of mind. Cambridge, MA: MIT Press.
Fónagy, I. (1961). Communication in poetry. Word, 17, 194-218.
Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. Perception & Psychophysics, 36, 359-368.
Fujisaki, H., & Kawashima, T. (1969). On the modes and mechanisms of speech perception. Annual Report of the Engineering Research Institute (Faculty of Engineering, University of Tokyo), 28, 67-73.
Fujisaki, H., & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. Annual Report of the Engineering Research Institute (Faculty of Engineering, University of Tokyo), 29, 207-214.
Ganong, W. F., III, & Zatorre, R. J. (1980). Measuring phoneme boundaries four ways. Journal of the Acoustical Society of America, 68, 431-439.
Gibson, J. J. (1966). The senses considered as perceptual systems. Boston: Houghton Mifflin.
Green, K. P., & Miller, J. L. (1985). On the role of visual rate information in phonetic perception. Perception & Psychophysics, 38, 269-276.
Grossberg, S. (in press). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), Pattern recognition by humans and machines (Vol. I). New York: Academic.

Hary, J. M., & Massaro, D. W. (1982). Categorical results do not imply categorical perception. Perception & Psychophysics, 32, 409-418.

Hayek, F. A. (1952). The sensory order. Chicago: University of Chicago Press.

Howell, P., & Rosen, S. (1983). Production and perception of rise time in the voiceless affricate/fricative distinction. Journal of the Acoustical Society of America, 73, 976-984.

Hrushovski, B. (1980). The meaning of sound patterns in poetry. Poetics Today, 2, 39-56.

Jusczyk, P. W. (1985). On characterizing the development of speech perception. In J. Mehler & R. Fox (Eds.), Neonate cognition: Beyond the blooming buzzing confusion (pp. 199-229). Hillsdale, NJ: Erlbaum.

Jusczyk, P. W. (1986). Toward a model of the development of speech perception. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability in speech processes (pp. 1-18). Hillsdale, NJ: Erlbaum.

Jusczyk, P. W. (in press). Implications from infant speech studies on the unit of perception. In M. E. H. Schouten (Ed.), The psychophysics of speech perception. The Hague: Martinus Nijhoff Publishers.

Kasuya, H., Takeuchi, S., Sato, S., & Kido, K. (1982). Articulatory parameters for the perception of bilabials. Phonetica, 39, 61-70.

Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E. L., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. Journal of the Acoustical Society of America, 77, 266-280.

Kent, R. D. (1973). The imitation of synthetic vowels and some implications for speech memory. Phonetica, 28, 1-25.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. Journal of Phonetics, 7, 279-312.

Klatt, D. H. (1982). Prediction of perceived phonetic distance from critical-band spectra: A first step. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Paris, France (pp. 1278-1281). New York: IEEE.

Klatt, D. H. (1986). Problem of variability in speech recognition and in models of speech perception. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability in speech processes (pp. 300-319). Hillsdale, NJ: Erlbaum.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. Journal of the Acoustical Society of America, 66, 1668-1679.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. Journal of the Acoustical Society of America, 70, 340-349.

Kuhl, P. K. (1985). Categorization of speech by infants. In J. Mehler & R. Fox (Eds.), Neonate cognition: Beyond the blooming buzzing confusion (pp. 231-262). Hillsdale, NJ: Erlbaum.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. Science, 218, 1138-1141.

Ladefoged, P., Harshman, R., Goldstein, L., & Rice, L. (1978). Generating vocal tract shapes from formant frequencies. Journal of the Acoustical Society of America, 64, 1027-1035.

Liberman, A. M. (1982). On finding that speech is special. American Psychologist, 37, 148-167.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21, 1-36.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), The production of speech (pp. 207-246). New York: Springer-Verlag.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & D. Dahl (Eds.), Explanations of linguistic universals (pp. 181-203). The Hague: Mouton.

Lindblom, B. E. F., & Sundberg, J. E. F. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. Journal of the Acoustical Society of America, 50, 1166-1179.

Linell, P. (1982). The concept of phonological form and the activities of speech production and speech perception. Journal of Phonetics, 10, 37-72.

Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D., & Mandler, R. (1977). On pushing the voice-onset-time (VOT) boundary about. Language and Speech, 20, 209-216.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. Applied Psycholinguistics, 2, 369-390.

Macmillan, N. A., Braida, L. D., Goldberg, R. F., & Khazatsky, V. (in press). Central and peripheral processes in the perception of speech and nonspeech sounds. In M. E. H. Schouten (Ed.), The psychophysics of speech perception. The Hague: Martinus Nijhoff Publishers.

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. (1977). The psychophysics of categorical perception. Psychological Review, 84, 452-471.

Marks, L. E. (1978). The unity of the senses. New York: Academic.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology, 10, 29-63.

Massaro, D. W. (in press-a). Categorical partition: A fuzzy logical model of categorization behavior. In S. N. Harnad (Ed.), Categorical perception. New York: Cambridge University Press.

Massaro, D. W. (in press-b). A commentary on research and theory in speech perception. In M. E. H. Schouten (Ed.), The psychophysics of speech perception. The Hague: Martinus Nijhoff Publishers.

Massaro, D. W., & Cohen, M. M. (1983a). Evaluation and integration of visual and auditory information in speech perception. Journal of Experimental Psychology: Human Perception and Performance, 9, 753-771.

Massaro, D. W., & Cohen, M. M. (1983b). Categorical or continuous speech perception: a new test. Speech Communication, 2, 15-35.

Massaro, D. W., & Oden, G. C. (1980a). Speech perception: A framework for research and theory. In N. J. Lass (Ed.), Speech and language: Advances in research and practice (Vol. 3, pp. 129-165). New York: Academic.

Massaro, D. W., & Oden, G. C. (1980b). Evaluation and integration of acoustic features in speech perception. Journal of the Acoustical Society of America, 67, 996-1013.

Mattingly, I. G. (1972). Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and eye: The relationships between speech and reading (pp. 133-148). Cambridge, MA: MIT Press.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.

Meltzoff, A. N., & Moore, M. K. (1985). Cognitive foundations and social functions of imitation and intermodal representation in infancy. In J. Mehler & R. Fox (Eds.), Neonate cognition: Beyond the blooming buzzing confusion (pp. 139-156). Hillsdale, NJ: Erlbaum.

Miller, G. A., & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 27, 338-352.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? Cognition, 7, 323-331.

Neisser, U. (1976). Cognition and reality: Principles and implications of cognitive psychology. San Francisco: Freeman.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Marek, R. A. Hendrick, & M. F. Miller (Eds.), Papers from the parasession in language and bahavior (pp. 178-203). Chicago: Chicago Linguistic Society.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), The production of speech (pp. 189-216). New York: Springer-Verlag.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception & Psychophysics, 13, 253-260.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. Journal of the Acoustical Society of America, 55, 328-333.

Remez, R. E. (in press). Units of organization and analysis in the perception of speech. In M. E. H. Schouten (Ed.), The psychophysics of speech perception. The Hague: Martinus Nijhoff Publishers.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. Science, 212, 947-950.

Repp, B. H. (1981). On levels of description in speech research. Journal of the Acoustical Society of America, 69, 1462-1464.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. Psychological Bulletin, 92, 81-110.

Repp, B. H. (1983). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. Speech Communication, 2, 341-362.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), Speech and language: Advances in research and practice (Vol. 10, pp. 243-335). New York: Academic.

Repp, B. H., & Liberman, A. M. (in press). Phonetic category boundaries are flexible. In S. N. Harnad (Ed.), Categorical perception. New York: Cambridge University Press.

Repp, B. H., & Williams, D. R. (1985). Categorical trends in vowel imitation: Preliminary observations from a replication experiment. Speech Communication, 4, 105-120.

Rosen, S., & Howell, P. (in press). Auditory, articulatory, and learning explanations of categorical perception in speech. In S. N. Harnad (Ed.), Categorical Perception. New York: Cambridge University Press.

Rosner, B. S. (1984). Perception of voice-onset-time continua: A signal detection analysis. Journal of the Acoustical Society of America, 75, 1231-1242.

Rubin, P., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. Journal of the Acoustical Society of America, 70, 321-328.

Sachs, R. M., & Grant, K. W. (1976). Stimulus correlates in the perception of voice onset time (VOT): II. Discrimination of speech with high and low stimulus uncertainty. Journal of the Acoustical Society of America, 60 (Suppl. No. 1), S91. (Abstract)

Samuel, A. G. (1977). The effect of discrimination training on speech perception: Noncategorical perception. Perception & Psychophysics, 22, 321-330.

Samuel, A. G. (1982). Phonetic prototypes. Perception & Psychophysics, 31, 307-314.

Schouten, M. E. H. (in press). Speech perception and the role of long-term memory. In M. E. H. Schouten (Ed.), The psychophysics of speech perception. The Hague: Martinus Nijhoff Publishers.

Schroeder, M. R., & Strube, H. W. (1979). Acoustic measurements of articulator motions. Phonetica, 36, 302-313.

Shepard, R. N. (1980). Psychophysical complementarity. In M. Kubovy & J. R. Pomerantz (Eds.), Perceptual organization (pp. 279-341). Hillsdale, NJ: Erlbaum.

Shepard, R. N. 1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. Psychological Review, 91, 417-447.

Sidwell, A., & Summerfield, Q. (1985). The effect of enhanced spectral contrast on the internal representation of vowel-shaped noise. Journal of the Acoustical Society of America, 78, 495-506.

Silman, S., Gelfand, S. A., & Silverman, K. E. A. (1984). Late-onset auditory deprivation: Effects of monaural versus binaural hearing aids. Journal of the Acoustical Society of America, 76, 1357-1362.

Soli, S. D. (1983). The role of spectral cues in discrimination of voice onset time differences. Journal of the Acoustical Society of America, 73, 2150-2165.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. Journal of the Acoustical Society of America, 64, 1358-1368.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), Perspectives in the study of speech (pp. 1-38). Hillsdale, NJ: Erlbaum.

Studdert-Kennedy, M. (1982). On the dissociation of auditory and phonetic perception. In R. Carlson & B. Granström (Eds.), The representation of speech in the peripheral auditory system (pp. 9-26). Amsterdam: Elsevier.

Studdert-Kennedy, M. (1985). Perceiving phonetic events. In W. H. Warren & R. E. Shaw (Eds.), Persistence and change: Proceedings of the first international conference on event perception (pp. 139-156). Hillsdale, NJ: Erlbaum.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. Psychological Review, 77, 234-249.

Summerfield, Q. (1979). Use of visual information for phonetic perception. Phonetica, 36, 314-331.

Summerfield, Q. (in press). Preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), Hearing by eye. Hilldale, NJ: Erlbaum.

Tallal, P., & Stark, R. (1981). Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. Journal of the Acoustical Society of America, 69, 568-574.

Tatham, M. A. A. (1980). Phonology and phonetics as part of the language encoding/decoding system. In N. J. Lass (Ed.), Speech and language: Advances in research and practice (Vol. 3, pp. 35-73). New York: Academic.

Titchener, E. B. (1909). Lectures on the elementary psychology of the thought-process.

Toulmin, S. (1972). Human understanding: The collective use and evolution of concepts. Princeton: Princeton University Press.

Traunmüller, H. (in press). Some aspects of the sound of speech sounds. In M. E. H. Schouten (Ed.), The psychophysics of speech perception. The Hague: Martinus Nijhoff Publishers.

Tuller, B., & Fowler, C. A. (1980). Some articulatory correlates of perceptual isochrony. Perception & Psychophysics, 27, 277-283.

Tyler, R. S., Summerfield, Q., Wood, E. J., & Fernandes, M. A. (1982). Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners. Journal of the Acoustical Society of America, 72, 740-752.

Vinegrad, M. D. (1972). A direct magnitude scaling method to investigate categorical versus continuous modes of speech perception. Language and Speech, 15, 114-121.

Warren, R. M. (1981). Chairman's comments. In T. Myers, J. Laver, & J. Anderson (Eds.), The cognitive representation of speech (pp. 34-37). Amsterdam: North-Holland.

Watson, C. S., Kewley-Port, D., & Foyle, D. C. (1985). Temporal acuity for speech and nonspeech sounds: The role of stimulus uncertainty. Journal of the Acoustical Society of America, 77 (Suppl. No. 1), S27. (Abstract)

Welsh, L. W., Welsh, J. J., & Healy, M. P. (1983). Effect of sound deprivation on central hearing. Laryngoscope, 93, 1569-1575.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. Child Development, 52, 349-355.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development, 7, 49-63.

Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. Journal of the Acoustical Society of America, 60, 1381-1389.

Yates, J. (1985). The content of awareness is a model of the world. Psychological Review, 92, 249-284.

SPECIALIZED PERCEIVING SYSTEMS FOR SPEECH AND OTHER BIOLOGICALLY SIGNIFICANT SOUNDS*

Ignatius G. Mattingly† and Alvin M. Liberman††

Abstract. Perception of speech rests on a specialized mode, narrowly adapted for the efficient production and perception of phonetic structures. This mode is similar in some of its properties to the specializations that underlie, for example, sound localization in the barn owl, echolocation in the bat, and song in the bird.

Our aim is to present a view of speech perception that runs counter to the conventional wisdom. Put so as to touch the point of this symposium, our unconventional view is that speech perception is to humans as sound localization is to barn owls. This is not merely to suggest that humans are preoccupied with listening to speech, much as owls are with homing in on the sound of prey. It is, rather, to offer a particular hypothesis: like sound localization, speech perception is a coherent system in its own right, specifically adapted to a narrowly restricted class of ecologically significant events. In this important respect, speech perception and sound localization are more similar to each other than is either to the processes that underlie the perception of such ecologically arbitrary events as squeaking doors, rattling chains, or whirring fans.

To develop the unconventional view, we will contrast it with its more conventional opposite, say why the less conventional view is nevertheless the more plausible, and describe several properties of the speech-perceiving system that the unconventional view reveals. We will compare speech perception with other specialized perceiving systems that also treat acoustic signals, including not only sound localization in the owl, but also song in the bird and echolocation in the bat. Where appropriate, we will develop the neurobiological implications, but we will not try here to fit them to the vast and diverse literature that pertains to the human case.

---

*To appear in G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), <u>Functions of the auditory system</u>. New York: Wiley.

†Also University of Connecticut
††Also University of Connecticut and Yale University

Acknowledgment. The writing of this paper was supported by a grant to Haskins Laboratories (NIH-NICHD-HD-01994). We are grateful to Harriet Magen and Nancy O'Brien for their help with references and to Alice Dadourian for invaluable editorial assistance and advice. We received shrewd comments and suggestions from Carol Fowler, Masakazu Konishi, Eric Knudsen, David Margoliash, Bruno Repp, Michael Studdert-Kennedy, Nobuo Suga, and Douglas Whalen. Some of these people have views very different from those expressed in this paper, but we value their criticisms all the more for that.

Through most of this paper we will construe speech, in the narrow sense, as referring only to consonants and vowels. Then, at the end, we will briefly say how our view of speech might nevertheless apply more broadly to sentences.

Following the instructions of our hosts, we will concern ourselves primarily with issues and principles. We will, however, offer the results of just a few experiments, not so much to prove our argument as to illuminate it.[1]

Two Views of Speech Perception: Generally Auditory vs. Specifically Phonetic

The conventional view derives from the common assumption that mental processes are not specific to the real-world events to which they are applied. Thus, perception of speech is taken to be in no important way different from perception of other sounds.[2] In all cases, it is as if the primitive auditory consequences of acoustic events were delivered to a common register (the primary auditory cortex?), from whence they would be taken for such cognitive treatment as might be necessary in order to categorize each ensemble of primitives as representative of squeaking doors, stop consonants, or some other class of acoustic events. On any view, there are, of course, specializations for each of the several auditory primitives that, together, make up the auditory modality, but there is surely no specialization for squeaking doors as such, and, on the conventional view, none for stop consonants, either.

Our view is different on all counts. Seen our way, speech perception takes place in a specialized phonetic mode, different from the general auditory mode and served, accordingly, by a different neurobiology. Contrary to the conventional assumption, there is, then, a specialization for consonants and vowels as such. This specialization yields only phonetic structures; it does not deliver to a common auditory register those sensory primitives that might, in arbitrarily different combinations, be cognitively categorized as any of a wide variety of ordinary acoustic events. Thus, specialization for perception of phonetic structures begins prior to such categorization and is independent of it.

The phonetic mode is not auditory, in our view, because the events it perceives are not acoustic. They are, rather, gestural. For example, the consonant [b] is a lip-closing gesture; [h] is a glottis-opening gesture. Combining lip-closing and glottis-opening yields [p]; combining lip-closing and velum-lowering yields [m], and so on. Despite their simplistic labels, the gestures are, in fact, quite complex: as we shall see, a gesture usually requires the movements of several articulators, and these movements are most often context-sensitive. A rigorous definition of a particular gesture has, therefore, to be fairly abstract. Nevertheless, it is the gestures that we take to be the primitives of speech perception, no less than of speech production. Phonetic structures are patterns of gestures, then, and it is just these that the speech system is specialized to perceive.

The Plausible Function of a Specially Phonetic Mode

But why should consonants and vowels be gestures, not sounds, and why should it take a specialized system to perceive them? To answer these questions, it is helpful to imagine the several ways in which phonetic communication might have been engineered.

Accepting that Nature had made a firm commitment to an acoustic medium, we can suppose that she might have defined the phonetic segments -- the consonants and vowels--in acoustic terms. This, surely, is what common sense suggests, and, indeed, what the conventional view assumes. The requirements that follow from this definition are simply that the acoustic signals be appropriate to the sensitivities of the ear, and that they provide the invariant basis for the correspondingly invariant auditory percept by which each phonetic segment is to be communicated. The first requirement is easy enough to satisfy, but the second is not. For if the sounds are to be produced by the organs of the vocal tract, then strings of acoustically defined segments require strings of discrete gestures. Such strings can be managed, of course, but only at unacceptably slow rates. Indeed, we know exactly how slow, because speaking so as to produce a segment of sound for each phonetic segment is what we do when we spell. Thus, to articulate the consonant-vowel syllables [di] and [du], for example, the speaker would have to say something like [də i] and [də u], converting each consonant and each vowel into a syllable. Listening to such spelled speech, letter by painful letter, is not only time-consuming, but also maddeningly hard.

Nature might have thought to get around this difficulty by abandoning the vocal tract in favor of a to-be-developed set of sound-producing devices, specifically adapted for creating the drumfire that communication via acoustic segments would require if speakers were to achieve the rates that characterize speech as we know it, rates that run at eight to ten segments per second, on average, and at double that for short stretches. But this would have defeated the ear, severely straining its capacity to identify the separate segments and keep their order straight.

Our view is that Nature solved the problems of rate by avoiding the acoustic strategy that gives rise to them. The alternative was to define the phonetic segments as gestures, letting the sound go pretty much as it might, so long as the acoustic consequences of the different gestures were distinct. On its face, this seems at least a reasonable way to begin, for it takes into account that phonetic structures are not really objects of the acoustic world anyway; they belong, rather, to a domain that is internal to the speaker, and it is the objects of this domain that need to be communicated to the listener. But the decisive consideration in favor of the gestural strategy is surely that it offers critical advantages for rate of communication, both in production and in perception. These advantages were not to be had, however, simply by appropriating movements that were already available--for example, those of eating and breathing. Rather, the phonetic gestures and their underlying controls had to be developed, presumably as part of the evolution of language. Thus, as we will argue later, speech production is as much a specialization as speech perception; as we will also argue, it is, indeed, the same specialization.

In production, the advantage of the gestural strategy is that, given the relative independence of the muscles and organs of the vocal tract and the development of appropriately specialized controls, gestures belonging to successive segments in the phonetic string can be executed simultaneously or with considerable overlap. Thus, the gesture for [d] is overlapped with component gestures for the following vowel, whether [i] or [u]. By just such coarticulation, speakers achieve the high rates at which phonetic structures are, in fact, transmitted, rates that would be impossible if the gestures had to be produced seriatim.

In perception, the advantage of the gestural strategy is that it provides the basis for evading the limit on rate that would otherwise have been set by the temporal resolving abilities of the auditory system. This, too, is a consequence of coarticulation. Information about several gestures is packed into a single segment of sound, thereby reducing the number of sound segments that must be dealt with per unit time.

But the gain for perception is not without cost, for if information about several gestures is transmitted at the same time, the relation between these gestures and their acoustic vehicles cannot be straightforward. It is, to be sure, systematic, but only in a way that has two special and related consequences. First, there is no one-to-one correspondence in segmentation between phonetic structure and signal; information about the consonant and the vowel can extend from one end of the acoustic syllable to the other. Second, the shape of the acoustic signal for each particular phonetic gesture varies according to the nature of the concomitant gestures and the rate at which they are produced. Thus, the cues on which the processes of speech perception must rely are context-conditioned. For example, the perceptually significant second-formant transition for [d] begins high in the spectrum and rises for [di], but begins low in the spectrum and falls for [du].

How might the complications of this unique relation have been managed? Consider, first, the possibility that no further specialization is provided, the burden being put, rather, on the perceptual and cognitive equipment with which the listener is already endowed. By this strategy, the listener uses ordinary auditory processes to convert the acoustic signals of speech to ordinary auditory percepts. But then, having perceived the sound, the listener must puzzle out the combination of coarticulated gestures that might have produced it, or, failing that, learn ad hoc to connect each context-conditioned and eccentrically segmented token to its proper phonetic type. However, the puzzle is so thorny as to have proved, so far, to be beyond the capacity of scientists to solve; and, given the large number of acoustic tokens for each phonetic type, ad hoc learning might well have been endless. Moreover, listening to speech would have been a disconcerting experience at best, for the listener would have been aware, not only of phonetic structure, but also of the auditory base from which phonetic structure would have had to be recovered. We gain some notion of what this experience would have been like when we hear, in isolation from their contexts, the second-formant transitions that cue [di] and [du]. As would be expected on psychoacoustic grounds, the transition for [di] sounds like a rising glissando on high pitches (or a high-pitched chirp); the transition for [du], like a falling glissando on low pitches (or a low-pitched chirp). If the second-formant transition is combined with the concomitant transitions of other formants, the percept becomes a "bleat" whose timbre depends on the nature of the component transitions. Fluent speech, should it be heard in this auditory way, would thus be a rapid sequence of qualitatively varying bleats. The plight of the listener who had to base a cognitive analysis of phonetic structure on such auditory percepts would have been like that of a radio operator trying to follow a rapid-fire sequence of Morse code dots and dashes, only worse, because, as we have seen, the "dots and dashes" of the speech code take as many different acoustic forms as there are variations in context and rate.

The other strategy for recovering phonetic structure from the sound--the one that must have prevailed--was to use an appropriate specialization. Happily, this specialization was already at hand in the form of those arrangements, previously referred to, that made it possible for speakers to articulate and coarticulate phonetic gestures. These must have incorporated in their architecture all the constraints of anatomy, physiology, and phonetics that organize the movements of the speech organs and govern their relation to the sound, so access to this architecture should have made it possible, in effect, to work the process in reverse--that is, to use the acoustic signal as a basis for computing the coarticulated gestures that caused it. It is just this kind of perception-production specialization that our view assumes. Recovering phonetic structure requires, then, no prodigies of conscious computation or arbitrary learning. To perceive speech, a person has only to listen, for the specialization yields the phonetic percept immediately. This is to say that there is no conscious mediation by an auditory base. Rather, the gestures for consonants and vowels, as perceived, are themselves the distal objects; they are not, like the dots and dashes of Morse code (or the squeak of the door), at one remove from it. But perception is immediate in this case (and in such similar cases as, for example, sound localization), not because the underlying processes are simple or direct, but only because they are well suited to their unique and complex task.

Some Properties of the Phonetic Mode Compared with Those of Other Perceptual Specializations

Every perceptual specialization must differ from every other in the nature of the distal events it is specialized for, as it must, too, in the relation between these events and the proximal stimuli that convey them. At some level of generality, however, there are properties of these specializations that invite comparison. Several of the properties that are common, perhaps, to all perceiving specializations--for example, "domain specificity," "mandatory operation," and "limited central access"--have been described by Fodor (1983: Part III), and claimed by us to be characteristic of the phonetic mode (Liberman & Mattingly, 1985). We do not review these here, but choose, rather, to put our attention on four properties of the phonetic mode that are not so widely shared and that may, therefore, define several subclasses.

Heteromorphy.

The phonetic mode, as we have conceived it, is "heteromorphic" in the sense that it is specialized to yield perceived objects whose dimensionalities are radically different from those of the proximal stimuli.[3] Thus, the synthetic formant transitions that are perceived homomorphically in the auditory mode as continuous glissandi are perceived heteromorphically in the phonetic mode as consonant or vowel gestures that have no glissando-like auditory qualities at all. But is it not so in sound localization, too? Surely, interaural disparities of time and intensity are perceived heteromorphically, as locations of sound sources, and not homomorphically, as disparities, unless the interaural differences are of such great magnitude that the sound-localizing specialization is not engaged. Thus, the heteromorphic relation between distal object and the display at the sense organ is not unique to phonetic perception. Indeed, it characterizes, not only sound localization, but also, perhaps, echolocation in the bat, if we can assume that, as Suga's (1984) neurobiological results imply, the bat perceives, not echo-time as such, but rather something more like the distance it measures.

If we look to vision for an example, we find an obvious one in stereopsis, where perception is not of two-dimensionally disparate images, but of third-dimensional depth.

To see more clearly what heteromorphy is, let us consider two striking and precisely opposite phenomena of speech perception, together with such parallels as may be found in sound localization. In one of these phenomena, two stimuli of radically different dimensionalities converge on a single, coherent percept; in the other, stimuli lying on a single physical dimension diverge into two different percepts. In neither case can the contributions of the disparate or common elements be detected.

<u>Convergence on a single percept: Equivalence of acoustic and optical stimuli.</u> The most extreme example of convergence in speech perception was discovered by McGurk and McDonald (1976). As slightly modified for our purpose, it takes the following form. Subjects are repeatedly presented with the acoustic syllable [ba] as they watch the optical syllables [bɛ], [vɛ], [ðɛ], and [dɛ] being silently articulated by a mouth shown on a video screen. (The acoustic and optical syllables are approximately coincident.) The compelling percepts that result are of the syllables [ba], [va], [ða], and [da]. Thus, the percepts combine acoustic information about the vowels with optical information about the consonants, yet subjects are not aware--indeed, they cannot become aware--of the bimodal nature of the percept.

This phenomenon is heteromorphy of the most profound kind, for if optical and acoustic contributions to the percept cannot be distinguished, then surely the percept belongs to neither of the modalities, visual or auditory, with which these classes of stimuli are normally associated. Recalling our claim that phonetic perception is not auditory, we add now that it is not visual, either. Rather, the phonetic mode accepts all information, acoustic or optical, that pertains in a natural way to the phonetic events it is specialized to perceive. Its processes are not bound to the modalities associated with the stimuli presented to the sense organs; rather, they are organized around the specific behavior they serve and thus to their own phonetic "modality."

An analogue to the convergence of acoustic and optical stimuli in phonetic perception is suggested by the finding of neural elements in the optic tectum of the barn owl that respond selectively, not only to sounds in different locations, but also to lights in those same locations (Knudsen, 1984). Do we dare assume that the owl can't really tell whether it heard the mouse or saw it? Perhaps not, but in any case, we might suppose that, as in phonetic perception, the processes are specific to the biologically important behavior. If so, then perhaps we should speak of a mouse-catching "modality."

Putting our attention once more on phonetic perception, we ask: where does the convergence occur? Conceivably, for the example we offered, "auditory" and "visual" processes succeed, separately, in extracting phonetic units. Thus, the consonant might have been visual, the vowel auditory. These would then be combined at some later stage and, perhaps, in some more cognitive fashion. Of course, such a possibility is not wholly in keeping with our claim that speech perception is a heteromorphic specialization nor, indeed, does it sit well with the facts now available. Evidence against a late-stage, cognitive interpretation is that the auditory and visual components cannot be distinguished phenomenally, and that convergence of the McGurk-McDonald type does not occur when printed letters, which are familiar but arbitrary indices of phonetic structure, are substituted for the naturally revealing movements of the silently articulating mouth. Additional and more direct evidence, showing that the convergence occurs at an early stage, before phonetic percepts are formed, is available from a recent experiment by Green and Miller (in press; and see also Summerfield, 1979). The particular point of this experiment was to test whether optically presented information about rate of articulation affects placement on an acoustic continuum of a boundary known to be rate-sensitive, such as the one between [bi] and [pi]. Before the experiment proper, it was determined that viewers could estimate rate of articulation from the visual information alone, but could not tell which syllable, [bi] or [pi], had been produced; we may suppose, therefore, that there was no categorical phonetic information in the optical display. Nevertheless, in the main part of the experiment, the optical information about rate did affect the acoustic boundary for the phonetic contrast; moreover, the effect was consistent with what happens when the information about rate is entirely acoustic. We should conclude, then, that the visual and auditory information converged at some early stage of processing, before anything like a phonetic category had been extracted. This is what we should expect of a thoroughly heteromorphic specialization to which acoustic and optical stimuli are both relevant, and it fits as well as may be with the discovery in the owl of bimodally sensitive elements in centers as low as the optic tectum.

<u>Convergence on a coherent percept: Equivalence of different dimensions of acoustic stimulation.</u> Having seen that optical and acoustic information can be indistinguishable when, in heteromorphic specialization, they specify the same distal object, we turn now to a less extreme and more common instance of convergence in speech perception: the convergence of the disparate acoustic consequnces of the same phonetic gesture, measured most commonly by the extent to which these can be "traded," one for another, in evoking the phonetic percept for which they are all cues. If, as such trading relations suggest, the several cues are truly indistinguishable, and therefore perceptually equivalent, we should be hard put, given their acoustic diversity, to find an explanation in auditory perception. Rather, we should suppose that they are equivalent only because the speech perceiving system is specialized to recognize them as products of the same phonetic gesture.

A particularly thorough exploration of such equivalence was made with two cues for the stop consonant [p] in the word <u>split</u> (Fitch, Halwes, Erickson, & Liberman, 1980). To produce the stop, and thus to distinguish <u>split</u> from <u>slit</u>, a speaker must close and then open his lips. The closure causes a period of silence between the noise of the [s] and the vocalic portion of the syllable; the opening produces particular formant transitions at the beginning of the vocalic portion. Each of these--the silence and the transition--is a sufficient cue for the perceived contrast between <u>split</u> and <u>slit</u>. Now, the acid test of their equivalence would be to show that the <u>split-slit</u> contrast produced by the one cue cannot be distinguished from the contrast produced by the other. Unfortunately, to show this would be to prove the null hypothesis. So equivalence was tested, somewhat less directly, by assuming that truly equivalent cues would either cancel each other or summate, depending on how they were combined. The silence and transition cues for <u>split-slit</u> passed the test: patterns that differed by two cues weighted in opposite phonetic directions (one biased for [p], the other against) were harder to discriminate than patterns that differed by the same two cues weighted in the same direction (both biased for [p]).

A similar experiment, done subsequently on the contrast between say and stay, (Best, Morrongiello, & Robson, 1981) yielded similar results, but with an important addition. In one part of this later experiment, the formants of the synthetic speech stimuli were replaced by sine waves made to follow the formant trajectories. As had been found previously, such sine-wave analogues are perceived under some conditions as complex nonspeech sounds--chords, glissandi, and the like--but under others as speech (Remez, Rubin, Pisoni, & Carrell, 1981). For those subjects who perceived the sine-wave analogues as speech, the discrimination functions were much as they had been in both experiments with the full-formant stimuli. But for subjects who perceived the patterns as nonspeech, the results were different: patterns that differed by two cues were about equally discriminable, regardless of the direction of a bias in the phonetic domain; and these two-cue patterns were both more discriminable than those differing by only one. Thus, the silence cue and the transition cue are equivalent only when they are perceived in the phonetic mode as cues for the same gesture.

If we seek parallels for such equivalence in the sound-locating faculty, we find one, perhaps, in data obtained with human beings. There, binaural differences in time and in intensity are both cues to location in azimuth, and there also it has been found that the two cues truly cancel each other, though not completely (Hafter, 1984).

We consider equivalences among stimuli--whether between stimuli belonging to different modalities, as traditionally defined, or between stimuli that lie on different dimensions of the same modality--to be of particular interest, not only because they testify to the existence of a heteromorphic specialization, but also because they provide a way to define its boundaries.

Divergence into two percepts: Nonequivalence of the same dimension of acoustic stimulation in two modes. We have remarked that a formant transition (taken as an example of a speech cue) can produce two radically different percepts: a glissando or chirp when the transition is perceived homomorphically in the auditory mode as an acoustic event, or a consonant, for example, when it is perceived heteromorphically in the phonetic mode as a gesture. But it will not have escaped notice that the acoustic context was different in the two cases--the chirp was produced by a transition in isolation, the consonant by the transition in a larger acoustic pattern--and the two percepts were, of course, not experienced at the same time. It would surely be a stronger argument for the existence of two neurobiologically distinct processes, and for the heteromorphic nature of one of them, if, with acoustic context held constant, a transition could be made to produce both percepts in the same brain and at the same time. Under normal conditions, such maladaptive "duplex" perception never occurs, of course, presumably because the underlying phonetic and auditory processes are so connected as to prevent it. (In a later section, we will consider the form this connection might take.) By resort to a most unnatural procedure, however, experimenters have managed to undo the normal connection and so produce a truly duplex percept (Rand, 1974; Liberman, 1979). Into one ear--it does not matter critically which one--the experimenter puts one or another of the third-formant transitions (called the "isolated transition") that lead listeners to perceive two otherwise identical formant patterns as [da] or [ga]. By themselves, these isolated transitions sound, of course, like chirps, and listeners are at chance when required to label them as [d] or [g] (Repp, Milburn, & Ashkenas, 1983). Into the other ear is put the remaining, constant portion of the pattern (called the "base"). By itself, the base sounds like a consonant-vowel syllable, ambiguous between [da] and [ga]. But now, if the two stimuli are presented dichotically and in approximately the proper temporal arrangement, then, in the ear stimulated by the base, listeners perceive [da] or [ga], depending on which isolated transition was presented, while in the other ear they perceive a chirp. The [da] or [ga] is not different from what is heard when the full pattern is presented binaurally, nor is the chirp different from what is heard when the transition is presented binaurally without the base.

It is, perhaps, not to be wondered at that the dichotically presented inputs fuse to form the "correct" consonant-vowel syllable, since there is a strong underlying coherence. What is remarkable is that the chirp continues to be perceived, though the ambiguous base syllable does not. This is to say that the percept is precisely duplex, not triplex. Listeners perceive in the only two modes available: the auditory mode, in which they perceive chirps, and the phonetic mode in which they perceive consonant-vowel syllables.

The sensitivities of these two modes are very different, even when stimulus variation is the same. This was shown with a stimulus display, appropriate for a duplex percept, in which the third-formant transition was the chirp and also the cue for the perceived difference between [da] and [ga] (Mann & Liberman, 1983). Putting their attention sometimes on the "speech" side and sometimes on the "chirp" side of the duplex percept, subjects discriminated various pairs of stimuli. The resulting discrimination functions were very different, though the transition cues had been presented in the same context, to the same brain, and at the same time: the function for the chirp side of the duplex percept was linear, implying a perceived continuum, while the function for the phonetic side rose to a high peak at the location of the phonetic boundary (as determined for binaurally presented syllables), implying a tendency to categorize the percepts as [da] or [ga].

These results with psychophysical measures of discriminability are of interest because they support our claim that heteromorphic perception in the phonetic mode is not a late-occurring interpretation (or match-to-prototype) of auditory percepts that were available in a common register. Apparently, heteromorphic perception goes deep.

The facts about heteromorphy reinforce the view, expressed earlier, that the underlying specialization must become distinct from the specializations of the homomorphic auditory system at a relatively peripheral stage. In this respect, speech perception in the human is like echolocation in the bat. Both are relatively late developments in the evolution of human and bat, respectively, and both apparently begin their processing independently of the final output of auditory specializations that are older.

Generative Detection

Since there are many other environmental signals in the same frequency range to which the speech-perceiving system must be sensitive, we should wonder how speech signals as a class are detected, and what keeps this system from being jammed by nonspeech signals that are physically similar. One possibility is that somewhere in the human brain there is a preliminary sorting mechanism that directs speech signals to the heteromorphic speech-perceiving system and other signals to the collection of homomorphic systems that deal with environmental sounds in general. Such a sorting mechanism would necessarily rely, not on the deep properties of the signal that are presumably used by the speech-perceiving system to determine phonetic structure, but rather on superficial properties like those that man-made speech-detection devices exploit: quasi-periodicity, characteristic spectral structure, and syllabic rhythm, for example.

The idea of a sorting mechanism is appealing because it would explain not only why the speech-perceiving system is not jammed, but, in addition, why speech is not also perceived as nonspeech--a problem to which we have already referred and to which we will return. Unfortunately, this notion is not easy to reconcile with the fact that speech is perceived as speech even when its characteristic superficial properties are masked or destroyed. Thus, speech can be high-pass filtered, low-pass filtered, infinitely clipped, spectrally inverted, or rate adjusted, and yet remain more or less intelligible. Even more remarkably, intelligible speech can be synthesized in very unnatural ways: for example, as already mentioned, with a set of frequency-modulated sinusoids whose trajectories follow those of the formants of some natural utterance. Evidently, information about all these signals reaches the speech-perceiving system and is processed by it, even though they lack some or all of the characteristic superficial properties on which the sorting mechanism we have been considering would have to depend.

The only explanation consistent with these facts is there _is_ no preliminary sorting mechanism; it is instead the speech-perceiving system itself that decides between speech and nonspeech, exploiting the phonetic properties that are intrinsic to the former and only fortuitously present in the latter. Presumably, distorted and unnatural signals like those we have referred to can be classified as speech because information about phonetic structure is spread redundantly across the speech spectrum and over time; thus, much of it is present in these signals even though the superficial acoustic marks of speech may be absent. On the other hand, isolated formant transitions, which have the appropriate acoustic marks but, out of context, no definite phonetic structure, are, as we have said, classified as nonspeech. In short, the signal is speech if and only if the pattern of articulatory gestures that must have produced it can be reconstructed. We call this property "generative detection," having in mind the analogous situation in the domain of sentence processing. There, superficial features cannot distinguish grammatical sentences from ungrammatical ones. The only way to determine the grammaticality of a sentence is to parse it--that is, to try to regenerate the syntactic structure intended by the speaker.

Is generative detection found in the specialized systems of other species? Consider, first, the moustached bat, whose echolocation system relies on biosonar signals (Suga, 1984). The bat has to be able to distinguish its own echolocation signals from the similar signals of conspecifics. Otherwise, not only would the processing of its own signals be jammed, but many of the objects it located would be illusory, because it would have subjected the conspecific signals to the same heteromorphic treatment it gives its own. According to Suga, the bat probably solves the problem in the following way. The harmonics of all the biosonar signals reach the CF-CF and FM-FM neurons that determine the delay between harmonics F2 and F3 of the emitted signals and their respective echoes. But these neurons operate only if F1 is also present. This harmonic is available to the cochlea of the emitting bat by bone conduction, but weak or absent in the radiated signal. Thus, the output of the CF-CF and FM-FM neurons reflects only the individual's own signals and not those of conspecifics. The point is that, as in the case of human speech detection, there is no preliminary sorting of the two classes of signals. Detection of the required signal is not a separate stage, but inherent in the signal analysis. However, the bat's method of signal detection cannot properly be called generative, because, unlike speech detection, it relies on a surface property of the input signal.

Generative detection is, perhaps, more likely to be found in the perception of song by birds. While, so far as we are aware, no one has suggested how song detection might work, it is known about the zebra finch that pure tones as well as actual song produce activity in the neurons of the song motor nucleus HVc (Williams, 1984; Williams & Nottebohm, 1985), a finding that argues against preliminary sorting and for detection in the course of signal analysis. Moreover, since the research just cited also provides evidence that the perception of song by the zebra finch is motoric, generative detection must be considered a possibility until and unless some superficial acoustic characteristic of a particular song is identified that would suffice to distinguish it from the songs of other avian species. Generative detection in birds seems the more likely, given that some species--the winter wren, for example--have hundreds of songs that a conspecific can apparently recognize correctly, even if it has never heard them before (Konishi, 1985). It is, therefore, tempting to speculate that the wren has a grammar that generates possible song patterns, and that the detection and parsing of conspecific songs are parts of the same perceptual process.

While generative detection may not be a very widespread property of specialized perceiving systems, what does seem to be generally true is that these systems do their own signal detection. Moreover, they do it by virtue of features that are also exploited in signal analysis, whether these features are simple, superficial characteristics of the signal, as in the case of echolocation in the bat, or complex reflections of distal events, as in the case of speech perception. This more general property might, perhaps, be added to those that Fodor (1983) has identified as common to all perceptual modules.

Preemptiveness

As we have already hinted, our proposal that there are no preliminary sorting mechanisms leads to a difficulty, for without such a mechanism, we might expect that the general-purpose, homomorphic auditory systems, being sensitive to the same dimensions of an acoustic signal as a specialized system, would also process special signals. This would mean that the bat would not only use its own biosonar signals for echolocation, but would also hear them as it presumably must hear the similar biosonar signals of other bats; the zebra finch would perceive conspecific song, not only as song, but also as an ordinary environmental sound; and human beings would hear chirps and glissandi as well as speech. We cannot be sure with nonhuman animals that such double processing of special-purpose signals does not, in fact, occur, but certainly it does not for speech, except under the extraordinary and thoroughly unecological conditions, described earlier, that induce "duplex" perception. We should suppose, however, that, except where complementary aspects of the same distal object or event are involved, as in the perception of color and shape, double processing would be maladaptive, for it would result in the perception of two distal events, one of which would be irrelevant or spurious. For example, almost any environmental sound may startle a bird, so if a conspecific song were perceived as if it were also something else, the listening bird might well be startled by it.

The general-purpose homomorphic systems themselves can have no way of defining the signals they should process in a way that excludes special signals, since the resulting set of signals would obviously not be a natural class. But suppose that the specialized systems are somehow able to preempt signal information relevant to the events that concern them, preventing it from reaching the general-purpose systems at all. The bat would then use its own biosonar signals to perceive the distal objects of its environment, but would not also hear them as it does the signals of other bats; the zebra-finch would hear song only as song; and human beings would hear speech as speech but not also as nonspeech.

An arrangement that would enable the preemptiveness of special-purpose systems is serial processing, with the specialized system preceding the general-purpose systems (Mattingly & Liberman, 1985). The specialized system would not only detect and process the signal information it requires, but would also provide an input to the general-purpose systems from which this information had been removed. In the case of the moustached bat, the mechanism proposed by Suga (1984) for the detection of the bat's own biosonar signals would also be sufficient to explain how the information in these signals, but not the similar information in conspecific signals, could be kept from the general-purpose system. Though doubtless more complicated, the arrangements in humans for isolating phonetic information and passing on nonphonetic information would have the same basic organization. We suggest that the speech-perceiving system not only recovers whatever phonetic structure it can, but also filters out those features of the signal that result from phonetic structure, passing on to the general-purpose systems all of the phonetically irrelevant residue. If the input signal includes no speech, the residue will represent all of the input. If the input signal includes speech as well as nonspeech, the residue will represent all of the input that was not speech, plus the laryngeal source signal (as modified by the effects of radiation from the head), the pattern of formant trajectories that results from the changing configuration of the vocal tract having been removed. Thus the perception, not only of nonspeech environmental sounds, but also of nonphonetic aspects of the speech signal, such as voice quality, is left to the general-purpose systems.

Serial processing appeals to us for three reasons. First, it is parsimonious. It accounts for the fact that speech is not also perceived as nonspeech, without assuming an additional mechanism and without complicating whatever account we may eventually be able to offer of speech perception itself. The same computations that are required to recover phonetic structure from the signal also suffice to remove all evidence of it from the signal information received by the general-purpose system.

Second, by placing the speech processing system ahead of the general-purpose systems, the hypothesis exploits the fact that while nonspeech signals have no specific defining properties at all, speech signals form a natural class, with specific, though deep, properties by virtue of which they can be reliably assigned to the class.

Third, serial processing permits us to understand how precedence can be guaranteed for a class of signals that has special biological significance. It is a matter of common experience that the sounds of bells, radiators, household appliances, and railroad trains can be mistaken for speech by the casual listener. On the other hand, mistaking a speech sound for an ordinary environmental sound is comparatively rare. This is just what we should expect on ethological grounds, for, as with other biologically significant signals, it is adaptive that the organism should put up with occasional false alarms rather than risk missing a genuine message. Now if speech perception were simply one more cognitive operation on auditory primitives, or if perception of nonspeech preceded it, the organism would have to learn to favor speech, and the degree of precedence would depend very much on its experience with acoustic signals generally. But if, as we suggest, speech precedes the general-purpose system, the system for perceiving speech need only be reasonably permissive as to which signals it processes completely for the precedence of speech to be insured.

Commonality Between the Specializations for Perception and Production

So far, we have been concerned primarily with speech perception, and we have argued that it is controlled by a system specialized to perceive phonetic gestures. But what of the system that controls the gestures? Is it specialized, too, and how does the answer to that question bear on the relation between perception and production?

A preliminary observation is that there is no logical necessity for speech production to be specialized merely because speech perception appears to be. Indeed, our commitment to an account of speech perception in which the invariants are motoric deprives us of an obvious argument for the specialness of production. For if the perceptual invariants were taken to be generally auditory, it would be easy to maintain that only a specialized motoric system could account for the ability of every normal human being to speak rapidly and yet to manipulate the articulators so as to produce just those acoustically invariant signals that the invariant auditory percepts would require. But if the invariants are motoric, as we claim, it could be that the articulators do not behave in speech production very differently from the way they do in their other functions. In that case, there would be nothing special about speech production, though a perceptual specialization might nevertheless have been necessary to deal with the complexity of the relation between articulatory configuration and acoustic signal. However, the perceptual system would then have been adapted very broadly to the acoustic consequences of the great variety of movements that are made in chewing, swallowing, moving food around in the mouth, whistling, licking the lips, and so on. There would have been few constraints to aid the perceptual system in recovering the gestures, and nothing to mark the result of its processing as belonging to an easily specifiable class of uniquely phonetic events. However, several facts about speech production strongly suggest that it is, instead, a specialized and highly constrained process.

It is relevant, first, that the inventory of gestures executed by a particular articulator in speech production is severely limited, both with respect to manner of articulation (i.e., the style of movement of the gesture) and place of articulation (i.e., the particular fixed surface of the vocal tract that is the apparent target of the gesture). Consider, for example, the tip of the tongue, which moves more or less independently of, but relative to, the tongue body. In nonphonetic movements of this articulator, there are wide variations in speed, style, and direction, variations that musicians, for example, learn to exploit. In speech, however, the gestures of the tongue tip, though it is, perhaps, the most phonetically versatile of the articulators, are restricted to a small number of manner categories: stops (e.g., [t] in *too*), flaps ([D] in *butter*), trills ([r] in Spanish *perro*), taps ([ɾ] in Spanish *pero*), fricatives ([θ] in *thigh*), central approximants ([ɹ] in *red*) and lateral approximants ([l] in *law*). Place of articulation for these gestures is also highly constrained, being limited to dental, alveolar, and immediately post-alveolar surfaces. (Ladefoged, 1971, Chapters 5, 6; Catford, 1977, Chapters 7-8). These restricted movements of the tongue tip in speech are not, in general, similar to those it executes in nonphonetic functions (though perhaps one could argue for a similarity between the articulation of the interdental fricative and the tongue-tip movement required to expel a grape seed from the mouth. But, as Sapir (1925, p. 34) observed about the similarity between an aspirated [w] and the blowing-out of a candle, these are "norms or types of entirely distinct series of variants"). Speech movements are, for the most part, peculiar to speech; they have no obvious nonspeech functions.

The peculiarity of phonetic gestures is further demonstrated in consequences of the fact that, in most cases, a gesture involves more than one articulator. Thus, the gestures we have just described, though nominally attributed to the tongue tip, actually also require the cooperation of the tongue body and the jaw to insure that the tip will be within easy striking distance of its target surface (Lindblom, 1983). The requirement arises because, owing to other demands on the tongue body and jaw, the tongue tip cannot be assumed to occupy a particular absolute rest position at the time a gesture is initiated. Cooperation between the articulators is also required, of course, in such nonphonetic gestures as swallowing, but the particular cooperative patterns of movement observed in speech are apparently unique, even though there may be nonspeech analogues for one or another of the components of such a pattern.

Observations analogous to these just made about the tongue tip could be made with respect to each of the other major articulators: the tongue body, the lips, the velum, and the larynx. That the phonetic gestures possible for each of these articulators form a very limited set that is drawn upon by all languages in the world has often been taken as evidence for a universal phonetics (e.g., Chomsky & Halle, 1968, pp. 4-6). (Indeed, if the gestures were not thus limited, a general notation for phonetic transcription would hardly be possible.) That the gestures are eccentric when considered in comparison with what the articulators are generally capable of--a fact less often remarked--is evidence that speech production does not merely exploit general tendencies for articulator movement, but depends rather on a system of controls specialized for language.

A further indication of the specialness of speech production is that certain of the limited and eccentric set of gestures executed by the tongue tip are paralleled by gestures executed by other major articulators. Thus, stops and fricatives can be produced not only by the tongue tip but also by the tongue blade, the tongue body, the lips, and the larynx, even though these various articulators are anatomically and physiologically very different from one another. Nor, to forestall an obvious objection, are these manner categories mere artifacts of the phonetician's taxonomy. They are truly natural classes that play a central role in the phonologies of the world's languages. If these categories were unreal, we should not find that in language x vowels always lengthen before all fricatives, that in language y all stops are regularly deleted after fricatives, or that in all languages the constraints on the sequences of sounds in a syllable are most readily described according to manner of articulation (Jespersen, 1920, pp. 190 ff.). And when the sound system of a language changes, the change is frequently a matter of systematically replacing sounds of one manner class by sounds of another manner class produced by the same articulators. Thus, the Indo-European stops [p],[t],[k],[q] were replaced in Primitive Germanic by the corresponding fricatives [f],[θ],[x],[X], ("Grimm's law").

Our final argument for the specialness of speech production depends on the fact of gestural overlap. Thus, in the syllable [du], the tongue-tip closure gesture for [d] overlaps the lip-rounding and tongue-body-backing gestures for [u]. Even more remarkably, two gestures made by the same articulator may overlap. Thus, in the syllable [gi], the tongue-body-closure gesture for [g] overlaps the tongue-body-fronting gesture for [i], so that the [g] closure occurs at a more forward point on the palate than would be the case for [g] in [gu]. As we have already suggested, it is gestural overlap, making possible relatively high rates of information transmission, that gives speech its adaptive value as a communication system. But if the strategy of overlapping gestures to gain speed is not to defeat itself, the gestures can hardly be allowed to overlap haphazardly. If there were no constraints on how the overlap could occur, the acoustic consequences of one gesture could mask the consequences of another. In a word such as twin, for instance, the silence resulting from the closure for the stop [t] could obscure the sound of the approximant [w]. Such accidents do not ordinarily occur in speech, because the gestures are apparently phased so to provide the maximum amount of overlap consistent with preservation of the acoustic information that specifies either of the gestures (Mattingly, 1981). This phasing is most strictly controlled at the beginnings and ends of syllables, where gestural overlap is greatest, and most variable in the center of the syllable, where less is going on (Tuller & Kelso, 1984). Thus, to borrow Fujimura's (1981) metaphor, the gestural timing patterns of consonants and consonant clusters are icebergs floating on a vocalic sea. Like the individual gestures themselves, these complex temporal patterns are peculiar to speech and could serve no other ecological purpose.

We would conclude, then, that speech production is specialized, just as speech perception is. But if this is so, we would argue, further, that these two processes are not two systems, but rather, modes of one and the same system. The premise of our argument is that because speech has a communicative function, what counts as phonetic structure for production must be the same as what counts as phonetic structure for perception. This truism holds regardless of what one takes phonetic structure to be, and any account of phonetic process has to be consistent with it. Thus, on the conventional account, it must be assumed that perception and production, being taken as distinct processes, are both guided by some cognitive representation of the structures that they deal with in common. On our account, however, no such cognitive representation can be assumed if the notion of a specialized system is not to be utterly trivialized. But if we are to do without cognitive mediation, what is to guarantee that at every stage of ontogenetic (and for that matter phylogenetic) development, the two systems will have identical definitions of phonetic structure? The only possibility is that they are directly linked. This, however, is tantamount to saying that they constitute a single system, in which we would expect representations and computational machinery not to be duplicated, but rather to coincide insofar as the asymmetry of the two modes permits.

To make this view more concrete, suppose, as we have elsewhere suggested (Liberman & Mattingly, 1985; Liberman, Mattingly, & Turvey, 1972; Mattingly & Liberman, 1969), that the speech production/perception system is, in effect, an articulatory synthesizer. In the production mode, the input to the synthesizer is some particular, abstractly specified gestural pattern, from which the synthesizer computes a representation of the contextually varying articulatory movements that will be required to realize the gestures, and then, from this articulatory representation, the muscle commands that will execute the actual movements, some form of "analysis by synthesis" being obviously required. In the perceptual mode, the input is the acoustic signal, from which the synthesizer computes--again by analysis by synthesis--the articulatory movements that could have produced the signal, and then, from this articulatory representation, the intended gestural pattern. The computation of the muscle commands from articulatory movement is peculiar to production, and the computation of articulatory movement from the signal is peculiar to perception. What is common to the two modes, and carried out by the same computations, is the working out of the relation between abstract gestural pattern and the corresponding articulatory movements.

We earlier alluded to a commonality between modes of another sort when we referred to the finding that the barn owl's auditory orientation processes use the same neural map as its visual orientation processes do. Now we would remark the further finding that this arrangement is quite one-sided: the neural map is laid out optically, so that sounds from sources in the center of the owl's visual field are more precisely located and more extensively represented on the map than are sounds from sources at the edges (Knudsen, 1984). This is of special relevance to our concerns, because, as we have several times implied, a similar one-sidedness seems to characterize the speech specialization: its communal arrangements are organized primarily with reference to the processes of production. We assume the dominance of production over perception because it was the ability of appropriately coordinated gestures to convey phonetic structures efficiently that determined their use as the invariant elements of speech. Thus, it must have been the gestures, and especially the processes associated with their expression, that shaped the development of a system specialized to perceive them.

More comparable, perhaps, to the commonality we see in the speech specialization are examples of commonality between perception and production in animal communication systems. Evidence for such commonality has been found for the tree frog (Gerhardt, 1978); the cricket (Hoy, Hahn, & Paul, 1977; Hoy & Paul, 1973); the zebra finch (Williams, 1984; Williams & Nottebohm, 1985); the white-crowned sparrow (Margoliash, 1983) and the canary (McCasland &

Konishi, 1983). Even if there were no such evidence, however, few students of animal communication would regard as sufficiently parsimonious the only alternative to commonality: that perception and production are mediated by cognitive representations. But if we reject this alternative in explaining the natural modes of nonhuman communication, it behooves us to be equally conservative in our attempt to explain language, the natural mode of communication in human beings. Just because language is central to so much that is uniquely human, we should not therefore assume that its underlying processes are necessarily cognitive.

The Speech Specialization and the Sentence

As a coda, we here consider, though only briefly, how our observations about perception of phonetic structure might bear, more broadly, on perception of sentences. Recalling, first, the conventional view of speech perception--that it is accomplished by processes of a generally auditory sort--we find its extension to sentence perception in the assumption that coping with syntax depends on a general faculty, too. Of course, this faculty is taken to be cognitive, not auditory, but, like the auditory faculty, it is supposed to be broader than the behavior it serves. Thus, it presumably underlies not just syntax, but all the apparently smart things people do. For an empiricist, this general faculty is a powerful ability to learn, and so to discover the syntax by induction. For a nativist, it is an intelligence that knows what to look for because syntax is a reflection of how the mind works. For both, perceiving syntax has nothing in common with perception of speech, or, a fortiori, with perception of other sounds, whether biologically significant or not. It is as if language, in its development, had simply appropriated auditory and cognitive processes that are themselves quite independent of language and, indeed, of each other.

The parallel in syntax to our view of speech is the assumption that sentence structures, no less than speech, are dealt with by processes narrowly specialized for the purpose. On this assumption, syntactic and phonetic specializations are related to each other as two components of the larger specialization for language. We should suppose, then, that the syntactic specialization might have important properties in common, not only with the phonetic specialization, but also with the specializations for biologically significant sounds that occupy the members of this symposium.

References

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. Perception & Psychophysics, 29, 191-211.

Catford, J. C. (1977) Fundamental problems in phonetics. Bloomington: Indiana University Press.

Chomsky, N., & Halle, M. (1968). The sound pattern of English. New York: Harper and Row.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. L. (1980). Perceptual equivalence of two acoustic cues for stop consonant manner. Perception & Psychophysics, 27, 343-350.

Fodor, J. (1983). The modularity of mind. Cambridge, MA: MIT Press.

Fujimura, O. (1981). Temporal organization of speech as a multi-dimensional structure. Phonetica, 38, 66-83.

Gerhardt, H. C. (1978). Temperature coupling in the vocal communication system of the gray tree frog hyla versicolor. Science, 199, 992-994.
Green, K. P., & Miller, J. L. (in press). On the role of visual rate information in phonetic perception. Perception & Psychophysics.
Hafter, E. R. (1984). Spatial hearing and the duplex theory: How viable is the model? In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), Dynamic aspects of neocortical function. New York: Wiley.
Hoy, R., Hahn, J., & Paul, R. C. (1977). Hybrid cricket auditory behavior: Evidence for genetic coupling in animal communication. Science, 195, 82-83.
Hoy, R., & Paul, R. C. (1973). Genetic control of song specificity in crickets. Science, 180, 82-83.
Jespersen, O. (1920). Lehrbuch der Phonetik. Leipzig: Teubner.
Knudsen, E. I. (1984). Synthesis of a neural map of auditory space in the owl. In G. M. Edelman, W. E. Gall, & W. M. Cowan, (Eds.), Dynamic aspects of neocortical function. New York: Wiley.
Knudsen, E. I., & Konishi, M. (1978). A neural map of auditory space in the owl. Science, 200, 795-797.
Konishi, M. (1985). Birdsong: From behavior to neuron. Annul Review of Neuroscience, 8, 125-170.
Ladefoged, P. (1971). Preliminaries to linguistic phonetics. Chicago: University of Chicago Press.
Liberman, A. M. (1979). Duplex perception and integration of cues: Evidence that speech is different from nonspeech and similar to language. In E. Fischer-Jorgensen, J. Rischel, & N. Thorsen (Eds.), Proceedings of the IXth International Congress of Phonetic Sciences. Copenhagen: University of Copenhagen.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.
Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21, 1-36.
Liberman, A. M., Mattingly, I. G., & Turvey, M. (1972). Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), Coding processes in human memory. Washington, DC: Winston.
Lindblom, B. (1983). Economy of speech gestures. In P. MacNeilage (Ed.), The production of speech. New York: Springer.
Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. Cognition, 14, 211-235.
Margoliash, D. (1983). Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. Journal of Neuroscience, 3, 1039-1057.
Mattingly, I. G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, & J. Anderson (Eds.), The cognitive representation of speech. Amsterdam: North-Holland.
Mattingly, I. G., & Liberman, A. M. (1969). The speech code and the physiology of language. In K. N. Leibovic (Ed.), Information processing in the nervous system. New York: Springer.
Mattingly, I. G., & Liberman, A. M. (1985). Verticality unparalleled. The Behavioral and Brain Sciences, 8, 24-26.
McCasland, J. S., & Konishi, M. (1983). Interaction between auditory and motor activities in an avian song control nucleus. Proceedings of the National Academy of Science, 78, 7815-7819.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.

Rand, T. C. (1974). Dichotic release from masking for speech. Journal of the Acoustical Society of America, 55, 678-680.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. Science, 212, 947-950.

Repp, B. H., Milburn, C., & Ashkenas, J. (1983). Duplex perception: Confirmation of fusion. Perception & Psychophysics, 33, 333-337.

Sapir, E. (1925). Sound patterns in language. Language, 1, 37-51. Reprinted in D. G. Mandelbaum (Ed.), Selected writings of Edward Sapir in language, culture and personality. Berkeley: University of California Press.

Suga, N. (1984). The extent to which bisonar information is represented in the bat auditory cortex. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), Dynamic aspects of neocortical function. New York: Wiley.

Summerfield, Q. (1979). Use of visual information for phonetic perception. Phonetica, 36, 314-331.

Tuller, B., & Kelso, J. A. S. (1984). The relative timing of articulatory gestures: Evidence for relational invariants. Journal of the Acoustical Society of America, 76, 1030-1036.

Williams, H. (1984). A motor theory of bird song perception. Unpublished doctoral dissertation, Rockefeller University.

Williams, H., & Nottebohm, F. N. (1985). Auditory responses in avian vocal motor neurons: a motor theory for song perception in birds. Science, 229, 279-282.

Yin, T. C. T., & Kuwada, S. (1984) Neuronal mechanisms and binaural interaction. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), Dynamic aspects of neocortical function. New York: Wiley.

Footnotes

[1]For full accounts of these experiments and many others that support the claims we will be making below, see Liberman, Cooper, Shankweiler and Studdert-Kennedy (1967), Liberman and Mattingly (1985), and the studies referred to therein.

[2]Not surprisingly, there are a number of variations on the "conventional view"; they are discussed in Liberman and Mattingly (1985).

[3]Our notion of heteromorphy as a property of one kind of perceiving specialization seems consistent with comments about sound localization by Knudsen and Konishi (1978, p. 797), who have observed that "[the barn owl's] map of auditory space is an emergent property of higher-order neurons, distinguishing it from all other sensory maps that are direct projections of the sensory surface...these space-related response properties and functional organization must be specifically generated through neuronal integration in the central nervous system..." Much the same point has been made by Yin and Kuwada (1984, p. 264), who say that "the cochlea is designed for frequency analysis and cannot encode the location of sound sources. Thus, the code for location of an auditory stimulus is not given by a 'labeled line' from the receptors, but must be the result of neural interactions within the central auditory system."

"VOICING" IN ENGLISH: A CATALOG OF ACOUSTIC FEATURES SIGNALING /b/ VERSUS /p/ IN TROCHEES

Leigh Lisker†

Abstract. The English category sets /b,d,g/ and /p,t,k/ are now usually referred to as voiced and voiceless stops, respectively, although it is recognized that membership in these sets is not entirely determined by whether, according to commonly accepted definitions, a given phonetic element is voiced or voiceless; nor need it even be described as a stop. What is true is that if a phonetic element is phonetically a voiced stop, then it will be assigned to the /b,d,g/ set, and if it is a voiceless stop, it may, but need not be, assigned to /p,t,k/. A context in which the stop members of the two phonological sets may be distinguished simply on the basis of voicing (as narrowly defined with respect to stop consonants) is between vowels, as for example in the pair *rabid-rapid*. Acoustically, however, as many as sixteen pattern properties can be counted that may play a role in determining whether a listener reports hearing one of these words rather than the other. In purely acoustic terms these properties are rather disparate, although most of them show variations that can plausibly be considered to be primarily the diverse effects of a relatively simple difference in the management of the larynx together with the closing and opening of the mouth. This diversity makes it difficult to rationalize a purely acoustic account of the *rabid-rapid* opposition, that is, one that makes no reference to the articulatory mechanisms and maneuvers by which the common linguistic effect of varying these acoustic properties might be explained.

Introduction

If the topic of voicing as a distinctive attribute of speech sounds continues to be a subject of lively interest to students of speech communication, it must be because it continues to provoke new questions or to refuse final answers to old ones. From a strictly phonetic viewpoint it is unclear why the subject of stop voicing should not be considered closed. The acoustic and articulatory bases of the voiced-voiceless difference are fairly well understood, though it is of course true that details of the aerodynamic, physiological and other aspects of the picture always remain to be clarified. A specified interval of speech signal is readily described as voiced or

*Language and Speech, (1986), in press.
†Also University of Pennsylvania.
Acknowledgment. Preparation of this paper was supported by NICHD Grant HD-01994 to Haskins Laboratories. I want also to express thanks to Arthur Abramson and Catherine Browman for helpful criticisms of an earlier draft.

voiceless on the basis of whether or not it exhibits harmonic patterning that can be attributed to vocal fold vibration. In addition, it is generally agreed that a given phonetic unit is voiced or voiceless depending on whether or not an interval of speech signal with which it is equated is in fact voiced. This raises the question of selecting the interval over which presence or absence of voicing shall determine whether the phonetic unit is described as voiced or voiceless. For stop consonants the diagnostic interval that linguists usually choose (e.g., the International Phonetic Association) coincides with the interval of articulatory closure. A stop is then "voiced" if the closure is marked by laryngeal buzz, and it is "voiceless" if that interval is devoid of such signal. Aside from the facts that a closure interval may be neither entirely buzzed nor entirely silent, and that auditory judgment and acoustic record may not always agree, it is otherwise not immediately obvious why the subject of stop voicing still draws the amount of attention devoted to it in recent years.

English /b/-/p/ ≠ [b]-[p]

Given the spelling conventions and the definition of stop voicing to which linguists appear generally to subscribe, a phonetic unit represented as [b] is a voiced stop, while [p] stands for its voiceless counterpart. Many languages make contrastive use of stop categories that consistently differ in voicing, for example, Dutch, Italian, and Hungarian. In these languages phonological sets represented as /b/ and /p/ are regularly [b] and [p],that is, they are characterized by voiced and voiceless stoppages of airflow through the vocal tract. But in certain languages, among them English, there are phonological categories, also represented as /b/ and /p/, whose relation to the phonetic categories [b] and [p] is not so straightforward. Since many linguists have long recognized that members of the English "voiced" set are not invariably voiced, that is, /b/ may be initially [p] (though more often it is spelled "phonetically [b̥], with no clear indication of whether the preference for the latter spelling is dictated by phonetic or phonological considerations), and prepausally as well its "voicelessness is marked," that is, readily detected by ear (Trager & Smith, 1951), it follows that the search for the acoustic properties cueing the /b/-/p/ contrast in English is not necessarily a search for cues to the phonetic feature of stop voicing. What is called the subject of stop voicing in English continues to hold the attention of speech researchers not because of the problematical nature of the acoustic correlates of a [±voiced] difference, but because the phonological analysis of English yields /p/ and /b/ categories that are phonetically variable in nature and cued by different acoustic properties in different contexts. The "problem" of English stop voicing resides largely in the fact that the observed variability of the /b/-/p/ distinction runs counter to our reasonable expectation that all phonetic elements similarly designated should have some acoustic properties in common.

Medial /b/-/p/ ≈ [b]-[p]

A context in which the contrast between stop members of English /b/ and /p/ seems most nearly to be one involving [b] vs. [p] is medially in words before an unstressed syllable, particularly where the signal preceding and following the closure is voiced. In this context, then, the acoustic features that distinguish the two stop categories can perhaps be said to serve as cues to the phonetic feature of voicing. This is to said that if phoneticians generally agree that, for example, rabid and rapid differ in stop voicing alone, then the acoustic properties affecting their identification by listeners can be called cues to stop voicing. As it happens, of the two other features that have traditionally figured in accounts of the English stops, [±aspirated] and [±fortis], there is general agreement that the first of these plays no significant role in differentiating rabid and rapid, at least in American if not in standard southern British English (Bronstein, 1960; Jones, 1956; Trager & Smith, 1951). As for the second, aside from its controversial nature as a phonetic feature on a par with the others (Lisker, 1963), it appears that linguists are not fully agreed that it applies. Thus, for Trager and Smith (1951) the [p] of rapid is fortis, while Heffner (1950) follows Jespersen in describing the American pronunciation of /p/ in words like rapid as lenis. Of course, if the durational differences in closure and preclosure intervals between rabid and rapid are construed as evidence of a [±fortis] distinction, then it must be granted that not all the acoustic cues to the lexical distinction can be, strictly speaking, cues to [±voiced]. Despite these strictures, I find it reasonable to believe that the phonetic basis for the rabid-rapid distinction is as close to being just a matter of closure voicing as can be found in the language.

Counting the Acoustic Feature Differences

Oddly enough, although in medial position the phonetic difference between English /b/ and /p/ may well be smaller than elsewhere, the number of readily isolated acoustic pattern properties whose variation might be expected to affect the identification of a stimulus as rabid or rapid is larger. (It far exceeds the six listed in Klatt, 1975, for word-initial but utterance-medial intervocalic position, and is in fact more, by two, than the fourteen listed by Edwards, 1981, for the same position, where the phonetic basis for the "voicing distinction" is possibly maximal.) However, this fact is remarkable only if we suppose that the number of phonetic features that differentiate the contrasting sets should directly determine the number of properties that we can isolate and manipulate to linguistic effect. Otherwise it is not so very surprising, since utterance-initial stops cannot be cued by properties of the interval preceding closure (except for the pre-speech silence), nor are they in English regularly cued by any property of the closure interval itself. Of some sixteen acoustic properties that cue, or can plausibly be supposed to cue the identification of a form as rabid or rapid, seven are to be found in the signal preceding the medial closure, three are closure properties, and the remainder are post-closure.

They are the following:

Closure

1) duration of closure
2) duration of glottal signal
3) intensity of glottal signal

Pre-closure

4) duration of vowel
5) duration of first-formant (F_1) transition
6) F_1 offset frequency
7) F_1 transition offset time (i.e., "F_1 cutback," or, more precisely, "F_1 cut forward")

8) timing of voice offset
9) fundamental frequency (F_0) contour
10) delay time of signal.

Post-closure

11) release burst intensity
12) timing of voice onset (VOT)
13) onset of F_1 transition ("F_1 "cutback")
14) F_1 onset frequency
15) F_1 transition duration
16) F_0 contour

This list does not fully exhaust the inventory of properties that possibly affect listeners' labeling behavior, for we might imagine that factors contributing to the "prominence" of the second syllable relative to the first (i.e., the stress contour attributed to the form) could have secondary effects on word identification. A pattern labeled rapid might, as a result of acoustic alterations effecting a stress shift, be perceived to include /b/ rather than /p/, since a natural token of the derivative of the first word, rapidity, calls for the voiceless aspirate [p^h], whereas a [p] would not be incompatible with an interpretation of the pattern as the word rabidity. Nor can we in principle exclude the possibility that still other isolable acoustic properties, for example, higher formants, may make contributions to lexical identity, even though such effects might not be readily explained (Lisker, 1975).

In the inventory just listed sixteen acoustic properties were enumerated, and several more suggested, but the precise number cannot be taken very seriously, since with respect to some of them it is difficult to decide whether we have one property or more. And while we may decide that we have more than one, at least for purposes of experimentation, they may not be acoustically distinct, to say nothing of whether or not they are subject to independent control by the operator of the human vocal tract. Thus, for example, property #12 might be analyzed as two properties, voice-onset time and aspiration (following Klatt, 1975), since a delay in voice onset can be accompanied by a silent interval (per ejective articulation) or by aspiration. On the other hand, items #2 and #8 are counted as two rather than one, not on an acoustic basis, but only because of a prior segmentation of the speech patterns whereby the test stimuli were partitioned into pre-closure, closure and post-closure intervals. (A similar segmentation underlies the common distinction drawn between the phonetic features of stop voicing and voiceless aspiration in English and some other languages, and their subsequent treatment as independent properties of stop consonants.)

Acoustic Properties as Context-variable Lexical Cues

Of the above-listed acoustic properties that might affect the identification of a signal as rabid or rapid, it is probably true that none is indispensible, while it is possible that several play no significant role in the perception of unedited naturally produced tokens of these words. Thus, a reported rabid need not mean that the medial closure was voiced (Lisker, 1957), while a long closure duration does not invariably elicit a rapid labeling response (Lisker, 1981). At present we may only say that some of the properties demonstrably affect word perception under certain conditions, and that the rest of them are "candidate cues," inasmuch as none has so far been shown to make no contribution to the perception of medial /b/ vs. /p/. To be sure, it cannot in principle be proven that any conceivable acoustic property of a speech or speechlike signal is incapable of affecting the perception of an acoustic signal as a particular linguistic message; on the other hand, we have no right to assume a principle of "once a cue, always a cue." Thus, for example, the linguistic irrelevance of the [±voiced] difference in the case of initial /b/ does not mean that the identification of a stop as /b/ is everywhere unaffected by whether its closure is voiced or voiceless (although it is just this non sequitur that underlies the assertion by Jakobson and Halle, 1956, that the "distinctive feature" distinguishing the category sets /b,d,g/ and /p,t,k/ is one of articulatory force and not voicing). The aim of most research into the processes of speech perception has been to uncover all the acoustic properties that can somewhere serve as cues, and not so much to specify the conditions under which any one of them does and does not serve that function, or to assess the likelihood that the conditions under which it is a cue inside the laboratory are met outside it.

A reading of the phonetic literature suggests that the conditions an acoustic property must satisfy in order to qualify as a "cue" do not involve a demonstrable conformity with nature--it is enough that patterns be devised so that manipulating the property effects a significant shift in listeners' word identification, for example, from rabid to rapid. There is no absolute requirement, it would seem, that either the constant properties of the test stimuli or the range of values assigned the variables be copied from nature. Thus property #1 listed above, the duration of the formantless interval corresponding to oral closure, serves as a cue to the rabid-rapid contrast only in the absence of glottal signal over most of that interval, and it may be decisive only when varied over values that exceed the range observed in nature. When glottal signal persists over much of the closure interval, varying the closure duration will have no effect on listeners' word-labeling behavior; at most some tokens of the reported rabid may strike the listeners as having an abnormally long /b/. Nor is the absence of closure voicing enough to ensure that varying closure duration will affect word identification. For example, a pattern synthesized with very low values of F_1 offset and onset frequencies (properties #6 and #14) is likely to be reported as rabid no matter how long the closure (cf. Repp, 1978). Under some conditions, then, closure duration operates as a cue to stop "voicing," that is, lexical identity; otherwise it is a temporal property that is evaluated temporally. It is likely that every other one of the sixteen properties enumerated is also of restricted usefulness as a cue to the listener in deciding on the interpretation of a signal as one or the other word.

Acoustic Properties as [±Voiced] Cues

The acoustic properties that can serve as cues to listeners in deciding whether an auditory stimulus is an instance of rabid or rapid are said to be cues to the /b/-/p/ contrast in medial position because the accepted phonological representations of these forms, /'ræbɪd and /'ræpɪd/, appear to attribute their phonetic distinctiveness to a phonological contrast between the medial stops. Does it follow, then, that they are cues to the voiced-voiceless distinction as precisely defined? A reasonable answer would be that, *if* the lexical distinction is equivalent perceptually to a /b/-/p/ difference and that in turn is a matter of closure voicing, then the acoustic cues are cues to the [±voiced] feature. However, it is by no means generally agreed that phonological representations do in themselves amount to claims about the perceptual nature of a phonetic distinction, and it can be argued that the phonetic spellings ['ɹæ:bɨd] and ['ɹæpɨd] more directly reflect linguists' judgments about its perceptual basis, that is, that the lexical decision is based on a combined difference of vowel duration and stop closure signal. Moreover, even if we choose to view the rabid-rapid distinction as equivalent to a difference in their stop consonants, it can be argued that in order to be counted as a cue to the voicing status of the stop, it is not enough that a given acoustic property should significantly determine a listener's lexical decision; it must affect a decision as to whether or not the medial closure was or was not accompanied by laryngeal buzz, that is, voicing. Thus, for example, a variation in the duration of the [æ] might possibly affect the lexical decision and thus which stop was reported, but need not determine the answer to a question about stop voicing, which involves a judgment that is both auditory and phonetic. It seems quite possible that, of the sixteen or more acoustic properties that may help determine the lexical decision, only the three closure characteristics are directly cues to the perceived voicing state of the closure, while the others are cues to that state only in a derivative sense. If the properties of the pre-closure interval are set at values compatible with a [±voiced] closure, they may induce listeners to report hearing a /b/, that is, the word rabid, but it cannot be presumed that they will also lead them to report hearing a voiced closure. They might indeed more consistently report that a stimulus pair, one labeled rabid and the other rapid, differ in their [æ] durations than in the [±voiced] nature of their medial stop closures. In such a case it would hardly seem appropriate to call the duration of the vowel a cue to the voicing of the stop. (The situation would be analogous to the celebrated cases of rider-writer and ladder-latter in varieties of American English).

Acoustic Cues → Articulatory Gesture

If many of the acoustic properties listed above can be considered the consequences of a laryngeal gesture (Abramson,1977; Goldstein & Browman, in preparation; Lisker & Abramson, 1971) executed in conjunction with labial closure and opening, we may reasonably decide that a speech signal is more simply described as an ensemble of articulatory rather than acoustic events. A signal identified as rapid, which can differ acoustically in many ways from one heard as rabid, may be said to differ essentially from the latter in that vocal fold vibration is halted for much of the interval of labial closure. Thus, for example, the fact that two pairs of acoustic patterns, one differing only in closure duration and the other only in release burst intensity, are both interpreted as rabid vs. rapid, may be explained by the claim that both differences are consequences of a single difference in laryngeal activity. This many-one relation of the acoustic and articulatory differences between rabid and rapid can be understood to support the view that speech is better described in articulatory than in acoustic terms, that is, that the "sounds of speech" as represented in a linguist's phonetic and phonological spellings are connected more directly with articulatory gestures and states than with acoustic properties. This is not to say that charting the connections between articulation and the phonetic features of speech is a trivial matter, only that it is easier than establishing those that relate the latter to the acoustic signal. Interesting evidence recently reported by Flege (1982) shows that the two kinds of English /b/ found initially are frequently produced with glottal closing gestures having the same temporal relation to the supraglottal articulation, and thus there is an articulatory invariant underlying the allophonic [±voiced] difference.

Articulatory Gestures → Acoustic Cue

Even if it is accepted that speech perception is special in that it involves awareness, not of the acoustic properties, but rather the articulatory gestures that the listener infers from them (Liberman & Mattingly, 1985), it does not follow that phonetic explanation never goes the other way, that is, that it never seeks to explain articulatory diversity by pointing to a single acoustic consequence. The matter of consonantal voicing provides what appears to be a compelling case, where articulatory gestures of various kinds have been explained as maneuvers all "designed" to produce either voiced or voiceless closures. Thus the longer [æ], as well as the lowered larynx, the raised velum, and the generally "laxed" articulation associated with /b/ as against /p/, have all been considered to facilitate the acoustic feature of voicing during closure (Bell-Berti, 1975; Halle & Stevens, 1967; Kent & Moll, 1969; Riordan, 1980; Westbury, 1983). Moreover, it does not appear that there is a single laryngeal devoicing gesture for /p/. since the same acoustically silent closure is produced either by abducting the vocal folds or by halting their vibratory movement without very much glottal opening, and indeed, in British English, with glottalization or "glottal reinforcement" (Roach, 1983). In the cases of both the voiced and the voiceless closures, then, it might be argued that the articulatory gestures are many, the "intended" acoustic outcome one.

Summary

The number of acoustic properties that can be manipulated so as to affect listeners' decision in judging an auditory stimulus as an instance of the English words rabid or rapid is considerably greater than the number of phonetic features customarily enumerated as the basis on which they are distinguished. At least sixteen, and quite possibly more, may serve as cues to the lexical distinction. Insofar as the phonetic feature held to be chiefly responsible for the auditory distinctiveness of the two forms is a simple difference in the nature of the signal emitted during the interval of oral closure, to that extent can the acoustic properties that serve as lexical cues be said to be cues to the contrast between the phonetic categories [b] and [p] and hence, by definition, as cues to the [±voiced] difference. It is reasonable to regard the lexical decision as being equivalent to deciding whether a /b/ or a /p/ was present in the signal, but it is by no means clear whether the lexical decision as between rabid and rapid is the same as a decision about the acoustic nature of the signal emitted during closure. We may adopt the hypothesis that most of the acoustic properties whose variation affects the rabid-rapid decision are the consequences of articulatory maneuvers "designed" either to inhibit or not to inhibit production of voice during the closure interval. If we confine our attention to the larynx as the articulator chiefly responsible for the [±voiced] difference, then those articulatory maneuvers are possibly fewer and more simply described than are the acoustic properties they generate. But if, on the other hand, the nature of the signal emitted during the closure is a major acoustic cue to the lexical distinction (and this seems quite likely so far as naturally produced speech is concerned), and if all the articulatory maneuvers said to be associated with voiced versus voiceless stops can be seen as factors determining the [±voiced] feature (i.e., adjustments of glottal area, vocal fold stiffness, larynx height, velar height, and cavity wall tensity), then surely it is the articulatory picture whose relative complexity is to be explained by the acoustic reference. Thus, at least with respect to stop

voicing, it does not seem possible to give an adequate account of all the phonetic facts by deciding that, as between its articulatory and acoustic aspects, we can choose one to the exclusion of the other. A purely articulatory account, and a purely acoustical one as well, may appear to gain the simplicity that passes for explanation, but it is at the expense of adequacy.

References

Abramson, A. S. (1977). Laryngeal timing in consonant distinctions. *Phonetica*, *34*, 295-303.

Bell-Berti, F. (1975). Control of pharyngeal cavity size for English voiced and voiceless stops. *Journal of the Acoustical Society of America*, *57*, 456-461.

Bronstein, A. J. (1960). *The pronunciation of American English*. New York: Appleton-Century-Crofts.

Edwards, T. J. (1981). Multiple features analysis of intervocalic English plosives. *Journal of the Acoustical Society of America*, *69*, 535-547.

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics*, *10*, 177-192.

Goldstein, L., & Browman, C. P. (in preparation). Representation of voicing constrasts using articulatory gestures.

Halle, M., & Stevens, K. N. (1967). On the mechanism of glottal vibration for vowels and consonants. *Quarterly Progress Report of the Research Laboratory of Elecronics, M.I.T.*, *85*, 267-270.

Heffner, R-M. S. (1950). *General phonetics*. Madison: University of Wisconsin Press.

Jakobson, R., & Halle, M. (1956). *Fundamentals of language*. The Hague: Mouton.

Jones, D. (1956). *An outline of English phonetics*. (8th ed.). Cambridge: Cambridge University Press.

Kent, R. D. & Moll, K. L. (1969). Vocal-tract characteristics of the stop cognates. *Journal of the Acoustical Society of America*, *46*, 1549-1555.

Klatt, D. H. (1975). Voice-onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, *18*, 687-703.

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1-36.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, *33*, 42-49.

Lisker, L. (1963). On Hultzn's "voiceless lenis stops in prevocalic clusters." *Word*, *19*, 376-387.

Lisker, L. (1975). Is it VOT or a first-formant transition detector. *Journal of the Acoustical Society of America*, *57*, 1547-1551.

Lisker, L. (1981) On generalizing the *rabid-rapid* distinction based on silent gap duration. *Haskins Laboratories Status Report on Speech Research*, *SR-65*, 251-259.

Lisker, L. & Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language*, *1971*, 767-785.

Repp, B. H. (1978). Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*, *24*, 471-485.

Riordan, C. J. (1980). Larynx height during English stop consonants. *Journal of Phonetics*, *8*, 353-360.

Roach, P. (1983). _English phonetics and phonology_. Cambridge: Cambridge University Press.

International Phonetic Association (1949). _The principles of the International Phonetic Association_. London: University College.

Trager, G. L. & Smith, H. L. Jr. (1951). _An outline of English structure_. (Studies in Linguistics: Occasional Papers, 3) Norman, OK: Battenburg Press.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. _Journal of the Acoustical Society of America_, _73_, 1322-1336.

CATEGORICAL TENDENCIES IN IMITATING SELF-PRODUCED ISOLATED VOWELS*

Bruno H. Repp and David R. Williams†

Abstract. An earlier experiment requiring literal imitation of synthetic isolated vowels from [u]-[i] and [i]-[æ] continua (Repp & Williams, 1985) was replicated using as stimuli vowels produced by the subjects themselves. Even though imitation accuracy was much improved, the responses deviated from the stimuli in ways similar to those observed previously with synthetic stimuli. That is, categorical tendencies (nonlinear stimulus-response mappings of formant frequencies, nonuniform response variability across each continuum, and peaks in formant frequency distributions) were obtained even with stimuli that matched the subjects' articulatory capabilities. This rules out one possible explanation of the observed categorical tendencies, viz., that they arise in the perceptual translation of synthetic stimuli into a talker's production space.

Introduction

In a recent study (Repp & Williams, 1985), we investigated the claim that subjects' vocal imitations of isolated, steady-state vowels follow a categorical pattern (Chistovich, Fant, de Serpa-Leitao, & Tjernlund, 1966; Kent, 1975). Two subjects (the authors) imitated synthetic vowels from 12-member [u]-[i] and [i]-[æ] continua at three different temporal delays, which had little effect on response patterns. The functions relating stimulus and (average) response formant frequencies across each vowel continuum exhibited local changes in slope, response standard deviations varied, and the distributions of response formant frequencies showed distinct peaks and valleys. The response patterns thus showed categorical tendencies, but few instances of strictly categorical responses (i.e., identical responses to different stimuli representing the same vowel category).

Where do these categorical tendencies in imitation come from? There are at least four independent (but not mutually exclusive) possibilities, some perhaps more plausible than others. The tendencies could originate either in the subjects' perception of the stimulus vowels or in their production of the imitations. On the perceptual side, there are two possibilities: (1) Perceptual nonlinearities might arise when the stimuli are synthetic and/or not well matched to the subject's production capabilities. An additional

---

*Speech Communication, in press

†Also at Department of Psychology, University of Connecticut

Acknowledgment. This research was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. Portions of the results were reported at the 109th meeting of the Acoustical Society of America in Austin, TX, April 1985.

stage of translation may be required between such stimuli and the vocal response, and certain irregularities could arise at that stage. (2) Phonetic categorization may intrude upon the internal representations of the stimuli, as it apparently does in vowel discrimination tasks (Pisoni, 1975; Repp, Healy, & Crowder, 1979). In other words, the imitation task may simply elicit the same quasi-categorical response pattern that is typically obtained in vowel experiments following the "categorical perception" paradigm. On the production side, there are two additional possibilities: (3) The observed stimulus-response nonlinearities may reflect articulatory constraints on vowel production that are either universal or acquired through experience with a particular language (notwithstanding the relative rarity of isolated vowels in everyday communication). This hypothesis was favored by Chistovich et al. (1962). (4) Finally, there is the possibility that the constraints are not articulatory but acoustic in nature, in that certain discontinuities in the transform from vocal tract shape to the output lead a speaker to favor certain formant patterns, as suggested by Stevens' "quantal theory" of vowel production (Stevens, 1972).

The first hypothesis seems perhaps less plausible than the others in view of the fact that Chistovich et al. (1966), in their original demonstration of categorical imitation, used synthetic stimuli that were modelled after the (single) subject's own productions. On the other hand, that hypothesis is the easiest one to test and deserves to be ruled out before the other possibilities are investigated more thoroughly. This was the purpose of the present study.

Acoustic analysis of the responses obtained in our earlier study (Repp & Williams, 1985) revealed a large variety of formant patterns, which made it possible to select a number of utterances that formed naturally produced vowel continua specific to each subject. With this assurance that each subject was physically able to produce a precise match for each stimulus, we proceeded to replicate the experiment. Subjects, design, and procedure were identical, and the reader is referred to our earlier report (Repp & Williams, 1985) for some methodological details and for results not reproduced here. (Figures 1-8 correspond to earlier figures with the same numbers.)

Even though there were only two subjects in this study (due to our method of stimulus selection, our desire to make a within-subject comparison with the earlier results, and our preference for experienced subjects), we expected to have sufficient evidence against the hypothesis under test if (1) each subject's imitation responses along a vowel continuum show significant, nonuniform deviations from the stimulus parameters, and (2) these deviations follow a pattern similar to that obtained in our earlier study.

Methods

Stimuli

Two 12-member vowel continua, one intended to range from [u] to [i] and the other from [i] to [æ], were selected from appropriate two-dimensional scatter plots of each subject's imitation responses in the first study (Repp & Williams, 1985). Each formant frequency plot included 36 responses to each of 12 members of a synthetic vowel continuum, either [u]-[i] or [i]-[æ], a total of 432 data points. From each of these plots we selected twelve tokens that were as equidistant as possible and followed a pre-determined path in the (linearly scaled) formant frequency space. The resulting natural [i]-[æ] continuum was selected to fall along a straight line in the F1-F2 plane, determined by linear regression of F2 on F1 in the scatterplot, whereas the [u]-[i] continuum was made to follow a curve in the F2-F3 plane, derived by eye from the central tendencies in the data. In addition, since it was not possible to vary other stimulus parameters systematically, it was attempted to hold F1 on the [u]-[i] continuum, and F3 on the [i]-[æ] continuum, as constant as possible by avoiding tokens with deviant values. Extreme values of fundamental frequency and duration were likewise excluded by listening to each continuum and by replacing tokens that "stuck out." The average formant frequencies of the stimuli selected, determined by LPC analysis, are listed in Table 1. Stimulus durations varied between 150 and 210 ms, average fundamental frequencies between 104 and 127 Hz (DW) and between 111 and 132 Hz (BR).[1]

Table 1

Average Formant Frequencies of Stimulus Vowels (Hz).

	DW			BR		
Stim	F1	F2	F3	F1	F2	F3
			[u]-[i] continuum			
1	310	1036	2071	310	979	2068
2	301	1164	2102	312	1101	2025
3	308	1296	2122	318	1232	1986
4	305	1431	2135	309	1330	1995
5	308	1538	2154	320	1466	2030
6	308	1620	2210	312	1564	2075
7	307	1694	2308	319	1673	2141
8	313	1778	2378	324	1782	2195
9	312	1858	2453	317	1877	2257
10	307	1917	2523	309	1966	2359
11	302	2022	2617	297	1996	2451
12	276	2089	2666	293	2027	2568
			[i]-[æ] continuum			
1	269	2124	2629	297	2069	2592
2	300	2082	2488	313	2038	2533
3	334	2035	2442	341	2014	2512
4	370	2001	2457	366	1985	2460
5	381	1963	2453	383	1934	2378
6	414	1911	2396	412	1895	2443
7	442	1877	2355	424	1860	2362
8	472	1837	2401	462	1807	2381
9	505	1791	2372	476	1783	2355
10	530	1731	2375	495	1760	2391
11	566	1692	2390	513	1732	2347
12	594	1657	2392	539	1681	2253

Subjects, Procedure, and Analysis

The two authors served as subjects. DW is a native speaker of American English, BR of German. Each subject listened to 9 randomized blocks of 48 stimuli (4 repetitions of the 12 stimuli along a continuum) for each of his two personal stimulus sets. Following the design of our earlier study, each stimulus was either preceded (-500 ms stimulus onset asynchrony) or followed (750 or 3000 ms) by a 100-ms, 1000-Hz tone, with three stimulus blocks assigned to each of these three conditions in a counterbalanced order. The subjects rapidly imitated the stimulus vowel after hearing the tone if the tone followed ("delayed" and "deferred" imitation conditions) or after hearing the stimulus if the tone preceded ("immediate" imitation condition).

In a separate test conducted several months later, each subject also identified the stimuli in his own [i]-[æ] set, using the phonemic labels /i,I,e,ε,æ /. This test consisted of 10 randomized blocks of the 12 stimuli (without accompanying tones).

The only design change from the earlier study was that, foregoing an absolute identification (numerical labeling) task (Repp, & Williams, 1985), each subject produced a series of isolated vowels by reading from a list containing the symbols /u,i,I,e,ε,æ / 36 times in random order. These productions were to serve as "prototypical" reference points in interpreting the imitation data.

The recorded imitation responses were digitized at 10 kHz, low-pass filtered at 4.9 kHz, and subjected to LPC analysis.[2] The formant frequency estimates were edited to eliminate spurious and missing values, and were averaged across the whole duration of each response vowel. Mean formant frequencies and standard deviations across repeated imitations of the same stimulus were determined, as well as the distributions of formant frequencies across all responses to a given continuum. The prototypical productions were analyzed similarly. Imitation response latencies were also measured and will be discussed first.

Results and Discussion

Latencies

Chistovich et al. (1966) observed that imitation latencies, unlike the latencies of phonetic labeling responses, did not vary systematically across an acoustic vowel continuum, regardless of response delay. Relative uncertainty about phonemic category membership thus did not seem to influence the speed of imitation. This finding, which suggests that imitation is not mediated by phonemic classification, was essentially replicated in our earlier study (Repp & Williams, 1985). The average response latencies from the present experiment are shown in Figure 1 as a function of subject (top vs. bottom panels), continuum (left vs. right panels), stimulus number (abscissa), and delay condition (three functions). Two findings are apparent. First, although reaction times varied somewhat across each continuum, there was no consistent pattern to this variation. In other words, there were no peaks in the latency functions associated with phonetic category boundaries. Second, subject DW showed markedly slower reaction times in the immediate imitation condition than in the delayed or deferred imitation conditions, whereas subject BR showed slower latencies in the immediate and deferred conditions than in the delayed condition. While slower reaction times in the immediate imitation condition are expected because of the subjects' incomplete articulatory preparation, only BR was affected by a 3-second response delay. This pattern of results is remarkably similar to that obtained in our earlier study with synthetic stimuli.[3]

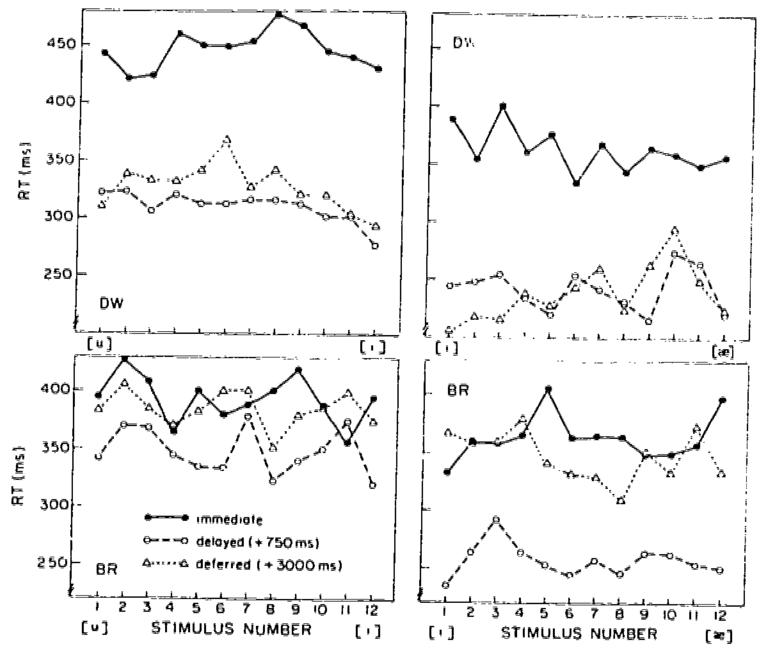


Figure 1. Average response latencies as a function of stimulus number and delay condition for two subjects (DW, BR) and two continua ([u]-[i], [i]-[æ]). Each data point represents 12 responses.

Separate repeated-measures analyses of variance with the factors Stimulus Number and Delay Condition were conducted on the average latencies for the three stimulus blocks of each continuum and each subject. The only significant effect involving Stimulus Number was a small main effect for DW on the [u]-[i] continuum, $F(11,72) = 2.11$, $p = .0304$, which was not readily interpretable; the other three main effects and the four interactions were nonsignificant, which suggests the absence of reliable peaks in the latency functions. The main effect of Delay Condition, however, was highly significant ($p < .0001$) in each of the four analyses.

Formant Frequencies

As in our earlier study, we found that the patterns of average response formant frequencies were extremely similar across the three delay conditions, so the data were collapsed across delays.[4] The mean values were thus based on 36 responses per stimulus. These means are plotted as a function of stimulus number in Figure 2 (solid lines); the dashed lines connect the stimulus formant frequencies.[5]

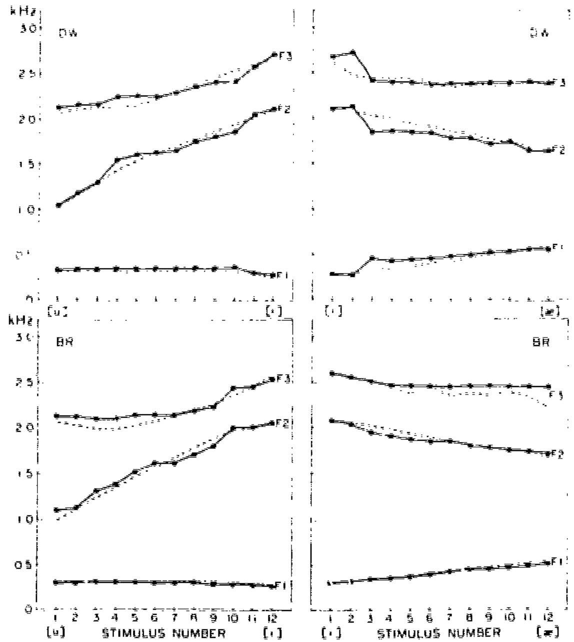


Figure 2. Average formant frequencies of the responses as a function of stimulus number (filled circles, solid lines). Each data point represents 36 responses. The stimulus formant frequencies are connected by the dashed lines.

Compared to our earlier results with synthetic stimuli, the response formant frequencies are much closer to those of the stimuli, as should be expected when subjects imitate their own vowels. Nevertheless, there appear to be systematic deviations that echo some of the response nonlinearities observed with synthetic stimuli. Many of these deviations are significant individually, since standard errors are small (one-sixth of the standard deviations displayed in Figure 5 below). They are also significant overall, as is clear from the results of analyses of variance on the deviations of the responses from the stimulus parameters. Four such analyses were conducted (two continua for each of two subjects) on three parameters (F1, F2, F3) considered jointly (using a multivariate statistic) and separately. Of the grand mean effects, which test the average stimulus-response difference on each continuum, all 4 multivariate and 11 of the 12 univariate F values were highly significant ($p < .0001$; exception: F2 for DW on the [u]-[i] continuum, which was nonsignificant). More importantly, all 16 stimulus number main effects, which test whether responses deviated nonuniformly from the stimuli across each continuum, were highly significant ($p < .0001$). Thus there is ample statistical support for stimulus-response nonlinearities in the data. These nonlinearities are examined more closely in the next two figures.

Figure 3 shows stimulus-response relations in F2-F3 space for the [u]-[i] continuum. DW's responses to stimuli 4-10 on this series tend to cluster together, though he was able to imitate their distinctive characteristics to some extent. A similar, but weaker tendency is exhibited by BR for stimuli 5-9; in addition, BR tended to respond categorically to the endpoint stimuli (1, 2, and 10, 11, 12, respectively). These tendencies are similar to those observed in our earlier study.

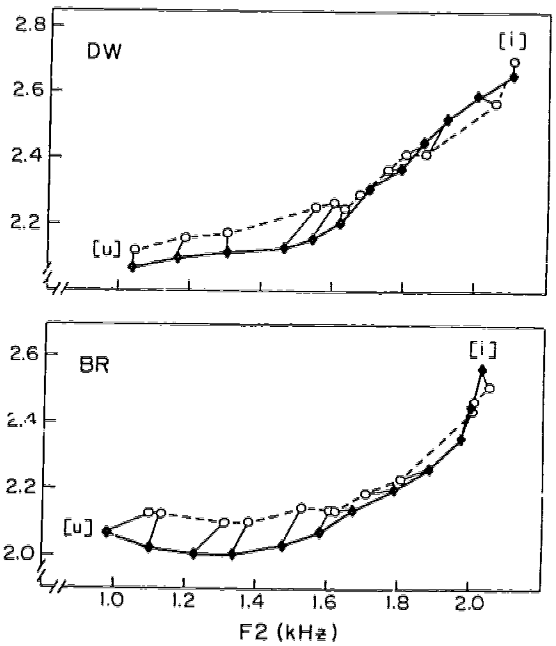


Figure 3. Average formant frequencies of responses to the [u]-[i] continuum in F2-F3 space (open circles, dashed line). Filled diamonds connected by a solid line represent the stimuli. Each stimulus is connected to its corresponding average response.

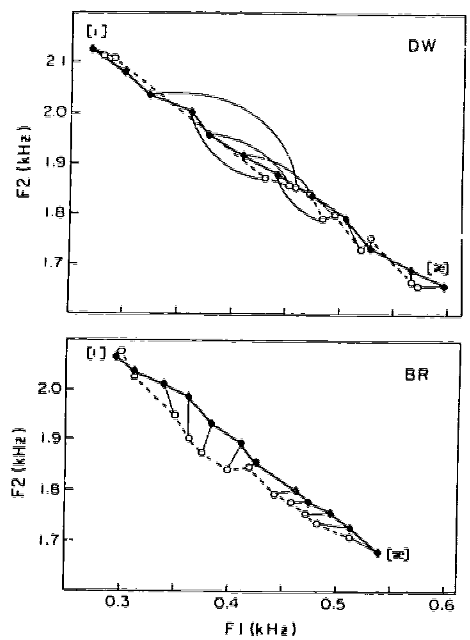


Figure 4. Average formant frequencies of responses to the [i]-[æ] continuum in F1-F2 space (open circles, dashed line). Filled diamonds connected by a solid line represent the stimuli. Each stimulus is connected to its corresponding average response; the curving connectors in the upper panel are necessitated by the large response shifts.

Figure 4 shows the stimulus-response mapping in F1-F2 space for the [i]-[æ] continuum. DW shows very dramatic deviations here. There is a huge gap between the responses to stimuli 2 and 3, and responses to stimuli 3-9 are transposed down along the F1-F2 regression line. (Note that the responses, like the stimuli, continue to observe this linear relationship despite the large discrepancies.) There is also evidence for some endpoint clustering (stimuli 1, 2, and 11, 12, respectively.) Subject BR, by contrast, shows relatively continuous responses to this continuum, although there is some contraction of the response space for stimuli 3-12. Once again, these patterns show similarities to those we have observed with synthetic stimuli. The similarities are difficult to quantify, however, because the stimuli in the two studies are not in one-to-one correspondence.

Standard Deviations

Another way to look for categorical tendencies is to examine the patterns of response variability. Response variability is expected to increase at category boundaries, if there are any. Standard deviations of formant frequencies, computed within but averaged across delay conditions, are shown in Figure 5. These patterns are remarkably similar to those observed with synthetic stimuli.

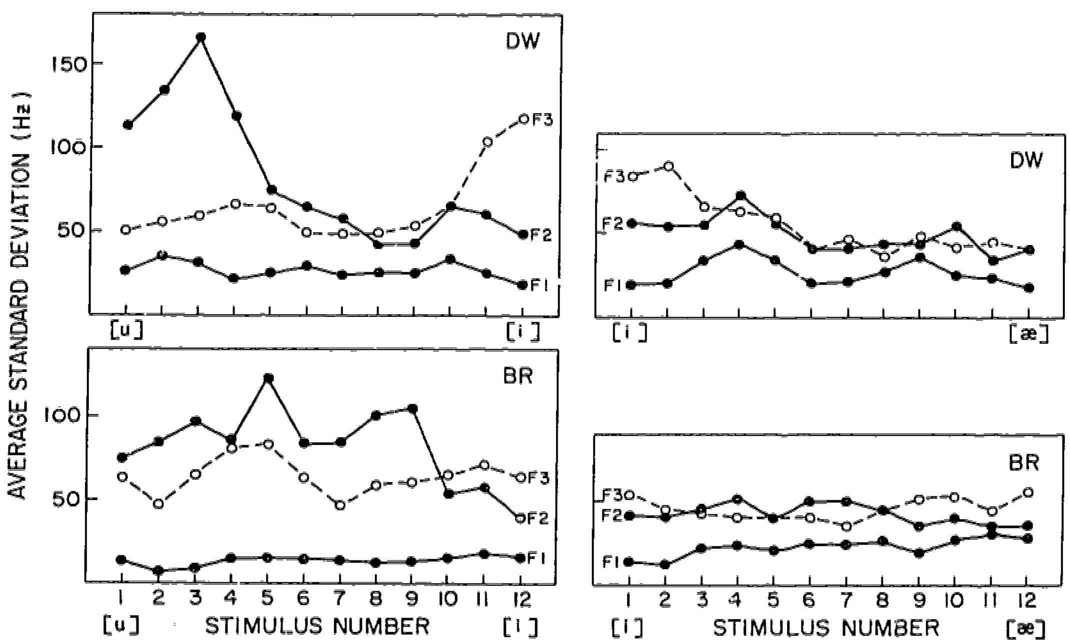


Figure 5. Average standard deviations of response formant frequencies.

Both subjects showed higher F2 variability along the [u]-[i] than along the [i]-[æ] continuum, except at the [i] end. For BR, F2 variability was elevated across most of the [u]-[i] continuum (stimuli 1-9), whereas DW showed elevated variability over a narrower region (stimuli 1-4), with a pronounced peak for stimulus 3. This peak corresponds to the gap in the formant frequency plot (Figure 3). Apart from this feature, there are no clear indications of a categorical structure in the standard deviations along the [u]-[i] continuum. Along the [i]-[æ] continuum, however, subject DW shows two peaks in both the F1 and F2 functions, which suggest a three-category structure. As in the earlier study, F1 and F2 standard deviations were correlated for DW ($r = 0.55$, $p < .05$) but not for BR ($r = 0.03$). For BR, therefore, the standard deviations do not reveal any obvious categorical tendencies. Individual differences aside, however, the point to be stressed is that the standard deviations follow the same pattern as in the earlier study, suggesting that the subjects responded similarly to synthetic and natural stimuli.

Formant Frequency Distributions

The best way to assess categorical response tendencies is to plot overall formant frequency distributions. Frequency histogram envelopes of the first three formants of the responses in all three delay conditions combined ($n = 432$ in each graph) are shown in Figures 6 and 7 (solid lines). For comparison, the histogram envelopes from our earlier study with synthetic stimuli are plotted alongside on the same scale (dashed lines). Significant similarities are evident.

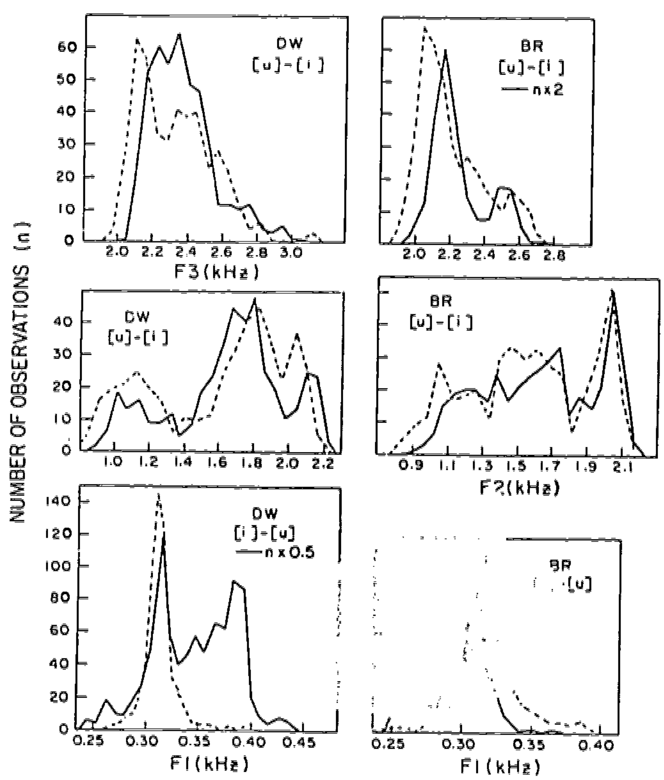


Figure 6. Histogram envelopes of response formant frequencies for the [u]-[i] continuum in the present study (solid lines) and in our earlier study using synthetic stimuli (dashed lines). Note that the plots for the three formants are not aligned with each other, and that the scale factor is altered for some individual functions to make the functions similar in height.

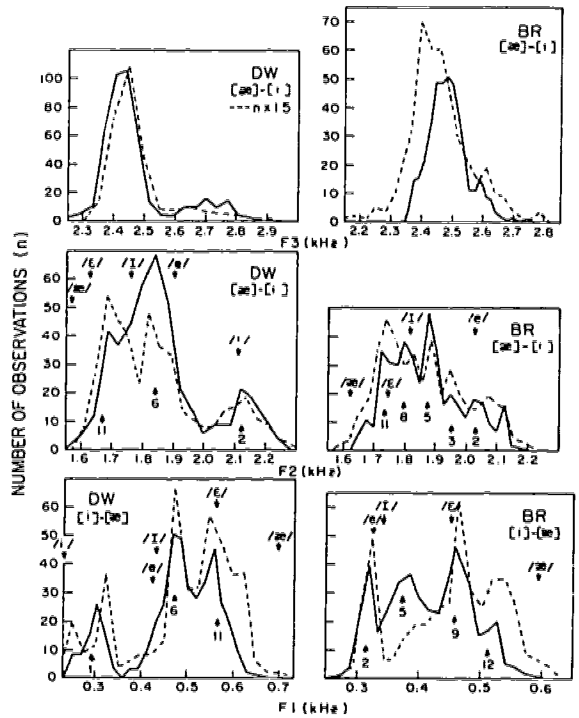


Figure 7. Histogram envelopes of response formant frequencies for the [i]-[æ] continuum in the present study (solid lines) and in our earlier study using synthetic stimuli (dashed lines). Note that the plots for the three formants are not aligned with each other, that the continuum is reversed for F2 and F3 with respect to F1, and that the scale factor is altered for the solid function in the upper left-hand panel. Arrows with numbers represent the stimuli whose average responses fell closest to histogram peaks. Arrows with phonetic symbols represent prototypical vowel productions.

On the [u]-[i] continuum (Figure 6), the only major discrepancy between the two sets of results is the presence of a second peak in DW's F1 distribution for natural speech stimuli. The cause for these unusually high F1 frequencies in many of DW's responses is unknown. (Stimulus F1 frequencies ranged from 276 to 313 Hz; see Table 1). BR has a single-peaked F1 function whose displacement with respect to the earlier study brings it in good agreement with the stimulus range and corrects a consistent F1 "overshoot" observed with synthetic stimuli. The F2 frequency distributions of both subjects are rather similar to those obtained with synthetic stimuli and show three major peaks, two probably representing the endpoint categories and the third a broad category of "unfamiliar" vowel sounds. The F3 distributions are essentially unimodal and shifted to the right with respect to the previous study, resulting in a better match of stimulus and response F3 ranges (cf. Table 1).

For the [i]-[æ] continuum (Figure 7), both F1 and F2 show highly irregular distributions indicative of categorical tendencies, whereas the F3 distribution is unimodal. For DW, both the F1 and F2 distributions are trimodal; moreover, the peaks (taking into account the reversal of the continuum along the F2 scale) are in fact aligned with each other. DW thus shows evidence for three categories along this continuum. For BR, the pattern is less clear. The F1 histogram shows four peaks, adding a new one to the three-peaked function for synthetic stimuli. The F2 function has multiple peaks--too many for any clear categorical structure to be inferred.

The main result of these comparisons is that individual response preferences are maintained to a considerable extent even when subjects imitate self-produced vowels. Clearly, few of the distributions are uniform, as they should be if formant frequencies were reproduced faithfully.

<u>Phonemic Identification</u>

The subjects labeled the stimuli along their own [i]-[æ] continua to provide a reference for the interpretation of categorical tendencies along that continuum. These classifications are plotted in Figure 8. It can be seen that DW used only three categories (/i,e,ε/) consistently; he used /æ/ interchangeably with /ε/, and /I/ not at all. That is, for him the stimulus continuum represented only three categories. BR, on the other hand, applied all five response categories to his vowels, although stimulus 12 still was only a weak /æ/ to him.

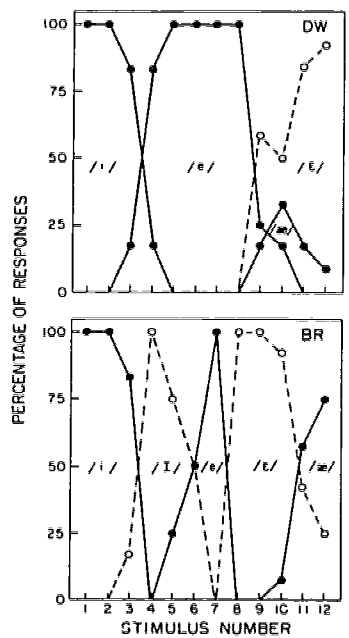


Figure 8. Labeling responses to the [i]-[æ] continuum.

To see whether these data are helpful in interpreting the histogram peaks, the ordinal numbers of the stimuli whose associated mean response formant frequencies were close to histogram peaks have been entered below arrows in Figure 7. For subject DW, the three major peaks in F1 and F2 are associated with responses to stimuli identified as /i/, /e/, and /ε/ (or /æ/), respectively. This correspondence is in agreement with that observed in our earlier study, except that we then interpreted the /e/ category as /I/. DW's categorical tendencies in imitation thus correspond well to his phonemic categories. For BR, the F1 and F2 peaks line up with stimuli labeled as /i/,

/I/, /ε/, and /ae/, respectively, although there seem to be two /i/ peaks in the F2 distribution. These alignments differ somewhat from those obtained in our earlier study and therefore must be regarded with caution. That BR, as a native speaker of German, should not have a well-defined /e/ category in imitation seems counterintuitive. For this subject, then, the imitation data are not clearly related to his (English) phonemic categories, perhaps because of his bilingualism.[6]

Prototypical Vowels

A new feature of the present study was the inclusion of "prototypical" productions representing the five English vowel categories along the [i]-[æ] continuum. The average frequencies of these productions have been entered above arrows in the F1 and F2 panels of Figure 7. Somewhat surprisingly, these values are not very helpful in interpreting the histogram peaks. The prototypical values for /æ/ generally fall outside the response ranges. Those for the other categories generally do not coincide with major peaks, although some tentative alignments can be made if small shifts in formant frequencies are allowed for. Clearly, the subjects did not simply produce their prototype vowels in the imitation task. Their responses definitely were more a function of the stimuli than of pre-established phonetic categories, although the categories may have exerted a certain "pull" on the responses.

To get a better idea of the locations of the prototype vowels in the formant frequency plane relative to the stimulus and response vowels, the subjects' responses to their [i]-[æ] continua have been replotted in Figure 9 together with the prototypes, with standard deviations represented as well.

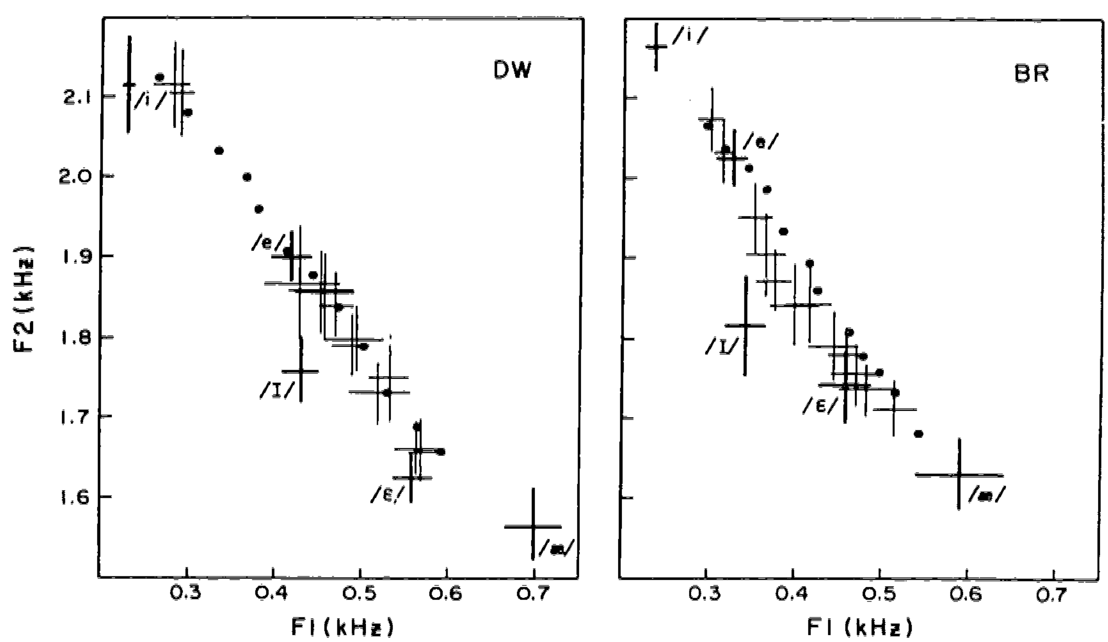


Figure 9. Average formant frequencies of responses along the [i]-[æ] continuum plus/minus one standard deviation (thin lines) and of prototypical vowel productions plus/minus one standard deviation (heavy lines) in F1-F2 space. The circles represent the stimuli.

The stimuli appear as filled circles. One interesting feature emerging from these plots is that, for both talkers, the five prototype vowels do not lie on a straight line in F1-F2 space, in contrast to the response (and stimulus) vowels. It seems that the subjects, being rather accurate imitators, fit their responses to the linear trajectory imposed by the stimuli, rather than gravitating toward their prototypical vowels. Prototypical /I/, in particular, lies outside the stimulus-response trajectory, and /æ/, as well as BR's /i/, is beyond the stimulus-response range. Most responses fall between prototypical /e/ and /ɛ/; only DW also produced some /i/-like vowels. The main difference between the two subjects is in the location of the /e/ prototype, which is closer to /i/ for BR and presumably reflects his native language. The absence of a prototype for DW in the same region may explain the large shifts in his responses to stimuli 3-5. Curiously, BR labeled stimuli as /i/ (Figure 8) that in fact were much closer to his prototypical /e/, and the stimuli he labeled as /e/ were closer to his prototypical /I/. DW's labeling responses are in much better agreement with the pattern of stimulus-prototype proximities shown in Figure 9.[7]

Fundamental Frequencies

We examined two additional stimulus-response relationships that we could not explore in our earlier study because of the constant fundamental frequency (F0) and duration of the synthetic stimuli. First, we compared the average fundamental frequencies (F0) of the stimuli and of the responses. For DW, there were no major trends in response F0 across either continuum; occasional deviations seemed to be related to stimulus F0. Stimulus-response correlations in the three delay conditions for each continuum ranged from 0.39 to 0.88 (4 out of 6 significant at $p < .01$), which indicated that DW unintentionally imitated stimulus F0. For BR, the correlations were lower but still positive, ranging from 0.25 to 0.62 (1 out of 6 significant at $p < .05$), and his response F0 tended to fall across both continua (from [u] to [i], and from [i] to [æ]), an effect that was apparently not induced by the stimuli. Stimulus-response correlations for both subjects tended to be lower in the immediate imitation condition. Delay conditions affected absolute F0, but these patterns varied between subjects and continua and were difficult to interpret.

Durations

Similarly, we examined stimulus and response durations along each continuum and found some very consistent patterns. The stimulus-response correlations were positive and surprisingly high in some instances. For DW, they ranged from 0.30 to 0.96 (5 out of 6 significant at $p < .01$); for BR, from 0.35 to 0.70 (4 out of 6 significant at $p < .05$, one of those at $p < .01$). Although it might be argued that a common articulatory or phonetic factor influenced stimulus and response durations alike, the pattern of durations across each continuum was sufficiently irregular (due to the method of stimulus selection) to suggest, rather, that both subjects unintentionally mimicked vowel durations. The stimulus-response correlations tended to be lower in the deferred imitation condition. In addition, there was a very pronounced effect of delay condition on the average duration of the responses: Response vowels were generally shorter in the immediate imitation condition.

Conclusions

On the whole, the present results replicate the findings of our first study (Repp & Williams, 1985). That is, categorical tendencies in vowel imitation are obtained even when the subjects are capable of producing the precise vowel they are to imitate. This rules out one possible explanation of the obtained stimulus-response nonlinearities, namely, that they arise in the translation of nonproducible stimuli into the subject's own production space. As pointed out in the Introduction, this hypothesis had limited plausibility to begin with; thus, a sample of two subjects seems sufficient for its dismissal. At the same time, the demonstration of similar nonlinearities with synthetic and natural stimuli confirms the robustness of these effects, as well as the presence of considerable individual differences in their pattern and magnitude (cf. also Kent, 1973).

One possible reason for the absence of very strong categorical effects in this study and its predecessors (Kent, 1973; Repp & Williams, 1985) is suggested by the relation of the subjects' prototypical vowels to the [i]-[æ] stimulus continuum. Our continuum derived from responses to a synthetic continuum (Repp & Williams, 1985), which we had copied from Kent (1973), who in turn had designed it to span the average male vowel formant frequencies for /i/ and /æ/ reported by Peterson and Barney (1952). These latter data derived from vowels in /h_d/ context and may not be representative of isolated vowel productions (especially /I/ and /ɛ/), for which normative English data are hard to come by in the literature. It is also possible that the present subjects were not representative of the average American male talker. In any case, it seems that the [i]-[æ] continua used by Kent and by us did not span the full space between /i/ and /æ/, and that they bypassed /I/. Chistovich et al. (1966) used a continuum that seems to have been more closely matched to their single subject's prototypes, and it remains to be seen whether their highly categorical results can be replicated with similarly constructed stimulus continua.

The question of the origin of categorical tendencies in vowel imitation needs to be addressed in further research. Perhaps the most interesting result to emerge from our studies and that of Chistovich et al. (1966) is that categorical tendencies in imitation appear regardless of response delay (up to 2 seconds) and with essentially constant reaction times. Imitation responses thus do not seem to be mediated by explicit phonemic decisions (which are slowed by stimulus ambiguity), nor do they depend on a rapidly decaying auditory memory (which plays a role in vowel discrimination, see Crowder, 1982a, 1982b; Pisoni, 1975). This suggests that the internal representation of perceived vowels is phonetic (or articulatory) but, at the same time, either noncategorical or only weakly categorical. If it is noncategorical, then the categorical tendencies must arise during the motor implementation of the imitations. Research is now in progress to examine this possibility.

References

Chistovich, L. A., Fant, G., de Serpa-Leitao, A., & Tjernlund, P. (1966). Mimicking and perception of synthetic vowels. Quarterly Progress and Status Report (Royal Technical University, Speech Transmission Laboratory, Stockholm), 2, 1-18.

Crowder, R. G. (1982). Decay of auditory information in vowel discrimination. Journal of Experimental Psychology: Human Learning, Memory, and Cognition, 8, 153-162.

Crowder, R. G. (1982). A common basis for auditory sensory storage in perception and immediate memory. Perception & Psychophysics, 31, 477-483.

Kent, R. D. (1973). The imitation of synthetic vowels and some implications for speech memory. Phonetica, 28, 1-25.

Peterson G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 24, 175-184.

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. Memory & Cognition, 3, 7-18.

Repp, B. H., Healy A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated, steady-state vowels. Journal of Experimental Psychology: Human Perception and Performance, 5, 129-145.

Repp B. H., & Williams, D. R. (1985). Categorical trends in vowel imitation: Preliminary observations from a replication experiment. Speech Communication, 4, 105-120.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), Human communication: A unified view. New York: McGraw-Hill.

Footnotes

[1]To simplify the analysis of stimulus-response relationships, acoustic parameter values were averaged across the whole duration of both stimulus and response vowels. The stimuli were not perfectly steady-state, however, although they represented imitations of truly stationary synthetic vowels. Formant measurements obtained at two specific points in each vowel--at onset and two-thirds into its duration--provided an indication of changes over time. These changes were relatively small and showed no orderly trends across the continua. In general, the frequencies of all formants and of the fundamental frequency declined through each vowel, except for F1 on the [i]-[æ] continua, which tended to rise. Most of the changes in F1 were less than 20 Hz; in F2 and F3, less than 100 Hz; in F0, less than 10 Hz. Only a few tokens exceeded these limits. Clearly, none of the stimulus vowels resembled diphthongs.

[2]The peak-picking algorithm used to estimate formant frequencies (part of the ILS package, Version 4.0, distributed by Signal Technology, Inc.) may produce artificial discontinuities when tracking formants in time-varying signals, due to certain limitations in the FFT routine. To make sure that the present, relatively steady-state vowels had been correctly analyzed, the data from DW's [u]-[i] condition were re-analyzed using the root-solving method included in ILS, which is more accurate but time-consuming. The results were practically identical to those obtained with the peak-picking method, except that F1 estimates were uniformly higher by about 10 Hz. The reason for this absolute difference is not known. The peak-picking algorithm thus seems to provide accurate results for relatively steady-state speech sounds.

[3]Only the absolute reaction times differed: Relative to the reaction times to synthetic stimuli, DW speeded up on the [i]-[æ] continuum, while BR slowed down on the [u]-[i] continuum. These changes are difficult to interpret and are of little theoretical interest.

[4]To justify this decision, analyses of variance were conducted on stimulus block mean values of F1, F2, and F3 for each subject and continuum, with the factors Stimulus Number and Delay. A significant interaction between these factors would indicate a change of formant pattern as a function of delay

condition. Of the twelve interactions tested, only one was significant, for F1 along the [u]-[i] continuum of subject BR, $\underline{F}(22,72) = 1.99$, $\underline{p} = .0157$, which is of little interest because responses to that continuum were analyzed primarily in F2-F3 space. The main effect of Delay was significant in several instances, indicating changes in absolute formant frequencies across delays without a concomitant change in stimulus-response relationships. The more striking of these included lower F1 frequencies (subject DW) and lower F2 frequencies (subject BR) in the immediate imitation of stimuli from the [i]-[æ] continuum.

[5]The responses, like the stimuli, were examined for changes in formant frequencies and F0 over time by comparing measurements taken at vowel onset and after two-thirds of its duration. This analysis revealed that the response vowels were monophthongal and, in fact, rather stationary. The mean response parameters exhibited a variety of systematic trends in within-vowel changes across each continuum, but the magnitudes of these changes were rather small (generally less than 25 Hz for F1, 65 Hz for F2, 40 Hz for F3, 16 Hz for F0). Of the 16 stimulus-response correlations of frequency changes (4 parameters, 2 continua, 2 subjects) 15 were positive, but only one was significant. Thus there was no strong evidence that the subjects imitated time-varying characteristics of the stimuli.

[6]Because of these puzzling results, we later repeated the identification task, also with the two subjects listening to each other's [i]-[æ] series. This replication revealed considerable inconsistency in the subjects' use of the /I/ category, and both subjects agreed that no very good instances of this vowel were present in either stimulus series. BR's data make more sense if stimulus 5, and the associated peaks in the F1 and F2 histograms, are taken to represent his /e/ category.

[7]See, however, footnote 6. BR's labeling data from the replication were in somewhat better agreement with his prototypes. Also, both subjects' productions of /I/ may have been anomalous; after all, this English vowel does not occur in isolation. As a matter of fact, both subjects' productions of all vowels deviate considerably from the Peterson-Barney norms (1952), which are based on vowels produced in /h_d/ context (not including /e/). It should also be mentioned that DW's prototypical productions, but not BR's, tended to be diphthongized. Both subjects' ability to identify their own and each other's prototypes was tested later. Scores ranged from 93 to 100 percent correct, with most confusions involving intended /I/ or /ε/.

AN ACOUSTIC ANALYSIS OF V-TO-C AND V-TO-V: COARTICULATORY EFFECTS IN CATALAN AND SPANISH VCV SEQUENCES

Daniel Recasens

Abstract. V-to-C and V-to-V coarticulatory effects in F_2 frequency are studied for Catalan and Spanish VCV sequences with vowels and consonants involving different degrees of articulatory constraint on tongue-dorsum activity. The findings reported in this paper indicate that coarticulatory effects decrease with the degree of articulatory constraint, for the following groups of consonants and vowels: [l]>[ɫ]; [ɾ]>[r]; [β], [ð]>[ɣ]; [a]>[i]. Differences in anticipatory vs. carryover coarticulation were also found to be strongly dependent on the degree of articulatory constraint associated with the intervening consonants and vowels. Overall, results suggest that coarticulatory effects are deeply related to the control mechanisms involved in the production of articulatory gestures.

Introduction

The main purpose of this paper is to show the need for a theory of coarticulation that accounts for coarticulatory effects in terms of the constraints (i.e., requirements) imposed on the articulators during the production of gestures for adjacent phonemes. According to gestural models of coarticulation, coarticulatory effects occur as long as the articulatory requirements for an ongoing gesture do not conflict with those for adjacent gestures (Öhman, 1966). In an effort to characterize the notion of articulatory conflict, evidence will be provided here in support of the hypothesis that the degree of compatibility between a given gesture and adjacent gestures decreases with the degree of articulatory constraint. Thus, highly constrained gestures ought to block coarticulatory effects to a larger extent than gestures specified for lesser degrees of articulatory constraint. Data from the literature support this view. For instance, Lubker and Gay (1982) have shown that the lip rounding gesture for [u:] allows lesser coarticulatory effects in Swedish than in American English because of being subject to higher articulatory requirements; thus, [u:] shows more lip protrusion and an earlier lip rounding onset in Swedish vs. American English in line with the fact that Swedish has more distinctive rounded vowels than English. Also, Recasens (1984a) showed for Catalan that coarticulatory effects on the degree of dorsopalatal contact for palatal, alveolopalatal, and

---

Acknowledgment. This research was supported by NICHD Grant HD-01994 and NINCDS Grant NS-13617 to Haskins Laboratories. I am grateful to Ignatius Mattingly and Michael Studdert-Kennedy for helpful comments on the subject investigated here.

[HASKINS LABORATORIES: Status Report on Speech Research SR-86/87 (1986)]

alveolar consonants vary inversely with the constriction degree; thus, coarticulatory effects decrease with an increase of the requirements imposed upon the tongue dorsum to make contact at the surface of the hard palate.

First, acoustic data will be presented that suggest that the degree of V-to-C and V-to-V coarticulation in VCV sequences is inversely related to the degree of articulatory constraint for the consonantal gesture. For this purpose, coarticulatory effects will be analyzed for the following consonants showing contrasting degrees of articulatory constraint on tongue-dorsum activity: (1) velarized apicoalveolar lateral [ɫ] vs. non-velarized apicoalveolar lateral [l]; (2) apicoalveolar trill [r] vs. apicoalveolar tap [ɾ]; (3) velar approximant [γ] vs. bilabial approximant [β] and dental approximant [ð]. Among these consonants tongue-dorsum activity is subject to a higher degree of articulatory constraint for [ɫ] vs. [l], [r] vs. [ɾ], and [γ] vs. [β] and [ð]: higher articulatory control over tongue-dorsum activity for [ɫ] vs. [l] results from the fact that, while the two realizations involve apicoalveolar contact, only [ɫ] is articulated with postdorsal constriction at the velopharyngeal region and predorsal lowering (Recasens, 1985); higher demands on tongue-dorsum activity for [r] vs. [ɾ] are reflected by some backing of the tongue dorsum and, presumably, some degree of dorsopharyngeal constriction for the trill (see, for Spanish, Navarro Tomás, 1970) to allow the execution of several apicoalveolar vibrations; finally, [γ] is subject to a higher degree of tongue-dorsum constraint than [β] and [ð] in accordance with the fact that, for [γ], the tongue dorsum is fully involved in the formation of a constriction at the palatovelar or velar regions.

Some data from the literature are relevant here. Acoustic data (F_2) for English show indeed that "dark" [l] is highly resistant to V-to-C effects during closure (American English: Lehiste, 1964; RP British English: Bladon & Al-Bamerni, 1976), more so than "clear" [l] (Bladon & Al-Bamerni, 1976). Also, larger V-to-V coarticulatory effects on tongue-dorsum activity have been reported across labial and alveolar consonants (Catalan: Recasens, 1984b; American English: Carney & Moll, 1971; Swedish: Öhman, 1966) than across velar consonants (German: Butcher & Weiher, 1976); these data are consistent with articulatory data (Catalan: Recasens, 1984a, 1984b) and acoustic data (American English: Lehiste, 1964; Stevens & House, 1963) showing that palatal consonants allow less coarticulatory effects on tongue-dorsum activity than labials and alveolars.

The issue as to whether vowel-dependent effects in VCV sequences can or cannot extend into the transconsonantal vowel is of interest here as well. Thus, while it was found in several early works that such effects do not extend beyond the period of consonantal closure (Gay, 1974, 1977) or the period of transconsonantal vowel transitions (Öhman, 1966; Carney & Moll, 1971), more recent acoustic evidence for English and other languages (Magen, 1984; Manuel & Krakow, 1984) reveals that vowel-dependent effects can also extend into the steady-state period of the transconsonantal vowel.

It will also be shown that the degree to which an F_2 difference between two vowels can be traced beyond consonantal closure or consonantal constriction depends on the degree of articulatory constraint for the transconsonantal vowel. Thus, some vowels have been reported to be more resistant than others to V-to-V effects. In a series of experiments, Gay (1974, 1977) found [i] to be more resistant than [a] to differences in jaw opening and tongue body height caused by contrasting consonants and vowels in

VCV sequences. Similarly, Carney and Moll (1971) found no anticipatory V-to-V effects in tongue-dorsum activity at the steady state period of V1=[i]. These articulatory data accord well with acoustic data. Thus, [i] has been found to be more resistant than [a] to V-to-V coarticulatory effects in F_2 frequency in Japanese (Magen, 1984), and in Swahili and Shona (Manuel & Krakow, 1984).

I will also analyze differences in nature between anticipatory and carryover coarticulatory effects. It is commonly accepted that carryover effects are more dependent than anticipatory effects on mechanical constraints, and that anticipatory effects are mainly timing effects resulting from articulatory preprogramming. Accordingly, contrasting V-to-V coarticulatory effects among consonants showing different degrees of articulatory constraint (such as the ones included here) ought to take place at the carryover level but less so--or not at all--at the anticipatory level. Thus, consonants subject to large degrees of constraint are expected to block V-to-V carryover coarticulation to a larger extent than consonants subject to less considerable articulatory requirements; on the other hand, a smaller contrast--or no contrast at all--between V-to-V coarticulatory effects for both sets of consonants is expected at the anticipatory level. In line with this hypothesis, differences in V-to-V coarticulation for Catalan consonants (palatals, alveolopalatals, and alveolars) involving different degrees of tongue-dorsum constraint were found to occur to a larger extent at the carryover level than at the anticipatory level (Recasens, 1984b).

Attention will also be paid to differences in magnitude between anticipatory and carryover effects. While carryover effects have generally been found to be larger than anticipatory effects in English (MacNeilage & DeClerk, 1969) and Catalan (Recasens, 1984a, 1984b), anticipatory effects have been shown to exceed carryover effects in Japanese (Magen, 1984), and in Swahili and Shona (Manuel & Krakow, 1984). This paper investigates the extent to which differences in the magnitude of anticipatory vs. carryover coarticulation follow from differences in the degree of constraint involved in the production of articulatory gestures.

Method

F_2 frequency data were collected for three sets of consonants, [l]-[ɫ], [ɾ]-[r], and [β]-[ð]-[ɣ], in all possible symmetrical and asymmetrical [V'CV] combinations with V=[i], [a]. Speakers of two different languages, Catalan and Spanish, were chosen in order to test coarticulatory effects for [l] vs. [ɫ], given the fact that the alveolar lateral consonant is known to be velarized ("dark") in Catalan ([ɫ]) and non-velarized ("clear") in Spanish ([l]) (Badia, 1951; Navarro Tomás, 1970). According to these two literature sources, the other phonetic categories tested in the experiment (the tap [ɾ], the trill [r], the approximants [β], [ð] and [ɣ], and the vowels [i] and [a]) show the same or highly similar articulatory characteristics in both languages.

All VCV sequences were embedded in Catalan and Spanish sentences about eight or nine syllables long and with the same stress pattern; in all cases the two vowels were adjacent to the stop consonant [t]. Each utterance was repeated ten times by two speakers of Eastern Catalan from the region of Barcelona and two speakers of Castilian Spanish from Madrid. Acoustic recordings were digitized at a sampling rate of 10 kHz, after preemphasis and low-pass filtering. An LPC (linear prediction coding) program included in an

ILS (Interactive Laboratory System) package was used for spectral analysis. F_2 measurements were taken at eleven equidistant points in time as detected visually on spectrographic displays of each VCV sequence:

(1) Onset of V1
(2) Equidistant point between (1) and (3)
(3) V1 midpoint, at half distance between (1) and (5)
(4) Equidistant point between (3) and (5)
(5) Onset of consonantal closure or constriction
(6) Equidistant point between (5) and (7)
(7) Offset of consonantal closure or constriction
(8) Equidistant point between (7) and (9)
(9) V2 midpoint, at half distance between (7) and (11)
(10) Equidistant point between (9) and (11)
(11) Offset of V2.

All consonants chosen for analysis allow airflow and, thus, display formant structure during the periods of closure or constriction. Measurements for points (5) and (7) were taken at the moment in time showing a sudden shift in F_2 frequency and intensity level (as determined on overall amplitude displays) from the endpoint of the V1 transitions into the consonant (point 5), and from the consonant into the V2 transitions (point 7).

Overall, 12,320 measurements were taken (28 sequences x 11 points in time x 10 repetitions x 4 speakers). Data were averaged across repetitions at each point in time, for each VCV sequence and for each speaker.

Results

1. Consonants [l] and [ɫ]

1.1 Coarticulatory effects during closure. Data on F_2 were collected at the closure period (measurement point (6)) of [l] and [ɫ] to test the following issues: (a) whether [ɫ] is articulated with more dorso-velopharyngeal constriction than [l]; (b) whether differences in the degree of constriction between the two consonants are inversely related to the degree of V-to-C coarticulation. It was predicted that a more considerable degree of dorsal constriction for [ɫ] than for [l] ought to cause a lower F_2 (Fant, 1960) since F_2 is inversely related to the degree of tongue backing. In addition, a more constricted tongue dorsum configuration for [ɫ] than for [l] ought to allow less V-to-C coarticulation and, thus, less vowel-dependent F_2 variability.

Figure 1 shows F_2 data at the midpoint of the closure period of intervocalic [l] and [ɫ] separately for each vocalic environment and for each speaker. The figure shows a lower F_2 for Catalan [ɫ] (speakers DR and PL) than for Spanish [l] (speakers FM and CA) in all four VCV environments. These data suggest that [ɫ] is produced with a more considerable degree of tongue-dorsum backing than [l] in all VCV contextual conditions, since F_2 is inversely dependent on the degree of tongue-backing and pharyngeal constriction (Fant, 1960).

The figure also shows a larger degree of vowel-dependent F_2 variability for [l] in Spanish than for [ɫ] in Catalan, thus indicating that the tongue dorsum is more resistant to changes in the articulatory configuration of the adjacent vowels during the production of [ɫ] vs. [l]. According to the figure, differences in F_2 between the two consonantal realizations increase as the number of adjacent high front vowels increases in the progression [iCi]>[iCa], [aCi]>[aCa]. This finding argues for different coarticulatory strategies during the production of adjacent [i] and [l] (Spanish) vs. adjacent [i] and [ɫ] (Catalan). On the one hand, tongue-dorsum activity for [l] is largely overridden by the tongue-dorsum fronting and raising gesture for [i], as suggested by the presence of a high F_2 (between 2000 and 2500 Hz) during closure in the sequence [ili]; on the other hand, the tongue-dorsum backing and lowering gesture for [ɫ] overrides the tongue-dorsum fronting and raising gesture for [i], as suggested by the presence of a low F_2 (about 1300-1500 Hz) during closure in the sequence [iɫi].

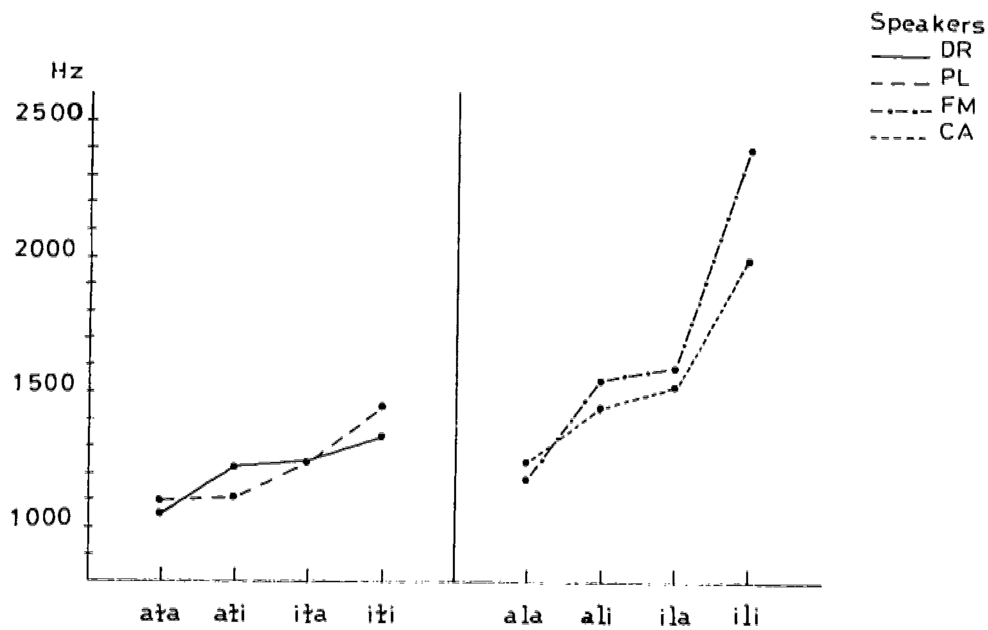


Figure 1. F_2 data at the midpoint of the closure period of [l] (right) and [ɫ] (left) in the vocalic environments [iCi], [iCa], [aCi] and [aCa]. Data are displayed for the Catalan speakers DR and PL ([ɫ]), and for the Spanish speakers FM and CA ([l]).

In summary, differences in the degree of vowel-dependent F_2 variability during consonantal closure (for [l]>[ɫ]) are inversely related to differences in the degree of tongue-dorsum constriction (for [ɫ]>[l]). These data suggest that [l] is more sensitive than [ɫ] to coarticulatory effects from the vocalic environment because the tongue-dorsum is less constrained to perform the velarization gesture.

1.2 <u>Coarticulatory effects over time</u>. Pairs of sequences were lined up for all eleven points in time (see Method section) to study V-to-V anticipatory and carryover effects. Anticipatory effects for the sequence pairs [iCi]-[iCa] and [aCa]-[aCi] were measured at points (1) through (8); carryover effects for the sequence pairs [iCi]-[aCi] and [aCa]-[iCa] were measured at points (4) through (11). Coarticulation was considered to occur when an observable difference between [i] and [a] in F_2 frequency caused an analogous difference to occur during the production of the consonant and the transconsonantal vowel. Coarticulatory effects in F_2 frequency at all intermediate points in time between (1) and (8) (anticipatory effects), and between (4) and (11) (carryover effects) were submitted to a t-test procedure; only significant effects at the p <0.01 level of significance were chosen for data interpretation.

Graph bars in Figure 2 show significant coarticulatory effects over time for [l] and [ɫ]. Anticipatory effects are plotted on each bar above the horizontal line for temporal frames 1 through 8, and carryover effects are plotted on each bar below the horizontal line for temporal frames 4 through 11; effects are displayed separately for consonants [l] and [ɫ], fixed vowels

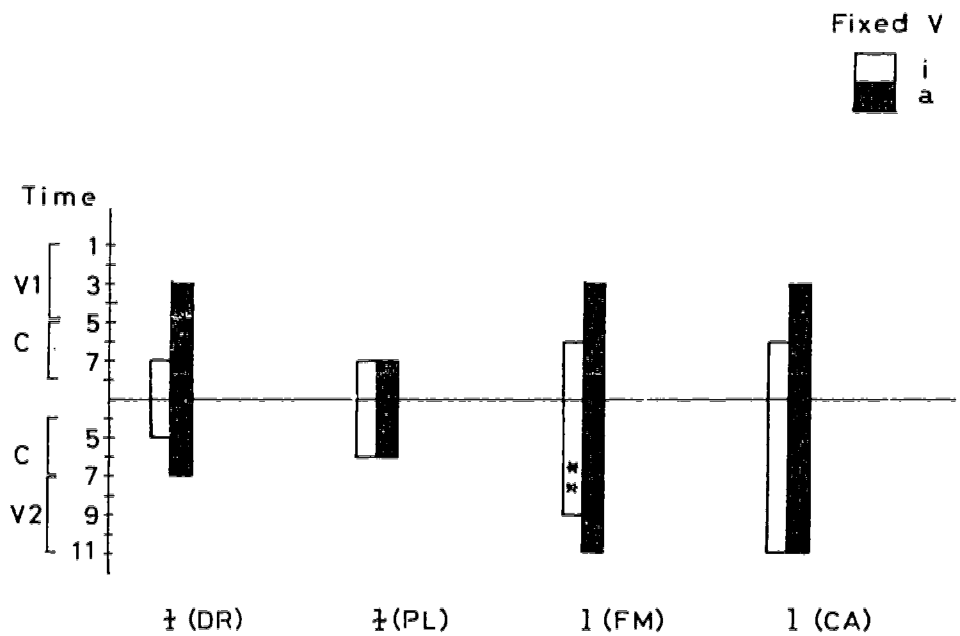


Figure 2. Significant V-to-V coarticulatory effects in F_2 frequency from [i] vs. [a] along the closure period of [l] and [ɫ], and the transconsonantal vowels [i] and [a]. Anticipatory effects have been plotted above the horizontal line along V1C (points in time 1 through 8); carryover effects have been plotted below the line along CV2 (points in time 4 through 11). Data are displayed separately for the consonants [l] and [ɫ], the fixed vowels [i] and [a], and the four speakers DR, PL, FM, and CA. Asterisks have been placed at intermediate temporal frames showing nonsignificant V-to-V coarticulatory effects.

[i] and [a], and different speakers. Let us consider, for example, the data for speaker DR. The onset of the V2-dependent anticipatory effects (above the line) for [iɫi] vs. [iɫa] occurs about the offset of closure (point in time 7); on the other hand, the onset of V2-dependent anticipatory effects for [aɫi] vs. [aɫa] occurs at V1 midpoint (point in time 3). At the carryover level (below the line), the offset of the V1-dependent carryover effects occurs later when V2=[a] (about offset of closure; point in time 7) than when V2=[i] (about onset of closure; point in time 5).

In general, significant V-to-V effects occurred continuously in time from point (8) back to onset of anticipatory coarticulation, and from point (4) until offset of carryover coarticulation. Occasionally, nonsignificant effects were found at intermediate time frames. Thus, data for speaker FM in the context [VCi] show significant carryover effects at frames 4 to 6 and at frame 9, but nonsignificant carryover effects at the intermediate frames 7 and 8. These two intermediate points in time showing nonsignificant V-to-V effects are indicated with asterisks in Figure 2 (see also Figure 6).

Figure 2 shows larger V-to-V effects for Spanish [l] than for Catalan [ɫ]. Carryover effects are consistently larger for [l] (speakers FM and CA) than for [ɫ] (speakers DR and PL); thus, while carryover effects for [l] usually last until V2 offset, those for [ɫ] do not extend into V2. Anticipatory effects, on the other hand, are usually somewhat larger for [l] (Spanish speakers) than for [ɫ] (Catalan speakers), but can show the same onset time for the two consonantal realizations (speakers DR, FM and CA; context [aCV]). Overall, [l] allows larger V-to-V effects than [ɫ], much more so at the carryover level than at the anticipatory level.

Larger significant V-to-V effects occur when the fixed vowel is [a] than when the fixed vowel is [i] for the two Spanish speakers and for the Catalan speaker DR; the Catalan speaker PL shows the same degree of coarticulation for the two fixed vowels. For those three speakers, anticipatory effects show an earlier onset when V1=[a] (at V1 midpoint) than when V1=[i] (during closure). Carryover effects, on the other hand, may show a later offset time when V2=[a] than when V2=[i] (speakers DR and FM) or the same offset time for the two fixed V2 (speakers CA and PL). Thus, overall, fixed [a] allows larger V-to-V effects than fixed [i], more so at the anticipatory level than at the carryover level.

Differences in magnitude between anticipatory and carryover effects appear to be mainly dependent on the degree of articulatory constraint for the intervocalic consonant. Thus, while "clear" [l] allows larger carryover than anticipatory effects, "dark" [ɫ] shows only a slight contrast between the extent in time of anticipatory vs. carryover coarticulatory trends.

In summary, the degree of V-to-V coarticulation appears to be inversely correlated, as for V-to-C effects during closure, with the degree of tongue-dorsum constraint for the consonant; coarticulatory differences between [l] and [ɫ] occur consistently at the carryover level but much less so at the anticipatory level. Also, carryover effects are manifestly larger than anticipatory effects for [l] but not for [ɫ]. Consistent differences in the temporal extent of V-to-V coarticulation occur for fixed [a] vs. fixed [i]; thus, fixed [a] allows larger V-to-V effects than fixed [i], more so at the anticipatory level than at the carryover level.

2. Consonants [ɾ] and [r]

2.1 Coarticulatory effects during closure. Figure 3 shows F_2 data at the midpoint of the closure period of intervocalic [ɾ] and [r] for each speaker. As for [l] and [ɫ], differences in F_2 frequency are plotted separately for the environments [iCi] (1), [iCa] (2), [aCi] (3) and [aCa] (4).

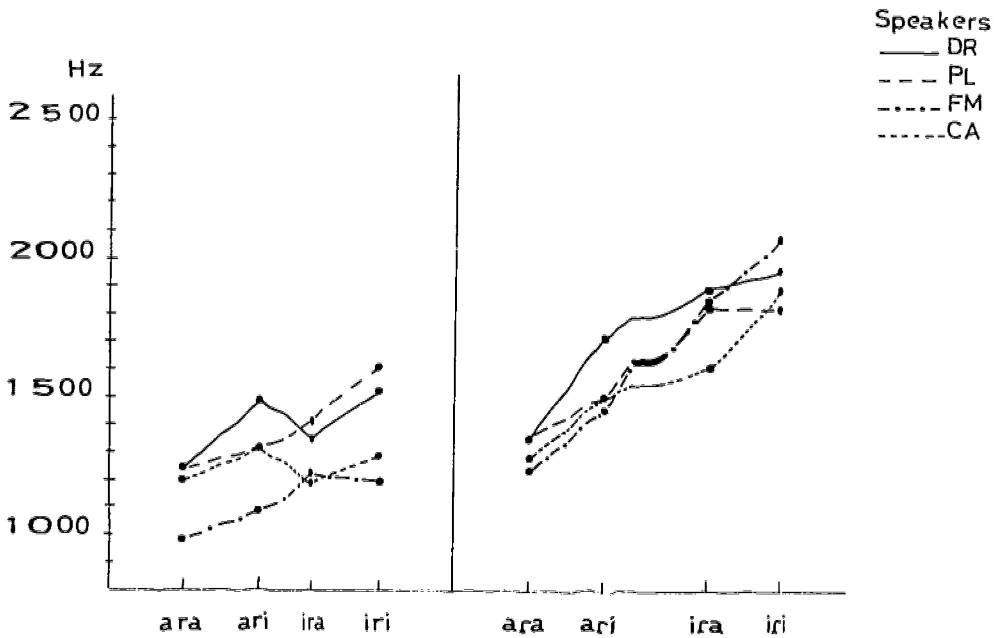


Figure 3. F_2 data at the midpoint of the closure period of [ɾ] (right) and [r] (left) in the vocalic environments [iCi], [iCa], [aCi], and [aCa]. Data are displayed separately for speakers DR, PL, FM, and CA.

The figure shows a lower F_2 for [r] than for [ɾ] in all four VCV environments for all speakers. As for [ɫ] vs. [l], this F_2 contrast is associated with more tongue-dorsum backing and predorsum lowering for [r] than for [ɾ] (see Introduction).

The figure also shows a larger degree of F_2 variability for [ɾ] than for [r] for all speakers. The tongue dorsum is, thus, less resistant to changes in the articulatory configuration of the adjacent vowels during the production of [ɾ] vs. [r]. Analogously to data for [l] and [ɫ], differences in F_2 between [ɾ] and [r] increase as the number of adjacent high front vowels increases in the progression [iCi]>[iCa], [aCi]>[aCa]. This finding argues for different coarticulatory strategies during the production of adjacent [i] and [ɾ] vs. adjacent [i] and [r]. On the one hand, as for [l], tongue-dorsum activity for [ɾ] is largely overridden by the tongue-dorsum fronting and raising gesture for [i], as suggested by the presence of a high F_2 (about 2000 Hz) during closure in the sequence [iɾi]; on the other hand, as for [ɫ], the tongue-dorsum backing and lowering gesture for [r] overrides the tongue-dorsum fronting and raising gesture for [i], as suggested by the presence of a low F_2 (below 1500 Hz) during closure in the sequence [iri].

In summary, differences in the degree of vowel-dependent F_2 variability during consonantal closure (for [ɾ]>[r]) are inversely related to differences in the degree of tongue-dorsum constraint for the consonant (for [r]>[ɾ]). Thus, [ɾ] is more sensitive than [r] to coarticulatory effects from the vocalic environment in line with differences in the degree of control over tongue-dorsum activity between the two consonants. Consonants [r] and [ɾ] require contrasting degrees of tongue-dorsum constraint, which may be associated with the execution of several vibrations for the trill as opposed to only one vibration for the tap.

2.2 <u>Coarticulatory effects over time</u>. Figure 4 displays significant coarticulatory effects over time for [ɾ] and [r]. The figure shows larger significant V-to-V effects for [ɾ] than for [r] for speakers DR, FM, and CA;

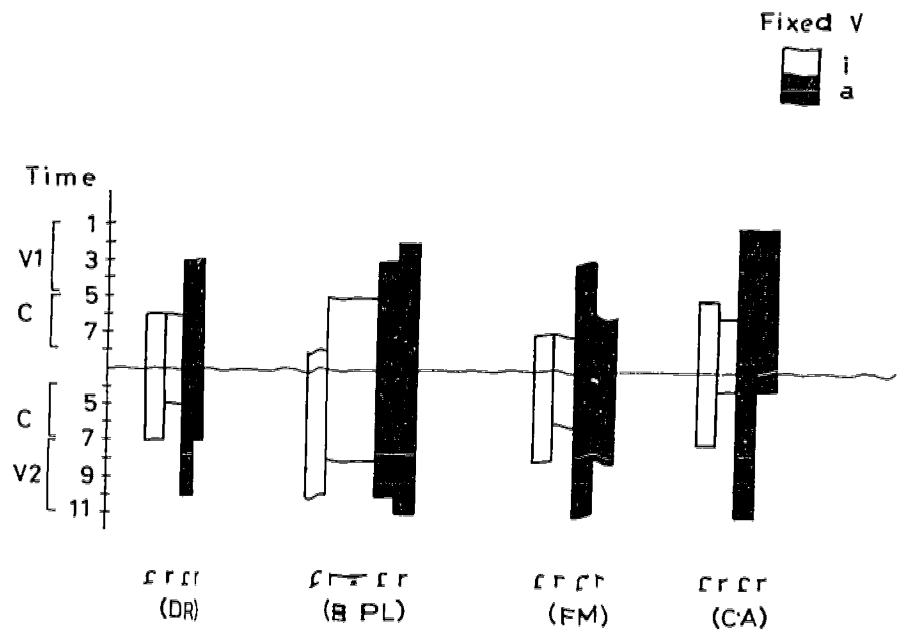


Figure 4. Significant V-to-V coarticulatory effects in F_2 frequency from [i] vs. [a] along the closure period of [ɾ] and [r], and the transconsonantal vowels [i] and [a]. Anticipatory and carryover effects are displayed analogously to those in Figure 2. Data are represented separately for speakers DR, PL, FM, and CA.

speaker PL, however, shows larger effects for [r] than for [ɾ]. Carryover effects are consistently larger for [ɾ] than for [r] for all speakers. However, differences in the extent of anticipatory coarticulation between the two consonants are highly asystematic: thus, while speakers DR and CA usually show an analogous onset time of anticipatory effects for [ɾ] and for [r], speaker FM may show an earlier onset time for [ɾ] than for [r] and speaker PL always shows an earlier onset time for [r] than for [ɾ]. Overall, [ɾ] allows larger V-to-V coarticulatory effects than [r], much more so at the carryover level than at the anticipatory level.

As for [l] and [ɫ], V-to-V effects are systematically larger for fixed [a] than for fixed [i] for all speakers. According to Figure 4, anticipatory effects show an earlier onset time when V1=[a] (about V1 midpoint or about V1 onset) than when V1=[i] (during C closure) for all speakers. Carryover effects may show a later offset time when V2=[a] than when V2=[i], or the same offset time for the two fixed V2. Thus, overall, fixed [a] allows larger V-to-V effects than fixed [i], more so at the anticipatory level than at the carryover level.

Differences in magnitude between anticipatory and carryover effects appear to be mainly dependent on the degree of articulatory constraint for the intervocalic consonant. All speakers show larger carryover effects than anticipatory effects when the intervocalic consonant is [ɾ]; as for [r], however, while speakers FM and PL show larger carryover effects than anticipatory effects, speakers DR and CA show larger anticipatory effects than carryover effects.

In summary, the degree of V-to-V coarticulation for consonants [ɾ] and [r] appears to be inversely correlated, as for V-to-C effects during closure, with the degree of tongue-dorsum constraint for the consonant; coarticulatory differences between the two consonants occur consistently at the carryover level but much less so at the anticipatory level. Also, carryover effects are manifestly larger than anticipatory effects for [ɾ] but not for [r]. Consistent differences in the temporal extent of V-to-V coarticulation occur for fixed [a] vs. fixed [i]; thus, fixed [a] allows larger V-to-V effects than fixed [i], more so at the anticipatory level than at the carryover level.

3. Consonants [β], [ð], and [γ]

3.1 Coarticulatory effects during consonantal constriction. Figure 5 shows F_2 data at the midpoint of the constriction period of intervocalic [β], [ð] and [γ] in the VCV environments [iCi] (1), [iCa] (2), [aCi] (3), and [aCa] (4). Data are displayed separately for each speaker. The figure shows a decrease in F_2 frequency in the progression [γ]>[ð]>[β], and a decrease in the degree of vowel-dependent F_2 variability in the progression [γ], [ð]>[β]. Obviously, such cross-consonantal differences in F_2 variability do not correspond to differences in the degree of tongue-dorsum constraint. Were that the case, the degree of vowel-dependent F_2 variability would decrease in the progression [β]>[ð]>[γ]; thus, little F_2 variability for [γ] ought to result from the fact that the tongue dorsum is fully involved in the constriction, and considerable F_2 variability for [β] ought to result from the fact that the tongue dorsum is left free to coarticulate with the phonetic environment. Instead, cross-consonantal differences in the degree of vowel-dependent F_2 variability reported in Figure 5 can be explained as follows.

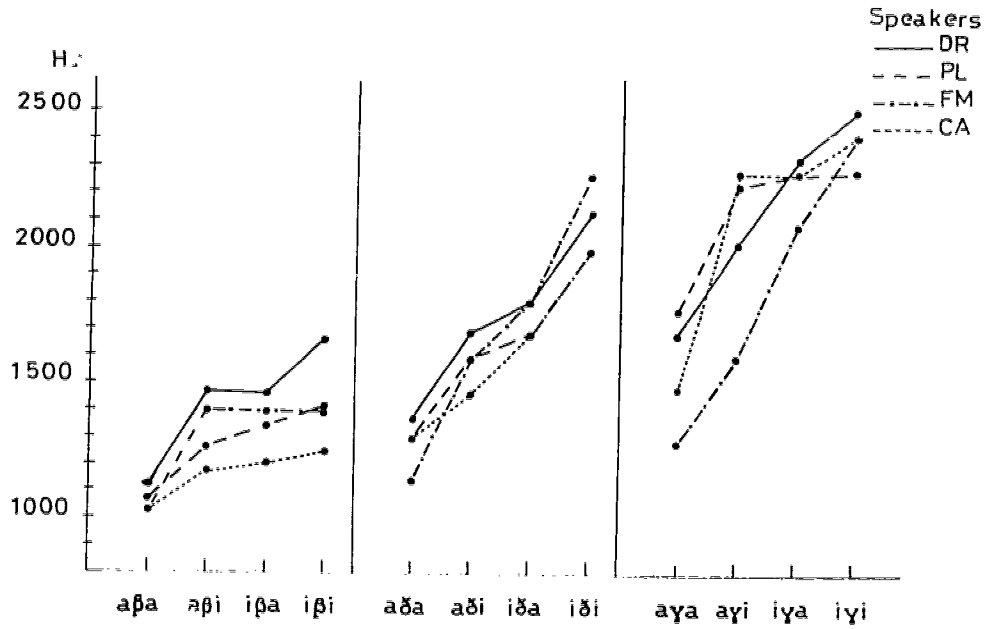


Figure 5. F_2 data at the midpoint of the constriction period of [β] (left), [ð] (center), and [γ] (right) in the vocalic environments [iCi], [iCa], [aCi], and [aCa]. Data are displayed separately for speakers DR, PL, FM, and CA.

For [ð], F_2 is dependent on the cavity behind the place of the dental constriction and, therefore, to a large extent, reflects changes in tongue-body configuration due to coarticulation with the adjacent vowels. For [γ], F_2 is particularly sensitive to changes in the place of the dorsal constriction: in the context [iCi], a high F_2 for palatovelars is inversely dependent on a small front cavity; in the context [aCa], a low F_2 for back velars is inversely dependent on a large front cavity. The articulatory differentiation between palatovelars and back velars is well documented in the literature (Catalan: Recasens, 1985; Swedish: Öhman, 1966; American English: Kent & Moll, 1972; German: Butcher & Weiher, 1976). As for [ð], changes in F_2 for [β] result from vowel-dependent coarticulatory effects on tongue-dorsum activity; however, lower F_2 variability for [β] than for [ð] is related to a highly constant lip closing gesture during the production of the bilabial consonant across vocalic environments. A smaller lip opening area for [β] than for [ð] and [γ] causes a lower F_2 frequency.

3.2 Coarticulatory effects over time. Figure 6 displays significant coarticulatory effects over time for [β], [ð] and [γ]. Asterisks have been placed at intermediate time frames showing nonsignificant V-to-V effects. Overall, significant effects decrease in the progression [ð]>[β]>[γ] for all speakers. The figure shows a clear trend for [γ] to allow shorter carryover effects than [β] and [ð]; thus, the offset time of carryover coarticulation occurs about V2 midpoint or about V2 offset for [β] and [ð], and, usually, at an earlier period in time for [γ]. On the other hand, the onset time of anticipatory effects occurs later for [γ] than for [β] and [ð]; however, such cross-consonantal differences in anticipatory coarticulation are less systematic than those observed at the carryover level. Overall, [β] and [ð] allow larger V-to-V coarticulatory effects than [γ], more so at the carryover level than at the anticipatory level.

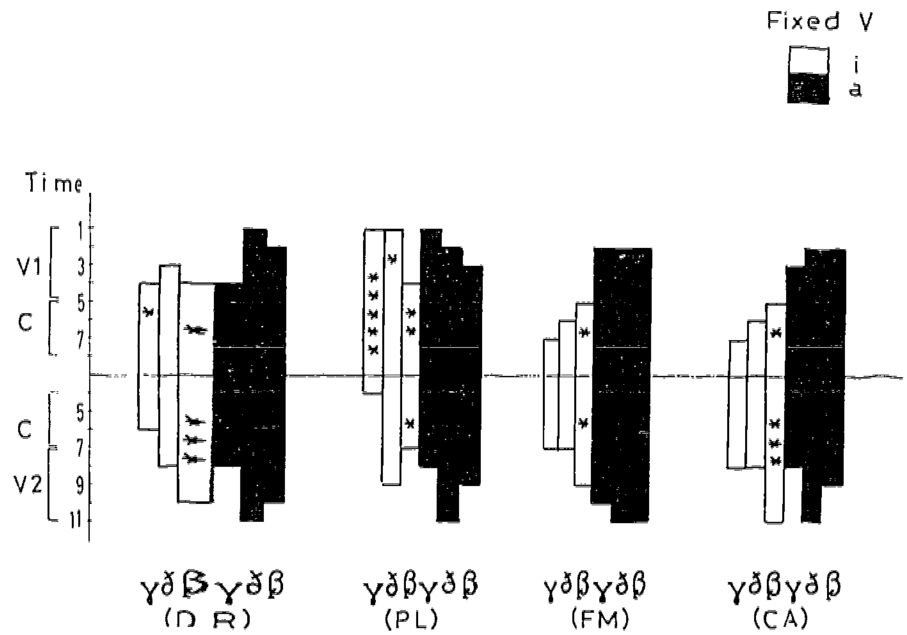


Figure 6. Significant V-to-V coarticulatory effects in F_2 frequency from [i] vs. [a] along the constriction period of [β], [ð] and [γ], and the transconsonantal vowels [i] and [a]. Anticipatory and carryover effects are displayed analogously to those in Figures 2 and 4. Data are represented separately for speakers DR, PL, FM, and CA. Asterisks have been placed at intermediate temporal frames showing nonsignificant V-to-V coarticulatory effects.

Larger V-to-V effects are systematically found for fixed [a] than for fixed [i] for all speakers. This trend occurs both at the anticipatory level and at the carryover level: while effects for fixed [i] are seldom found before V1 offset (anticipatory) and after V2 onset (carryover), effects for fixed [a] often reach V1 onset (anticipatory) and V2 offset (carryover). Thus, fixed [a] allows larger V-to-V effects than fixed [i], both at the carryover level and at the anticipatory level.

Small differences in magnitude between anticipatory and carryover effects occur as a function of the intervocalic consonant ([γ] vs. [β] and [ð]). Therefore, while anticipatory effects are usually larger than carryover effects when C=[γ], carryover effects are usually larger than anticipatory effects when C=[β] and [ð]. Also, while speakers DR and PL favor anticipatory over carryover coarticulation, speakers FM and CA show larger carryover than anticipatory effects.

In summary, [γ] is more resistant than [β] and [ð] to V-to-V effects, more so at the carryover level than at the anticipatory level. Moreover, contrary to [β] and [ð], [γ] does not favor carryover effects over anticipatory effects. Fixed vowel [a] allows larger V-to-V effects than fixed vowel [i], analogously to data reported for [l], [ɫ], [ɾ] and [r].

Summary and Conclusions

Data reported in the Results section reveal that the degree of V-to-V coarticulation in F_2 frequency varies inversely with the degree of articulatory constraint on tongue-dorsum activity for the intervocalic consonant. Thus, effects decrease inversely with the degree of tongue-dorsum constraint, for [l]>[ɫ], [ɾ]>[r] and [β], [ð]>[γ]. Moreover, V-to-C coarticulatory effects during the periods of consonantal closure and constriction have also been found to decrease for [l]>[ɫ] and [ɾ]>[r]. As suggested in the Introduction, a velarization gesture for [ɫ] (vs. [l]) and for [r] (vs. [ɾ]), and a tongue-dorsum raising gesture for [γ] (vs. [β] and [ð]), cause a high degree of resistance to V-to-V effects in F_2 frequency.

The degree of V-to-V coarticulation appears to be related to the articulatory characteristics of the fixed vowel as well. Thus, fixed [i] allows smaller effects than fixed [a] from transconsonantal [a] vs. [i]. These results indicate that [i] is more resistant that [a] to changes in tongue height and jaw opening.

This paper shows that transconsonantal anticipatory effects can extend all the way back to V1 onset and that transconsonantal carryover effects can last uninterruptedly until V2 offset. Differences in the degree of carryover vs. anticipatory coarticulation appear to be largely dependent on the nature of the articulatory gestures involved in the production of the VCV sequence. Separate trends have been found in this respect for contrasting intervocalic consonants and for contrasting fixed vowels:

(1) Differences in V-to-V coarticulation among consonants subject to different degrees of tongue-dorsum constraint are larger at the carryover level than at the anticipatory level. This finding confirms the view that V-to-V carryover effects are more dependent than V-to-V anticipatory effects on the mechanical constraints involved during the production of the intervocalic consonant. It also accords well with the fact that while unconstrained [l], [ɾ], [β], and [ð] allow larger carryover than anticipatory effects, highly constrained [ɫ], [r] and [γ] show asystematic differences between anticipatory and carryover trends.

(2) On the other hand, differences in V-to-V coarticulation between fixed [i] and [a] can be larger at the anticipatory level than at the carryover level. A closer look at differences in the temporal extent of coarticulation reveals that carryover effects for fixed V2=[i] and [a] can either last until consonantal closure or constriction, or extend into V2; on the other hand, while anticipatory effects for fixed V1=[i] usually start during consonantal closure or constriction, anticipatory effects for fixed

V1=[a] usually start at V1. Thus, while the onset of anticipatory coarticulation appears to be dependent on the articulatory characteristics of V1, the offset of carryover coarticulation is, to a large extent, independent of the articulatory nature of V2.

In summary, all these findings suggest that coarticulatory effects are deeply related to the control mechanisms involved during the production of adjacent articulatory gestures. The F_2 data presented here allow us to formulate the following model in order to explain coarticulatory effects on tongue-dorsum activity along VCV sequences.

At the carryover level, V1-dependent effects are found to vary inversely with the degree of tongue-dorsum constraint involved during the production of the CV2 sequence; thus, for example, V1-dependent effects do not extend beyond consonantal closure in [Vɫi] and [Vri] sequences, but usually extend until V2 offset in [Vla] and [Vɾa] sequences. On the one hand, after closure, little carryover V-to-V coarticulation is allowed by consonants requiring a high degree of articulatory constraint (e.g., as for [ɫ] and [r]); moreover, the temporal extent of V1-dependent transconsonantal coarticulatory effects is blocked even more if V2 is highly constrained (e.g., as for [i]). On the other hand, for CV2 syllables showing a low degree of tongue-dorsum constraint (e.g., as in the sequences [Vla] and [Vɾa]), V1-dependent effects are allowed to last until V2 offset.

Anticipatory effects appear to be temporal effects in so far as their onset is programmed to occur before the periods of consonantal closure or consonantal constriction. To a large extent, they are independent of the degree of tongue-dorsum constraint associated with the intervocalic consonant. However, the onset of V2-dependent anticipatory coarticulation is highly dependent on the degree of tongue-dorsum constraint exerted upon V1; thus, anticipatory effects have been consistently found to begin during closure when V1=[i], but at V1 when V1=[a]. In summary, the onset time of anticipatory effects for a given gesture is programmed to occur at V1 unless V1 entails conflicting articulatory requirements. In that respect, anticipatory effects in tongue-dorsum opening for V2=[a] show a late onset time if V1 requires a highly resistant tongue-dorsum raising gesture (i.e., when V1=[i]); on the other hand, however, tongue-dorsum raising for V2=[i] shows an early onset time when V1=[a], since the articulatory gesture for [a] is not resistant to tongue-dorsum raising for [i].

In addition to other findings reported in the literature (see Introduction), these data suggest that speakers use different degrees of constraint for different gestures, and that the extent to which the articulatory activity for adjacent gestures overlaps in running speech follows from those differences in degree of constraint. Thus, a theory of coarticulation that makes high predictions about the nature of coarticulatory effects ought to be based on appropriate notions about the degree of constraint required by phonemic gestures. Differences in articulatory constraint operate differently at the carryover and anticipatory levels: while carryover effects appear to be inversely related to the degree of articulatory constraint for the entire CV2 sequence, anticipatory effects are dependent on the degree of constraint for V1, but are largely independent of the degree of constraint for the intervocalic consonant.

Overall, the view according to which V-to-V coarticulation in VCV sequences is possible because adjacent consonants and vowels involve different classes of gestures (Öhman, 1966; Fowler, 1980) is far too simple. According to this view, V-to-V coarticulation occurs because vowels entail articulatory control over the positioning of the entire tongue body, while consonants involve articulatory control over the tongue articulator on which closure or constriction depend. Instead, it seems that V-to-V coarticulation proceeds according to contrasting degrees of constraint associated with gestures for adjacent phonemes; thus, for example, no carryover V-to-V effects are expected to occur for a highly constrained CV2 sequence. This view needs to be tested with further data from a good sample of different consonants and vowels, as well as different speakers and languages.

References

Badia, A. M. (1951). Gramática histórica catalana. Barcelona: Editorial Noguer.

Bladon, R. A. W., & Al-Bamerni, A. (1976). Coarticulatory resistance in English /l/. Journal of Phonetics, 4, 137-150

Butcher, A., & Weiher E. (1976). An electropalatographic investigation of coarticulation in VCV sequences. Journal of Phonetics, 4, 59-74.

Carney, P. J., & Moll, K. L. (1971). A cinefluorographic investigation of fricative consonant-vowel coarticulation. Phonetica, 23, 193-202.

Fant, G. (1960). Acoustic theory of speech production. The Hague: Mouton.

Fowler, C. (1980). Coarticulation and theories of extrinsic timing. Journal of Phonetics, 8, 113-133.

Gay, T. (1974). A cinefluorographic study of vowel production. Journal of Phonetics, 2, 255-266.

Gay, T. (1977). Articulatory movements in VCV sequences. Journal of the Acoustical Society of America, 62, 183-193.

Kent, R. D., & Moll, K. L. (1972). Cinefluorographic analyses of selected lingual consonants. Journal of Speech and Hearing Research, 15, 453-473.

Lehiste, I. (1964). Some acoustic characteristics of selected English consonants. Research Center in Anthropology, Folklore, and Linguistics, Indiana University, 34.

Lubker, J., & Gay, T. (1982). Anticipatory labial coarticulation: Experimental, biological, and linguistic variables. Journal of the Acoustical Society of America, 71, 437-448.

MacNeilage, P., & DeClerk, J. L. (1969). On the motor control of coarticulation in CVC monosyllables. Journal of the Acoustical Society of America, 45, 1217-1233.

Magen, H. (1984). Vowel-to-vowel coarticulation in English and Japanese. Paper presented at the 107th Meeting of the Acoustical Society of America.

Manuel, S. Y., & Krakow, R. A. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. Haskins Laboratories Status Report on Speech Research, SR-77/78, 69-78.

Navarro Tomás, T. (1970). Manual de pronunciación española (15th ed.). Madrid: Consejo Superior de Investigaciones Científicas.

Öhman, S. (1966). Coarticulation in VCV sequences: Spectrographic measurements. Journal of the Acoustical Society of America, 39, 151-168.

Recasens, D. (1984b). V-to-C coarticulation in Catalan VCV sequences. An articulatory and acoustical study. Journal of Phonetics, 12, 61-73.

Recasens, D. (1984b). V-to-V coarticulation in Catalan VCV sequences. Journal of the Acoustical Society of America, 76, 1624-1635.

Recasens, D. (1985). Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. Language and Speech 28, 97-114.

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. Journal of Speech and Hearing Research, 6, 111-128.

THE SOUND OF TWO HANDS CLAPPING: AN EXPLORATORY STUDY*

Bruno H. Repp

Abstract. Clapping is a little-studied human activity that may be viewed either as a form of communicative group behavior (applause) or as an individual sound-generating activity involving two "articulators"--the hands. The latter aspect was explored in this pilot study by means of acoustical analyses and perceptual experiments. Principal components analysis of 20 subjects' average clap spectra yielded several dimensions of interindividual variation that were related to observed hand configuration. This relationship emerged even more clearly in a similar analysis of a single clapper's deliberately varied productions. In perception experiments, subjects proved sensitive to spectral properties of claps: For a single clapper, at least, listeners were able to judge hand configuration with good accuracy. Besides providing some general information on individual variations in clapping, the present results support the general hypothesis that sound emanating from a natural source informs listeners about the changing states of the source mechanism.

Introduction

Clapping, the production of sound by striking the hands together, is perhaps the most common audible activity of humans that is (a) intended to be heard by others and (b) does not involve either the vocal tract or a musical instrument. It is practiced by virtually all individuals from an early age and, probably, in all cultures. Its most frequent function, at least in Western society, is to signal approval, in which case it is a rhythmic, repetitive activity maintained for at least several seconds, often collectively in a group. Given the widespread occurrence and the communicative function of clapping, it is surprising that scientific studies of this activity are difficult to find.

While research on clapping may not be of the highest priority, the topic offers a surprising variety of aspects to investigators who, prompted by curiosity, might wish to explore a little-studied human behavior. Thus

---

*Journal of the Acoustical Society of America, in press.
Acknowledgment. This research was supported by NICHD Grant HD-01994 to Haskins Laboratories. Results were reported at the 111th Meeting of the Acoustical Society of America in Cleveland, OH, May 1986. I would like to thank Cathe Browman, Leigh Lisker, Susan Nittrouer, Patrick Nye, Lawrence Rosenblum, Robert W. Young, and an anonymous reviewer for helpful comments on an earlier draft of this manuscript, Hwei-Bing Lin for assistance with the second perceptual experiment, Vin Gulisano for taking the photographs of my hands, and all my colleagues at Haskins who donated their time as subjects.

sociologists and historians might be interested in the role of clapping in different cultures and in the evolution of conventional applause in Western society (see Jenniches, 1969; Victoroff, 1959). Musicologists might want to explore the use of clapping in various kinds of folk music. Acousticians might be challenged to explain the generation of clapping sounds by applying acoustical theory. Students of motor behavior might wish to study clapping as a skill requiring precision, bimanual coordination, and auditory feedback.

For psychologists (represented by the author) two different aspects of clapping behavior seem of interest. The first, more obvious one, is the communicative function of clapping. Thus it might be asked how people convey their degree of enthusiasm for a performance, how their clapping behavior varies as a function of the stimulus and their state of mind, how a performer judges an audience's reaction from the applause, etc. While these topics are worthy of study, they are not the ones explored in the present investigation. This study, rather, pursues questions that arise when clapping is viewed as an individual articulatory activity, not unlike certain events occurring in the course of speaking.

To be sure, clapping and speaking have only few things in common. Communicative aspects of clapping may have certain parallels in paralinguistic features of speech, conveyed by parameters such as rate and loudness, which modulate the basic articulatory activity. Here we are concerned with another commonality: In both activities, the sound produced at any instant in time reflects the configuration of adjustable articulators that are part of the human body: the two hands in one case, and the various parts of the vocal tract in the other. The analogy is closest when brief transients in speech are considered, such as stop consonant release bursts or clicks, whose durations are similar to those of claps (see, e.g., Ladefoged & Traill, 1984; Repp, 1983; Fre Woldu, 1985). The dependency of sound properties on the configuration of the source mechanism follows from acoustical theory: Variations in the configuration will have systematic acoustical consequences. To the student of perception, be it of clapping or of speech, this means that the sound carries information about the momentary state of the articulators (as well as about their dynamic change, if the brief signal permits it) that can be apprehended by listeners who have (innate or acquired) knowledge of the constraints under which the source mechanism operates (cf. Gibson, 1966; Liberman & Mattingly, 1985; Neisser, 1976). Human listeners almost certainly have such knowledge available about the vocal tract and about a variety of environmental events (Jenkins, 1985); human hands should be no exception. Just as stop consonant release bursts convey information about vocal tract size (presumably) and configuration (e.g., Blumstein & Stevens, 1980), so claps may convey information about hand size and configuration.

This idea provided a useful point of departure for this preliminary investigation of the production and perception of claps. More specifically, the questions addressed were: What sorts of sounds are claps? What different ways are there of producing them? How much information do their acoustical properties contain about hand size and configuration? How sensitive are listeners to that information in the acoustic signal? Answers to these questions would not only increase our knowledge about a little-studied human activity but also would be relevant to the theoretical notion that there are general principles of perception-production relationships that extend across both speech and nonspeech domains.

Being a first exploration, the present study was fairly broad in scope but crude in some aspects of execution. The focus was on spectral properties of claps; rate and intensity (which are of much greater relevance to social communication) were considered only in passing. Analyses of clap spectra were conducted to determine how, and how consistently, information about hand size and about different hand configurations is acoustically represented. As a number of subjects were employed, the question of individual differences in clapping style necessarily entered the picture. A computer classification was conducted to explore the extent of intra- versus inter-individual variability in clap spectra, and two subsequent perceptual studies tested human listeners' ability to extract from claps information about hand size and hand configuration, respectively.

I. Production Study

A. Methods

1. Subjects

The subjects were 10 male and 10 female individuals between the ages of 25 and 45, all researchers, graduate students, or technicians at Haskins Laboratories.

2. Recording Procedure

Subjects were seated, one at a time, in a sound-insulated booth, with their hands about 60 cm from a Sennheiser microphone.[1] An Otari MX5050 tape recorder with peak indicator lights was located in an adjacent booth. Care was taken to set the recording level so that no peak distortion occurred. Each subject was asked to clap at his or her most comfortable rate, "the way you would normally clap after an average concert or theater performance," for about 10 seconds. The length and width of the subject's left hand were then measured with a ruler, from the wrist to the tip of the middle finger and across the palm above the thumb, respectively, and notes were taken on the hand configuration observed during clapping.

3. Acoustical Measurement Procedures

All recordings were digitized at a sampling rate of 10 kHz, with low-pass filtering at 4.9 kHz. From each subject's recording, a sequence of 10 consecutive claps was excerpted, starting a few claps into the series. Clap onsets were located using an automatic thresholding procedure, and onset-to-onset intervals (OOIs) were measured. The mean OOI and its standard deviation within a series provided measures of a subject's clapping speed and rhythmicity, respectively.

The FFT spectrum of each individual clap was calculated from the first 10 ms of the waveform, which generally occupied about 20 ms.[2] Subsequently, the spectra (each quantized in computer memory as a series of levels in 20-Hz bands) were averaged arithmetically over the 10 claps in a series to yield a subject's average clap spectrum. These average spectra were subjected to further analysis, as described below.

The relative amplitudes of the individual claps were estimated by the following rough procedure: A 20-ms Hamming window was moved in 10-ms steps across each subject's file of 10 digitized claps, and the maxima in the resulting series of dB values were taken to represent the clap amplitudes. Since some individuals were recorded on different days, and distance from the microphone was not precisely controlled, these amplitudes did not accurately reflect individual differences in clap intensity but merely represented the relative intensities of the claps as recorded (and as played back to the subjects in the perceptual experiments). The mean amplitude and its standard deviation within a series provided measures of a subject's recorded clapping strength and regularity, respectively.

B. Results and Discussion

1. Rate and Amplitude Measurements

Although rate and amplitude measures were not of primary interest, they are reported here for the sake of completeness and because they played a role in the perceptual experiments. The average "comfortable" rate of clapping was 4/s (mean OOI = 250 ms). Individual rates ranged from 2.7/s (OOI = 366 ms) to 5.1/s (OOI = 196 ms). There was a nonsignificant tendency for males (OOI = 265 ms) to clap slower than females (OOI = 236 ms), $t(18) = 1.59$, $p < .10$. If real, this difference could either be due to the fact that males, because of their generally larger arms and hands, have a larger mass to move in clapping, or it could represent a sex difference that is independent of size. The male subjects indeed had substantially larger hands (length x width = 162 cm^2 on the average) than the female subjects (126 cm^2), $t(18) = 6.89$, $p < .001$. The overall correlation between hand size and OOI reached significance ($r = 0.44$, $p < .05$). Computed separately for each sex, however, the correlation tended to hold up only for males ($r = 0.55$, $p < .10$), not for females ($r = 0.09$). In any case, only a small fraction of the individual differences in rate was accounted for by this factor.

Temporal variability was 6.8 ms on the average (range: 2.8 to 13.6 ms), or 2.7 percent of the mean OOI (range: about 1 to 5 percent). It should be noted that the subjects had not been instructed explicitly to clap as regularly as possible, and greater regularity could probably be achieved by most subjects under more controlled conditions. Even so, the lowest standard deviations probably are close to the maximum regularity attainable in clapping. Temporal variability showed neither any significant difference between males and females nor any relation to hand size.[3]

Clap amplitudes as recorded did not differ significantly between males and females. Amplitude standard deviations within a series ranged from 0.7 to 5.2 dB across subjects. They showed no sex difference and did not correlate with temporal variability ($r = -0.02$).

2. Spectral Analysis

The average clap spectra of the 20 subjects are shown in Figure 1. Whereas the averages are quite representative of the individual clap spectra (see section I.B.4), there is considerable variability in spectral shapes across individuals. In the figure, the spectra are arranged roughly according to visual similarity. The shapes range from a rather flat, rising type to those with a pronounced mid-frequency peak (between 2 and 3 kHz), those with

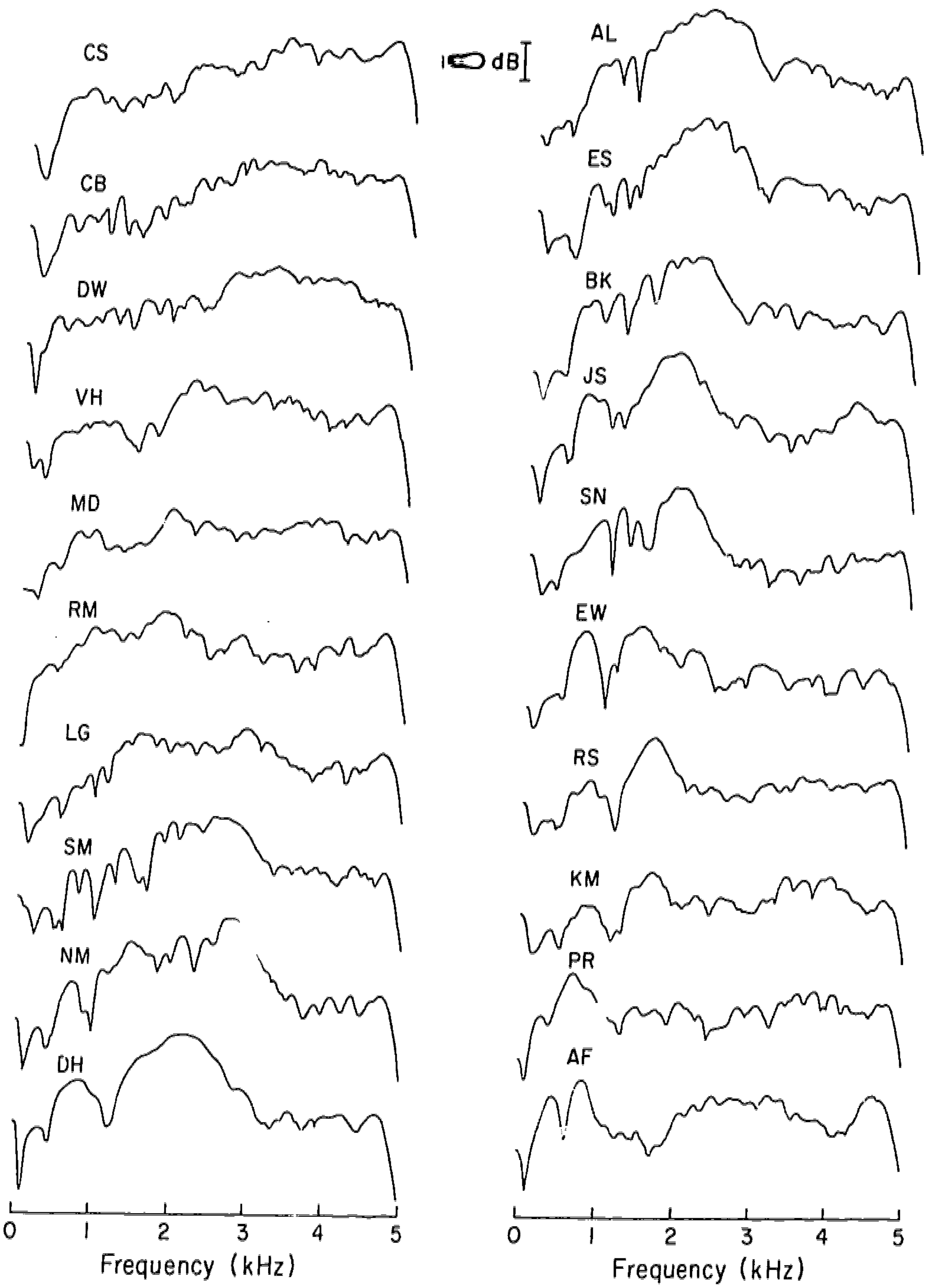


Figure 1. Average FFT spectra of claps from 20 subjects. Each spectrum is the arithmetic average of the spectra (levels in dB) of 10 individual claps, computed over a 10-ms window starting at clap onset. The spectra have been amplitude normalized and include high frequency pre-emphasis. They are arranged roughly according to visual similarity.

an emerging second peak below 1 kHz, and finally some with only this low-frequency peak.

For purposes of statistical analysis, it was desirable to quantify spectral shape in some way. A principal components factor analysis with Varimax rotation (which maximizes the variance of factor loadings for each input spectrum; see Harman, 1967) was conducted for this purpose. The input to the analysis was the set of 20 average spectra, each represented by 256 numbers (levels in 20-Hz bands). The 20 x 20 intercorrelation matrix was computed, and its linear decomposition yielded four significant factors (i.e., with eigenvalues greater than 1), which together accounted for 88 percent of the variance among subjects' clap spectra. These factors represent prototypical spectral shapes whose linear combinations (weighted by the factor loadings specific to each subject) approximate the 20 input spectra.[4]

The spectral shapes of the four factors are plotted in Figure 2, and the factor loadings of the 20 input spectra (i.e., their correlations with the factors) are listed in Table 1 in the order corresponding to Figure 1. The first factor, which accounts for 39 percent of the variance, is characterized

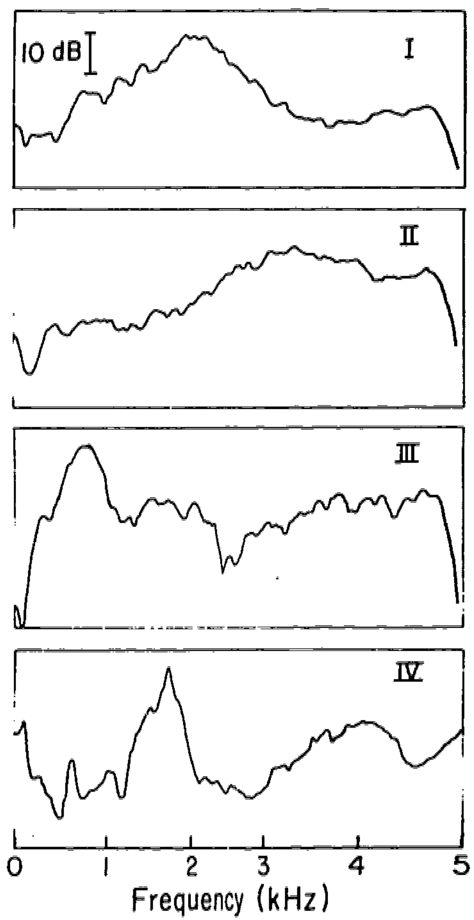


Figure 2. Spectral representation of the four principal factors, obtained by converting the (standardized) factor scores into levels (dB).

Table 1

Factor loadings (I-IV) of the 20 subjects' average clap spectra, with observed hand configuration (Hands) and listeners' hand configuration judgments (Ratings). See text for explanation.

Subject	Sex	I	II	III	IV	Hands	Ratings
CS	F	0.068	0.929	0.216	0.026	A2	2.86
CB	F	0.080	0.959	-0.020	-0.039	A2	2.82
DW	M	0.154	0.919	0.100	-0.020	A3	1.95
VH	F	0.530	0.725	0.052	-0.290	P2.5	1.95
MD	F	0.435	0.738	0.348	0.119	a2	2.36
RM	M	0.683	0.172	0.519	0.087	A2.5	1.95
LG	M	0.686	0.599	0.014	0.146	A2.5	2.09
SM	F	0.772	0.504	-0.177	-0.126	A3	2.23
NM	F	0.828	0.346	-0.074	-0.086	P2	2.73
DH	M	0.934	0.060	0.127	-0.047	A3	2.23
AL	F	0.889	0.358	0.089	0.045	P3	1.77
ES	M	0.890	0.341	-0.040	0.129	A2.5	2.45
BK	M	0.846	0.280	0.286	0.175	a3	1.91
JS	M	0.798	0.046	0.456	0.304	A3	1.09
SN	F	0.795	-0.210	0.297	0.182	A2	2.00
EW	M	0.610	-0.035	0.591	0.147	A2	1.18
RS	F	0.478	0.369	0.354	0.649	a2	2.59
KM	M	0.301	0.663	0.234	0.533	A3	2.41
PR	M	-0.001	0.294	0.903	-0.089	a1	1.18
AF	F	0.093	0.500	0.347	-0.702	A1	1.05

by a broad spectral peak in the vicinity of 2 kHz. More than half of the input spectra have substantial loadings in this factor, with subject DH being the closest match (cf. Fig. 1). The second factor, which accounts for 29 percent of the variance, represents spectral upward tilt, or strong high-frequency components without any pronounced peaks. A number of spectra have high loadings in this factor, with subject CB being the closest match. Some spectra, such as that of subject LG, represent a mixture of the two factors. The third factor, which accounts for 12 percent of the variance, represents a narrow peak below 1 kHz together with a notch around 2.5 kHz. Only one spectrum, that of subject PR, has a high loading on this factor; several others have moderate loadings. Some spectra, such as that of EW, constitute mixtures of factors one and three. Note that not all spectra with peaks below 1 kHz load on the third factor, only those without a pronounced mid-frequency peak. Finally, the fourth factor, which accounts for 8 percent of the variance, represents a narrow peak below 2 kHz and a broader peak around 4 kHz. There are no clear instances of this pattern among the input spectra, but several spectra have moderate loadings, including one (subject AF) with a negative loading (i.e., an inverted pattern). Subject RS has the most eclectic pattern, with moderate loadings in all four factors. (Note that the Varimax rotation, which aims for "simple structure," minimized the occurrence of such cases.) The individual spectrum with the smallest amount of variance accounted for by the four factors (74 percent) is that of subject EW.

The factors extracted, especially the first three, provide a useful framework for characterizing the shapes of clap spectra. In addition, they furnish numerical indices (the factor loadings) of the degree to which individual spectra resemble the factor prototypes. This quantification of spectral features permits statistical analyses to be conducted that would otherwise be impossible. Thus a multivariate analysis of variance was performed on the factor loadings to determine whether spectral shapes differed between males and females. There was no significant sex effect overall or for any of the four factors individually. This implies not only that males and females clapped similarly, but also that hand size had no important influence on the clap spectrum.

3. The Relation of Clap Spectra to Hand Configuration

The absence of a sex difference in clap spectra suggests that hand configuration, rather than hand size, is the most important determinant of the sound pattern and accounts for the individual differences observed. As a first step toward a better understanding of this variable, the author recorded himself clapping in eight different ways ("modes"), which are illustrated in Figure 3. Modes P1-P3 kept the hands parallel and flat but changed their vertical alignment from palm-to-palm (P1) to fingers-to-palm (P3), with P2 halfway between these extremes (i.e., with the right hand lowered by about 4 cm). Modes A1-A3 varied alignment in a similar way, but with the hands held at an angle. (Note that modes P1 and A1 differ in that the fingers of the two hands strike each other in P1 but not in A1. Modes P3 and A3 are more similar to each other.) Since the hands automatically tended to be more relaxed (slightly cupped) in the A modes than in the P modes, two additional versions of A1 were recorded, with the hands either very cupped (A1+) or flat (A1-), so as to examine the effect of this variable. Three parameters were thus manipulated in a semi-independent fashion: hand alignment, angle, and curvature.

All recordings were digitized, 10 consecutive claps were excerpted from each, and average spectra were calculated, which are shown in Figure 4. The spectral variation observed was somewhat smaller than expected, but nevertheless informative. Mode P1 yielded a rather flat spectrum, but a mid-frequency peak started to emerge, and low-frequency energy decreased, as the parallel hands became increasingly misaligned (modes P2 and P3). Similarly, displacement of the hands held at an angle (going from A1 to A3) led to a relative increase in mid-frequency energy and to a decrease of low-frequency energy. The palm-to-palm claps (P1, A1, A1-, A1+) all showed peaks below 1 kHz but no mid-frequency peak. Extreme cupping (A1+) or stretching of the hands (A1-) had relatively little effect on the spectrum.

These visual impressions were confirmed by entering the eight average clapping mode spectra together with the four factor shapes from the earlier analysis into another principal components analysis, in which the earlier (orthogonal) factors served as "marker variables." The factor loadings that emerged from this analysis are listed in Table 2. Again, four factors accounted for 88 percent of the variance. As can be seen from the factor loadings of the marker variables, the original factor I was second in the present analysis, the original factor III came out first, and the original factor II was third. The reason for these shifts in relative importance was the absence of very strong mid-frequency peaks (factor I) in the author's clap spectra, whereas low-frequency peaks (factor III) were very consistently

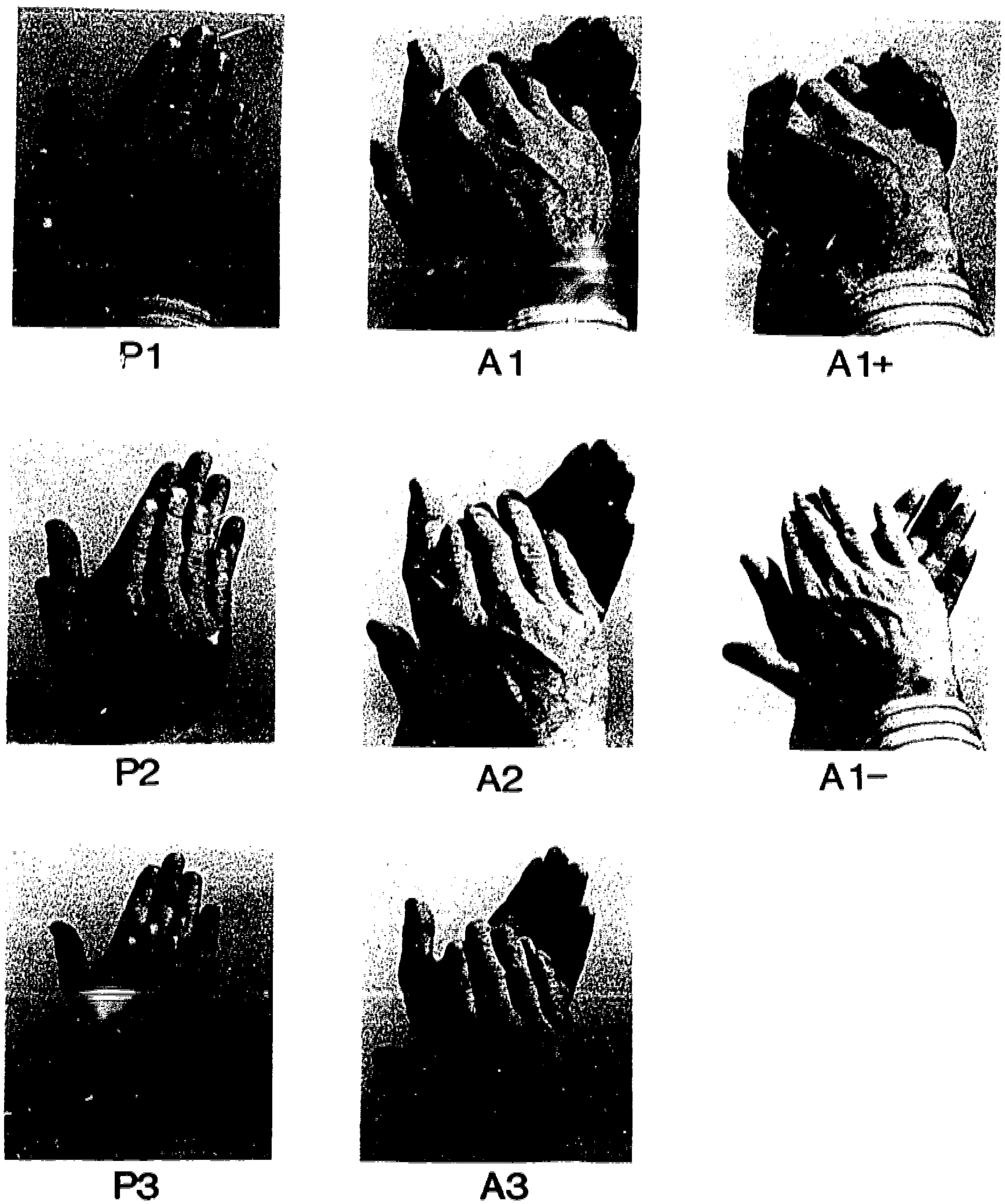


Figure 3. Eight clapping modes (see text). These still photographs were posed after the recording session.

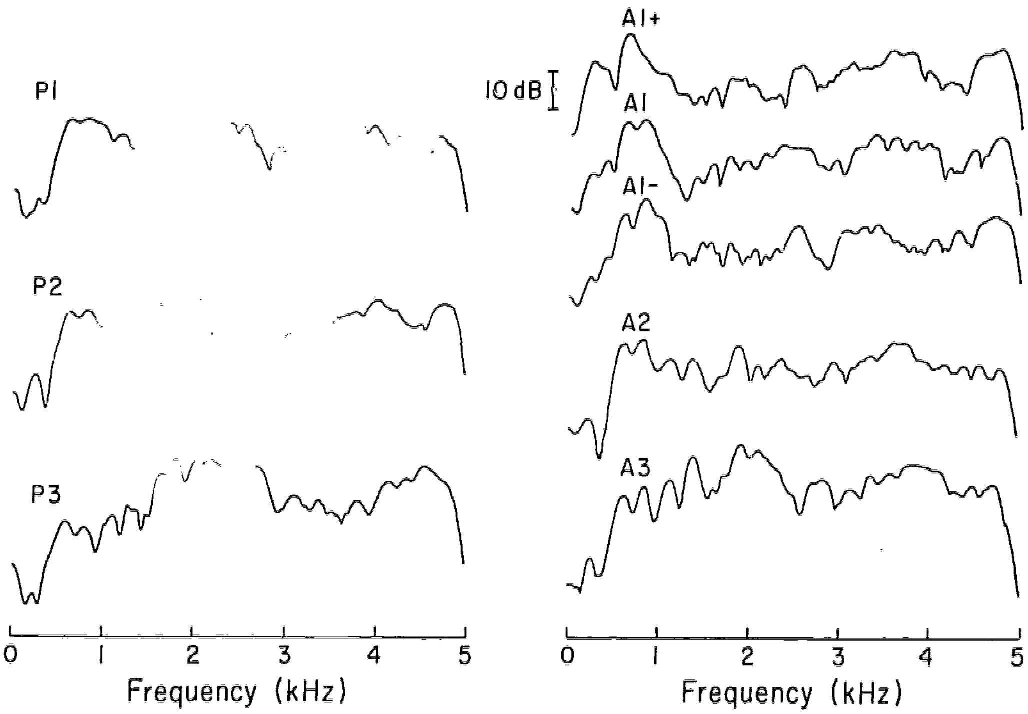


Figure 4. Average amplitude-normalized FFT spectra of the author's claps in eight different clapping modes (see Fig. 3).

Table 2

Factor loadings (I-IV) of author's clap spectra from eight different clapping modes. Factors from earlier analysis (Table 1) serve as marker variables (FI-FIV). Also shown are subjects' hand configuration judgments (Ratings).

Mode	III	I	II	IV	Ratings
P1	0.710	0.495	0.305	0.104	2.32
P2	0.475	0.614	0.164	0.553	2.36
P3	0.077	0.762	0.500	0.108	2.95
A1	0.861	0.107	0.262	-0.180	1.55
A2	0.676	0.536	0.240	0.185	1.91
A3	0.362	0.766	0.294	0.307	2.64
A1-	0.820	-0.272	0.223	-0.155	2.00
A1+	0.840	0.149	0.199	-0.085	1.00
FI	-0.121	0.939	-0.132	-0.119	
FII	0.255	0.172	0.926	0.047	
FIII	0.856	0.122	-0.317	0.208	
FIV	-0.182	0.042	0.021	0.947	

present. (The original numbering of the factors has been maintained in the table to avoid confusion.) The modes with high loadings in the low-frequency peak factor (III) were P1, A1, A1-, and A1+--those in which the two palms struck each other. Modes P2 and A2, with partial contact between the palms, had moderate loadings in this factor, and modes P3 and A3, where the palms did not touch, had the smallest loadings. These latter modes, however, had the highest loadings on the mid-frequency peak factor (I); modes P2 and A2, in which there was partial contact between the fingers of the right hand and the palm of the left hand, correlated moderately with this factor, and so did mode P1. No modes had high loadings on factors II and IV; moderate loadings were exhibited by modes P3 and P2, respectively.

This analysis leads to the conclusion that the low-frequency peak represents the palm-to-palm resonance, and the mid-frequency peak represents the fingers-to-palm resonance. The interpretation of the other two factors is less clear. The spectral upward tilt factor may simply represent a failure to achieve strong resonances due to insufficient force or lack of a sufficient seal around the hand contact areas, which is most likely to occur at intermediate hand alignments. It may also represent a fingers-to-fingers resonance.

We may now return to the 20 subjects' data and examine whether the same relation between factor loadings and hand configuration holds for them. Table 1 presents, following the factor loadings, a rough classification of the subjects' hand configurations, as observed at the time of recording. (Lower-case "a" denotes a small angle, and 2.5 a position close to 3.) The correlations between the factor loadings and the numerical hand position scores (neglecting hand angle) were, in order of magnitude: I ($r = 0.57$, $p < .01$), III ($r = -0.54$, $p < .01$), IV ($r = 0.38$, $p < .10$), II ($r = -0.01$). Thus fingers-to-palm clappers tended to show mid-frequency peaks (factor I) but not low-frequency peaks (factor III), as predicted from the analysis of the author's clapping modes. Because of other sources of variability, the relationship was less tight in this group of subjects. In a stepwise multiple regression analysis of the same data, factor I accounted for 33 percent of the variance, and factor II, though initially uncorrelated with the hand position scores, accounted for an additional 20 percent, while factors III and IV made no further contribution. Factor II thus seems to represent an aspect of hand configuration that is independent of factors I and III, whose loadings tend to be negatively correlated.

On the whole, it appears that the observed variations in hand configuration are responsible for about half of the spectral variability among individuals. The unexplained variation may derive from such factors as hand curvature and stiffness, fleshiness of the palms, tightness of the fingers, precision, and striking force, that could not be assessed accurately in this exploratory study. A more careful assessment of the roles of hand angle and finger contact also remains to be conducted.

4. Automatic Classification of Clap Spectra

The foregoing analyses were conducted on the subjects' average clap spectra. No attempt was made to assess quantitatively the amount of intra-individual spectral variation. Nevertheless, it seemed important to determine whether subjects were sufficiently consistent from one clap to the next to maintain distinctive individual characteristics. For that purpose, the correlations between the 200 individual clap spectra and the 20 average spectra were computed. Whenever an individual clap's spectrum was most highly correlated with the same subject's average spectrum, this was considered a correct identification. The computer thus simulated the "clapper identification" performance of an ideal human listener who is thoroughly familiar with each subject's characteristic way of clapping. Of the 200 claps, 181 or 90.5 percent were classified correctly in this way. No two individuals were consistently confused; the errors that occurred did not follow any particular pattern. This must be considered a remarkably high success rate, indicating that subjects maintained distinctive individual characteristics in their clapping, despite a certain amount of variability from one clap to the next, and despite often similar hand configurations across individuals. In the present sample of 20 subjects, at least, no individual made exactly the same sounds as any other.

II. Perception Studies

A. Perception of Hand Size, and Self-recognition

In contrast to the computer of the foregoing simulation, humans generally know little about each other's ways of clapping, so they cannot be expected to recognize individuals from their clapping sounds. If the following experiment was nevertheless presented to the subjects as one of individual clapper identification, it was primarily for the subjects' amusement. The primary purpose of the study was to determine whether subjects could extract some information about the clappers' sex and thus about their hand size. (The experiment was conducted before the results of the acoustical analyses became available, which suggested that there is little hand size information in the spectrum.) In addition to spectral information, the present listeners also had rate and loudness available as possible (but probably unreliable) cues to a clapper's physical size. A secondary purpose of the experiment was to find out whether listeners could recognize their own clapping.

1. Methods

Eighteen of the 20 subjects used in the production study served as listeners; two females (CS, NM) who were unavailable were replaced by Haskins colleagues of the same sex. All subjects were known to each other, with one exception (CS), who did not participate as a listener. The stimuli consisted of the 20 clapping excerpts (10 successive claps each) in random sequence, with 5 seconds of silence in between. The subjects were seated individually in a sound-insulated booth and listened to the test tape monaurally (right ear) over THD-39 headphones at a comfortable intensity. Each subject first listened to the whole stimulus sequence without responding, for purposes of familiarization. Then the tape was presented a second time, and subjects were asked to guess who had been clapping by writing down the initials of three different individuals for each excerpt, in order of confidence. An alphabetic list of the names of the 20 clappers was provided on the answer sheet. Subjects were permitted to use each name as a response as often as they liked or not at all; in fact, however, they tended to be fairly even-handed in their response choices.

2. Results and Discussion

In the analysis of the data, three points were assigned to a correct first guess, two to a correct second guess, and one to a correct third guess. Thus overall percent correct scores were calculated with respect to a possible maximum score of 60. Chance performance was at 5 percent correct.

Overall, clapper recognition was 11 percent correct with self-recognition scores excluded (13 percent correct otherwise), which is poor but significantly above chance ($t(19) = 3.74$, $p < .001$). Self-recognition, however, was much higher: 46 percent correct. That almost half of the 18 relevant subjects were able to recognize their own clapping among 20 excerpts indicates that clapping does convey stable individual characteristics, as did also the automatic classification exercise described earlier. Memory for their specific behavior during the recording session may have aided some subjects.

The question of primary interest was whether subjects were able to determine the clappers' sex and thus the size of the hands that produced the sounds. For this purpose the data were rescored in terms of "male" responses, disregarding the specific initials put down. The chance level for this score is 50 percent correct. The obtained score, with self-judgments excluded, was 54 percent correct (56 percent correct otherwise), which is barely above chance.[5] The correlation of average judged masculinity with clappers' measured hand size ($r = 0.36$, $p < .10$) fell short of significance.

The low sex recognition scores might suggest that subjects' responses were largely random. This was not the case, however. Subjects were very consistent in thinking that certain clappers were either male or female, though they were often wrong. The most striking instance was the clapping of the smallest female in the group (AF), which was judged as "male" 99 percent of the time. What variables influenced the subjects' responses?

To answer this question, the average percentages of "male" judgments for the 20 clappers were entered into a stepwise multiple regression analysis together with eight independent variables: average OOI, temporal variability, average amplitude, amplitude variability, and the factor loadings on the four spectral shape factors (I-IV). Four of these variables made a significant contribution to the regression equation and together accounted for 85 percent of the variance. OOI emerged as the most significant factor, accounting for 44 percent of the variance ($r = 0.67$). Subjects thus expected males to clap slower than females--an expectation that, however, was only weakly supported by the actual temporal measurements (hence the low accuracy of sex recognition). Second in importance, accounting for an additional 14 percent of the variance, was amplitude: Louder claps were considered more "male." (In fact, there was no such sex difference in the recordings.) The variable third in importance was factor IV, whose inclusion in the regression equation increased the variance accounted for by another 15 percent. This effect was probably due largely to AF's clapping which, it will be recalled, had the highest (negative) loading on factor IV and was overwhelmingly identified as "male." Finally, factor III added another 11 percent to the variance accounted for, indicating a tendency of subjects to consider low-frequency resonances as "male." It will be recalled that loadings in this factor did not differ between male and female clappers.

All these response trends may reflect general sex stereotypes (males: slow, loud, low-pitched; females: fast, soft, high-pitched) rather than tacit or explicit knowledge of sex differences in clapping behavior, of which there was no evidence in the present subject sample. It is conceivable, of course, that this sample was not representative, and that subjects' judgments do reflect expectations based on actual differences in clapping behavior in the population-at-large. All that can be concluded from the present data is that listeners are sensitive to a variety of physical parameters of claps, not only rate and intensity but also spectral aspects.

B. Perception of Hand Configuration

In hindsight, after the acoustical analyses revealed no effects of hand size on the clap spectrum, the poor recognition performance of the subjects in the preceding experiment is not surprising. The demonstration that subjects are sensitive to physical parameters of claps, however, leads to the question of whether subjects can judge hand configuration from the sound of claps, since that variable was a major determinant of the spectrum. A second perception experiment was conducted for this purpose.

1. Methods

Twenty-two new subjects participated in this study, partially Yale student volunteers (for whom the brief test was tacked on to the end of a paid experimental session) and partially Haskins researchers who were unfamiliar with the previous clapping experiments. The same stimulus sequence as in the preceding experiment was used. In addition, however, the author's eight clapping mode excerpts were recorded in two different randomizations. The first of these served as familiarization, without any responses being required. For the second randomization, and for each of the following 20 excerpts, the subjects judged which hand configuration was used by choosing from the numbers "1,2,3." The three configurations corresponding to these judgments were illustrated by photographs of the author's hands in modes A1, A2, and A3, respectively (see Figure 3), which remained visible to the subjects throughout the experiment. The subjects were told that the first 8 excerpts represented a single person clapping in different ways, whereas the following 20 excerpts derived from different people, each clapping in his or her most comfortable way. The instructions also mentioned specifically that hand configuration affects the sound of claps, but not in which way.

2. Results and Discussion

The data were reduced by computing the average rating of each excerpt by the 22 subjects. An average score of 1.0 thus means that all subjects judged these claps as having been produced in a palm-to-palm position, a score of 3.0 means complete agreement on a fingers-to-palm position, and intermediate scores represent either agreement on an intermediate position or various amounts of disagreement among subjects. In fact, subjects' judgments were quite systematic, and while there was some variability, no excerpt received a bimodal response distribution (i.e., more "1" and "3" judgments than "2" judgments).

Let us consider first the responses to the author's eight clapping modes. The average ratings are shown in the last column of Table 2. It is evident that the subjects were able to recognize the different hand configurations.

They seemed more accurate with modes A1-A3 than with modes P1-P3, which all sounded more like fingers-to-palm to them, perhaps because of the greater flatness of the hands and the added finger contact in P1 and P2. (This may also have been an artifact of illustrating the hand positions with photographs of modes A1-A3.) Subjects were also able to distinguish the three versions of the A1 mode, despite the relatively small spectral differences among them (see Figure 4), by translating degree of cupping into hand configuration estimates.

Analysis of variance confirmed these impressions. In one repeated-measures analysis, modes A1+ and A1- were omitted, and hand angle and position were the two crossed factors. There were highly significant main effects for both angle, $F(1,21) = 43.35$, $p < .0001$, and position, $F(2,42) = 22.94$, $p < .0001$, but no significant interaction, $F(2,42) = 1.59$, $p = .2158$. Thus, although it seemed that subjects were better at distinguishing hand configurations when the hands were held at an angle, this tendency was not reliable. In a second analysis, the three degrees of hand cupping for the A1 mode (A1+, A1, A1-) were compared. The main effect of this variable was highly significant also, $F(2,42) = 22.48$, $p < .0001$.

The average hand position ratings were entered into a stepwise multiple regression analysis, with the loadings in the four spectral factors (Table 2) as independent variables. Factor III alone accounted for 73 percent of the variance in subjects' judgments, with high factor loadings corresponding to low (palm-to-palm) hand configuration ratings ($r = -0.85$). None of the other three factors made a significant additional contribution, even though the loadings in each of them correlated positively with subjects' ratings. The principal determinant of subjects' judgments, then, seemed to be the presence and extent of low-frequency peaks in the spectrum.

For the ratings of the 20 subjects' excerpts a similar regression analysis was conducted, with OOI, temporal variability, amplitude, and amplitude variability as additional independent variables.[6] Although these variables were not considered relevant to the judgment of hand position, they were included because of their perceptual salience, and also to make the analysis comparable to that conducted earlier on the masculinity scores. Three variables made a significant contribution, explaining 72 percent of the variance. Surprisingly, OOI came out first, explaining 44 per-cent of the variance (longer OOIs, or slower rates, leading to more palm-to-palm judgments); factor III accounted for a further 16 percent, and factor IV for another 12 percent. These results resemble those of the immediately preceding analysis in that they reveal a significant influence of the low-frequency peak factor (III) on subjects' judgments. However, they also resemble the results of the analysis of the masculinity scores, with the main difference being the total absence of any correlation of hand position ratings with amplitude.

The last-mentioned similarities raise the questions of whether masculinity and hand configuration judgments were related, and whether OOI had any true relation to hand configuration. Indeed, the correlation between the two types of judgments was high ($r = -0.82$, $p < .001$), which confirms that the listeners (different groups in the two tests) relied largely on the same acoustical information in judging sex (hand size) and hand configuration. The spectral information did reflect hand configuration, at least in part, whereas it had no obvious relation to hand size. Average OOI, however, was not related to either clappers' sex or hand size. Actual hand configuration (derived from the "Hands" column of Table 1) was likewise uncorrelated with

OOI (r = 0.05). Unfortunately, actual hand configuration was also uncorrelated (r = 0.19) with judged hand configuration. Therefore, it is not clear whether the listeners were really able to perceive or infer what the 20 clappers did with their hands. The large variations in the irrelevant rate parameter may have diverted subjects' attention from the relevant spectral properties. Subjects' success in the preceding test based on the author's clapping modes suggests that they would perform more accurately if irrelevant variation were reduced.

III. General Discussion

A. Methodological Shortcomings

The present study was a first exploration of a hitherto little-studied subject, and it was conducted under time constraints. As such, it suffers from a number of methodological weaknesses that need to be improved upon in a more thorough follow-up study. These weaknesses shall be acknowledged before proceeding to the conclusions.

First, the recording procedure was far from optimal. Future studies will have to avoid reverberation by using a sufficiently large or anechoic chamber, and distance from the microphone will have to be controlled more carefully. The data, however, provide no indications of serious artifacts due to these factors.

Second, the spectral analysis was based on low-pass filtered signals. (See also Note 2.) Future analyses may reveal that there is additional spectral information in frequencies above 5 kHz.

Third, the registration of subjects' hand configuration was casual and possibly inaccurate (except for the author's own clapping modes). More precise ways will have to be found for recording hand position (as well as angle, degree of cupping, etc.) by means of measurements *in situ*, from still photographs, or from video tapes.

Fourth, by asking a number of subjects to clap in their most comfortable ways, differences in hand configuration were confounded with a variety of other individual differences. It would be desirable to separate these aspects in a future study by asking each individual to clap in different, precisely specified "modes," as was done here with a single subject (the author).

Finally, subjects' ability to infer hand configuration from the sound of claps was probably impaired by the presence of irrelevant but salient variations in rate and loudness, as well as by the elimination of the higher frequencies in the spectrum. To test subjects' full ability, it would be desirable to present high-quality recordings in which rate and loudness variations are neutralized.

B. Conclusions and Further Questions

With these caveats, then, what conclusions can be drawn from this pilot study, and what questions do they raise or perhaps even answer?

First, it is evident that different individuals clap in different ways. This simple fact raises interesting questions about the origin of these individual differences--questions that the present study could not even begin to address, but that are worth listing here: To what extent are individual differences in clapping anatomically conditioned, and to what extent to they represent learned behavior patterns? If an individual's preferred hand configuration, in particular, is learned, when and how did this learning take place? How consistently do individuals employ a particular way of clapping, and to what extent do they vary their behavior across different situations? The assumption here has been that situational factors lead primarily to adjustments in clapping rate and loudness--parameters that are relevant to the ordinary communicative function of applause--but not to changes in characteristic hand configuration. There may be some people, however, who do vary their hand configuration systematically or randomly, so that they could not be said to have a characteristic way of clapping at all. It is also possible that adjustments in hand position are contingent on large changes in rate (see Note 3) and loudness.

Second, apart from variations in rate and loudness, which are of secondary interest here, different individuals produce different clapping sounds. A considerable part of that spectral variability appears to be due to differences in hand configuration. Other factors must contribute to the spectral shapes, however, or else it would not have been possible to classify over 90 percent of individual clap spectra correctly by computer. What these factors are is not clear at present. The success of the computer classification analysis suggests that individuals may have a "clap signature"--a characteristic spectrum that distinguishes them from many other individuals. To support this suggestion, however, it will be necessary to assess intra-individual variability over a wider range than merely a train of 10 consecutive claps, and also to eliminate possible artifactual contributions from variations in recording conditions.

Third, no sex differences in clapping were evident in the present group of subjects. While sex differences as such were not of particular interest here, the finding does contradict popular opinion that "ladies clap differently from gentlemen." The present subjects, all Ph.D.'s or graduate students, did not seem to fit these traditional categories of social demeanor. It remains to be seen whether a sample drawn from the population-at-large will show the differences that are often attributed to the sexes, and/or whether such differences emerge only in real-life situations. More to the point of the present study, however, it appears that hand size--which exhibits clear sexual dimorphism--does not have any influence on the sound of claps. This is an unexpected finding.

Fourth, the spectral differences among claps, as well as their rate and loudness, were readily discriminated by listeners and were systematically related to their judgments of clappers' presumable sex and hand configuration. The most salient parameter was rate: Slower rates were considered to represent a male clapper and a palm-to-palm hand position, even though rate was in fact unrelated to both sex and hand configuration in the present sample of subjects. Thus the listeners relied on expectations or stereotypes that linked these variables. Spectral properties of claps, which were correlated with actual hand configurations, also contributed to listeners' judgments. In the case of a single clapper (the author), it was quite clear that subjects were able to determine hand configuration from the sound of claps. In the case of the more heterogeneous sample of 20 clappers, the evidence was not conclusive.

C. Theoretical and Practical Issues

At the theoretical level, the results of the present study give some support to the hypothesis that sound emanating from a natural source, particularly one involving parts of the human body, conveys perceptible information about the configuration of that source. The prime example of the principle embodied in this hypothesis is speech, whose sounds convey the changing states of the vocal tract. In the case of listening to continuous speech, there is little awareness of the pure sound qualities (the proximal stimulus), and perception is focused on the distal events. It has been argued that the distal speech events are perceived directly, without mediation by an auditory representation of the input (Fowler, 1986; Liberman & Mattingly, 1985). This argument is less convincing, however, when applied to the common laboratory situation of individual speech sounds (e.g., fricative noises or stop consonant release bursts) that are removed from their context and presented in isolation (e.g., Blumstein & Stevens, 1980; Repp, 1981). Listeners then do perceive characteristic auditory qualities as well as the articulatory information behind them, so the former could, in principle, be used to infer the latter. Listening to claps is like listening to isolated stop release bursts in that auditory, pitch-like qualities are perceived together with, presumably, the "place of articulation" on the clapper's hands. It is a moot point whether listeners arrive at judgments of hand configuration from claps directly, as it were, or _via_ an inferential process based on perceived sound quality. Actually, this question becomes unnecessary if perception itself is viewed as involving unconscious inference (Rock, 1983). It seems plausible to assume that perception of isolated speech sounds differs from clap perception only in the availability of well-established phonetic categories to classify speech stimuli. The perceiver's tacit knowledge of the constraints under which parts of the body operate, and the consequent possibility of deriving articulatory information even from static spectral properties (cf. Stevens & Blumstein, 1981), may be similar in the two cases. Of course, when it comes to longer stretches of (time-varying) speech, the information to be perceived becomes much more complex than that in isolated sounds.

It is more difficult to say anything convincing about the practical utility of the present research. After all, it focused precisely on those parameters of clapping that presumably play no role in the communicative function of applause. Two aspects, however, may be of slight interest to the pragmatist. The possibility of an individual "clap signature," though it is in need of much stronger empirical support, may be of interest to those concerned with automatic recognition of individuals from acoustic signals. Devices are on the market now that are said to respond to claps, and it might be suggested that they could be tuned to respond selectively to different individuals or to different hand configurations of the same individual. Another possible application of knowledge gained from a study of clapping might be in music performance. The hands might be considered as a percussion instrument with the capability of producing two or more timbres, and while this is not an impressive range, the instrument is cheap, portable, easy to maintain, and readily mastered. Apart from the universal use of clapping for purely rhythmic purposes, the capability of the hands to produce different timbres may in fact already have been discovered by some folk musicians.[7] If so, more detailed knowledge about the production and perception of clapping may help in analyzing such existing practices, and also may lead to their deliberate introduction into some contemporary art music as a welcome humanizing element.

References

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. Journal of the Acoustical Society of America, 67, 648-662.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics, 14, 3-28.

Fre Woldu, K. (1985). The perception and production of Tigrinya stops. RUUL 13. Uppsala, Sweden: Uppsala University, Department of Linguistics.

Gibson, J. J. (1966). The senses considered as perceptual systems. Boston, MA: Houghton Mifflin.

Harman, H. H. (1967). Modern factor analysis. Chicago, IL: University of Chicago Press.

Jenkins, J. J. (1985). Acoustic information for objects, places, and events. In W. H. Warren & R. E. Shaw (Eds.), Persistence and change. Proceedings of the First International Conference on Event Perception. Hillsdale, NJ: Erlbaum.

Jenniches, K. M. (1969). Der Beifall als Kommunikationsmuster im Theater. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 21, 569-584.

Ladefoged, P., & Traill, A. (1984). Linguistic phonetic descriptions of clicks. Language, 60, 1-20.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21, 1-36.

Neisser, U. (1976). Cognition and reality. San Francisco, CA: Freeman.

Repp, B. H. (1981). Two strategies in fricative discrimination. Perception & Psychophysics, 30, 217-227.

Repp, B. H. (1983). Coarticulation in sequences of two nonhomorganic stop consonants: Perceptual and acoustic evidence. Journal of the Acoustical Society of America, 74, 420-427.

Rock, I. (1983). The logic of perception. Cambridge, MA: MIT Press.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), Perspectives in the study of speech. Hillsdale, NJ: Erlbaum.

Victoroff, D. (1959). El aplauso, una conducta social. Revista Mexicana de Sociologia, 21, 703-739.

Zahorian, S. A., & Rothenberg, M. (1981). Principal-components analysis for low-redundancy encoding of speech spectra. Journal of the Acoustical Society of America, 69, 832-845.

Footnotes

[1]The recording environment and procedure were not optimal but were deemed adequate for this pilot study. Distance from the microphone was not controlled precisely, and some reverberation was present.

[2]A short window was used to exclude reverberation as much as possible. The FDI program of the ILS package (Version 4.0, Signal Technology Inc.) was used to compute the spectrum. This program employs a fixed window of 25.6 ms duration and fills the unused portion with silence. The program also uses a Hamming window by default, which was maintained (unnecessarily) in the present

analyses. Reanalysis of several claps without the Hamming window and/or using a window of longer duration revealed only minimal changes in the spectrum.

[3]A temporal analysis was also conducted of each subject clapping as fast as possible. The clapping rates achieved under these instructions ranged from 5.4/s (OOI = 184 ms) to 8.1/s (OOI = 123 ms), with an average of 6.6/s (OOI = 152 ms). Although the instructions requested that the hand configuration remain the same, many subjects stiffened their hands and reduced hand excursion to an extent that would rarely be encountered in natural applause. There was no difference between the fast clapping rates of males (OOI = 149 ms) and females (OOI = 154 ms), nor was there any relation to hand size (*r* = -0.28), even though limitations imposed by the mass of the limbs might have been expected to be revealed more clearly in this extreme situation. The average variability of fast clapping was 6.2 ms (4.1 percent), with no sex difference and no significant correlation with OOI (*r* = 0.23). The correlations between normal and fast clapping rates (*r* = 0.35) and between variability measures at normal and fast rates (*r* = 0.30) were nonsignificant.

[4]It should be noted that this analysis differs from the type of principal components analysis commonly conducted on speech spectra (e.g., Zahorian & Rothenberg, 1981), in which the correlations are computed for all pairs of frequency bands across a number of different spectra. (In the present case, this would have resulted in a 256 x 256 intercorrelation matrix.) The factors emerging from such an analysis represent spectral components such as formant peaks, whereas the present factors represent full spectra that instantiate the types of spectral shapes observed for a group of subjects. In other words, the more common analysis is meant to uncover dimensions underlying spectral shape, whereas the present analysis was employed primarily as a data reduction procedure.

[5]A significance test of the difference from chance becomes meaningless in view of the enormous variation of scores (from 1 to 97 percent correct) across stimuli, to be discussed below.

[6]These variables were not analyzed for the author's clapping modes. Although subjects had them available in that test also, their range of variation was much more restricted.

[7]The author has not yet come across any relevant recordings or literature and would welcome pertinent information, also about any other literature on clapping that may exist.

AN AEROACOUSTICS APPROACH TO PHONATION: SOME EXPERIMENTAL AND THEORETICAL OBSERVATIONS*

Richard S. McGowan

Abstract. We examine the sources of sound during phonation using an aeroacoustic formulation. Some sources of sound during phonation are found to have a dipole character. The most important of these in the low frequency limit is the result of vorticity-velocity interaction. This picture is in contrast to the usual picture of the voice source as a monopole that can be modeled as a piston in a tube. A fluid mechanical approach to voice modeling is promoted in this note.

Introduction

The characterization of the voice source in terms of fluid dynamic variables is a part of the subject called aeroacoustics. Aeroacoustic theories are formulated from the conservation equations of fluid mechanics, so that such questions as the amount of total fluid energy that is converted into the fluid energy of acoustic motion during phonation can be answered. To date, no satisfactory partition of energy has been proposed, as noted by Teager and Teager (1983) and Kaiser (1983). They have suggested a fluid mechanical approach to phonation, and it is hoped that this note will contribute in that direction.

During phonation, there is modulated fluid movement in the region near the glottis. It is known that fluid motion can be decomposed into two kinds: solenoidal and irrotational, where the latter can support acoustic oscillation, and the former cannot (Batchelor, 1970). In the standard model of the voice source, the entire oscillatory field in the glottal region is treated as acoustic: the volume velocity at the glottis is the input to the one-dimensional analog circuit of the vocal tract. This picture of the voice source treats the glottal source as a piston in a tube, which can be classified as a monopole source. Here we will argue that this is not the correct model of the source of phonation.

Some of the general results from the aeroacoustics literature will be considered along with a discussion of their application to the voice source. The main results of this discussion can be summarized as follows. The solenoidal field, which contains the rotational motion of the fluid, and hence vorticity, is important for creating sound. The solenoidal field creates the

---

Acknowledgment. The author thanks Vin Gulisano for his photography, and Ed Wiley, Dick Sharkany, and Don Hailey for help with experimental hardware. Thanks goes to Professor K. R. Sreenivasan for his help with flow visualization.

necessary potential energy for acoustic motion through dynamic pressure fluctuations near the folds. This type of source can be classified as a dipole type source. As a result, the fraction of total oscillatory fluid kinetic energy that is converted into acoustic energy is small, but of course, not insignificant in acoustic terms.

Along with the theoretical discussion given the voice source, we have a few measurements to support some of the ideas presented here. The measurements are taken from an oscillatory jet that mimics the air flow of the glottal region, with two exceptions: there are no moving surfaces near the jet, and the jet exits into a nearly unbounded region. However, this situation allows us to view some important aspects of sound production, like vortex formation. Further, it allows us to observe the sound produced by vortex formation, abstracted from that produced by moving boundaries. These measurements are incidental to the main point of urging an aeroacoustics approach to vocal tract acoustics in the future.

Sound Production

The production of sound by the interaction of fluid with itself and solid surfaces falls within the field of aeroacoustics, whose modern beginnings came with the work of M. J. Lighthill in 1952. Lighthill derived a nonlinear wave equation from the equations of motion for a Newtonian fluid. He wrote it so the familiar linear wave operator on the left when applied to density is set equal to nonlinear terms on the right. The right-hand side terms are to be identified with "sources" for the acoustic propagation of the left-hand side, density perturbation. This identification is known as Lighthill's acoustic analogy. Largely motivated by engineering problems, much work has been done on understanding this and other similar equations by solving them in various geometries and by rewriting the source terms. The analogy is difficult to test directly because the source terms usually have not been measured.

The work reported here relies on the simplification of the source terms from the work of Powell (1964). For low Mach number flows without entropy spottiness, Powell shows that the dominant source of sound involves the nonlinear interaction of vorticity and velocity. In fact, the wave equation appears as:

$$\frac{\partial^2 \rho}{\partial t^2} - c_0^2 \nabla^2 \rho = \rho_0 \mathrm{div}\,(\underline{J} \wedge \underline{v}) + \rho_0 \nabla^2 |\underline{v}|^2$$

where:

$\underline{v}$ = fluid particle velocity	$\underline{J}$ = vorticity = $\nabla \wedge \underline{v}$
ρ = perturbation fluid density	ρ_0 = ambient fluid density
	c_0 = ambient speed of sound

Since vorticity is a quantity that appears in the source term, we attempt to determine whether vorticity is a part of the vocal tract flow in the next section. Later, a formal solution to this equation will be exhibited.

Vorticity

First, it is argued that vorticity occurs as part of the flow from the glottis. The geometry of the vocal tract in the region of the glottis can be idealized as that of one cylindrical pipe of relatively small diameter emptying into a cylindrical pipe of relatively large diameter. The ratio of the areas is taken as approximately ten. Air flowing from the small pipe into the larger pipe, or even into an unbounded region, forms a j[illegible] jet is a region of shear flow, which is necessary to meet the boundary conditions at the walls of the larger pipe, or at infinity in the case of an unbounded region.

We follow Batchelor (1970) in arguing that vorticity is formed in the special case of a periodically modulated jet, which occurs above the glottis. While Batchelor considers the steady case, we will make the quasi-steady assumption.

Initially, the jet is in a region where the flow from the smaller pipe is mixing with the fluid in the larger pipe. The initial mixing region extends a distance Δx in the direction of the pipe axis, after which the jet fills the entire pipe. We use V to denote particle velocity in the axial direction, p to denote pressure, and A to denote cross-sectional area. The subscript $_1$ denotes the smaller pipe, and the subscript $_2$ the larger pipe. If f is the frequency of motion under consideration, the quasi-steady assumption can be made if:

$$f \ll \frac{\text{unsteady part of } V_1}{\Delta x}$$

This assumption can be seen to be approximately valid above the glottis for $f < 1000$ Hz, the unsteady part of $V_1 = 4,000$ cm/sec and $\Delta x < 2$ cm. (From van den Berg's experiments [van den Berg, Zantema, & Doornenbal, 1957] the assumption that $\Delta x < 2$ cm appears well founded, because all the loss of pressure head appears to occur before their final transducer. From their figure 1, the distance from the jet exit to the final transducer is apparently under 2 cm.) We do not argue the validity of the common quasi-steady assumption in real voice, but use it knowing the limitations.

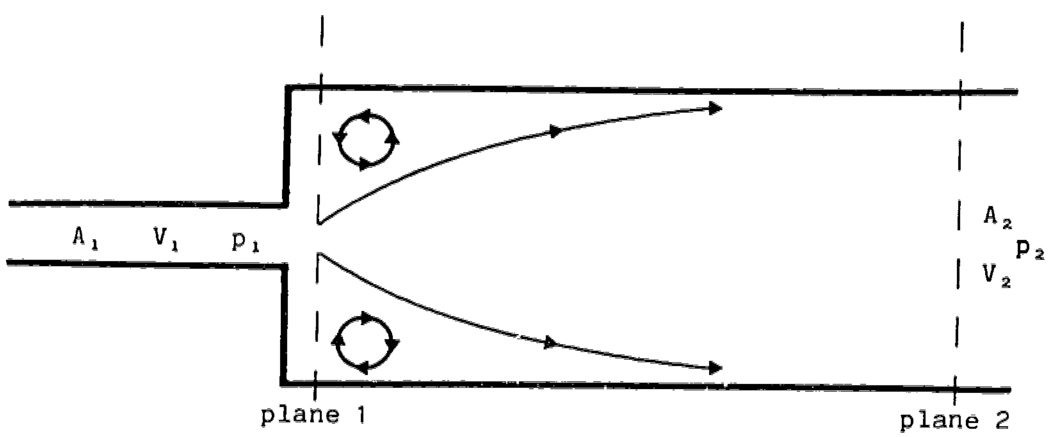


Figure 1. Jet exiting into cylinder.

The equations for mass and momentum conservation can be written in integral form, using control planes at the exit and after the mixing region (see Figure 1). Irrotational motion before and after the mixing region is assumed. Mass conservation gives:

$$V_1 A_1 = V_2 A_2$$

Momentum conservation gives:

$$p_1 A_2 + \rho_0 V_1^2 A_1 = p_2 A_2 + \rho_0 V_2^2 A_2$$

Solving for pressure:

$$p_2 = p_1 + \rho_0 V_1^2 \ (A_1/A_2)(1-A_1/A_2)$$

If we were to assume irrotational flow in the mixing region, Bernoulli's relation under the quasi-steady assumption gives:

$$p_2^* = p_1 + \rho_0 V_1^2/2 \ (1-(A_1/A_2)^2)$$

The difference is:

$$p_2^* - p_2 = +\rho_0 V_1^2/2 \ (1-A_1/A_2)^2$$

The difference in pressure implies that energy must be going into rotational motion, and then into heat, and, possibly sound.

The rotational motion discussed in the above paragraph may be in the form of a vortex ring. As the jet enters the larger tube, fluid is pulled along side of the jet by viscous action. Farther from the jet, fluid must return because of mass conservation. Taking a cross-section in plane containing the cylinders' axis, the following picture of the velocity field may be made (see Figure 1). It can be seen that because the jet has cylindrical symmetry, the rotational flow field is toroidal, that is, a vortex ring. In a real vocal tract, the vortex will no longer be toroidal, but it may be topologically equivalent to a toroid.

In the experimental situation to be described, we verified the existence of vorticity in an indirect way. Instead of the larger tube, a nearly unbounded region (a large box) was used, where the boundaries can be considered to be at infinity. If vorticity is found in our experiment, then it will be found in the case of two tubes. The effect of the cylindrical wall of the larger tube is to increase shear, and hence vorticity. So this experiment is less favorable to the formation of vorticity than the case of two tubes.

Our apparatus was primarily a flow visualization apparatus using smoke. The modulated air flow was produced using an air compressor from an electric tire pump. The volume displacement of the compressor piston was 2.57 cm^3, and it was equipped with a valve to block flow into the piston cylinder during the downstroke. This resulted in a series of pulses of air with the mean level of flow on the same order of magnitude as the peak flow. The speed of the compressor was varied using the voltage control on a D.C. power supply. The output of the compressor was fed by rubber hose to a cylindrical brass nozzle with inner diameter .9 cm. The brass nozzle exhausted air into a cardboard box lined with flat-black paper, with a slot for photography cut in the side.

Smoke was supplied using a cigarette in a plastic tube, with one end connected to a ducted fan and the other connected through an intervenous needle to the hose between the compressor and the nozzle. Lighting was provided at the floor of the box using a General Radio 1531-A Strobatac, which was wired through a General Radio 1531-P2 delay. Using an electric eye, the strobe was triggered from the drive shaft of the electric motor for synchrony.

A 35mm Konica single lens reflex camera with a close-up lens and tri-X, 400 ASA film was used to photograph this nozzle region. The image plane was approximately 9 inches from the jet. The compressor was operated at a frequency of about 20 Hz, in order to use the high intensity setting on the Strobatac. The best exposures occurred for shutter speeds of between 1/8 and 1/4 sec and an f-stop of 2.0. (We did not try lower f-stops for high intensity exposure.) The resulting photographs, one of which is shown below, showed signs of an oscillating vortex ring. There are bands of smoke perpendicular to the jet, which indicates rotational motion in a vortex ring. More detail could be seen if we had stroboscopically luminated the jet in a cross-section parallel to the jet axis, and used a higher density smoke.

Figure 2. Vorticity in an oscillating jet.

Although we did not obtain a quantitative measure of vorticity, we do have good reason to suggest that vorticity of appreciable strength may be generated near the glottis. Also, this vorticity may be modeled as a vortex ring.

General Properties of Integral Solutions

Having discussed some aspects of the acoustic source, we now exhibit a solution for the acoustic quantities in the far-field (i.e., at distances large compared to the wavelength). The integral solution to the wave equation shown above was proposed by Powell (1964). We will discuss Powell's solution applied to the vocal tract and to our experimental situation, described below. Powell assumes that the Mach number of the flow is small and the product of the Mach number with the source compactness parameter is also small. The source compactness parameter is the product of the source length scale and the typical wavenumber of the acoustic wave. Because we are considering only low frequencies, the mixing region is presumed short (i.e. $\Delta x < 2$ cm), and the Mach numbers small, Powell's assumptions appear to be valid for the vocal tract. If we draw a control volume that contains the great majority of the vorticity, the observer position $\underline{x}$, and with surfaces along solid boundaries or within the fluid where acoustic relations are valid, then the solution in the far-field can be written (Powell, 1964):

$$\rho(\underline{x}) \cong \frac{-\rho_0}{4\pi x c_0} \frac{\partial}{\partial t} \int_{V_0} (\underline{J} \wedge \underline{v}) \cdot \frac{\underline{x}}{|x|} \, dV_0^*(\underline{y}) \; -$$

$$\frac{1}{4\pi x c_0} \frac{\partial}{\partial t} \int_{S_0} (p + \tfrac{1}{2}\rho_0 |\underline{v}|^2)\underline{n} \cdot \frac{\underline{x}}{|x|} \, dS_0^*(\underline{y}) - \frac{\rho_0}{4\pi |\underline{x}|} \frac{\partial}{\partial t} \int_{S_0} \underline{v} \cdot \underline{n} \, dS_0^*(\underline{y})$$

* denotes acoustic time delay
$\underline{x}$ = far-field coordinate
$\underline{y}$ = source coordinate
$\underline{n}$ = normal to surface pointing away from control volume

where S_0 denotes the part of the surface of the control volume V_0, which coincides with a solid surface. (The surface integrals appear because we used the free-space fundamental solution. Future research should include finding the Green's function suitable for vocal tract geometry. Loosely, Green's functions are to boundary value problems what impulse responses are to initial value problems.)

For the vocal tract, we take the control volume bounded by the lungs, trachea, glottis, pharynx, mouth, and a large sphere outside the mouth. The first and second integrals apparently provide dipole sources, and the third, apparently, a monopole source.

Because there are solid boundaries present, there can be a net energy exchange between the fluid and the solid, and because $\rho_0 \underline{J} \wedge \underline{v}$ is proportional to the time rate of change in the momentum in the fluid, the first integral can be approximated without considering acoustic time delays. Therefore, the first integral truly provides a dipole source. This integral will be nonzero in the region just above the glottis, where we assert the existence of a strong oscillating vortex ring with an axis of symmetry coincident with that of the dipole. This term is associated with the loss of pressure head discussed in the section on vortex formation, and we call it the vorticity-velocity interaction term. It was seen that the loss of head was an order one multiple of $(\rho_0/2)\ (Vg)^2$, where Vg is the glottal fluid particle velocity. This can be taken as the order of magnitude of the acoustic pressure provided by such a source.

The second integral also provides for dipole sources of sound with axes normal to the interior surfaces of the vocal tract. Because of the direction of the axes, this term should contribute little to the propagation of sound, except perhaps in the region of the vocal folds. The quantity $p + (\rho_0/2)\ |\underline{V}|^2$ is equal to $-\rho_0 \int \partial \underline{V}/\partial t \cdot d\underline{y}$ on the surface of the folds, so that an order-of-magnitude comparison between the first and second integrals can be carried out. The ratio of the second to the first is on the order of magnitude: $(r \cdot f)/Vg$ where r is the radius of the vocal tract and f is the frequency of sound under consideration. In a low-frequency approximation, consistent with the quasi-steady approximation used earlier, the first integral dominates the second.

The final integral involves the movement of the folds themselves. This integral appears to provide for a monopole source of sound. However, the integral is identically zero when acoustic time delays are neglected, because the folds do not change volume as they oscillate. Further, this integral is negligible in relation to the first, especially at low frequencies, because the peak velocity of the folds is so much less than that of the particle fluid velocity at the glottis.

This is only one possible form for an integral solution in the aeroacoustics literature. There are others that show the boundary forcing more explicitly, but without explicit reference to vorticity-velocity interaction (Goldstein, 1976). Also, a formulation by Howe (1975) combines the effects of the first and second integral and exhibits the vorticity-velocity interaction explicitly.

We have argued that the three integrals above provide dipole sources in the region of the glottis. The first integral, which is the result of energy transfer between the solid surface and the fluid, is arguably the largest of the three in the low frequency limit. This is not the whole story, because there is time varying motion of the fluid above the vorticity producing mixing region, which is required by mass conservation. This may provide an acoustic signal beyond what we have discussed so far, perhaps obeying nonlinear propagation laws. We are not prepared to compare the amplitude of this wave with that produced by the terms already discussed. The aeroacoustic formulation is not complete as we have discussed it here, but we have identified sources that have not been considered previously.

It should be noted that since the acoustic pressure fluctuations provided by the first integral are on the order of $(\rho_0/2)\ (Vg)^2$, this source is inefficient. The ratio of the acoustic intensity radiated by this term to the flux of fluid kinetic energy density in the glottal region is on the order of the square of the peak Mach number of the oscillatory part of the glottal flow.

We performed an experiment with the modulated jet to determine the importance of the vorticity-velocity interaction as a sound source. Using the same compressor described in the section on vorticity, we attached the hose from the compressor to a lamp post wrapped in packing foam to minimize its effect on the field. The nozzle thus was oriented horizontally, about 2 ft. above the floor. We used a B+K Sound Level Meter, with a wind shield, at 1 ft. 10 in. from the nozzle and in the horizontal plane of the nozzle. We ran the compressor at 80 Hz, and used no band-pass filter for measuring the intensity. Measurements were taken at 45° intervals from -90° to 90° to the centerline of the nozzle.

We model this situation with a control volume consisting of the interior of the tube down to the piston connected with a large sphere outside the tube. The sphere has a cylindrical section removed, which contains the tube (see Figure 3). Thus, part of the bounding surface of the control volume contains the piston of the compressor. Because there is an oscillatory change in volume we have a monopole source, which is represented by the third integral in the integral solution. However, because there is oscillatory vortex shedding from the tube exit, there is some dipole component of sound seen in the far field.

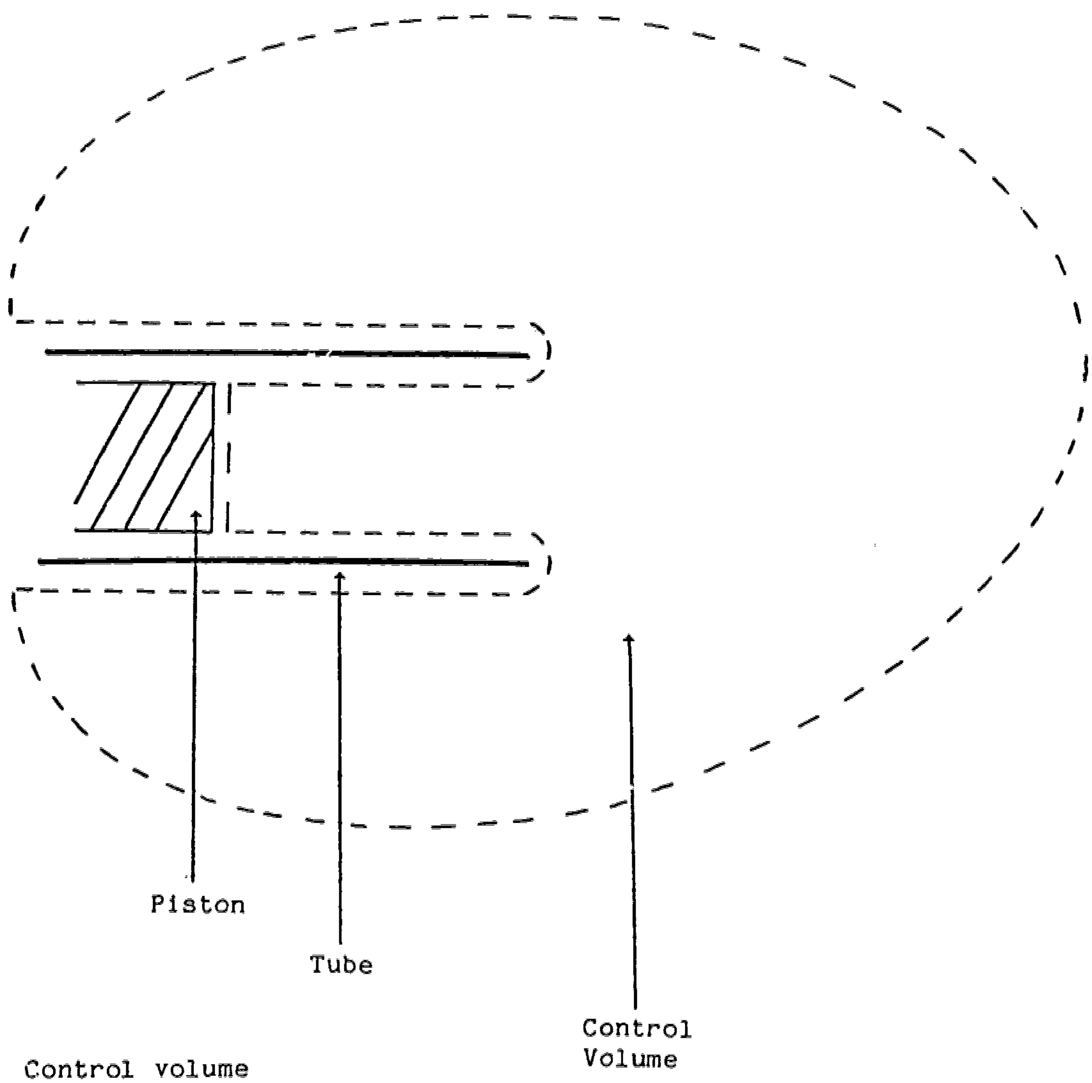


Figure 3. Control volume

The experimental results show a directivity pattern that is not omnidirectional and shows a large dipole component. These results are summarized in the table below.

Table

SPL directivity (background subtracted)

angle	SPL
90°	58 dB
45°	65 dB
0°	67 dB
-45°	64 dB
-90°	55 dB

Indeed, if we take 58 dB to be the intensity of the monopole source at the distance the measurements took place, then the main lobe of the dipole field adds 9 dB. This indicates that a great deal of fluid energy is in the form of vorticity, or the result of boundary forcing, which goes into making the dipole source. Only a small fraction of acoustic energy can be associated with the monopole source. (Covective amplification, where vorticity is convected in a mean flow, is known to alter the directivity of the sound field. However, since the mean flow Mach number is small, this effect cannot account for the present deviation from an omni-directional pattern.)

Conclusion

In this note we have found good reason for supposing the existence of vorticity above the glottis. This nonacoustic motion can be shown to produce sound via the mechanism of a fluctuating pressure head near the folds. This fluctuating pressure head is the result of an exchange of energy between the solid and fluid, which is realized in the fluid as vorticity-velocity interaction. This source is in addition to any oscillating fluid motion that may be considered to be acoustic above the region of strong vorticity-velocity interaction. We have not accounted for this latter wave in this presentation, so that our application of the aeroacoustic formulation may yet be incomplete.

The picture presented here stands in contrast to the standard picture of a piston source. In the picture presented here, a large amount of fluid energy goes into rotational motion near the folds, with only a small fraction of this energy being converted to sound. This process cannot be accounted for by a piston in a tube.

Future research should include finding the Green's function appropriate to the vocal tract, and experimental measurements and theoretical predictions of the oscillatory fluid field just above the glottis and the forcing of the fluid by the vocal folds. These ingredients, difficult to obtain, will completely characterize the acoustic field produced during phonation.

References

Batchelor, G. K. (1970). _An introduction to fluid dynamics._ Cambridge: Cambridge University Press.

Howe, M. S. (1975). Contributions to the theory of aerodynamic sound with applications to excess jet noise and theory of the flute. _Journal of Fluid Mechanics, 71_, 625-673.

Goldstein, M. E. (1976). _Aeroacoustics._ New York: McGraw-Hill.

Kaiser, J. (1983). Some observations on vocal tract operation from a fluid flow point of view. In I. R. Titze & R. C. Scherer (Eds.), _Vocal fold physiology._ Denver: Denver Center for Performing Arts.

Lighthill, M. J. (1952). On sound generated aerodynamically, I. General theory. _Proceedings of the Royal Society, A211_, 564-587.

Powell, A. (1964). Theory of vortex sound. _Journal of the Acoustical Society of America, 36_, 177-195.

Teager, H., & Teager, S. (1983). Active fluid dynamic voice production models, or there is a unicorn in the garden. In I. R. Titze & R. C. Scherer (Eds.), _Vocal fold physiology_ (pp. 387-401). Denver: Denver Center for Performing Arts.

van den Berg, J., Zantema, J. T., & Doornenbal, P., Jr. (1957). On the air resistance and the Bernoulli effect of the human larynx. _Journal of the Acoustical Society of America, 29_, 626-631.

PATTERN FORMATION IN SPEECH AND LIMB MOVEMENTS INVOLVING MANY DEGREES OF FREEDOM*

J. A. S. Kelso†

Abstract. An important task for neuroscience is to understand how task-specific ensembles of neuromuscular elements (coordinative structures) are formed to produce coherent spatiotemporal behavior. Using tools and concepts of synergetics (which deals with cooperative phenomena in nonequilibrium, open systems) and nonlinear dynamics (which provides low dimensional descriptions of forms of motion that are produced by high dimensional systems), four related themes are addressed: (1) Cooperativity, that is, the nature of the unitary organization formed by an ensemble of neuromuscular components; (2) Control, that is, the kind of dynamic control structure (based on the fundamental notion of elastic deformation) that is capable of generating a diversity of movement patterns; (3) Stability, that is, the informational basis (defined in terms of critical phase angles) that may underlie certain invariances in movement patterning; and (4) Change, that is, how new (or different) modes of spatiotemporal behavior may arise under the influence of parameter scaling and system nonlinearities. Under each theme, relevant data are discussed, theoretical conclusions drawn, and hints for further research are provided.

1. Introduction

There are over 792 muscles and 100 joints in the human body. And, according to my elder son's biology textbooks, the elephant's trunk contains over 40,000 muscles and tendons. Thus, any activity of the human body or the elephant's trunk involves the cooperative effort of very many degrees of freedom. But what form do principles of cooperation in multivariable movements take? For some years now, my colleagues and I have viewed this question as continuous with the general issue of understanding the emergence of order and regularity in complex systems (see e.g., Yates, 1979, for defining characteristics of complexity). The core idea that we have pursued is that the collective action among multiple neuromuscular components is fundamentally task-related, that the significant units of control and coordination are functional groupings of muscles and joints, which we call coordinative structures or functional synergies (e.g., Fowler, Rubin, Remez, & Turvey, 1980; Kelso, Southard, & Goodman, 1979; Kelso & Tuller, 1984a; Kugler,

---

*In H. Heuer & C. Fromm (Eds.), Generation and modulation of action patterns (Experimental Brain Research Series; 15, pp. 105-128). Berlin: Springer-Verlag, 1986.

†Also Center for Complex Systems and Department of Psychology, Florida Atlantic University, Boca Raton.

Acknowledgment. This work was supported in part by NIH Grant NS-13617, Biomedical Research Support Grant RR-05596 and Contract No. N0014-83-K-0083 from the U. S. Office of Naval Research. Comments by Bruce Kay, Kevin Munhall, Elliot Saltzman and Betty Tuller were much appreciated.

Kelso, & Turvey, 1980; Saltzman & Kelso, in press; Turvey, 1977). The hallmark of a coordinative structure is the temporary marshalling of several articulators into a task-specific pattern.

This notion of functional units of action, or coordinative structures, differs in significant ways from conventional treatments of movement control that are based in either the neurophysiological notion of a central pattern generator or the information processing notion of a motor program. First, unlike the notion of a hard "pre-wired" central pattern generator, the coordinative structure construct underscores the soft or flexible nature of action units that are functionally-specific, not anatomically-specific. One of the goals of this paper is to buttress this claim using examples from recent research on the motor control of speech and limb movements. Second, contrary to the motor program formulation that relies on symbol-string manipulation familiar to computer technology, the coordinative structure construct highlights the analytic tools of qualitative (nonlinear) dynamics (e.g., Kelso, Holt, Rubin, & Kugler, 1981; Kelso, V.-Bateson, Saltzman, & Kay, 1985; Saltzman & Kelso, in press) and the physical principles of cooperative phenomena (e.g., Kelso & Tuller, 1984a, 1984b; Kugler et al., 1980; Kugler, Kelso, & Turvey, 1982). Thus, the problem of pattern formation for skilled actions is couched as a specific aspect of the more general topic of cooperative phenomena in nonlinear, open systems (see e.g., Haken, 1975, 1977, 1983). Such systems display ordered states that are not imposed by programs, but that actively evolve from the dynamic interplay of processes, in a so-called "self-organized" fashion. Although the present theoretical approach is in preliminary form as far as biological movements are concerned (see Kelso, 1981a; Kelso & Tuller, 1984a; Kugler et al., 1980, 1982) this paper attempts to convey the flavor of the approach, not only theoretically, but in terms of the kinds of experiments that it motivates. In the following sections I shall address briefly four questions drawing from our own and others' experimental work on multidegree of freedom movements of limb and speech articulators:

(i) The cooperativity question. What kind of unitary organization is formed by an ensemble of neuromuscular components?

(ii) The control question. What kind of control structure underlies the generation of certain movement patterns? What are the essential control parameters and how are parameter values specified?

(iii) The stability question. What characterizes the stability of a movement pattern, and what is the informational basis of the stability? Colloquially speaking, what holds a pattern together?

(iv) The change question. What are the necessary and sufficient conditions that give rise to change in articulatory pattern?

2. The Cooperativity Question

2.1 The Concept of Coordinative Structure

Do relatively independent articulators (muscles, joints) function as a unitary ensemble, and, if so, what kind of ensemble is it? Consider the act of speaking. Even a simple speech gesture involves cooperation among very many degrees of freedom operating at respiratory, laryngeal, and supralaryngeal levels. Yet in spite of (or perhaps because of) such a large number of neuromuscular elements, speech emerges as a coherent and organized activity. An attractive hypothesis proposed by Bernstein (1928/1967) and developed by his colleagues (e.g., Gelfand, Gurfinkel, Fomin, & Tsetlin, 1971) is that the central nervous system, rather than controlling each degree of freedom separately, organizes them into "collectives," "linkages," or "synergies" that then behave, from the perspective of control, as a single degree of freedom.

Of course, as emphasized earlier, our notion of synergy or coordinative structure is unlike that of Sherrington (1906) or Easton (1972) (see Kelso & Tuller, 1984a) in that the collective action among multiple muscles or kinematic components is not rigid or machine-like, but fundamentally task or functionally-specific. In this Darwinian-like hypothesis, function dictates the form of cooperativity observed in an aggregate of neuromuscular components, not anatomical connections. But how might this notion be tested, and what evidence exists in its favor for complex actions?

A window into the behavior of a complex system possessing large numbers of active, interacting components can be gained by perturbing it dynamically during an activity and examining how the system reconfigures itself (e.g., with respect to response latencies, magnitudes, etc.). Thus a group of potentially independent articulators could be said to behave in a unitary fashion if it were shown that a disruption to one (or more) members of the group was responded to by other members of the group at a site remote from the challenge. By the concept of coordinative structure, the response of the articulatory ensemble would not be stereotypic; rather it would be adapted quickly and precisely to accomplish the task. In general, the components of the neuromuscular system would cooperate in such a way as to preserve the performer's intent. Some evidence of so-called "remote compensation" phenomena that support a motor system design based on coordinative structures exists in both the speech and limb movement behavior literatures. These data are considered below.

2.2 Coordinative Structures in Multidegree of Freedom Movements

Although the speech literature contains a number of observations that are consistent with a coordinative structure mode of articulator organization, few experiments have been designed to test the notion explicitly. In one experiment by Folkins and Abbs (1975) the jaw was occasionally loaded during the closure movement for the initial /p/ in the utterance "a /hæ pæp/ again." Lip closure was attained in all cases apparently by exaggerated displacement and velocities of the lip closing gestures, particularly by the upper lip. Similarly, Folkins and Zimmermann (1982) used electrical stimulation to produce an unexpected depression of the lower lip prior to, and during, bilabial closure. Compensatory changes in the jaw and upper lip were observed to effect the bilabial closure. Abbs and colleagues (see Abbs, Gracco, & Cole, 1984, for review) report that both autogenic (that is, lower lip) and remote (upper lip) effects occur when a 40 g load is applied to the lower lip unexpectedly 30 ms before the onset of the phasic EMG burst in orbicularis oris inferior. They interpret these remote effects, after Houk and Rymer (1981), as evidence for open-loop, feedforward control in which "...a precise, experience-based representation of the relationship between afferent signals from one movement (from which a potential error is detected) and the motor output of a parallel synergistic movement (where the adjustment is implemented)." Autogenic compensations made by the perturbed structure are slower and thought to be under closed-loop feedback control.

Although these findings are consistent with the coordinative structure concept, it is not clear whether, in fact, the patterns of articulator coupling following perturbations are in any sense standardized (as one might predict if they were completely preprogrammed or a result of fixed input-output loops) or whether they are indeed "functionally organized," that is, directed to the stable production of the intended utterance. If the former, the pattern of response to a given perturbation should be the same regardless of the utterance. If the latter, different patterns of articulator cooperation (coordinative structures) should occur, tailored to the particular phonetic requirements.

Direct evidence that speech articulators (lip, tongue, jaw) do make functionally specific, near-immediate compensations to unexpected perturbations at sites remote from the locus of perturbation comes from our recent work (Kelso, Tuller, & Fowler, 1982; Kelso, Tuller, V.-Bateson, & Fowler, 1984). An unexpected constant force load (5.88 Newtons) applied during upward motion for final /b/ closure in /bæb/ revealed near-immediate changes in upper and lower lip muscles and movements (15-30 ms), but no changes in tongue muscle activity. The same perturbation applied during the utterance /bæz/ evoked rapid and increased tongue muscle activity (genioglossus) for /z/ frication, but no active lip compensation. Although the jaw perturbation represented a threat to both utterances, no perceptible distortion of speech occurred. That a challenge to one member of a group of potentially independent articulators was met--on the very first perturbation experience--by remotely linked members of the group provides preliminary support for coordinative structures. Further anecdotal support for coordinative structures is mentioned in a review paper by Abbs and Gracco (1983). They report that for the utterance /aba/, upper and lower lips compensate when the lower lip is loaded in order to preserve bilabial closure. In contrast, for /afa/ which in theory does not require upper-lip movement, only lower lip compensatory responses to a lower lip perturbation occur.

Analogous results emerge from recent studies of human posture (e.g., Cordo & Nashner, 1982; Marsden, Merton, & Morton, 1983). For example, in response to a perturbation applied to the thumb, which was performing a tracking task, Marsden et al. observed reactions in muscles remote from the prime mover (e.g., in pectoralis major of the same limb; in triceps of the opposite limb; in the opposite thumb when it served to stabilize motion, etc.). These distant reactions are much faster than typical reaction time responses; indeed they are sometimes faster (e.g., 40 ms in pectoralis) than the local, autogenetic reflex in the structure perturbed. But most interesting for the coordinative structure hypothesis is that postural responses occur only if they perform a useful function and they are flexibly tuned to that function. For example, postural responses in triceps disappear if the hand is not exerting a firm grip on an object. If, instead of holding a table top, the non-tracking hand holds a cup of tea, the responses in triceps reverse, which is precisely what they have to do to prevent the tea from spilling. Marsden et al. (1983) conclude that these rapid, remote effects "...constitute a distinct, and apparently new, class of motor reaction" (p. 645) that has led them to abandon an account based on stretch reflexes. Such remarks, however, reflect a strong Western bias. For example, Russian studies done in the 60's reveal similar interactions between posture and voluntary movement (see Gelfand, Gurfinkel, Fomin, & Tsetlin, 1971). Moreover, Bernstein (1967) refers to his published experimental work (in Russian) in the early 1920's that afford the conclusion that "Movements react to changes in one single detail with a whole series of others which are sometimes very far removed from the former both in space and in time..." (Bernstein, 1967, p. 69).

The microscopic workings of a coordinative structure can be further explored by varying the phase of the jaw perturbation during bilabial consonant production. For example, recent work has asked: does perturbing the jaw during the opening phase of the utterances /bæb/ and /bæp/ induce a remote reaction in the upper lip? If the cooperativity between oral structures is functionally-based, remote effects are predicted only when a jaw perturbation occurs in the closing phase (that is, during the transition out of the vowel into the final consonant), when the upper lip is actively involved in producing consonantal closure. On the other hand, if the form of interarticulator coupling is in any sense rigid, remote reactions should be seen regardless of when the jaw is perturbed. In fact, the data support the former hypothesis. Remote reactions in the upper lip were observed only when the jaw (V.-Bateson & Kelso, 1984; Kelso et al., 1984) or lower lip (Munhall & Kelso, 1985) was perturbed during the closing phase of motion, that is, when the reactions were necessary to preserve the identity of the spoken utterance.

The phase-specific patterning observed in speech shares a likeness to recent work in other motor systems. For example, in cat locomotion (cf. Forssberg, 1982, for review), when light touch or weak electrical shock is applied to a cat's paw during the flexion phase of the locomotor cycle, an abrupt withdrawal response occurs as if the cat were trying to lift its leg over an obstacle. When the same stimulus is applied during the stance phase of the cycle, the flexion response (which would make the animal fall over) is inhibited, and the cat responds with added extension (Forssberg, Grillner, & Rossignol, 1975). This "stumble corrective reaction" is present in intact and spinal animals and, like speech compensation, occurs remarkably quickly. The earliest flexor burst in response to a tactile stimulus applied during the swing phase, for example, occurs with a latency of 10 ms. Just as the foregoing data on articulatory reactions to perturbation appear specific to the spoken utterance, so also do the data on cat locomotion reveal reactions that are non-stereotypic and functionally suited to the phase-dependent requirements of locomotion.

In summary, the evidence presented in this section in support of task-specific action units poses a challenge not only to the neuroscientist but to anyone who seeks to understand the relation between an organism's structure and its function. The adaptive reactions discussed here could certainly be described as reflexive because of their speed. Their mutability, on the other hand, speaks against any hypothesis about fixed reflex connections or rigidly constructed servomechanisms. Similarly, it is not parsimonious to assume that the computation is preprogrammed in such a way that the articulatory ensemble produces precisely those movements that accomplish the task. The problem is exacerbated when unexpected environmental challenges are introduced whose dimensions (e.g., magnitude, duration, locus) are potentially manifold. The main message that emerges is that the multiple components of the motor system are "softly" assembled and flexible in function, not machinelike and rigid--in either the hard-wired language of central pattern generators or the hard-algorithmed language of computers which are the source of the motor program idea.

3. The Control Question

What are the essential control structures that govern the patterning of articulator motion in space and time? Although this question is of much interest to many in the field of motor control in general, the movements of speech articulators will be the primary focus here. However, the kinematic relationships that we shall identify and focus upon are not unique to speech at all, a fact that is quite appealing in that it suggests a common vocabulary might exist to describe the underlying control structure of speech and other actions.

Obviously there are many surface features of a movement that one might propose as significant candidates for controlled variables. What then, fashions the constraints on the choices one makes? Is the selection among controlled variables really like a multiple choice exam (cf. Stein, 1982)? Or, might a "deep structure" for motor control exist, that can be recognized in the face of much surface variability? And, if so, on what principle(s) is it based? Below, the idea that a *dynamic* control regime governs movement patterns is developed.

After Maxwell (1877), dynamics can be viewed as the simplest and most abstract description of the motion of a system. The relations among, and the values of, dynamic parameters (e.g., mass, stiffness, damping) can produce a wide variety of kinematic consequences (e.g., position, velocity). Thus, kinematics provides a surface description of the movements of a system that are generated from a given type of dynamical organization. Note that the dynamics referred to here is not to be interpreted as local and concrete, or to be equated with pure biomechanics. Rather the branch of dynamics emphasized here, nonlinear dynamics, is concerned with the underlying, abstract basis of *forms of motion* or pattern formation in complex, multidegree of freedom systems (e.g., Abraham & Shaw, 1982; Haken, 1983). These forms of motion are specified, roughly, by the qualitative shapes observed in phase portraits of a system's behavior (see below). For example, the muscles, joints, and neuronal structures that cooperate to produce a walking pattern involve literally thousands of degrees of freedom, but the pattern itself represents a low dimensional form--a cyclical motion of the limbs--which can be operated by low dimensional control (see Garfinkel, 1983). In fact, changes in gait in the decerebrate cat can be manipulated experimentally by a single parameter--the intensity of electrical stimulation delivered to the midbrain (Shik, Severin, & Orlovskii, 1966).

Such low dimensional forms are called *attractors* and represent the asymptotic stable behavior of a whole family of trajectories. As a simple example a damped mass spring system can have many trajectories depending on its initial conditions and its parameter values (mass, stiffness, damping). Such a system is called a *point attractor*, a generic dynamical category that reflects the fact that all trajectories converge to an asymptotic, static equilibrium state. Importantly, however, a *multi*degree of freedom system whose trajectories likewise converge to a single rest position can also be described as a point attractor. Thus, a point attractor is a low dimensional description of a potentially high dimensional state space and exhibits the property of equifinality--the tendency to achieve an equilibrium position regardless of initial conditions. Though the language and the concepts of nonlinear dynamics may be unfamiliar (but see Kelso & Kay, in press, for a tutorial), the intent here is to show that one can apply this framework (combined with a quantitative treatment of articulator trajectories) to the analysis of speech production and other biological activities.

The advantages of a dynamical approach to control are several (see Kelso & Kay, in press, for details; also Saltzman, in press). Among these, hinted at above are: 1) Generativity--an invariant dynamic structure can give rise to much surface kinematic variability; 2) No explicit representation or pointwise control of the system's planned trajectory need exist in a dynamical system; 3) Different dynamic regimes (e.g., point attractor, periodic attractor) can serve to categorize different tasks (see Kelso & Tuller, 1984b; Saltzman, in press; Saltzman & Kelso, in press). For example, recent work in the motor control field--especially on voluntary limb and finger movements--indicates that discrete and rhythmical movements can be modeled as a damped mass-spring, point attractor (e.g., Bizzi et al., 1976; Cooke, 1980; Fel'dman, 1966; Houk, 1978; Kelso, 1977; Kelso & Holt, 1980; Schmidt & McGown, 1980) or limit cycle, periodic attractor system, respectively (Fel'dman, 1980; Kelso et al., 1981). These control structures are characterized by sets of invariant dynamic parameters (e.g., damping, stiffness, and equilibrium length), and kinematic variations (e.g., position, velocity, acceleration over time) can be viewed as consequences of these underlying patterns of dynamic control parameters. A final, related advantage is that the abstract task level of description and the description of muscle-joint properties are entirely commensurate. That is, a dynamical description applies at all levels. The problem becomes one of relating dynamics that operate on different time scales.

In complex movements like speech, however, we seldom make direct measurements of the dynamic parameters themselves, for example, the mass, damping, and stiffness values for an organization of neuromuscular elements. In our ongoing work, we measure and compute articulatory kinematics during the production of simple syllables and use the relations among these kinematic variables to infer the underlying functionally-defined dynamic control regimes. One main paradigm involves reiterant speech, in which subjects are required to substitute a simple syllable (e.g., /ba/ or /ma/) for the real syllable in an utterance, yet still maintain the utterance's normal prosodic structure. The benefit of the reiterant technique for production studies is that the removal of segmental factors (that is, the different consonants and vowels of real speech), besides having minimal effects on the timing/metrical pattern, allows one to measure movements of those supralaryngeal articulators that are consistently active over the entire utterance, in this case the lips and jaw involved in /ba/ or /ma/.

Kelso et al. (1985) employed a phase plane analysis (a continuous plot of articulator position versus velocity) of lip and jaw movement trajectories followed by a quantitative kinematic analysis of opening and closing gestures. Several interesting kinematic results were obtained (see Figure 1). First, largely unimodal velocity patterns of jaw and lips occurred for opening and closing gestures at both slow and fast speaking rates (See Figure 1, Right); Second, a given gesture's peak velocity (Vp) covaried with its displacement (d). Regression analyses of the data showed not only a strong relation between Vp and d, but also that the slope of the relation changed depending on the reiterant syllable's stress and rate. As shown in Figure 1, shorter amplitude motions corresponding to unstressed gestures and faster speaking rates had steeper slopes than stressed gestures spoken at a normal rate.

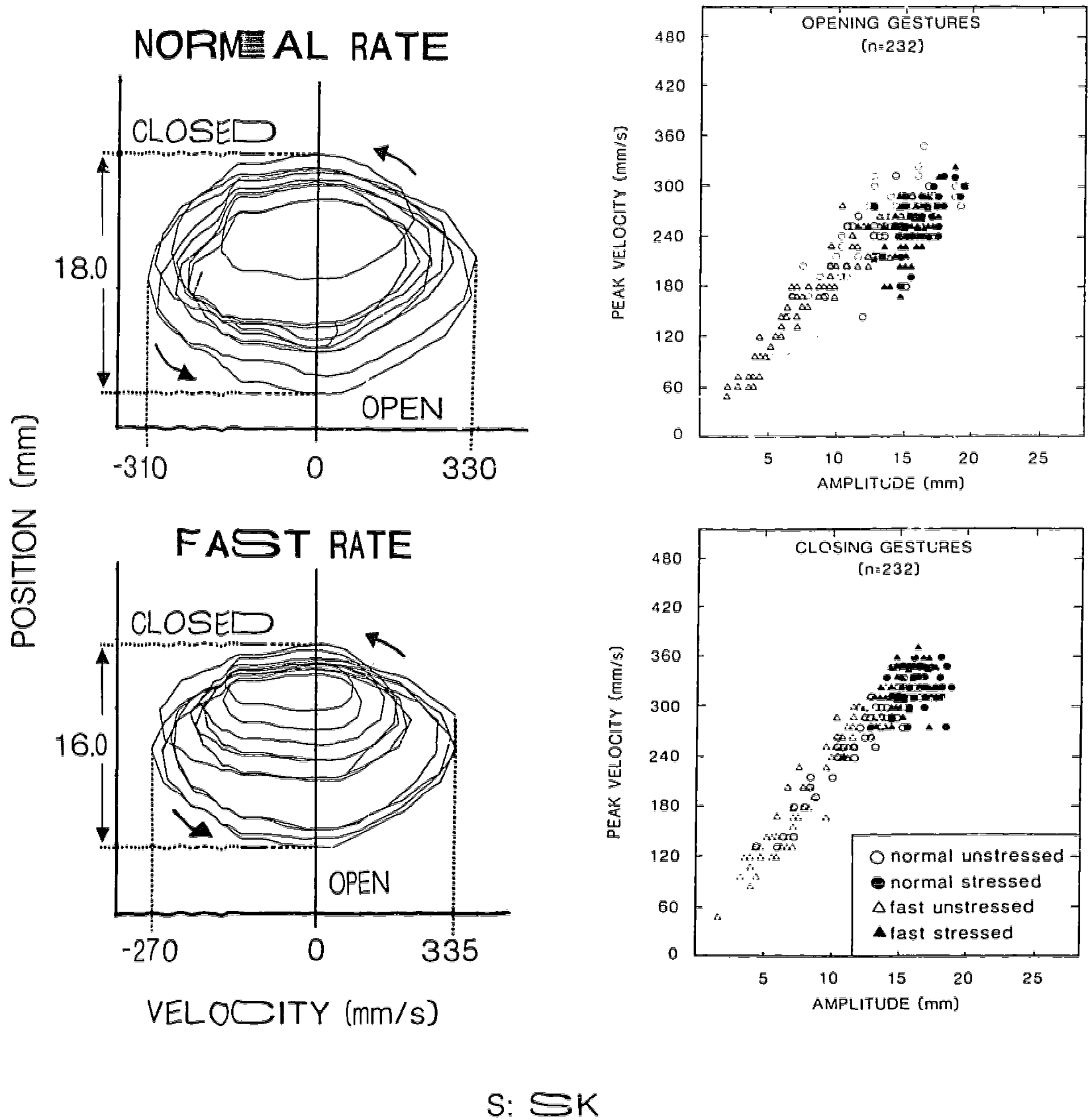


Figure 1. Left: Phase plane trajectories of lower lip plus jaw (that is, from a sensor placed on the lower lip) for reiterant speech spoken at a normal (top) and fast (bottom) rate with /ba/ as the reiterant syllable. Right: Scatter plot of peak velocity versus displacement (lower lip plus jaw) of a subject's opening gestures associated with the consonant-vowel portion of the syllable (top) and closing gestures associated with the vowel-consonant portion of the syllable. The legend specifies conditions [from Kelso et al., 1985].

The impressive scaling relation between Vp and d is not unique to speech where it has been reported before, often as an incidental result (e.g., Kent & Moll, 1975; Sussman, MacNeilage, & Hanson, 1973). An inventory of other activities, ranging from natural reaching movements to tongue movements (see Kelso & Kay, in press, for review and Viviani, this volume) to infant kicking (Thelen, Skala, & Kelso, 1985) shows the same relationship. *Thus, this lawful regularity is observed not only in different material structures but also in activities involving multiple degrees of freedom.*

What kind of dynamical control structure could give rise to such kinematic relations? Consider the relationship "ut tensio sic vis; that is the power of any spring is in the same proportion with the tension thereof" (Hooke, 1678). By "spring," Hooke meant any springy body and by "tension," what we would now call "extension" or more generally, strain. This linear relationship is called Hooke's Law ($F = -kx$), where F is the restoring force, k is a proportionality constant representing spring stiffness, and x is displacement. The elementary equation of motion can be derived from Newton's Second Law, $F = m\ddot{x}$. That is, $F = -kx = m\ddot{x}$; therefore, $m\ddot{x} + kx = 0$, where m is mass and $\ddot{x}$ is acceleration. This last equation describes the motion of a simple harmonic oscillator with a given mass and stiffness and no damping. On the phase portrait, all concentric trajectories of the oscillator have the same shape with the same periodicity for a given set of dynamic parameters. Note importantly, that any changes in initial conditions (x, $\dot{x}$) are precisely accommodated by changes in peak velocity. Thus, the Vp-d scaling relationship is *specified* by this particular dynamical system. The peak velocity-displacement relation reflects the stiffness of the system, since $\omega_0 A$ is the peak velocity of simple harmonic motion, and the slope of $\omega_0 A$ versus A is ω_0 (where A is cycle amplitude and $\omega_0 [= (k/m)^{\frac{1}{2}}]$ is the angular frequency of motion). Assuming constant mass, the slope of the Vp/d relationship is proportional to $k^{\frac{1}{2}}$. Changing stiffness changes the eccentricity of the phase plane trajectories (which is what Kelso et al., 1985, observed) and increases the slope of the peak velocity-displacement relation[1] (see also Cooke, 1980; Ostry & Munhall, 1985).

This simple model strongly suggests that the stiffness or elasticity of the system (in an abstract sense) is an important control parameter for skilled actions. In concluding this section, the potential (and more generalized) theoretical significance of this claim is addressed. To do this, we need to develop briefly a perspective based on elasticity theory (see Landau & Lifshitz, 1981; Love, 1927; Timoschenko, 1953).

The most general form of Hooke's Law, that is, beyond a simple force-displacement description, is that over a wide range of applied stresses, the measured strain increases in the same proportion. The proportionality linking stresses to strains is the elastic constant, k. Thus, Hooke's Law is fundamentally a description of *elastic deformation processes*. This generalization, though entirely consistent with recent work demonstrating stiffness or impedance control (e.g., Hogan, 1984) offers a very different image for movement control. It characterizes movement fundamentally as *form*: solid bodies (limbs, jaws, tongues) can be made to change their size and shape, that is, their *configuration*, by the application of suitable forces (stresses). In this view, any new configuration is expressed by the specification of strains. Note that displacement is only a *measure*, often on a single plane of motion, of strain or deformation. Strains themselves are changes in the *relative* positions (or configuration) of a body. They usually require a tensorial description (e.g., Love, 1927). In Kelso et al. (1985), changes in movement duration and displacement that occurred when speaking rate and stress changed were characterized as consequences of the dynamic parameters of stiffness and equilibrium position. This formulation can now be recast into an equivalent, but more conceptually meaningful form, one that affords insights into the regulation of multiple muscles during action, not simply an agonist-antagonist pair (e.g., Bizzi, Accornero, Chapple, & Hogan, 1982; Cooke, 1980).

When an effector system, say the jaw-lip complex, moves from one configuration to another, the system in general does some work. A way to envisage the system specification of equilibrium position and stiffness is to express the work done as a potential strain or energy function. The latter specifies the macroscopic relation between stresses and strains. In Figure 2A, a linear force-displacement relation is mapped onto a strain-energy surface in which the potential energy is a quadratic function of the strain components (in this case simply displacement). The corresponding phase portrait is also shown. For comparison purposes, the case in which stiffness changes nonlinearly as a function of displacement (the so-called "soft" spring, cf. Jordan & Smith, 1977; Kelso, Putnam, & Goodman, 1983) is illustrated in Figure 2B.

It is apparent from Figure 2 that the amount of potential energy is proportional to displacement (or more generally, configuration) and that the slope of the force-displacement function specifies stiffness. In this view, the system's "endpoints" or "targets" correspond to minima of potential energy functions whose gradients define spring force. As Kugler et al. (1980) emphasize, to produce a movement is to effect a change in the underlying geometry of the dynamics, captured as a potential field.[2] Recently, Hogan (1984) has elaborated this framework for the trajectories of multijoint movements. Successive target locations are specified by means of a time-varying potential field with stable equilibria at the "target" locations.

There are two main points that arise from the perspective advanced here. First, because we are dealing with potential energy functions, only scalar quantities are involved. Several advantages for control accrue immediately. Since energy is a scalar quantity, unlike force (which is a vector), energy is invariant under coordinate transformations. Thus the coordinate system can be chosen to simplify the problem (see Marion, 1970). Also, it is often impossible to define exactly what the forces are (e.g., in a multimuscle system), whereas it is often possible to express the kinetic and potential energies. The latter are intrinsic to the system under study, whereas the standard force description places its emphasis on an outside agency acting on a body. Relatedly, because scalar potentials may be superimposed, the overall effect of multiple muscle activity can be obtained by addition of the potential functions (Marion, 1970). This characterization may offer considerable advantages for a compact description of control in multidegree of freedom movements.

But the second main feature of the present perspective is that the potential or strain-energy function (Love, 1927) can be properly conceived as an elastic field. [As an aside, Asada (1982) has recently demonstrated how elastic fields can be used for planning stable grasp in a robot manipulator.] The notion that movement involves deformation of an elastic field may ground one of Bernstein's most interesting intuitions, namely that movement is a

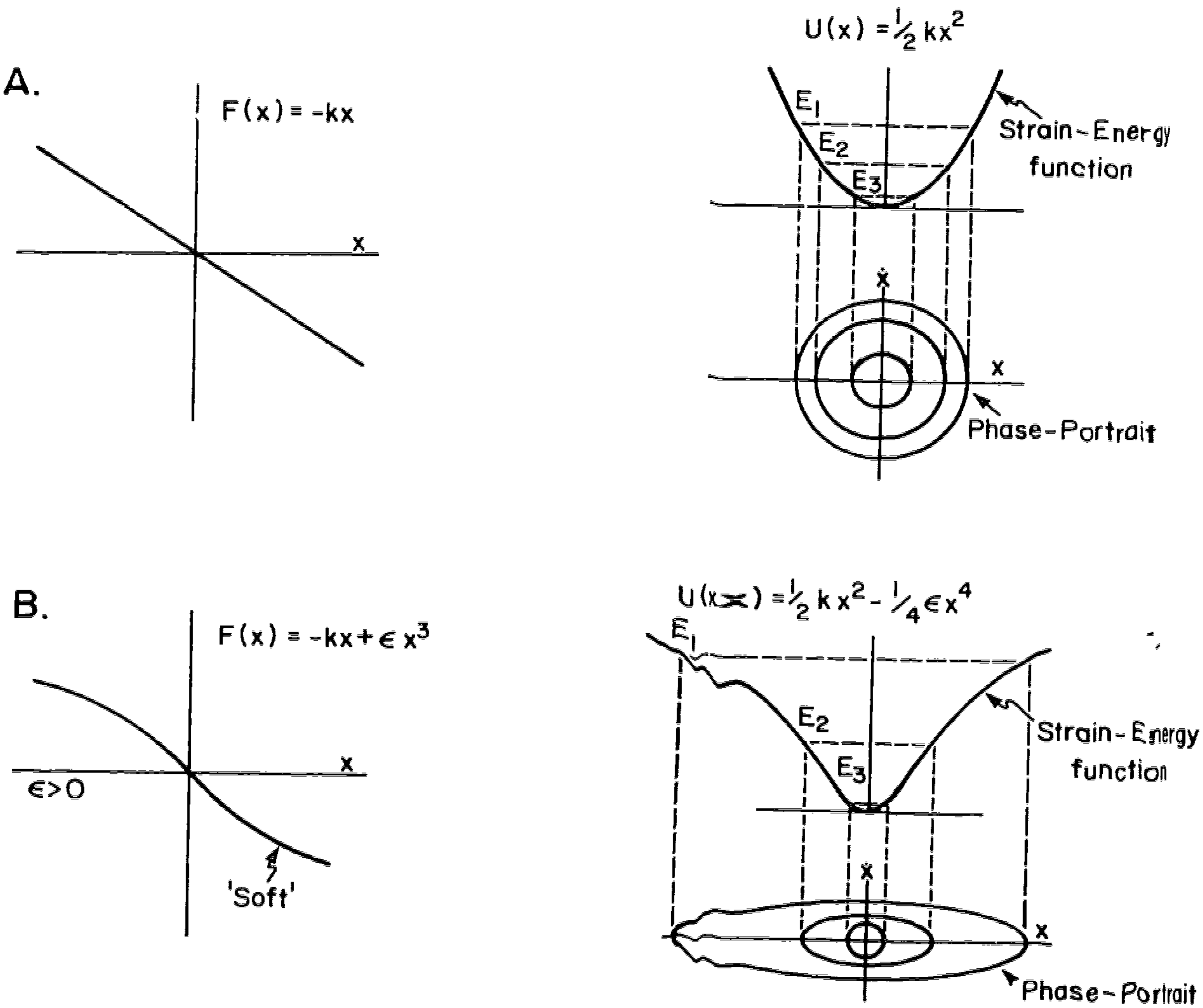


Figure 2. A. Left: A graph illustrating Hooke's Law. The deviation of the force (F) from linearity is symmetrical about the equilibrium position (x=0). Right: The potential (or strain-) energy function corresponding to the linear force-displacement relation and its associated phase diagram in $x, \dot{x}$ coordinates. Three phase paths are drawn, corresponding to the three values of total energy, E, indicated by dotted lines in the potential function. B. The form of the above relationships when the force is less than the linear term alone and the system is said to possess a "soft" nonlinearity.

morphological object. The ubiquity of the peak velocity-displacement relation, then, may offer a window into processes that form and deform the configuration of the body. If correct, an ancient theme for "static" forms--that a few simple rules can fashion some very intricate products (e.g., Gould, 1980; Stevens, 1974)--may apply equally well to the forms of patterned motion that interest us here.

4. The Stability Question

It is now quite well-established that as an articulator ensemble performs its task at different speeds and forces, the relative timing of muscle contractions and/or articulator motions is preserved invariantly[3] (e.g., Boylls, 1975; Kelso et al., 1979; Kugler et al., 1980; Schmidt, 1982; Shapiro 1978; Shapiro et al., 1981, for reviews). Such results have been taken as evidence that the central program determines when pulses of muscular force are applied to the limbs and their durations and relative sizes. Thus, according to Schmidt (1980) the determination of time (emphasis his) of contractions and relaxations appears to be directly controlled (see Footnote 1).

In the previous section, it was argued that although muscles contract and relax and though movements flow in time, a movement's temporal structure may be a consequence of the system's dynamic parameterization. Here I want to show how stable relative timing among gestures may be understood without recourse to an extrinsically-imposed timing program (see Kelso & Tuller, 1985, in press). But if timing is not controlled extrinsically in such a fashion, what processes might underlie the observed temporal stability? How, in a complex system of articulators, does a given gesture/articulator "know" when it should be activated in relation to other gestures/articulators? With respect to our relative timing data in speech, for example, what information is needed for the upper lip (a remote, non-mechanically linked articulator associated with a consonantal gesture) to move in appropriate temporal relation to the vocalic movement cycle of the jaw? As we shall see, different views of relative timing emerge when the articulator motions are examined in different coordinative spaces.

Consider first a very simple, but paradigmatic case in which the delay (in ms) of onset of upper lip motion for a medial consonant is measured relative to the interval (in ms) between onsets of jaw motion for flanking vowels. Figure 3, taken from Tuller and Kelso (1984) plots these events for one of four speakers who produced the utterances /babab/, /bapab/, and /bawab/, at two speaking rates and with emphatic stress placed on either the first or second syllable. The data for all four subjects were very similar. This figure shows that over changes in speaking rate and stress, the measured intervals change considerably, as do the magnitude of the events themselves, but the function relating these events is linear. That is, the metrics (amplitude, velocity, duration) of the events change, but the relative timing does not. Note that this is a strictly temporal description relating discrete movement events. Like most, if not all of the work on relative timing, measurements are confined to the onsets and offsets of articulator movement (see e.g., Schmidt, 1982).

A very different view of articulatory "timing" emerges when a re-analysis of the movements using phase plane trajectories is employed (Kelso & Tuller, 1985). Figure 4 illustrates the mapping from time domain to phase plane trajectories. On the left, hypothetical jaw and upper lip motions (position

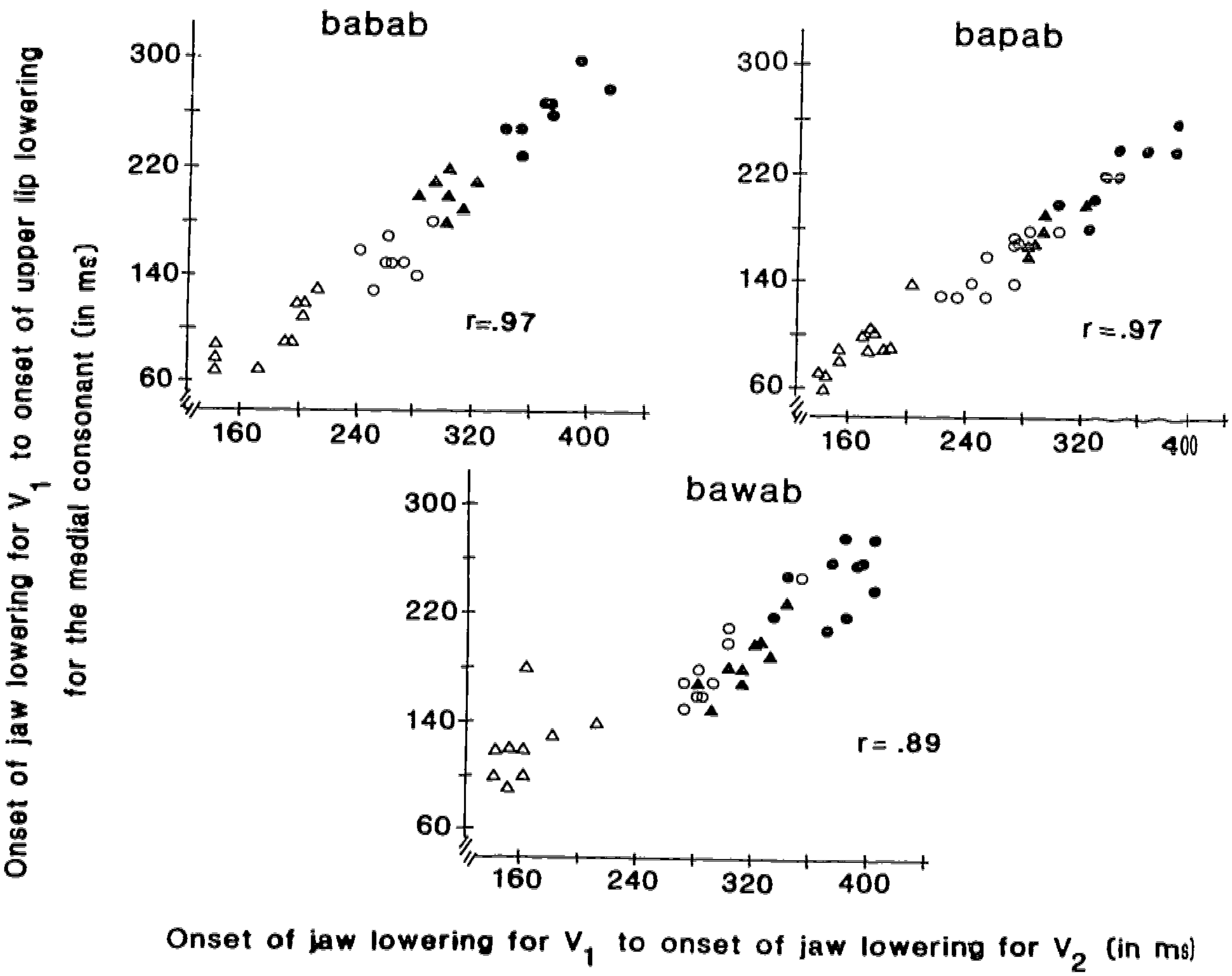


Figure 3. Timing of upper lip lowering for medial consonant articulation as a function of vowel-to-vowel period for one subject's production of the indicated utterances. Each point represents a single token of the utterance. (●) primary stress on the first syllable spoken at a conversational rate; (○) primary stress on the second syllable (conversational rate); (▲) and (△) primary stress on the first and second syllables, respectively, spoken at a faster rate [from Tuller & Kelso, 1984].

129

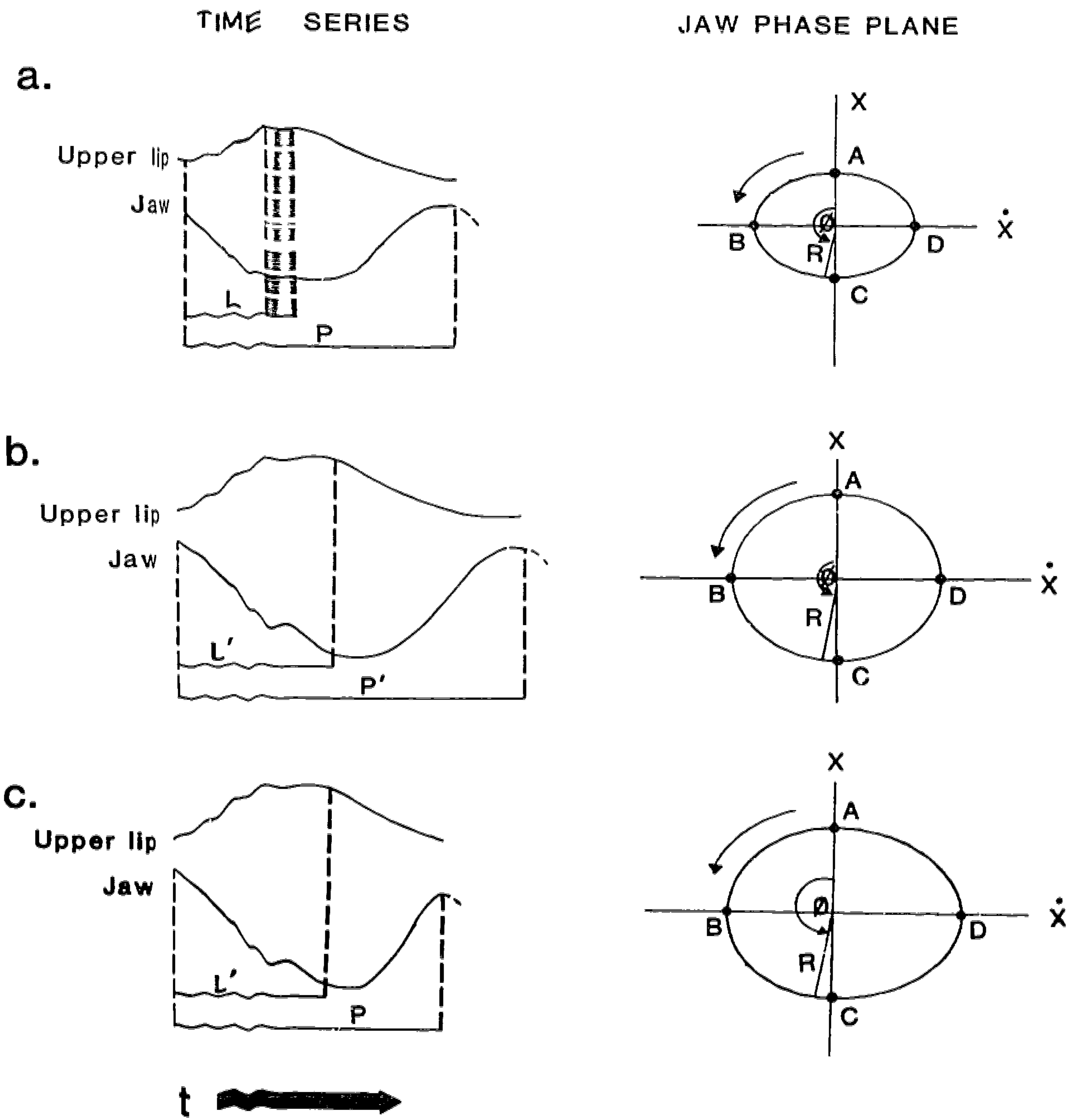


Figure 4. Left: Time series representations of idealized utterances. Right: Corresponding jaw motions displayed on the 'functional' phase plane, that is, position (x) on the vertical axis and velocity ($\dot{x}$) on the horizontal axis. Parts a, b, and c represent three tokens with vowel-to-vowel periods (P and P') and consonant latencies (L and L') that are not linearly related. Phase position (ϕ) of upper lip movement onset relative to the jaw cycle is indicated (see text; from Kelso & Tuller, 1985/in press).

as a function of time) are shown for an unstressed /bab/ (top left) and a stressed /bab/ (bottom left). On the right are shown the corresponding idealized phase plane trajectories. In this figure we have reversed the typical orientation of the phase plane so that displacement is shown on the vertical axis and velocity on the horizontal axis. Thus, downward movements of the jaw are displayed as downward movements of the phase path. The vertical crosshair indicates all points of zero velocity and the horizontal crosshair indicates zero position (midway between minimum and maximum displacement). As the jaw moves from its highest to its lowest point (from A to C) velocity increases to a local maximum (B), then decreases to zero when the jaw changes direction of movement (C). Similarly, as the jaw is raised from the low vowel /a/ into the following consonant constriction, velocity peaks approximately midway through the gesture (D) then returns to zero (A). It is useful to transform the Cartesian position-velocity coordinates into equivalent polar coordinates, namely, a phase angle, $\phi = \tan^{-1}[\dot{x}/x]$ and a radial amplitude, $R = [x^2 + \dot{x}^2]^{\frac{1}{2}}$. The phase angle is a key concept in the re-analysis of interarticulator timing because it signifies position on a cycle of states.

Notice in Figure 4 that the phase plane trajectory preserves some important differences between stressed and unstressed syllables. For example, maximum displacement of the jaw for the unstressed vowel is less than displacement for the stressed vowel and maximum articulator velocity differs noticeably between these two orbits. In contrast, note that the different durations taken to traverse the orbit as a function of stress are not represented in this description. Time, although implicit and recoverable from the phase plane description, does not appear explicitly. Jaw cycles of different durations are characterized as single orbits on the plane and they are topologically equivalent.

Now one can pose the question of how the upper lip "knows" when to begin its movement for the medial consonant by asking where on the cycle of jaw phase angles the lip motion for medial consonant production begins. One possibility is that upper lip motion begins at the same phase angle of the jaw across different jaw motion trajectories (that is, across rate and stress). In other words, the information for timing of a remote articulator, such as the upper lip, would not be time itself, nor absolute position of another articulator (e.g., the jaw), but rather a relationship defined over the position-velocity state (or, in polar coordinates, the phase angle) of the other articulator. It is also important to recognize that the motion need not be perfectly sinusoidal in order to apply a phase angle analysis. In fact, the jaw motions actually observed are usually not sinusoidal; the displacements at zero velocity are affected by the stress and rate characteristics of the surrounding vowels. For this reason, we normalize each jaw cycle's amplitude and peak velocity to unity.

When the original Tuller and Kelso (1984) data were reanalyzed in this fashion, the result was that phase angle was indeed constant across both rate and stress variations. The complete statistical analysis is presented in Kelso, Saltzman, and Tuller (1986). The mean phase position of the upper lip relative to the jaw was found to be constant and the standard error of the mean tiny. It should be emphasized that a critical phase angle description in no way entails, or is predicted by, the relative timing results. Instead, it constitutes an alternative description of the data set. For example, two utterances that have identical vowel-to-vowel periods and consonant latencies can nonetheless show very different phase positions for upper lip movement onset relative to the cycle of jaw states. Specifically, the phase angle analysis incorporates the full space-time trajectory of motion; the relative timing analysis ignores trajectory, once movement has begun.

There are at least two empirical advantages of this result over our relative timing description and that of others. First, in the relative timing analysis, the overall correlations across rate and stress conditions are very high, but the within-condition slopes tend to vary somewhat. In the phase analysis, on the other hand, the mean phase angle is the same across conditions. Second, although the relative timing scenario is described by two parameters, a slope and an intercept (Figure 3), the phase description requires only a single parameter (phase angle). Thus, if nothing else, the phase description is more parsimonious.

The phase angle conceptualization also has a number of theoretical advantages over our original relative timing analysis. First, once articulatory motions are represented geometrically on the phase plane, duration is normalized across stress and speaking rate. Strictly speaking, the system's topology is unaffected by durational changes. Second, neither absolute nor relative durations have to be extrinsically monitored or controlled in this formulation. There is no need to posit a timing program. This fact potentially provides a grounding for, and a principled analysis of, so-called intrinsic timing theories of speech production (e.g., Fowler et al., 1980; see also Kelso & Tuller, 1985, in press). The present view is bolstered indirectly by demonstrations in the articulatory structures themselves of afferent bases for phase angle information (e.g., position and velocity sensitivities of muscle spindle and joint structures), but not for time-keeping information (e.g., time receptors; cf. Kelso, 1978). It might well be the case that certain critical phase angles provide information for orchestrating the temporal flow of activity among articulators (beyond those considered here) and/or vocal tract configurations.[4] Such phase angles would serve as natural, that is, dynamically specified, information sources for guaranteeing the stability of coordination in the face of scalar (metrical) changes. As in a candle (which provides a metric for time by a change in its length) or a water clock (where the metric is number of drops), the units of time for speech production might be defined entirely in terms of the state variables of the system. Thus, according to the present analysis, it is gestural phase angle (a space-time description) not gestural time (a purely temporal description) that captures the stable cooperative relation among articulators. This essential parameter, phase angle, will take on added significance in the final section.

5. The Change Question

The previous sections focused on the stability characteristics within and among coordinative structures--a description of what remains stable in articulatory ensembles as the metrics of the activity are systematically scaled. Here the other side of the coin is addressed: How do new (or different) forms of spatiotemporal behavior come about? Although the invariance aspect of coordinative structures has been emphasized, it is nevertheless clear that such organizations are not strictly invariant, but change over time according to different time scales. Over the relatively long time span of early childhood, skills are acquired: the forms of motion that emerge must, in turn, be adapted to slow changes in body morphology that accompany growth. As adults, we can learn new skills (within limits), such as tennis and juggling, given sufficient practice. And finally, certain activities involve coordinative structures whose forms change swiftly and dramatically within the performance of particular skill, as with gait transitions in locomotion (e.g., Hoyt & Taylor, 1981). It is this faster kind of change that I want to address here. The reasons are as follows: First, one can design experiments to examine the necessary and sufficient conditions that may underlie such rapid changes in organization (see below). In contrast, slower kinds of change that occur with learning and development often require longitudinal studies and intervening variables can play a significant role. Second, it seems possible that fast and slow changes in perception-action systems follow similar kinds of principles, except that the time-scales are very different. Just as evolution may occur in qualitative "jumps" (see Eldredge & Gould, 1972) so also may skill learning and development. What constitutes a jump at one time scale, however, can be nearly continuous or quasi-static in another. Nevertheless, principles of change may transcend the particular time-scales involved. Third, rapid changes in spatiotemporal behavior may provide a test field for comparing the motor program/central pattern generator account of motor control, with the "movement as cooperative phenomenon" approach promoted here. A fundamental prediction of this approach is that movement patterns, like other cooperative phenomena (see e.g., Haken, 1975; Prigogine, 1980) exhibit qualitatively new modes of organization when certain parameters are scaled past critical bounds. Unlike the motor program construct, however, no a priori prescription exists before the new mode of organization appears (see Kelso, 1981a; Kugler et al., 1980).

The concept of mode is quite crucial here: modes are macroscopic descriptors for collective behavior in systems with many degrees of freedom. Modal descriptions are distinct from those at a microscopic level. For example, an oscillating string made up of 10^{22} atoms is described by "macro" quantities like wavelength and amplitude that are entirely different from the atomistic description (see Haken, 1977). Analogously, in certain biological activities the relative phase among movement components serves as a macroscopic description of the spatiotemporal order, say, among the limbs during the act of locomoting, or the articulators during speech. Thus, particular phasings among the legs of a quadruped correspond to particular modes or locomotory gaits. A microscopic description, on the other hand, requires minimally an identification of the ensemble's neuromuscular elements, their membrane and synaptic properties and all the connections among them. Though it is commonplace for the neuroscientist to talk of neural circuits controlling behavior, it has proved difficult--even in the simplest neural networks--to relate specific patterns of electrical activity to behavioral action. Indeed, if one were to manipulate the parameters of a central pattern generator experimentally, one would be confronted (by some limited estimates) with a space that contains forty-six parameters (Bullock, 1976). Clearly, some other principles--continuous perhaps with the treatment of cooperative phenomena in other natural systems--are needed to guide the selection of relevant parameters.

As Haken, Kelso, and Bunz (1985) note, this problem of relating neuronal events to global behavioral patterns--say, abrupt changes in phase and other characteristic indices of a movement--is reminiscent of problems faced by physicists 50 years ago (and in many cases today as well). Even though the microscopic properties of atoms were thought to be theoretically understood, it still proved difficult to derive the system's macroscopic behavior from its microscopic features. In the field of synergetics, for example, which deals with the formation of order in open, nonequilibrium systems (e.g., Haken, 1975), it has been shown that the behavior of complex systems can be successfully modeled by means of a few macroscopic quantities--called order parameters--in those situations where the system's behavior changes qualitatively.

Elsewhere, we have presented numerous examples--drawn largely from Haken and Prigogine's work--of dissipative or synergetic structures in physics, chemistry, and biology (Kelso & Tuller, 1984a; see also Kelso et al., 1980, and Kugler et al., 1980, for empirical and theoretical treatment of such structures in the realm of action systems). The mechanism common to all these systems is that the values of one or more order parameters become unstable and undergo sudden discontinuous changes when control parameters are scaled (usually under experimental manipulation). The observed bifurcation results from the competition, as it were, between the "forces" or inputs that are systematically scaled (e.g., by increasing the velocity of a treadmill and forcing an animal to move faster), and the "forces" holding the system together (e.g., the order parameter describing, say, a synergistic modal pattern or locomotory gait). Thus, under the influence of continuous scaling, a given mode may suddenly become dominant, and capture or slave (in Haken's terms) the other modes. The significant, and universal feature of such critical behavior is that around transition regions, where stability is lost, the behavior of the system is governed by the order parameters alone. This implies a tremendous reduction in the degrees of freedom since the behavior of all the subsystems is now governed by a single order parameter.

These kinds of sharp, discontinuous behaviors are omnipresent in the action system when system-sensitive parameters are appropriately scaled, e.g., in voluntary limb movements (Kelso, 1981b, 1984), speech (Kelso & Tuller, 1984a), locomotion (e.g., Hoyt & Taylor, 1981; Kugler et al., 1980) and posture (e.g., Nashner & McCollum, 1985; Saltzman & Kelso, 1985). For example, in recent work on bimanual activities, Kelso (1981b, 1984) had subjects move their right and left hands together at a comfortable rate in both an out-of-phase (180 degrees phase difference) and in-phase (zero degrees phase difference) modal pattern, and either with or without an added frictional resistance. The preferred frequencies and amplitudes of each hand were measured under the two resistance conditions. Subjects then attempted to perform the out-of-phase rhythmic movement at steadily increasing frequencies. Of special interest was the critical frequency at which the out-of-phase movements could no longer be sustained, and the rhythmic organization abruptly became in-phase. Although this critical phase transition frequency was different for subjects, when expressed in units of each subject's preferred frequency, the same dimensionless number was obtained. As in many physical and biological systems, new "modes" or spatiotemporal orderings were observed when the system was scaled beyond equilibrium. Continuous scaling on frequency in Kelso's experiments resulted in the initial out-of-phase modal pattern (or phase relation) becoming unstable, until, at a critical point, bifurcation occurred and a different modal pattern appeared. Although not given a bifurcation interpretation, similar results have been obtained by Cohen (1971), MacKenzie and Patla (1983) and Baldissera, Cavallari, and Civaschi (1982).

Recently, Haken et al. (1985) have modeled these bimanual phase transitions, using some of the central concepts and mathematical tools of synergetics and nonlinear oscillator theory. Using relative phase as an order parameter[5] they first specified a potential function corresponding to the layout of modal attractor states (that is, the stable in-phase and out-of-phase patterns), and showed how that layout was altered as a control parameter (driving frequency) was scaled. From the behavior of the potential function they then derived the equations of motion for each hand, and the nonlinear coupling between the hands. Analytic derivations and consequent numerical simulation revealed that if the system was "prepared" in the out-of-phase mode (that is, by instruction to the subject), and driving frequency was increased slowly, the oscillation remained in that mode until the solution of the coupled equations of motion became unstable. At this point, a jump occurred and the only stable stationary solution produced by the system corresponded to the in-phase mode (see Haken et al., 1985, for more details). Ongoing empirical and theoretical work (Kelso & Scholz, 1985; Schöner, Haken, & Kelso, 1986) has revealed that the nonlinear coupling strength as well as fluctuations (both intrinsically generated due to noise in system parameters and extrinsically generated due to an added random forcing function) play an important role in effecting the modal transitions between the hands.

Although it is tempting to ascribe transitions in phasing among the limbs to "switches" or (in the case of gait) a "gait selection process" (Gallistel, 1980), such an account possesses a Kiplingesque "just so" quality. To assign a phenomenon, switching--an abrupt shift in spatiotemporal order--to a device or a mechanism that is said to perform the duty of explaining the phenomenon, is a questionable strategy at best. Yet modal shifts in coordination are often "explained" in this fashion, e.g., by motor programs (cf. Schmidt, 1982, p. 316). The synergetic framework offered here asks instead: What are the necessary and sufficient conditions giving rise to order in biological activities? It is antithetical to views that try to account for complex behaviors by devices that embody (or represent) these behaviors. A principled account of new spatiotemporal patterns should not rest on the introduction of special mechanisms, even when such "mechanisms" are borrowed from current computer technology.

6. Epilogue (after Kelso & Tuller, 1984a)

Unlike machines that are designed by people to exhibit special structures and functions, the structures and functions discussed here develop in a self-organized fashion.[6] Often a new mode emerges when a random event occurs in an unstable region of the system's parameter space and the fluctuation becomes amplified. Such is the case, one suspects, in the gait of a quadruped or in the bimanual experiments described here (see Kelso & Scholz, 1985, for a more complete treatment of critical fluctuations in the bimanual case). Near the unstable region--where it is energetically expensive to maintain a given mode--a small change in speed produces dramatic effects: a new mode arises. Literally, a phase transition occurs.

Throughout the present paper the emphasis has been on similarities--in terms of dynamical behavior--exhibited by articulatory systems that vary widely in their material composition. Common to all of them is their intrinsically nonlinear and dissipative nature, and the fact that they possess many degrees of freedom. These are features that the perception-action system shares with many other natural systems. The focus in this paper has been on the discovery and elaboration of principles that embrace cooperative phenomena, regardless of any particular structural embodiment. From such principles it may be possible to generate an account of the emergence and stability of movement patterns without hermeneutic devices that prescribe such patterns.

References

Abbs, J. H., & Gracco, V. L. (1983). Sensorimotor actions in the control of multimovement speech gestures. Trends in Neuroscience, 6, 393-395.

Abbs, J. H., Gracco, V. L., & Cole, K. J. (1984). Control of multimovement coordination: Sensorimotor mechanisms in speech motor programming. Journal of Motor Behavior, 16.

Abraham, R. H., & Shaw, C. D. (1982). Dynamics-The geometry of behavior. Santa Cruz, CA: Aerial Press.

Asada, H. (1982). A geometrical representation of manipulator dynamics and its application to arm design. In W. J. Book (Ed.), Robotics research and advanced applications (pp. 1-8). New York: American Society of Mechanical Engineers.

Baldissera, F., Cavallari, P., & Civaschi, P. (1982). Preferential coupling between voluntary movements of ipsilateral limbs. Neuroscience Letters, 34, 95-100.

Bernstein, N. A. (1967). The coordination and regulation of movements. London: Pergamon Press.

Bizzi, E., Accornero, N., Chapple, W., & Hogan, N. (1982). Arm trajectory formation in monkeys. Experimental Brain Research, 46, 139-143.

Bizzi, E., Polit, A., & Morasso, P. (1976). Mechanisms underlying achievement of final head position. Journal of Neurophysiology, 39, 435-444.

Boylls, C. C. (1975). A theory of cerebellar function with applications to locomotion. II. The relation of anterior lobe climbing fiber function to locomotor behavior in the cat (COINS Technical Report 76-1). Amherst, MA: University of Massachusetts.

Bullock, T. H. (1976). In search of principles of neural integration. In J. D. Fentress (Ed.), Simple networks and behavior. Sunderland, MA: Sinauer Associates.

Cohen, L. (1971). Synchronous bimanual movements performed by homologous and non-homologous muscles. Perceptual Motor Skills, 32, 639-644

Cooke, J. D. (1980). The organization of simple, skilled movements. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North Holland.

Cordo, P. J., & Nashner, L. M. (1982). Properties of postural adjustments associated with rapid arm movements. Journal of Neurophysiology, 47, 287-302.

Easton, T. A. (1972). On the normal use of reflexes. American Scientist, 60, 591-599.

Eldredge, N., & Gould, S. J. (1972). Punctuated equilibria: An alternative to physical gradualism. In T. J. M. Schopf (Ed.), Models in paleobiology (pp. 82-115). San Francisco: Freeman.

Fel'dman, A. G. (1966). Functional tuning of the nervous system with control of movement or maintenance of a steady posture. III. Mechanographic analysis of execution by man of the simplest motor tasks. Biophysics, 11, 766-775.

Fel'dman, A. G. (1980). Superposition of motor programs. I. Rhythmic forearm movements in man. Neuroscience, 5, 81-90.

Folkins, J. W., & Abbs, J. H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. Journal of Speech and Hearing Research, 18, 207-220.

Folkins, J. W., & Zimmermann, G. N. (1982). Lip and jaw interaction during speech: Responses to perturbation of lower-lip movement prior to bilabial closure. Journal of the Acoustical Society of America, 71, 1225-1233.

Forssberg, H. (1982). Spinal locomotion function and descending control. In B. Sjölund & A. Björkland (Eds.), Brainstem control of spinal mechanisms. New York: Ferström Foundation Series.

Forssberg, H., Grillner, S., & Rossignol, S. (1975). Phase dependent reflex reversal during walking in chronic spinal cats. Brain Research, 55, 247-304.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), Language production. New York: Academic Press.

Fowler, C. A., & Turvey, M. T. (1978). Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In G. Stelmach (Ed.), Information processing in motor control and learning. New York: Academic Press.

Fox, J. L. (1984). The brain's dynamic way of keeping in touch. Science, 225, 820-821.

Gallistel, C. R. (1980). The organization of action: A new synthesis. New York: LEA.

Garfinkel, A. (1983). A mathematics for physiology. American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 245, R4555-R466.

Gelfand, I. M., Gurfinkel, V. S., Fomin, S. V., & Tsetlin, M. D. (Eds.). (1971). Models of the structural-functional organization of certain biological systems. Cambridge, MA: MIT Press.

Gould, S. J. (1980). The evolutionary pursuit of constraint. Daedalus, 109, 39-52.

Greene, P. H., & Boylls, C. C., Jr. (1984). Introduction: Bernstein's significance today. In H. T. A. Whiting (Ed.), Human motor actions: Bernstein reassessed (pp. xix-xxxv). Amsterdam: North-Holland.

Haken, H. (1975). Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems. Review of Modern Physics, 47, 67-121.

Haken, H. (1977). Synergetics: An introduction. Heidelberg: Springer Verlag.

Haken, H. (1983). Advanced synergetics. Heidelberg: Spring-Verlag.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. Biological Cybernetics, 51, 347-356.

Hogan, N. (1981). Impedance control of a robotic manipulator. Paper presented at Winter Annual Meeting of the American Society of Mechanical Engineers, Washington DC.

Hogan, N. (1984). Impedance control: An approach to manipulation. Part II: Implementation. Journal of Dynamic Systems, Measurement and Control.

Hooke, R. (1678). De potentia restitutiva ("of spring"). Cited in S. P. Timoschenko (see below).

Houk, J. C. (1978). Participation of reflex mechanisms and reaction time processes in compensatory adjustments to mechanical disturbances. In J. Desmedt (Ed.), Cerebral motor control in man: Long loop mechanisms. Basel: Karger.

Houk, J. C., & Rymer, W. (1981). Handbook of physiology, Section 1, Vol. II, Motor control, Part 1. In V. B. Brooks (Ed.), (pp. 257-323). Bethesda, MD: American Physiology Society.
Hoyt, D. F., & Taylor, C. R. (1981). Gait and the energetics of locomotion in horses. Nature, 292, 239-240.
Jordan, D. W., & Smith, P. (1977). Nonlinear ordinary differential equations. Oxford: Clarendon Press.
Kelso, J. A. S. (1977). Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 3, 529-543.
Kelso, J. A. S. (1978). Joint receptors do not provide a satisfactory basis for motor timing and positioning. Psychological Review, 85, 474-481.
Kelso, J. A. S. (1981). Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance IX. Hillsdale, NJ: LEA.
Kelso, J. A. S. (1981). On the oscillatory basis of movement. Bulletin of the Psychonomic Society, 18, 63(A).
Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 15, R1000-R1004.
Kelso, J. A. S., & Holt, K. G. (1980). Exploring a vibratory systems analysis of human movement production. Journal of Neurophysiology, 43, 1183-1196.
Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: II. Empirical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 49-70). New York: North-Holland.
Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of nonlinear limit cycle oscillatory processes: Theory and data. Journal of Motor Behavior, 13, 226-261.
Kelso, J. A. S., & Kay, B. A. (in press). Information and control: A macroscopic analysis of perception-action coupling. In H. Heuer & A. F. Sanders (Eds.), Tutorials in perception and action. Amsterdam: North-Holland.
Kelso, J. A. S., Putnam, C. A., & Goodman, D. (1983). On the space-time structure of human interlimb coordination. Quarterly Journal of Experimental Psychology, 35A, 347-376.
Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. Journal of Phonetics, 14, 29-59.
Kelso, J. A. S., & Scholz, J. P. (1985). Cooperative phenomena in biological motion. In H. Haken (Ed.), Complex systems: Operational approaches in neurobiology, physical systems and computers. Berlin: Springer-Verlag.
Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979). On the nature of human interlimb coordination. Science, 203, 1029-1031.
Kelso, J. A. S., & Tuller, B. (1984). Converging evidence in support of common dynamical principles for speech and movement coordination. American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 15, R928-R935.
Kelso, J. A. S., & Tuller, B. (1984). A dynamical basis for action systems. In M. S. Gazzaniga (Ed.), Handbook of cognitive neuroscience (pp. 321-356). New York: Plenum.
Kelso ,J. A. S., & Tuller, B. (in press). Intrinsic time in speech production: Theory, methodology, and preliminary observations. In

E. Keller & M. Gopnik (Eds.), Motor and sensory processes of language. Hillsdale, NJ: Erlbaum.

Kelso, J. A. S., Tuller, B., & Fowler, C. A. (1982). The functional specificity of articulatory control and coordination. Journal of the Acoustical Society of America, 72, S103.

Kelso, J. A. S., Tuller, B., V.-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. Journal of Experimental Psychology: Human Perception and Performance, 10, 812-832.

Kelso, J. A. S., V.-Bateson, E., Saltzman, E., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. Journal of the Acoustical Society of America, 77, 266-280.

Kent, R. D., & Moll, K. (1975). Articulatory timing in selected consonant sequences. Brain and Language, 2, 304-323.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 3-47). New York: North-Holland.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1982). On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), The development of movement control and coordination (pp. 5-78). Chichester: John Wiley.

Landau, L. D., & Lifshitz, E. M. (1981). Theory of elasticity. Oxford: Pergamon.

Love, A. E. H. (1927). A treatise on the mathematical theory of elasticity. New York, 1944: Dover Reprint.

Mackenzie, C. L., & Patla, A. E. (1983). Breakdown in bimanual finger tapping as a function of orientation and phasing. Society for Neuroscience, (Abstract).

Marion, J. B. (1970). Classical dynamics of particles and systems. New York: Academic.

Marsden, C. D., Merton, P. A., & Morton, H. B. (1983). Rapid postural reactions to mechanical displacement of the hand in man. In J. E. Desmedt (Ed.), Motor control mechanisms in health and disease (pp. 645-659). New York: Raven.

Maxwell, J. C. (1877). Matter in motion. New York: Dover Press.

Nashner, L. M., & McCollum, G. (1985). The organization of human postural movements: A formal basis and experimental synthesis. The Behavioral and Brain Sciences, 8, 135-172.

Ostry, D. J., & Munhall, K. (1985). Control of rate and duration of speech movements. Journal of the Acoustical Society of America, 77, 640-648.

Prigogine, I. (1980). From being to becoming: Time and complexity in the physical sciences. San Francisco: W. H. Freeman & Co.

Saltzman, E. L. (in press). Task dynamic coordination of the speech articulators: A preliminary model. Experimental Brain Research Supplementum.

Saltzman, E. L., & Kelso, J. A. S. (1985). Synergies: Stabilities, instabilities, and modes. The Behavioral and Brain Sciences, 8, 161-163.

Saltzman, E., & Kelso, J. A. S. (in press). Skilled actions: A task dynamic approach. Psychological Review.

Schmidt, R. A. (1980). On the theoretical status of time in motor program representations. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland.

Schmidt, R. A. (1982). Motor control and learning: A behavioral emphasis. Champaign, IL: Human Kinetics.
Schmidt, R. A., & McGown, C. (1980). Terminal accuracy of unexpectedly loaded rapid movements: Evidence for a mass-spring mechanism in programming. Journal of Motor Behavior, 12, 149-161.
Schöner, G., Haken, H., & Kelso, J. A. S. (1986). A stochastic theory of phase transitions in human hand movement. Biological Cybernetics.
Shapiro, D. C. (1978). The learning of generalized motor programs. Unpublished doctoral dissertation, University of Southern California.
Shapiro, D. C., Zernicke, R. F., Gregor, R. J., & Diestal, J. D. (1981). Evidence for generalized motor programs using gait-pattern analysis. Journal of Motor Behavior, 13, 33-47.
Sherrington, C. S. (1906). The integrative action of the nervous system. London: Constable.
Shik, M. L., Severin, F. V., & Orlovskii, G. N. (1966). Control of walking and running by means of electrical stimulation of the midbrain. Biophysics, 11, 1011.
Stein, R. B. (1982). What muscle variables does the central nervous system control? The Behavioral and Brain Sciences, 5, 535-577.
Stevens, P. S. (1974). Patterns in nature. Boston: Little and Brown.
Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. Journal of Speech and Hearing Research, 16, 397-420.
Thelen, E., Skala, K. D., & Kelso, J. A. S. (1985). Very young infants' kicking. Haskins Laboratories Status Report on Speech Research, SR-81, 305-313. Also Developmental Psychology, in press.
Timoschenko, S. P. (1953). History of strength of materials. New York: Dover Reprint (1983). 305-313.
Tuller, B., & Kelso, J. A. S. (1984). The timing of articulatory gestures: Evidence for relational invariants. Journal of the Acoustical Society of America, 76, 1030-1036.
Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing: Toward an ecological psychology. Hillsdale, NJ: LEA.
V.-Bateson, E., & Kelso, J. A. S. (1984). Remote and autogenic articulatory adaptation to jaw perturbations during speech: more on functional synergies. Journal of the Acoustical Society of America, 76, S23-S24.
Yates, F. E. (1979). Physical biology: A basis for modeling living systems. Journal of Cybernetics and Information Science, 2, 57-70.

Footnotes

[1]Note that time or duration *per se* plays no explicit role here as a controlled variable. Rather, spatiotemporal pattern arises as a consequence of a dynamic regime in which--at worst--only two system parameters, stiffness and rest length are specified according to task requirements. Movement certainly *evolves* in time, but time is not directly controlled or metered out by a central executive or time keeper in this scenario (cf. Schmidt, 1980, 1982).

[2]Some years ago, this language was not common in the field of motor control. However, it is interesting to quote Greene and Boylls' (1984) assessment of trends in the field--post Bernstein--"that bear watching ...[It] seems likely that the theory of impedance or endpoint control will soon be

recast in terms of potential functions (with endpoints as extrema of such functions to be 'sought,' gradient-fashion, by the state of the skeletomuscular system" (p. xxiii). See Kugler et al., 1980, pp. 34-40; and Hogan, 1981, for applications to robotic motion.

[3]A point that came up during the conference was what these invariances tell us about motor control. Many view an identified invariant as indicating a relevant control parameter. Our position has been exactly the opposite (e.g., Fowler & Turvey, 1978; Kelso et al., 1979): namely, that invariance represents a system constraint, a 'freezing' of degrees of freedom. That is, an invariance tells the investigator what does <u>not</u> have to be directly controlled.

[4]Note that the statement that lip motion starts at a particular phase angle of the jaw is equivalent to saying that it occurs when the rate of dilation of the jaw (that is, $\dot{x}/x$) reaches a particular value. Lee's work (see this volume for references) shows similarly how the inverse of the rate of dilation of the optic image of a surface specifies time-to-contact τ (t) with that surface. The critical phase angle may be the <u>proprioceptive flow field</u> analogue of Lee's optic flow field variable so crucial to the visual guidance of action.

[5]There are several criteria for the identification of an order parameter. A main one is that the order parameter changes much more slowly than the subsystems it is said to govern. Relative phase fits this criterion well. Remember (see Section 4.0) it is the phasing structure of many different activities that remains stable across scalar transformations. Thus, in the bimanual experiments, relative phase changes much more slowly than the kinematic variables describing the motion of each hand.

[6]In fact, neuroscience is beginning to talk this way. A recent report has described systematic changes in topographic maps of sensorimotor cortex that occur due to finger ablation and cortical tissue removal, as evidence that the brain ..."has embedded processes...that make it self-organizing..." And that ..."The dominant view of the nervous system [as] a machine with static properties...[is] incorrect" (Fox, 1984, quoting Merzenich and colleagues' work). Times, it seems are achanging.

THE SPACE-TIME BEHAVIOR OF SINGLE AND BIMANUAL RHYTHMICAL MOVEMENTS: DATA AND MODEL*

B. A. Kay,† J. A. S. Kelso,†† E. L. Saltzman,† and G. Schöner††

Abstract. How do space and time relate in rhythmical tasks that require the limbs to move singly or together in various modes of coordination? And what kind of minimal theoretical model could account for the observed data? Earlier findings from human cyclical movements were consistent with a nonlinear, limit cycle oscillator model (Kelso, Holt, Rubin, & Kugler, 1981), although no detailed modeling was performed at that time. In the present study, kinematic data were sampled at 200 samples/second and a detailed analysis of movement amplitude, frequency, peak velocity, and relative phase (for the bimanual modes, in-phase, and anti-phase) performed. As frequency was scaled from 1 to 6 Hz (in steps of 1 Hz) using a pacing metronome, amplitude dropped inversely and peak velocity increased. Within a frequency condition, the movement's amplitude scaled directly with its peak velocity. These diverse kinematic behaviors were modeled explicitly in terms of low-dimensional (nonlinear) dissipative dynamics with linear stiffness as the *only* control parameter. Data and model are shown to compare favorably. The abstract, dynamical model offers a unified treatment of a number of fundamental aspects of movement, including 1) the postural steady state (when the linear damping coefficient, α, is positive); 2) the onset of movement (when the sign of α becomes negative); 3) the persistence and stability of rhythmic oscillation [guaranteed by a balance between excitation (via $\alpha\dot{x}$, $\alpha < 0$) and dissipation (as indexed by the nonlinear dissipative terms, $\beta\dot{x}^3$ and $\gamma x^2\dot{x}$. This balance determines the limit cycle, a periodic attractor to which all paths in the phase plane $(x,\dot{x})$ converge]; 4) frequency and phase-locking between the hands; and 5) switching among coordinative modes (the latter properties due to a nonlinear coupling structure, see Haken, Kelso, & Bunz, 1985). In short, we show how a rather simple dynamical control structure requiring variations in only one system parameter can describe the spatiotemporal behavior of the limbs moving singly and together. The model is open to further empirical tests, which are underway.

---

*Journal of Experimental Psychology: Human Perception and Performance, in press.

†Also University of Connecticut

††Also Center for Complex Systems, Florida Atlantic University

Acknowledgment. Work on this paper was supported by NINCDS Grant NS-13617, BRS grant RR-05596, and Contract No. N0014-83-K-0083 from the U. S. Office of Naval Research. G. Schöner was supported by a Forschungsstipendium of the Deutsche Forschungsgemeinschaft, Bonn. Thanks to David Ostry, John Scholz, Howard Zelaznik, and three anonymous reviewers for comments.

1. Introduction

How do space and time relate in rhythmical tasks that require the hands to move singly or together in various modes of coordination? And what kind of minimal theoretical model could account for the observed data? The present paper addresses these fundamental questions, which are of longstanding interest to experimental psychology and movement science (e.g., von Holst, 1937/1973; Scripture, 1899; Stetson & Bouman, 1935). It is well known, for example, that discrete and repetitive movements of different amplitude vary systematically in movement duration (provided accuracy requirements are held constant, e.g., Craik, 1947). This and related facts were later formalized into Fitts's Law (1954), a relationship between movement time, movement amplitude, and target accuracy whose underpinnings have been extensively studied (and debated upon) quite recently (e.g., Meyer, Smith, & Wright, 1982; Schmidt, Zelaznik, Hawkins, Frank, & Quinn, 1979).

In the present study, the accuracy of movement is neither fixed nor manipulated as in many investigations of Fitts's Law: only frequency is scaled systematically and amplitude allowed to vary in a natural way. Surprisingly, there has been little research on movements performed under these particular experimental conditions (see Freund, 1983). Fel'dman (1980) reports data from a subject who attempted to keep a maximum amplitude (elbow angular displacement) as frequency was gradually increased to a limiting value (7.1 Hz). An inverse relationship was observed, accompanied by an increasing tonic coactivation of antagonistic muscles. In addition, the slope of the so-called "invariant characteristic" (see also Asatryan & Fel'dman, 1965; Davis & Kelso, 1982)--a plot of joint torque versus joint angle--increased with rhythmical rate, suggesting that natural frequency (or its dynamic equivalent, stiffness) was a controllable parameter. Other studies have scaled frequency, but fixed movement amplitude. Similar to Fel'dman's conclusions, frequency changes over a range were accounted for by an increase in system stiffness (e.g., Viviani, Soechting, & Terzuolo, 1976).

A rather different paradigm that has explored spatiotemporal relationships in cyclic movement patterns has been employed by Brooks and colleagues (e.g., Conrad & Brooks, 1974; see Brooks, 1979, for review). In several studies, monkeys produced rapid elbow flexions/extensions as they slammed a manipulandum back and forth between mechanical stops (thus allowing no variation in amplitude). After a training period, the movement amplitudes were shortened artificially by bringing the stops closer together. The monkeys, however, continued to exert muscular control for the "same" length of time, pressing the handle against the stops when they would normally have produced larger amplitude movements. Since the original rhythm of rapid alterations established during training was maintained in the closer-stop condition, "the rhythm...or some correlate of it" (Brooks, 1979, p. 23) was deemed to be centrally programmed. However, it is not at all clear how these findings or conclusions relate to situations in which subjects are not prevented from adjusting movement amplitude voluntarily in response to scalar increases in rate (see Schmidt, 1985).

Turning to less confined experimental paradigms in which speech and handwriting have been studied, several interesting results have come to light. As speaking rate is increased, for example, the displacement of observed articulator movements is reduced (e.g., Kelso, V.-Bateson, Saltzman, & Kay, 1985; Kent & Moll, 1972; Ostry & Munhall, 1985). The precise nature of the function relating these variables, however, is not known because only a few speaking rates have been employed in such experiments. In handwriting, it is well known that when the amplitude of the produced letter is increased, movement duration remains approximately constant (e.g., Hollerbach, 1981; Katz, 1948; Viviani & Terzuolo, 1980). This handwriting result is theoretically interesting in at least two respects. First, many interacting degrees of freedom are involved in writing a letter, be it large or small, yet quite simple kinematic relations are reproducibly observed at the end effector. Second, because the anatomy and biomechanics are entirely different between writing on notepaper and on a blackboard, a rather abstract control structure is implicated.

In the present paper we offer a dynamical model that is entirely consistent with such an abstract control structure and that is shown to reproduce observed space-time relations of limbs operating singly or together (in two specific modes of coordination) quite nicely. Moreover, exactly the same model can be applied to transitions among coordinative modes of hand movement (see below). The present dynamical model is not tied locally and concretely to the biomechanics of the musculoskeletal periphery. Rather, the approach is consistent with an older view of dynamics, namely, that it is the simplest and most abstract description of the motion of a system (Maxwell, 1877, p. 1). It is possible to use such abstract dynamics in complex multidegree of freedom systems when structure or patterned forms of motion arise (e.g., Haken, 1975, 1983). Such patterned regularities in space and time are characterized by low-dimensional dynamics whose variables are called order parameters. One can imagine, for example, the high dimensionality involved in a simple finger movement were one to include a description of participating neurons, muscles, vascular processes, etc., and their interconnections. Yet in tasks such as pointing a finger, the whole ensemble cooperates such that it can be described by a simple, damped mass-spring dynamics for the end effector position. Thus, under the particular boundary conditions set by the pointing task, end position and velocity are the order parameters that fully specify the cooperative behavior of the ensemble. Such "compression," from a microscopic basis of huge dimensionality to a macroscopic, low-dimensional structure, is a general and predominant feature of nonequilibrium, open systems (e.g., Haken, 1983). In the context of movement, it is characteristic of a coordinative structure, viz., a functional grouping of many neuromuscular components that is flexibly assembled as a single, functional unit (e.g. Kelso, Tuller, V.-Bateson, & Fowler, 1984).

In earlier work (e.g., Kelso, Holt, Kugler, & Turvey, 1980; Kugler, Kelso, & Turvey, 1980), we have identified such unitary ensembles--following Fel'dman (1966)--with the qualitative behavior of a damped mass-spring system. Such systems possess a point attractor, that is, all trajectories converge to an asymptotic, static equilibrium state. Thus, the property of equifinality is exhibited, namely, a tendency to achieve an equilibrium state regardless of initial conditions. The control structure for such motion can be characterized by a set of time-independent dynamic parameters (e.g., stiffness, damping, equilibrium position) with kinematic variations (e.g., position, velocity, acceleration over time) emerging as a consequence. This dynamical model has received a broad base of empirical support from studies of single, discrete head (Bizzi, Polit, & Morasso, 1976), limb (e.g., Cooke, 1980; Polit & Bizzi, 1978; Schmidt & McGown, 1980) and finger movement targeting tasks (Kelso, 1977; Kelso & Holt, 1980). In addition, point attractor dynamics can be shown to apply not only to the muscle-joint level but to the abstract, task-level of description as well (see Saltzman & Kelso, in press). That is, a dynamical description is appropriate at more than one "level." Striking support for this notion has been recently accumulated by Hogan and colleagues (see Hogan, 1985). In their work on postural maintenance of the upper extremity, the well known "spring-like" behavior of a single muscle was shown to be a property of the entire neuromuscular system. As Hogan (1985) notes "...despite the evident complexity of the neuromuscular system, coordinative structures...go to some length to preserve the simple 'spring-like' behavior of the single muscle at the level of the complete neuromuscular system" (p. 166).

It is important to emphasize that point attractor dynamics provide a single account of both posture and targeting movements. Hence, a shift in the equilibrium position (corresponding to a given postural configuration) gives rise to movement (see, e.g., Fel'dman, in press). What then of rhythmical movement, our major concern here? It is easy to see, in principle, how a dynamical description might be elaborated to include this case. For example, a single movement to a target may be underdamped, overdamped, or critically damped depending on the system's parameter values (for example, see Kelso & Holt, 1980). A simple way to make the system oscillate would be to change the sign of the damping coefficient to a negative value. This amounts to inserting "energy"[1] into the system. However, for the motion to be bounded, an additional dissipative mechanism must be present in order to balance the energy input and produce stable limit cycle motion. This combination of linear negative damping and nonlinear dissipative components comprise an escapement function for the system that is autonomous in the conventional mathematical sense of a time-independent forcing function.

In the present research we adopt this autonomous description of rhythmical movement, though we do not exclude--on empirical grounds alone--the possibility that forcing may occur in a time-dependent fashion. Oscillator theory tells us that nonlinear autonomous systems can possess a so-called periodic attractor or limit cycle, that is, all trajectories converge to a single cyclic orbit in the phase plane $(x,\dot{x})$. Thus, a non-trivial correspondence between periodic attractor dynamics and rhythmical movement (entirely analogous to the foregoing discussion of point attractor dynamics and discrete movement) is stability in spite of perturbations and different initial conditions.

In a set of experiments several years ago, we demonstrated such orbital stability (along with other behaviors such as mutual and sub-harmonic entrainment) in studies of human cyclical movements (Kelso, Holt, Rubin, & Kugler, 1981). Although our data were consistent with a nonlinear limit cycle oscillator model for both single and coupled rhythmic behavior, no explicit attempt to model the results was made at that time. More recently, however, Haken, Kelso, and Bunz (1985) have successfully modeled the circumstances under which observed transitions occur between two modes of coupling the hands, namely antiphase motion of relative phase ≈ 180 degrees, that involves nonhomologous muscle groups, and in-phase motion of relative phase ≈ 0 deg, in which homologous muscles are used. The Haken et al. (1985) nonlinearly-coupled nonlinear oscillator model was able to reproduce the phase transition, that is, the change in qualitative behavior from antiphase to inphase coordination that occurs at a critical driving frequency, as the driving frequency (ω) was continuously scaled (see Kelso, 1981, 1984; MacKenzie & Patla, 1983). This model has been further extended in a quantitative fashion to reveal the crucial role of phase fluctuations in provoking observed changes in behavioral pattern between the hands and to further identify the phenomenon as a nonequilibrium phase transition (Schöner, Haken, & Kelso, 1986). Remarkably good agreement between Schöner et al.'s (1986) stochastic theory and experiments conducted by Kelso and Scholz (1985) has been found.

In the present work we provide quantitative experimental results pertinent to the foregoing modeling work of Haken et al. (1985) and Schöner et al. (1986). For example, although the Haken et al. (1985) model provided a qualitative account of decreases in hand movement amplitudes with increasing frequency, the actual function relating these variables was not empirically measured in earlier experiments nor was any fit of parameters performed. A goal of this research is to show how a rather simple dynamical model ("control structure")--requiring variations in only one system parameter--can account for the spatiotemporal behavior of the limbs acting singly and together. The experimental strategy was to have subjects perform cyclical movements in response to a metronome whose frequency was manipulated (in 1 Hz steps) between 1 and 6 Hz. The data reveal a reciprocal relationship between cycling frequency and amplitude for both single and bimanual movements that is stable and reproducible. This constraint between the spatial and temporal aspects of movement patterns invokes immediately a nonlinear dynamical model (linear systems exhibit no such constraint), the particular parameters of which can be specified according to kinematic observables (e.g., frequency, amplitude, maximum velocity). Though we make no claims for the uniqueness of the present model, we do show that other models can be excluded by the data as well as suggest explicit ways in which uniqueness may be sought.

2. Methods

2.1 Subjects

The subjects were four right-handed male volunteers, none of whom were paid for their services. They participated individually in two experimental sessions, the sessions being separated by a week. Each session consisted of approximately one hour of actual data collection.

2.2 Apparatus

The apparatus was a modification of one described in detail on previous occasions (Kelso & Holt, 1980; Kelso et al., 1981). Essentially it consisted of two freely rotating hand manipulanda, which allowed flexion and extension about the wrist (radiocarpal) joint in the horizontal plane. Angular displacement of the hands was measured by two DC potentiometers riding the shafts of the wrist positioners. The outputs of the potentiometers and a pacing metronome (see below) were recorded with a 16-track FM tape recorder (EMI SE-7000).

2.3 Procedure

Subjects were placed in a dentist's chair, their forearms rigidly placed in the wrist-positioning device such that the wrist joint axes were directly in line with the positioners' vertical axes. Motion of the two hands was thus solely in the horizontal plane. Vision of the hands was not excluded.

Each experimental session was divided into two sub-sessions. In the first session, single-handed movements were recorded, followed by two-handed movements; this was reversed for the second session. Within each sub-session, preferred movements were recorded, followed by metronome-paced movements. For the preferred trials, subjects were told to move their wrists cyclically "at a comfortable rate." On the paced trials, subjects were told to follow the "beeps" of an audio metronome to produce one full cycle of motion for each beep. Pacing was provided for six different frequencies, 1, 2, 3, 4, 5, and 6 Hz, presented in random order. For both the preferred and paced conditions, subjects were not explicitly instructed concerning the amplitude of movement, e.g., were not told to move their wrists maximally.

For the single-hand subsession there were, therefore, 14 conditions, one preferred and six paced data sets being collected for each hand. For the two-handed trials, there were also 14 conditions, one preferred and six paced data sets being collected for each of two different movement patterns. These bimanual patterns consisted of a mirror, symmetric mode that involved the simultaneous activation of homologous muscles and a parallel, asymmetric mode that involved simultaneous activation of nonhomologous muscle groups (see, e.g., Kelso, 1984). Two trials of data were collected for each condition in each session. For the preferred trials, 30 seconds of data were collected, while 20 seconds were collected at the pacing frequencies of one to four Hz, and six to eight seconds at five and six Hz, to minimize fatigue effects.

2.4 Data Reduction and Dependent Measures

Following the experimental sessions, the movement signals were digitized at 200 samples/second and smoothed with a 35 ms triangular window. Instantaneous angular velocity was computed from the smoothed displacement data via the two-point central difference algorithm, and smoothed with the same triangular window (see Kay, Munhall, V.-Bateson, & Kelso, 1985, for details of the signal processing steps involved). A cycle was defined by the occurrence of two (adjacent) peak extension events, which, along with peak flexions, were identified by a peak-picking algorithm. Peak velocity was measured using the same peak-picker on the velocity data; the values reported here are summaries across both positive and negative velocity peaks. Cycle frequency (in Hz) was defined as the inverse of the time between two peak extensions, and cycle amplitude (peak-to-peak, in deg) as the average of the extension-flexion, flexion-extension half-cycle excursions. For the two-handed trials, the relative phase (or phase difference) between the two hands was also computed on a cycle-by-cycle basis, using Yamanishi, Kawato, and Suzuki's (1979) definition. This is a purely temporal measure, and is not computed from a motion's phase plane trajectory (Kelso & Tuller, 1985). The measurement is based on the temporal location of a left peak extension within a cycle of right hand movement as defined above. In our convention, for the mirror mode, phase differences less than zero deg indicate that the left hand leads the right, and vice versa for positive values. For the parallel, asymmetric mode, values less than 180 deg indicate that the left hand leads the right (i.e., the left peak extension event is reached prior to exactly 180 deg); values greater than 180 deg indicate that the right hand leads. For qualitative comparisons between model-generated simulations and data, phase plane trajectories were also examined. These were created by simultaneously plotting transduced angular position against the derived instantaneous velocity.

After obtaining these measures for each cycle, measures of central tendency (means) and variability across all cycles of each trial were obtained. Coefficients of variation (CVs) were used as variability measures for frequency, amplitude, and peak velocity, in order to remove the effects of the frequency scaling on the mean data and to compare variability data validly across the observed frequency range. The standard deviation was used as the phase variability measure, because coefficients of variation would be clearly inappropriate in comparing the two patterns of movement, whose mean phase differences were always around zero and 180 deg. These within-trial summary data are reported in the following results section because of the large number of cycles collected. In under 1 percent of the trials, a trial was lost due to experimenter error. Thus, for statistical purposes, means across trials within each experimental condition were used.

3. Results

The means and variability measures of frequency (in Hz), amplitude (in deg), peak velocity (in deg/sec) and relative phase (for the two-handed conditions) are presented in Tables 1 to 4, collapsed across trials, sessions, and subjects. Both preferred and paced data are included in these tables.

Table 1

Mean frequency, amplitude, and peak velocity for single-handed trials, collapsed across trial, sessions, and subjects. Average within-trial, cross-cycle coefficients of variation (in percent).

	Frequency (Hz)		Amplitude (Degrees)		Peak Velocity (Degs/sec)	
	L	R	L	R	L	R
Preferred:	2.04	2.04	46.87	46.88	311.91	307.08
	3.8	3.3	7.2	6.4	6.5	6.1
Paced:						
1 Hz	1.00	1.00	51.17	53.54	194.04	187.40
	6.9	4.9	5.8	7.0	8.5	8.7
2 Hz	2.00	2.00	43.11	46.01	291.19	298.62
	3.7	3.3	7.6	7.7	8.2	7.8
3 Hz	3.00	3.00	37.64	40.50	358.17	380.45
	4.7	4.0	10.7	8.1	9.4	7.0
4 Hz	4.02	4.04	38.64	33.54	463.31	416.85
	6.5	4.8	10.7	10.7	9.0	8.6
5 Hz	5.19	5.14	32.82	33.35	540.37	522.10
	7.8	4.9	13.7	9.6	9.8	7.6
6 Hz	6.33	6.01	26.81	27.83	516.89	499.33
	6.9	6.6	21.8	12.9	10.9	10.7

Table 2

Mean frequency, amplitude, and peak velocity for homologous (mirror) two hand trials, collapsed across trial, sessions, and subjects, for the stable data only. Average within-trial, cross-cycle coefficients of variation (in percent).

	Frequency (Hz)		Amplitude (Degrees)		Peak Velocity (Degs/sec)	
	L	R	L	R	L	R
Preferred:	1.90	1.90	41.49	47.05	252.93	280.72
	7.3	6.6	4.0	3.7	7.3	6.6
Paced:						
1 Hz	1.00	1.00	52.71	56.85	188.30	196.60
	3.9	4.0	6.2	6.0	8.6	8.2
2 Hz	2.00	2.00	38.80	42.20	260.85	280.91
	3.5	3.3	9.6	8.1	9.4	7.5
3 Hz	3.01	3.00	33.15	35.85	318.45	345.51
	5.3	4.0	11.0	9.6	9.4	8.1
4 Hz	4.08	4.08	30.50	32.95	387.18	415.44
	8.1	5.7	14.1	11.6	9.5	9.0
5 Hz	5.29	5.25	26.12	29.64	430.64	474.90
	9.7	5.5	17.6	13.5	12.4	11.2

Table 3

Mean frequency, amplitude, and peak velocity for nonhomologous (parallel) two hand trials, collapsed across trials, sessions, and subjects, for the stable data only. Average within-trial, cross-cycle coefficients of variation (in percent).

	Frequency (Hz)		Amplitude (Degrees)		Peak Velocity (Degs/sec)	
	L	R	L	R	L	R
Preferred:	1.56	1.56	52.30	57.50	288.57	314.39
	3.8	4.1	5.7	4.7	6.8	4.9
Paced:						
1 Hz	1.01	1.01	53.22	54.79	196.21	201.96
	4.2	3.9	6.5	5.7	9.3	7.7
2 Hz	2.02	2.00	46.41	48.21	316.15	325.46
	4.4	3.8	9.3	7.7	7.8	7.3

Table 4

Mean phase for homologous (mirror) and nonhomologous (parallel) two hand trials, collapsed across trials, sessions, and subjects. Average within-trial, cross-cycle standard deviations, in parentheses.

	Phase (Degrees) Homologous	Nonhomologous
Preferred:	6.46	185.28
	(11.36)	(11.09)
Paced:		
1 Hz	3.60	177.75
	(6.75)	(9.54)
2 Hz	10.44	185.99
	(10.84)	(16.65)
3 Hz	6.19	188.82
	(18.00)	(52.49)
4 Hz	4.00	193.64
	(26.36)	(93.46)
5 Hz	-5.81	181.68
	(42.53)	(104.02)
6 Hz	5.33	168.88
	(51.91)	(110.38)

3.1 Preferred Conditions

3.1.1 Frequency, Amplitude, and Peak Velocity

For both single and bimanual preferred movements, repeated-measures ANOVAs were performed on the within-trial means and variability measures obtained for frequency, amplitude, and peak velocity. The design was a 2×3×2 factorial, with hand (left, right), movement condition (single, mirror, and parallel), and session as factors.

Mean data: Looking first at frequency means, the only effect found was for movement condition, $F(2,6) = 9.14$, $p < .05$. Post-hoc Scheffé tests show that the single (2.04 Hz) and mirror (1.90 Hz) mode preferred frequencies were similar to each other but higher than the parallel mode frequency (1.56 Hz). The two hands did not differ in preferred frequency in any of the three movement conditions. Turning to amplitude means, a main effect for hand, $F(1,3) = 14.16$, $p < .05$, and a hand by mode interaction, $F(2,6) = 5.81$, $p < .05$, occurred. There was no significant movement condition effect, suggesting that the three movement conditions assumed the same amplitude in the preferred case. However, the interaction indicated that the amplitude means for the single conditions were identical for the two hands, but differed in both bimanual conditions, the left hand assuming a lower amplitude than the right in each case. No significant main effects or interactions were found for the preferred peak velocity data.

Variability data: ANOVAs performed on the frequency and peak velocity within-trial coefficients of variation revealed no effects. For the amplitude CVs, however, there was a significant effect for movement condition, $F(2,6) = 5.17$, $p < .05$. Post-hoc tests showed that single hand amplitudes were more variable than parallel amplitudes, which were more variable than those for mirror movements.

3.1.2 Relative Phase

For the bimanual movement conditions, repeated-measures ANOVAs were performed on the within-trial means and standard deviations of the relative phase between the two hands. The design was a 2×2 factorial, coordinative mode (mirror and parallel) by session. The only effect observed for phase was mode, $F(1,3) = 13756.6$, $p < .0001$, showing that the subjects were indeed performing the task properly, producing two distinct phase relations between the hands. The 95 percent confidence interval for the mirror mode was 6.56 ± 11.34 deg, and for the parallel mode, 185.28 ± 9.93 deg; the intervals overlap with the "pure" modes of zero and 180 deg, respectively (although in both modes the right hand tends to lead the left). There were no effects or interactions for phase variability in the preferred conditions.

3.2 Metronome-paced Conditions

As can be seen in Tables 1-4, the manipulation of movement frequency had a profound effect on almost all the measured observables. With increasing frequency, amplitude decreased, while peak velocity and all variability measures appeared to increase. There were some apparent differences among the three movement conditions as well, although the two hands behaved quite similarly. Valid comparisons among the experimental conditions on the kinematic variables of frequency, amplitude, and peak velocity can only be made, however, when it is established that subjects are actually performing the bimanual tasks in a stable fashion. Looking at Table 4, one can see that the phase variability of the two modes increased quite rapidly with increasing frequency.

In a 6×2×2 factorial design, with pacing frequency (1-6 Hz in one Hz steps), coordinative mode (mirror and parallel), and session as factors, the only effect observed on the mean relative phase data was mode, $F(1,3) = 233.01$, $p < .001$, and the means observed across all pacing frequencies were 4.21 and 182.93 deg in the mirror and parallel modes, respectively. Apparently the two criterion phase angles are approximated, on the average, within trials. However, effects for pacing frequency, $F(5,15) = 124.91$, $p < .0001$, mode, $F(1,3) = 265.75$, $p < .001$, and their interaction, $F(5,15) = 18.24$, $p < .001$, were found on the within-trial relative phase standard deviations. The interaction was consistent with both main effects: variability in phase increased with increasing frequency for both modes, but the parallel mode's variability increased much faster than the mirror mode's. Note, in Table 4, the order of magnitude increase in phase variability in the parallel mode between two Hz and three Hz. A comparable degree of phase variability in the mirror mode is not evident until the six Hz pacing condition. This result is consistent with other findings (e.g., Kelso, 1984; Kelso & Scholz, 1985) that the parallel mode is highly unstable between two and three Hz for similar movements, and a transition to the mirror mode is frequently observed above that frequency.

The foregoing pattern of phase variability suggests, therefore, that we perform two separate analyses on the remainder of the paced data, in order to make comparisons only within the stable regions of behavior. A reasonable criterion for phase stability is ±45 deg. Thus, we now report a) the analyses comparing mirror mode and single hand behavior from one to five Hz and b) the analyses on all three movement conditions for one and two Hz.

3.2.1 Single Hand Versus Mirror Mode Movements, One to Five Hz

For single hand and mirror mode paced movements, repeated-measures ANOVAs were performed on the within-trial means and variability measures obtained for frequency, amplitude, and peak velocity. The design was a 5×2×2×2 factorial, with pacing frequency (1 to 5 Hz in one Hz steps), hand (left, right), movement condition (single and mirror) and session as factors.

Mean data: Looking at the observed frequency means, the pacing frequency was, as expected, a highly significant effect, $F(4,12) = 1117.76$ $p < .0001$. The only other effect present was a weak three-way interaction, session by hand by pacing frequency, $F(4,12) = 4.51$ $p < .05$, indicating some very minor fluctuations in observed frequency. The main feature of this interaction is a simple effect for mode at the three Hz pacing frequency, $F(2,6) = 9.02$, $p < .02$, which was observed for none of the other pacing frequencies.

For the amplitude means, the main effect of pacing frequency, $F(4,12) = 9.51$, $p < .005$, shows that amplitude decreased with increasing frequency. Three of the four subjects' linear correlations between amplitude and frequency were significant, (Pearson *r*s = -.50, -.86, and -.87, *p*s < .001), while the fourth subject's amplitude trend, although decreasing, failed to reach significance ($r = -.18$, $p = .12$). The only other effect on amplitude was a weak three-way interaction, mode by hand by pacing frequency, $F(4,12) = 3.30$, $p < .05$, chiefly the result of the left hand amplitude in the single case at 5 Hz being slightly higher than the rest of the data at that frequency. Otherwise no differences were found, the two movement conditions exhibiting much the same amplitude across the entire frequency range. Pacing frequency, $F(4,12) = 8.26$, $p < .005$, was the only significant effect on the peak velocity means; the latter increased with increasing frequency for both movement conditions.

The main effect of pacing frequency found for both amplitude and peak velocity indicates that each covaries with frequency of movement, but an interesting relationship exists between the two: looking at the means across each pacing frequency, amplitude and peak velocity exhibited an inverse relation (see Figure 1) for both the single hand and mirror movements ($r = -.986$ for the single hands, $r = -.958$ for the mirror movements, on the overall means; $N = 5$ and $p < .01$ for both correlations). At first blush, this result seems to contradict the wealth of findings on this relationship that showed that peak velocity scales directly with movement amplitude (see Kelso & Kay, in press, for a review). However, an analysis of the individual trial data within a given pacing frequency condition indicates that peak velocity and amplitude do indeed scale directly with each other (see Figure 1). Pearson *r* correlations for each of the movement frequencies are listed in Table 5, and range from .772 to .997 ($p < .01$ in all cases). Slopes of the lines of best fit for peak velocity as a function of amplitude are also reported; none of the intercepts were significantly different from zero.

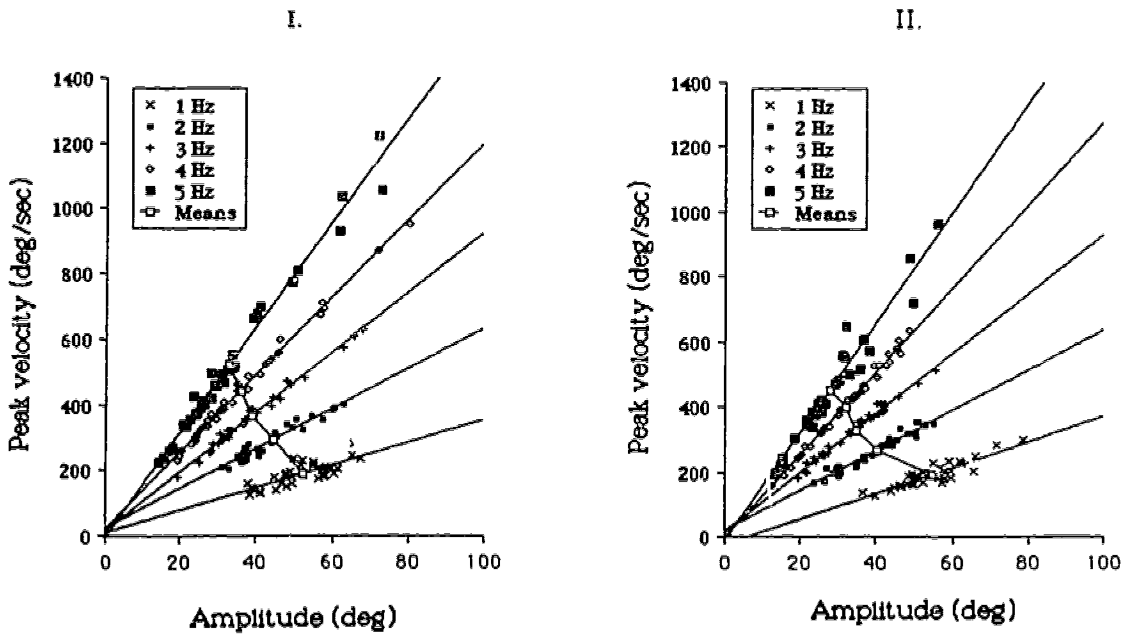


Figure 1. Amplitude (in deg) and peak-velocity (in deg/sec) individual trial data for the 1 to 5 Hz pacing frequencies, and means within each frequency. I. Single hand movements. II. Mirror mode movements.

Table 5

Correlations of amplitude and peak velocity, within each pacing frequency, for stable frequencies. Pearson _r_, slope (_m_) of the line of best fit (peak velocity as a function of amplitude), and number of trials for each correlation are presented.

	Single			Mirror			Parallel		
	_r_	_m_	_N_	_r_	_m_	_N_	_r_	_m_	_N_
1 Hz	.772	3.44	32	.903	3.98	30	.733	4.62	26
2 Hz	.970	6.08	32	.972	6.19	32	.967	6.58	32
3 Hz	.995	9.09	32	.992	9.15	32	-	-	-
4 Hz	.997	11.77	33	.996	12.82	36	-	-	-
5 Hz	.991	15.94	34	.975	16.86	28	-	-	-

Variability data: The within-trial coefficients of variation (CVs) for observed frequency showed significant effects of pacing frequency, $F(4,12) = 13.68$, $p < .0005$, hand, $F(1,3) = 12.59$, $p < .05$, and the pacing frequency by mode interaction, $F(4,12) = 5.92$, $p < .01$. Overall, the left hand was more variable in frequency than the right (CVs of 6.0% and 4.4%, respectively).

Analysis of simple main effects showed that pacing frequency was a significant effect for both single hand and mirror movements, $\underline{F}(4,12) = 3.989$, $\underline{p} < .05$, and $\underline{F}(4,12) = 33.24$, $\underline{p} < .0001$, respectively, but that the only difference between the two movement conditions occurred at three Hz, $\underline{F}(1,3) = 20.18$, $\underline{p} < .05$. At that pacing frequency, the mirror mode was slightly more variable than the single hand movements.

The only significant effect on amplitude CVs was pacing frequency, $\underline{F}(4,12) = 29.10$, $\underline{p} < .0001$. Amplitude variability increased very consistently with increasing movement frequency (see also Figure 1, which shows the cross-trial variability in amplitude as well as in peak velocity). For the peak velocity CVs, session, $\underline{F}(1,3) = 13.10$, $\underline{p} < .05$, and pacing frequency, $\underline{F}(4,12) = 3.51$, $\underline{p} < .05$, were significant effects; the second session's variability was lower than the first's (the only clear-cut practice effect in the experiment), and higher frequency movements were consistently more variable on this measure.

3.2.2 Comparison of All Three Movement Conditions at One and Two Hz

For all three movement conditions, repeated measures ANOVAs were performed on the within-trial means and variability measures obtained for frequency, amplitude, and peak velocity. The design was a 2×2×3×2 factorial, with pacing frequency (one and two Hz), hand (left, right), movement condition (single, mirror, parallel), and session as factors.

Mean data: For the observed frequency, pacing frequency, $\underline{F}(1,3) = 32708.6$, $\underline{p} < .0001$, and mode, $\underline{F}(1,3) = 6.64$, $\underline{p} < .05$, were significant effects, with the parallel mode being slightly faster than the other two movement conditions overall. The difference, however, was less than one percent of the pacing frequency. For amplitude, no main effects or interactions were found; the three movement conditions assumed a single overall amplitude, and amplitude differences were not apparent across the two observed frequencies. For peak velocity, pacing frequency, $\underline{F}(1,3) = 19.32$, $\underline{p} < .05$, and its interactions with movement condition, $\underline{F}(2,6) = 5.92$, $\underline{p} < .05$, and hand, $\underline{F}(1,3) = 15.18$, $\underline{p} < .05$, were significant. A simple main effects analysis for the first of these interactions indicated that the pacing frequency effect was significant for the single and parallel movements, but not for the mirror mode. In addition, the movement conditions differed at two Hz (order from least to greatest peak velocity: mirror, single, parallel) but not at one Hz. The second interaction was consistent with the associated main effects--the pacing frequency effect was significant for both hands, and no simple effects for hand appeared. However, at two Hz the right hand showed slightly greater peak velocities than the left. As observed for single hand and mirror movements (see above), amplitude and peak velocity covaried directly in the parallel movements, within each pacing frequency (see Table 5).

Variability data: For observed frequency, no main effects or interactions were found for the within-trial coefficients of variation (CVs). For amplitude CVs, the movement condition by hand interaction was significant, $\underline{F}(2,6) = 13.51$, $\underline{p} < .05$, yet no simple main effects were found at any level of the two independent variables. However, for the left hand, both bimanual conditions were more variable than single hand movements, while the reverse was true for the right. For peak velocity CVs, the only effect was a weak three-way interaction of movement condition, hand, and frequency, $\underline{F}(2,6) = 7.87$, $\underline{p} < .05$.

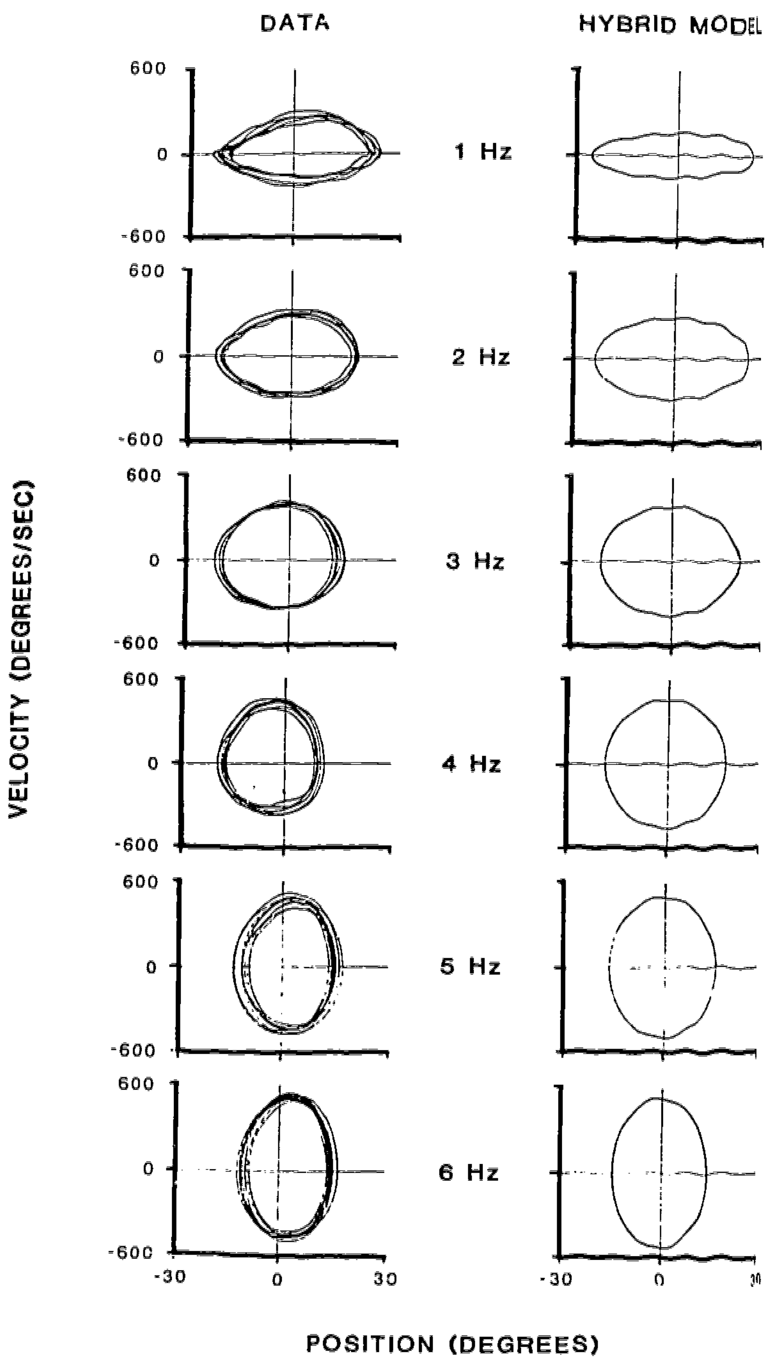


Figure 2. Phase plane trajectories from 1 to 6 Hz. Left: representative examples from the collected data set of one subject. Right: trajectories of the hybrid model (Eq. 4.5), simulated on digital computer.

Analysis of simple main effects showed that pacing frequency was a significant effect for both single hand and mirror movements, $\underline{F}(4,12) = 3.989$, $\underline{p} < .05$, and $\underline{F}(4,12) = 33.24$, $\underline{p} < .0001$, respectively, but that the only difference between the two movement conditions occurred at three Hz, $\underline{F}(1,3) = 20.18$, $\underline{p} < .05$. At that pacing frequency, the mirror mode was slightly more variable than the single hand movements.

The only significant effect on amplitude CVs was pacing frequency, $\underline{F}(4,12) = 29.10$, $\underline{p} < .0001$. Amplitude variability increased very consistently with increasing movement frequency (see also Figure 1, which shows the cross-trial variability in amplitude as well as in peak velocity). For the peak velocity CVs, session, $\underline{F}(1,3) = 13.10$, $\underline{p} < .05$, and pacing frequency, $\underline{F}(4,12) = 3.51$, $\underline{p} < .05$, were significant effects; the second session's variability was lower than the first's (the only clear-cut practice effect in the experiment), and higher frequency movements were consistently more variable on this measure.

3.2.2 Comparison of All Three Movement Conditions at One and Two Hz

For all three movement conditions, repeated measures ANOVAs were performed on the within-trial means and variability measures obtained for frequency, amplitude, and peak velocity. The design was a 2×2×3×2 factorial, with pacing frequency (one and two Hz), hand (left, right), movement condition (single, mirror, parallel), and session as factors.

Mean data: For the observed frequency, pacing frequency, $\underline{F}(1,3) = 32708.6$, $\underline{p} < .0001$, and mode, $\underline{F}(1,3) = 6.64$, $\underline{p} < .05$, were significant effects, with the parallel mode being slightly faster than the other two movement conditions overall. The difference, however, was less than one percent of the pacing frequency. For amplitude, no main effects or interactions were found; the three movement conditions assumed a single overall amplitude, and amplitude differences were not apparent across the two observed frequencies. For peak velocity, pacing frequency, $\underline{F}(1,3) = 19.32$, $\underline{p} < .05$, and its interactions with movement condition, $\underline{F}(2,6) = 5.92$, $\underline{p} < .05$, and hand, $\underline{F}(1,3) = 15.18$, $\underline{p} < .05$, were significant. A simple main effects analysis for the first of these interactions indicated that the pacing frequency effect was significant for the single and parallel movements, but not for the mirror mode. In addition, the movement conditions differed at two Hz (order from least to greatest peak velocity: mirror, single, parallel) but not at one Hz. The second interaction was consistent with the associated main effects--the pacing frequency effect was significant for both hands, and no simple effects for hand appeared. However, at two Hz the right hand showed slightly greater peak velocities than the left. As observed for single hand and mirror movements (see above), amplitude and peak velocity covaried directly in the parallel movements, within each pacing frequency (see Table 5).

Variability data: For observed frequency, no main effects or interactions were found for the within-trial coefficients of variation (CVs). For amplitude CVs, the movement condition by hand interaction was significant, $\underline{F}(2,6) = 13.51$, $\underline{p} < .05$, yet no simple main effects were found at any level of the two independent variables. However, for the left hand, both bimanual conditions were more variable than single hand movements, while the reverse was true for the right. For peak velocity CVs, the only effect was a weak three-way interaction of movement condition, hand, and frequency, $\underline{F}(2,6) = 7.87$, $\underline{p} < .05$.

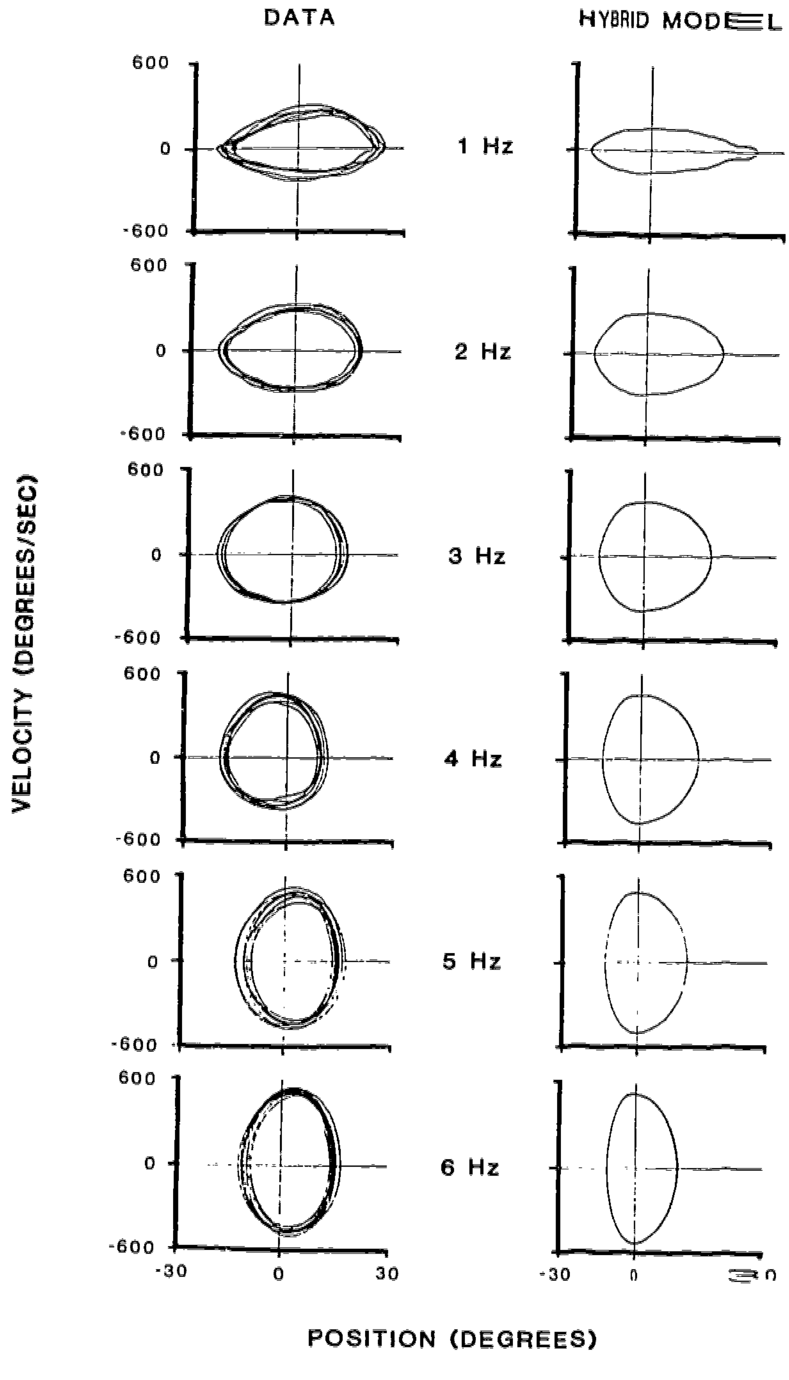


Figure 2. Phase plane trajectories from 1 to 6 Hz. Left: representative examples from the collected data set of one subject. Right: trajectories of the hybrid model (Eq. 4.5), simulated on digital computer.

3.3 Qualitative Results--Examples of Phase Portraits

The shapes of the limit cycle trajectories can be very informative of the underlying dynamics. Figure 2 shows typical phase plane trajectories for single hand movements; a section of one trial is displayed for each of the pacing frequencies from one to six Hz, along with the trajectories of the model (see Section 4) at the same frequencies. As shown in the figure, trajectory shape varies with movement frequency: higher frequency movements appear to be somewhat more sinusoidal (i.e., more elliptical on the phase plane) than lower frequency ones. This was especially apparent in going from one to two Hz. Some subjects showed this tendency less than others, but the shapes of the trajectories did not appear to differ among the three movement conditions. Note also that the velocity profiles are unimodal in these rhythmical movements, a result also observed in recent speech (Kelso et al., 1985) and discrete arm movements (e.g., Bizzi & Abend, 1982; Cooke, 1980; Viviani & McCollum, 1983).

4. Limit cycle models

In this section we first present a limit cycle model that accounts for a number of observed kinematic characteristics of rhythmical hand movements, including the observed amplitude-frequency and peak velocity-frequency relations across conditions, as well as the peak velocity-amplitude relationship within a given pacing condition. In addition, an adequate generalization of the limit cycle model to coordinated rhythmic hand movements is presented (Haken et al., 1985), and conclusions drawn from comparisons with the experimental data. A discussion of the assumptions that are implicit in our modeling strategy is deferred to the General Discussion.

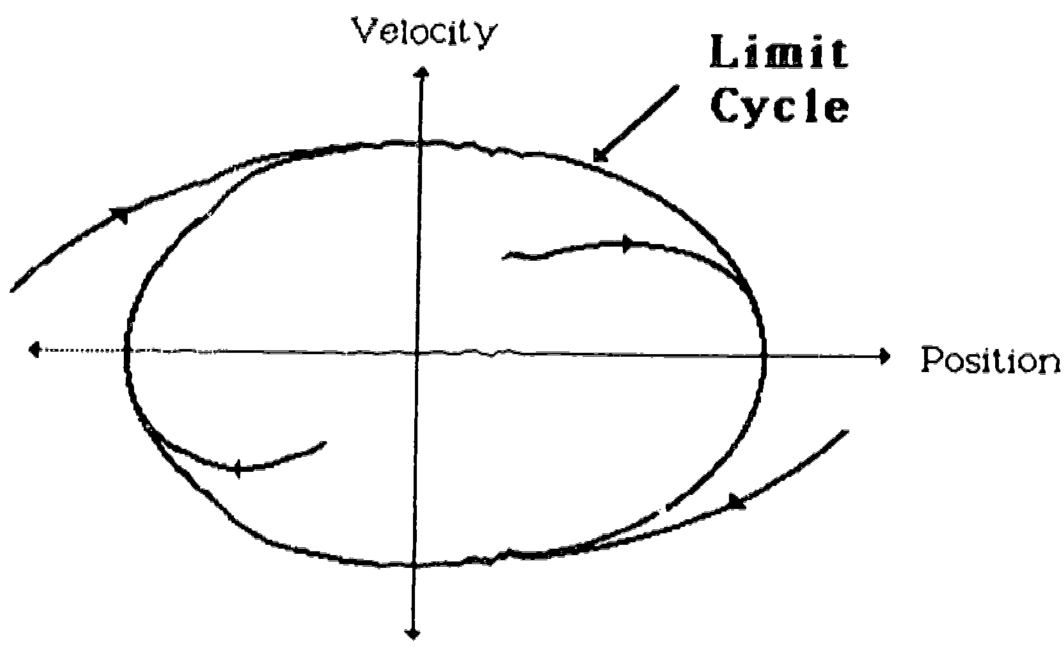


Figure 3. Examples of phase plane trajectories for a limit cycle (see text for details).

As noted earlier by Haken et al. (1985), a combination of two well-known limit cycle oscillators is a strong candidate to model the observed monotonous decrease of amplitude as a function of frequency. These two oscillators are the van der Pol (van der Pol, 1927) and the Rayleigh oscillator (Rayleigh, 1894). The first is described by an equation of motion of the form:

$$\ddot{x} + \alpha\dot{x} + \gamma x^2\dot{x} + \omega^2 x = 0 \qquad (4.1)$$

where α, γ and ω^2 are constants. For $\alpha < 0$ and $\gamma > 0$ this equation has a limit cycle attractor. In a phase portrait in the $(x,\dot{x})$-plane this means that there is a closed curve, on which the system rotates (the limit cycle) and to which all trajectories are attracted after a sufficiently long transient time. For $|\alpha| << \omega$ the frequency of oscillation on and near the limit cycle is, to a good approximation, just ω (see Minorsky, 1962, Sect. 10.6). Figure 3 illustrates this situation schematically. An analytic description of the limit cycle can be given if the slowly varying amplitude and rotating wave approximations are used (Haken et al., 1985; see Appendix 1 for a brief summary of the methods and the results). The amplitude of the limit cycle, which in this approximation is a harmonic oscillation, is found to be:

$$A = 2\sqrt{|\alpha|/\gamma} \qquad (4.2)$$

and is independent of the frequency ω. Thus the van der Pol oscillator can account for the intercept of the amplitude-frequency relation but not for its monotonic decrease. The Rayleigh oscillator has the equation of motion

$$\ddot{x} + \alpha\dot{x} + \beta\dot{x}^3 + \omega^2 x = 0 \qquad (4.3)$$

and possesses a limit cycle attractor for $\alpha < 0$, $\beta > 0$, again with an oscillation frequency ω as long as $|\alpha| << \omega$. Using again the two above-mentioned approximations we obtain the amplitude of this limit cycle as (see Haken et al., 1985):

$$A = (2/\omega)\sqrt{|\alpha|/3\beta} \qquad (4.4)$$

The decrease of amplitude with frequency observed in the data is captured by this expression, although the divergence of (4.4) at small frequency is clearly non-physical.

It is easy to imagine that a combination of both types of oscillators may provide a more accurate account of the experimental results. Therefore, let us consider the following model:

$$\ddot{x} + \alpha\dot{x} + \beta\dot{x}^3 + \gamma x^2\dot{x} + \omega^2 x = 0 \qquad (4.5)$$

which we refer to from now on as the "hybrid" oscillator. For $\beta, \gamma > 0$, $\alpha < 0$ this yields again a limit cycle attractor of frequency ω (for $|\alpha| << \omega$) with amplitude (again in the approximations of Appendix 1):

$$A = 2\sqrt{|\alpha|/(3\beta\omega^2 + \gamma)} \qquad (4.6)$$

This function exhibits both a hyperbolic decrease in amplitude as well as a finite intercept at zero frequency and accounts qualitatively for the experimental data. In Figure 4 we have plotted the amplitude A of the hybrid model together with the experimental data as a function of frequency. The two parameters β and γ were fitted (using a least squares fit, see Footnote 2) while α was chosen as $\alpha = -0.05 * w_{preferred}$ (= .641 Hz) without a further attempt to minimize deviations from the data. (The values for β and γ were: $\beta = .007095\ Hz^3$, $\gamma = 12.457$ Hz, where A was taken to be of the same scale as the experimental degree values.) The choice of α is consistent with the slowly varying amplitude approximation (for which we need $|\alpha| << \omega$; see Appendix 1) and amounts to assuming that the nonlinearity is weak (see Appendix 2 and General Discussion below). For illustrative purposes the corresponding least-squares fits for the van der Pol and the Rayleigh oscillators are also shown in Figure 4.

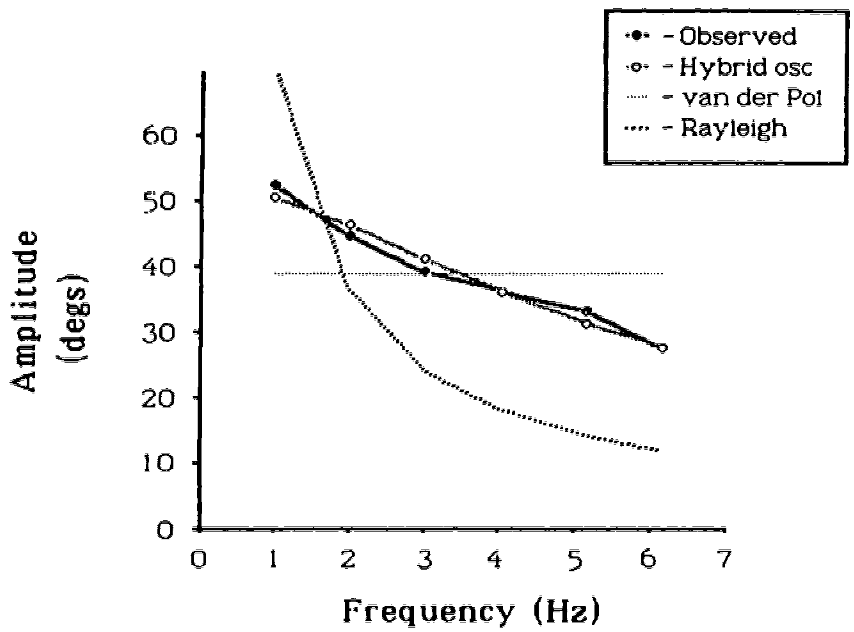


Figure 4. Frequency (in Hz) versus amplitude (in deg) for the single hand data and the curves of best fit for the van der Pol, Rayleigh, and hybrid oscillators (see text). The observed data are the mean values at each pacing frequency.

Note that only one fit parameter, β or γ respectively, was used for these fits. It is obvious how the two foregoing models each account for only one aspect of the experimental observations, and the hybrid accounts for both. In summary, the model parameters were determined by: a) identifying the pacing frequency with ω (which is a good approximation for $|\alpha| << \omega$); b) choosing $\alpha = -0.05 * w_{preferred}$; and c) finding β and γ by a least squares fit of the amplitude-frequency relation. A more stringent evaluation of the parameters is possible if more experimental information is available (see the discussion of the assumptions in General Discussion below). Note, however, that even on this level of sophistication the model accommodates several further features of the data. For example the peak velocity-amplitude relation given by the limit cycle model is the simple relation:

$$V_p = \omega A \tag{4.7}$$

This relation holds whenever the trajectory is close to the limit cycle. Thus if trajectories fluctuate around the limit cycle (due to ever-present small perturbations), we expect the scatter of the peak velocity-amplitude data to lie on a straight line of slope ω. Moreover, this same relation is shown to hold in the situation where amplitude varies <u>across</u> trials (see Figure 1 and Table 5). Note that peak-to-peak amplitude equals 2A so that the slopes reported in Table 5 are ω/2 = π*frequency. An additional piece of experimental information concerns the peak velocity-frequency relation (see Table 1 and Figure 5), the theoretical prediction for which results if we insert (4.6) into (4.7) as follows:

$$V_p = 2\omega\sqrt{|\alpha| \;/\; (3\beta\omega^2 + \gamma)} \tag{4.8}$$

This theoretical curve is also included in Figure 5. It is important to emphasize that all parameters have been fixed previously. Clearly, the match between model and experiment is quite close.

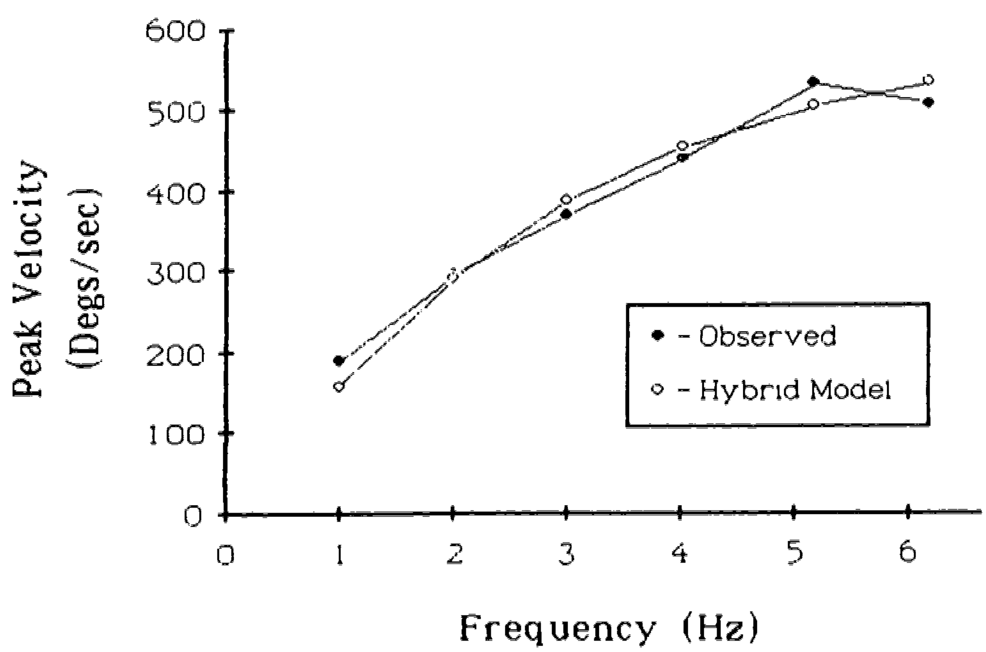


Figure 5. Frequency (in Hz) versus peak velocity (in deg/sec) for the single hand data and the corresponding function for the hybrid model (see Eq. 4.8), as derived from the amplitude-frequency data. The observed data are the mean values at each pacing frequency.

We now turn to the modeling of the two-handed movements. The essential idea is to couple two single hand oscillators of type (4.5) together. Assuming symmetry of the two hands, Haken et al. (1985) have established a coupling structure that accounts for both the in-phase (symmetric/mirror) and the anti-phase (asymmetric/parallel) coordinative modes as well as the transition from an asymmetric to symmetric organization as frequency is scaled (see Introduction). This coupling structure has the following explicit form:

$$\ddot{x}_1 + g(x_1,\dot{x}_1) = (\dot{x}_1 - \dot{x}_2)[a + b(x_1 - x_2)^2] \quad (4.9)$$

$$\ddot{x}_2 + g(x_2,\dot{x}_2) = (\dot{x}_2 - \dot{x}_1)[a + b(x_2 - x_1)^2] \quad (4.10)$$

where

$$g(x,\dot{x}) = \alpha\dot{x} + \beta\dot{x}^3 + \gamma x^2\dot{x} + \omega^2 x \quad (4.11)$$

and a and b are coupling constants. Using again the approximations of Appendix 1 (see Haken et al., 1985, for the calculations), one obtains the amplitudes

$$A_1 = A_2 = 2\sqrt{\frac{|\alpha| + a(1 - \cos\phi)}{3\beta\omega^2 + \gamma - 3b + 4b\cos\phi - b\cos 2\phi}} \quad (4.12)$$

In this expression $\phi = \phi_2 - \phi_1$ is the relative phase of the two oscillators, which is $\phi = \pm 180$ deg for the asymmetric motion and $\phi = 0$ deg for the symmetric motion. Note that for $a = b = 0$ we recover the amplitude of the single hybrid oscillator (see equation (4.6). Indeed, the experimental observation that the amplitudes of the two-handed modes of movement did not differ significantly from the single hand amplitudes (see Sect. 2.1.1) leads us to the conclusion that the coupling is weak in the sense that $a \ll \alpha$ and $b \ll \gamma$. This is an interesting result in that it shows that even when the coupling is much weaker than the corresponding dissipative terms of the single hand oscillators (which guarantee a stable amplitude-frequency relation), phase locking and transitions within phase locking can occur. This may rationalize, to some degree, the ubiquity of phase locking in the rhythmical movements of animals and people and is worthy of much more investigation.

A final remark concerns the preferred frequencies chosen by subjects in the single hand condition compared with the two coordinative modes. The observation was that the preferred frequency was always lower in the asymmetric mode than in either the symmetric mode or the single hand movement conditions, which were roughly equal (see Sect. 2.1.1). As mentioned before, a transition takes place from the asymmetric mode to the symmetric mode as frequency is scaled beyond a certain critical value. The coupled oscillator model accounts for that transition in the sense that the stationary state $\phi = \pm 180$ deg for the relative phase becomes unstable (Haken et al., 1985). In fact, the stability of that state decreases when frequency increases, as exhibited by the relaxation rate of this state (see Schöner et al., 1986, and General Discussion). A simple analysis reveals that the preferred frequency in the asymmetric mode is shifted such that the stability of the relative phase is larger than it would be if the preferred frequency of the single hand oscillation was maintained. This observation may well be important for a fuller understanding of the preferred frequencies, in terms, perhaps, of variational principles such as minimization of energy (see Hoyt & Taylor, 1981; Kelso, 1984).

5. General Discussion

In this paper we have shown how a low-dimensional description in terms of dissipative dynamics can account--in a unified manner--for a number of observed facts. First, the present "hybrid" model includes the well-known mass-spring characteristic of postural tasks (see Introduction). That is, when the linear damping coefficient, α, is positive, the model exhibits a stable equilibrium position in the resting state ($x = 0$, $\dot{x} = 0$ is a point attractor). Second, when the sign of the linear damping coefficient is negative, this equilibrium point is unstable, and an oscillatory solution with a frequency determined by the linear restoring force, $\omega^2 x$, is stable and attracting. The persistence of the oscillation and its stability is guaranteed by a balance between excitation (via $\alpha\dot{x}$ with negative damping coefficient, $\alpha < 0$), and dissipation (as indexed by the nonlinear dissipative terms, $\beta\dot{x}^3$ and $\gamma x^2\dot{x}$). This balance determines the limit cycle, a periodic attractor to which all paths in the phase plane $(x,\dot{x})$ converge from both the inside and the outside. For example, if $\dot{x}$ or x are large, corresponding to a condition outside the limit cycle, the dissipative terms dominate and amplitude will decrease. If, on the other hand, $\dot{x}$ and x are small, the linear excitation term dominates and amplitude will increase (see Figure 3). Third, oscillatory behavior is systematically modified by specific parameterizations, such as those created by a pacing manipulation. The model accounts for the amplitude-frequency and peak velocity-frequency relations with a simple change in one parameter, the linear stiffness ω^2 (for unit mass). Further support for the latter control parameter comes from the direct scaling relation (observed within a pacing condition) of peak velocity and amplitude--a relationship that is now well-established in a variety of tasks (e.g., Cooke, 1980; Jeannerod, 1984; Kelso, Southard, & Goodman, 1979; Kelso et al., 1985; Ostry & Munhall, 1985; Viviani & McCollum, 1983). Thus, a number of kinematic characteristics and their relations emerge from the model's dynamic structure and parameterization. Fourth, and we believe importantly, the same oscillator model for the individual limb behavior can be generalized to the case of coordinated rhythmic action. A suitable coupling of limit cycle (hybrid) oscillators gives rise to transitions among modes of coordination when the pacing frequency reaches a critical value (Haken et al., 1985; Kelso & Scholz, 1985; Schöner et al., 1986).

In summary, the model offers a synthesis of a variety of quite different movement behaviors that we have simulated explicitly on a digital computer (see Figure 2). That is, a successful implementation of the model has been effected that is now subject to further controlled experimentation. One appealing aspect of the model is that it formalizes and extends some of Fel'dman's (1966) early but influential work (see, e.g., Bizzi et al., 1976; Cooke, 1980; Kelso, 1977; Ostry & Munhall, 1985; Schmidt & McGown, 1980). Fel'dman (1966) presented observations on the execution of rhythmic movement that strongly suggested that the nervous system was capable of controlling the natural frequency of the joint using the so-called invariant characteristics--a plot of joint angle versus torque (see also Berkenblit, Fel'dman, & Fukson, in press; Davis & Kelso, 1982). But Fel'dman also recognized that "...a certain mechanism to counteract damping in the muscles and the joint..." must be brought into play, in order to "...make good the energy losses from friction in the system" (1966, p. 774). Our model shows--in an abstract sense--how excitation and dissipation balance each other so that stable rhythmic oscillations may be produced.

On the other hand, in modeling movement in terms of low-dimensional, nonlinear dynamics, we have made certain assumptions that will now be addressed, as they require additional experimental test. For reasons of clarity we list these modeling assumptions systematically:

1) Equifinality. This is a pivotal issue of the entire approach. The very fact that the oscillatory movement pattern can be reached reproducibly from uncontrolled initial conditions indicates--as far as the theory is concerned--that (a) a description of the system dynamics in terms of a single variable (a displacement angle about a single rotation axis) and its derivative is sufficient, that is, there are no hidden dynamical variables that influence the movement outcome and (b) the modeling in terms of a low dimensional description must be dissipative in nature (allowing for attractor sets that are reached independent of initial conditions). An experimental test of the equifinality property consists of studying the stability of the movement pattern under perturbations. Although such stability was observed in earlier studies (Kelso et al., 1981), a much more systematic investigation is now required.

2) Autonomy. A further reduction in the number of relevant variables is possible through the assumption of autonomous dynamics. Nonautonomous forcing--as mentioned in the introduction--essentially represents one additional variable, namely time itself. Apart from the conceptual advantages discussed in the introduction there are experimental ways to test this assumption. One such method consists of studying phase resetting curves, in perturbation experiments (Winfree, 1980). For example, in a system driven by a time-dependent forcing function (e.g., a driven damped harmonic oscillator), perturbations will not introduce a permanent phase shift. On the other hand, if consistent phase shifts are observed in the data, the rhythm cannot be due fundamentally to a nonautonomous driving element.

A strong line of empirical support for the autonomy assumption comes from the transition behavior in the bimanual case, as frequency is scaled (Kelso, 1981, 1984; Kelso & Scholz, 1985). Here autonomous dynamics were able to account for the transition behavior in some detail (Haken et al., 1985; Schöner et al., 1986). Note also that during the transition one or both of the hands must make a shift in phase, a result that would require a not easily understood change in the periodic forcing function(s). That is, one or both "timing programs" would have to alter in unknown ways to accomplish the transition.

3) Minimality. The effective number of system degrees of freedom can be further limited by the requirement that the model be minimal in the following sense: the attractor layout (i.e., the attractors possible for varying model parameters) should include only attractors of the observed type. In the present single hand case, for example, the model should not contain more than a (mono-stable) limit cycle and a single fixed point (corresponding to posture). This limits the dynamics to those of second order: Higher orders would allow, for example, quasiperiodic or chaotic solutions, (e.g., Haken, 1983), which have not been observed thus far.

The above considerations (equifinality, autonomy, minimality) thus constrain the number of possible models considerably. Explicitly, the most general form of the model given these constraints is:

$$\ddot{x} + f(x,\dot{x}) = 0 \tag{5.1}$$

We can illustrate the relation of the hybrid model to the general case (5.1) by expanding f in a Taylor series (assuming symmetry under the operation $x \rightarrow -x$, as inferred to be a good approximation from the phase portraits (Figure 2)), as follows:

$$\ddot{x} = \omega^2 x + \alpha\dot{x} + \beta\dot{x}^3 + \gamma x^2\dot{x} + \delta x\dot{x}^2 + \varepsilon x^3 + O(\dot{x}^5, x\dot{x}^4) \tag{5.2}$$

The hybrid model (4.5) then results from putting $\delta = \varepsilon = 0$.

Our discussion of modeling assumptions can be drawn to a close by remarking that more detailed information about the system dynamics can now be gained by asking experimental questions that are motivated by the theory. For example, in the model the system's relaxation time (i.e., the time taken to return to the limit cycle after a perturbation) is approximately the inverse of α (see Appendix 1), which a simple dimensional analysis reveals to be related to the strength of the nonlinearity (see Appendix 2). Thus, relaxation time measurements can give important information about how and by how much the system supplies and dissipates "energy" in its oscillatory behavior (where energy is to be understood as the integral along x of the right hand side of equation 5.2, see Jordan & Smith, 1977, and Footnote 1). In another vein, it should be recognized that the model's dynamics are entirely deterministic in their present form. Stochastic processes, which have been shown quite recently to play a crucial role in effecting movement transitions (Kelso & Scholz, 1985: Schöner et al., 1986), have not been considered. However, these processes are probably present, as evidenced, for example, in the scatter of amplitudes at a given oscillation frequency. Stochastic properties of rhythmic movement patterns may be explored independent of perturbation experiments by appropriate spectral analysis of the time-series data (see, e.g., Kelso & Scholz, 1985). Elaboration of the model to incorporate stochastic aspects is warranted and is a goal of further research.

A final comment concerns the physiological underpinnings of our behavioral results. With respect to the present model such underpinnings are obscure at the moment. Just as there are many mechanisms that can achieve macroscopic ends, so too there are many mechanisms that can instantiate limit cycle behavior (for a brief discussion, see Kelso & Tuller, 1984, pp. 334-338). The aim here has been to create a model that can realize the stability and reproducibility of certain so-called "simple" movement behaviors. Whatever the physiological bases of the latter our argument is that they must be consistent with low-dimensional dissipative dynamics. There is not necessarily a dichotomy between the present macroscopic account that stresses kinematic properties as emergent consequences of dynamics, and a more reductionistic approach that seeks to explain macrophenomena on the basis of microscopic properties. The basis for explanation of a complex phenomenon like movement may be the same (i.e., dynamical) at all levels within the system, operative, perhaps, at different time scales.

References

Asatryan, D. G., & Fel'dman, A. G. (1965). Functional tuning of the nervous system with control of movement or maintenance of a steady posture - I. Mechanographic analysis on the work of the joint on execution of a postural task. Biophysics, 10, 925-935.

Berkenblit, M. B., Fel'dman, A. G., & Fukson, O. I. (in press). Adaptability of innate motor patterns and motor control mechanisms. The Behavioral and Brain Sciences.

Bizzi, E., & Abend, W. (1982). Posture control and trajectory formation in single and multiple joint arm movements. In J. E. Desmedt (Ed.), Brain and spinal mechanisms of movement control in man. New York: Raven.

Bizzi, E., Polit, A., & Morasso, P. (1976). Mechanisms underlying achievement of final head position. Journal of Neurophysiology, 39, 435-444.

Brooks, V. B. (1979). Motor programs revisited. In R. E. Talbott & D. R. Humphrey (Eds.), Posture and movement (pp. 13-49). New York: Raven.

Conrad, B., & Brooks, V. B. (1974). Effects of dentate cooling on rapid alternating arm movements. Journal of Neurophysiology, 37, 792-804.

Cooke, J. D. (1980). The organization of simple, skilled movements. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland.

Craik, K. J. W. (1947). Theory of the human operator in control systems. I. The operator as an engineering system. British Journal of Psychology, 38, 56-61. II. Man as an element in a control system. British Journal of Psychology, 38, 142-148.

Davis, W. E., & Kelso, J. A. S. (1982). Analysis of invariant characteristics in the motor control of Down's Syndrome and normal subjects. Journal of Motor Behavior, 14, 194-212.

Fel'dman, A. G. (1966). Functional tuning of the nervous system with control of movement or maintenance of a steady posture. III. Mechanographic analysis of execution by man of the simplest motor tasks. Biophysics, 11, 766-775.

Fel'dman, A. G. (1980). Superposition of motor programs. I. Rhythmic forearm movements in man. Neuroscience, 5, 81-90.

Fel'dman, A. G. (in press). Once more on the equilibrium point hypothesis. Journal of Motor Behavior.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 47, 381-391.

Freund, H.-J. (1983). Motor unit and muscle activity in voluntary motor control. Physiological Reviews, 63, 387-436.

Haken, H. (1975). Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems. Review of Modern Physics, 47, 67-121.

Haken, H. (1983). Advanced synergetics. Heidelberg: Springer-Verlag.

Haken, H. (1985). Laser light dynamics. Amsterdam: North-Holland.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. Biological Cybernetics, 51, 347-356.

Hollerbach, J. (1981). An oscillator theory of handwriting. Biological Cybernetics, 39, 139-156.

Hogan, N. (1985). Control strategies for complex movements derived from physical systems theory. In H. Haken (Ed.), Complex systems: Operational approaches in neurobiology, physics, and computers (pp. 156-168). Berlin: Springer-Verlag.

Holst, E. von (1937/1973). On the nature of order in the central nervous system. In The behavioral physiology of animals and man: The collected papers of Erich von Holst. Coral Gables, FL: Univ. of Miami Press.

Hoyt, D. F., & Taylor, C. R. (1981). Gait and the energetics of locomotion in horses. Nature, 292, 239-240.

Jeannerod, M. (1984). The timing of natural prehensile movements. Journal of Motor Behavior, 16, 235-254.

Jordan, D. W., & Smith, P. (1977). Nonlinear ordinary differential equations. Oxford: Clarendon Press.

Katz, D. (1948). Gestaltpsychologie (pp. 124-129). Basel: Schwabe.

Kay, B., Munhall, K. G., V.-Bateson, E., & Kelso, J. A. S. (1985). A note on processing kinematic data: Sampling, filtering, and differentiation. Haskins Laboratories Status Report on Speech Research, SR-81, 291-303.

Kelso, J. A. S. (1977). Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 3, 529-543.

Kelso, J. A. S. (1981). On the oscillatory basis of movement. Bulletin of the Psychonomic Society, 18, 63.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. American Journal of Physiology: Regulatory, Integrative, and Comparative, 246, R1000-R1004.

Kelso, J. A. S., & Holt, K. G. (1980). Evidence for a mass-spring model of human neuromuscular control. In C. H. Nadeau, W.R. Halliwell, K. M. Newell, & G. C. Roberts (Eds.), Psychology of motor behavior and sport. Champaign, IL: Human Kinetics, 408-417.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: II. Empirical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 49-70). New York: North-Holland.

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of nonlinear limit cycle oscillatory processes: Theory and data. Journal of Motor Behavior, 13, 226-261.

Kelso, J. A. S., & Kay, B. (in press). Information and control: A macroscopic basis for perception-action coupling. To appear in H. Heuer & A. F. Sanders (Eds.), Tutorials in perception and action. Amsterdam: North-Holland.

Kelso, J. A. S., & Scholz, J. P. (1985). Cooperative phenomena in biological motion. In H. Haken (Ed.), Complex systems: Operational approaches in neurobiology, physics, and computers (pp. 124-149). New York: Springer-Verlag.

Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979). On the coordination of two-handed movements. Journal of Experimental Psychology: Human Perception and Performance, 5, 229-238.

Kelso, J. A. S., & Tuller, B. (1984). A dynamical basis for action systems. In M. S. Gazzaniga (Ed.), Handbook of cognitive neuroscience (pp. 321-356). New York: Plenum.

Kelso, J. A. S., & Tuller, B. (1985). Intrinsic time in speech production: Theory, methodology, and preliminary observations. Haskins Laboratories Status Report on Speech Research, SR-81, 23-39.

Kelso, J. A. S., Tuller, B., V.-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbation during speech: Evidence for coordinative structures. Journal of Experimental Psychology: Human Perception and Performance, 10, 812-832.

Kelso, J. A. S., V.-Bateson, E., Saltzman, E. L., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. Journal of the Acoustical Society of America, 77, 266-280.

Kent, R. D., & Moll, K. L. (1972). Cinefluorographic analyses of selected lingual consonants. Journal of Speech and Hearing Research, 15, 453-473.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 3-47). New York: North-Holland.

MacKenzie, C. L., & Patla, A. E. (1983). Breakdown in rapid bimanual finger tapping as a function of orientation and phasing. Society for Neuroscience. (Abstract).

Maxwell, J. C. (1877). Matter and motion. New York: Dover Press (1952 reprint).

Meyer, D. E., Smith, J. E., & Wright, C. E. (1982). Models for speed and accuracy of aimed movements. Psychological Review, 89, 449-482.

Minorsky, N. (1962). Nonlinear oscillations. Princeton, NJ: Van Nostrand.

Ostry, D. J., & Munhall, K. (1985). Control of rate and duration in speech. Journal of the Acoustical Society of America, 77, 640-648.

Polit, A., & Bizzi, E. (1978). Processes controlling arm movements in monkeys. Science, 201, 1235-1237.

Rayleigh, J. W. S., 3rd Baron (1894). Theory of sound (Vol. 1). London.

Saltzman, E. L., & Kelso, J. A. S. (in press). Skilled actions: A task dynamic approach. Psychological Review.

Schmidt, R. A. (1985). "Motor" and "action" perspectives on motor behavior: Some important differences, mainly common ground. Paper presented at the conference entitled "Perspectives on motor behavior and control" at Zentrum für interdiszplinare Forschung (Center for Interdisciplinary Research), Universitat Bielefeld, 4800 Bielefeld, West Germany, November.

Schmidt, R. A., & McGown, C. (1980). Terminal accuracy of unexpectedly loaded rapid movements: Evidence for a mass-spring mechanism in programming. Journal of Motor Behavior, 12, 149-161.

Schmidt, R. A., Zelaznik, H. N., Hawkins, B., Frank, J. S., & Quinn, J. T. Jr. (1979). Motor-output variability: A theory for the accuracy of rapid motor acts. Psychological Review, 86, 415-451.

Schöner, G, Haken, H., & Kelso, J. A. S. (1986). A stochastic theory of phase transitions in human hand movements. Biological Cybernetics, 53, 247-257.

Scripture, E. W. (1899). Observations of rhythmic action. Studies from the Yale Psychological Laboratory, 7, 102-108.

Stetson, R. H., & Bouman, H. D. (1935). The coordination of simple skilled movements. Archief Neerderlandica Physiology, 20, 179-254.

Viviani, P., & McCollum, G. (1983). The relation between linear extent and velocity in drawing movements. Neuroscience, 10, 211-218.

Viviani, P., Soechting, J. F., & Terzuolo, C. A. (1976). Influence of mechanical properties on the relation between EMG activity and torque. Journal of Physiology, Paris, 72, 45-52.

Viviani, P., & Terzuolo, V. (1980). Space-time invariance in learned motor skills. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland.

van der Pol, B. (1927). Forced oscillations in a circuit with nonlinear resistance (reception with reactive triode). Philosophical Magazine (7), 3, 65-80.

Winfree, A. T. (1980). The geometry of biological time. New York: Springer-Verlag.

Yamanishi, J., Kawato, M., & Suzuki, R. (1979). Studies on human finger tapping neural networks by phase transition curves. Biological Cybernetics, 33, 199-208.

Footnotes

[1]It is important to emphasize here that we use terms like "energy" and "dissipation" in the abstract sense of dynamical systems theory (cf. Jordan & Smith, 1977; Minorsky, 1962). These need not correspond to any observable biomechanical quantities.

[2]The parameters β and γ were found via a pseudo Gauss-Newton search for the parameters, using the single hand observed frequency and amplitude trial data (N=192). The least-squares criterion was the minimization of squared residuals from the model amplitude-frequency function stated in Equation 3.6. The overall fit was found to be significant, $F(2,190) = 35.314$, $p < .0001$, and the overall R-squared was .2748; standard deviations for β and γ were .001025 Hz^3 and 1.0129 Hz, respectively.

Appendix 1

In this appendix we illustrate some of the basic tools employed in the model calculations in terms of the van der Pol oscillator. For an introduction to such techniques see, e.g., Haken, 1983; Jordan & Smith, 1977; Minorsky, 1962.

The equation of motion of the van der Pol oscillator is again

$$\ddot{x} + \alpha\dot{x} + \gamma x^2\dot{x} + \omega^2 x = 0 \qquad \text{(A1.1)}$$

For small nonlinearity this is very close to a simple harmonic oscillator of frequency ω. The idea here is that the nonlinearity stabilizes the oscillation at a frequency not too different from ω. This suggests a transformation from $x(t)$ and $\dot{x}(t)$ to new variables, namely, an amplitude $r(t)$ and phase $\phi(t)$ ($x(t) = 2r(t)\cos(\omega t\phi(t))$). For ease of computation, we adopt complex notation:

$$x = B(t)e^{i\omega t} + B^*(t)e^{-i\omega t} \qquad \text{(A1.2)}$$

where B is a complex time dependent amplitude, and B^* is its complex conjugate. In this new coordinate system we can define two important approximations to the exact solution (which is unobtainable analytically). The slowly varying amplitude approximation amounts to assuming $|\dot{B}| \ll \omega B$ and is used in a self-consistent manner (see below). The rotating wave approximation (RWA) consists of neglecting terms higher in frequency than the fundamental, such as $e^{3i\omega t}$, $e^{-3i\omega t}$, etc. This means that the anharmonicity of the solution is neglected (this is why the RWA is sometimes also called the harmonic balance approximation). See, for example, Haken (1985) for a physical interpretation of these approximations. Using (A1.2) and these two approximations we obtain for (A1.1):

$$\dot{B} = -\frac{\alpha B}{2} - \frac{\gamma|B|^2 B}{2} \qquad \text{(A1.3)}$$

Introducing polar coordinates in the complex plane,

$$B(t) = r(t)e^{i\phi(t)} \qquad \text{(A1.4)}$$

and separating real and imaginary parts we find:

$$\dot{r} = -\frac{\alpha r}{2} - \frac{\gamma r^3}{2} \qquad \text{(A1.5)}$$

$$\dot{\phi} = 0 \qquad \text{(A1.6)}$$

Equation (A1.5) for the radius r of the limit cycle (which here is a limit circle in the complex plane due to the RWA) has a form that makes visualization of its solutions very simple, namely, it corresponds to the overdamped movement of a particle in the potential:

$$V(r) = \frac{\alpha r^2}{4} + \frac{\gamma r^4}{8} \qquad (A1.7)$$

This potential is illustrated in Figure 6 for $\alpha > 0$ and for $\alpha < 0$, while $\gamma > 0$ in both cases.

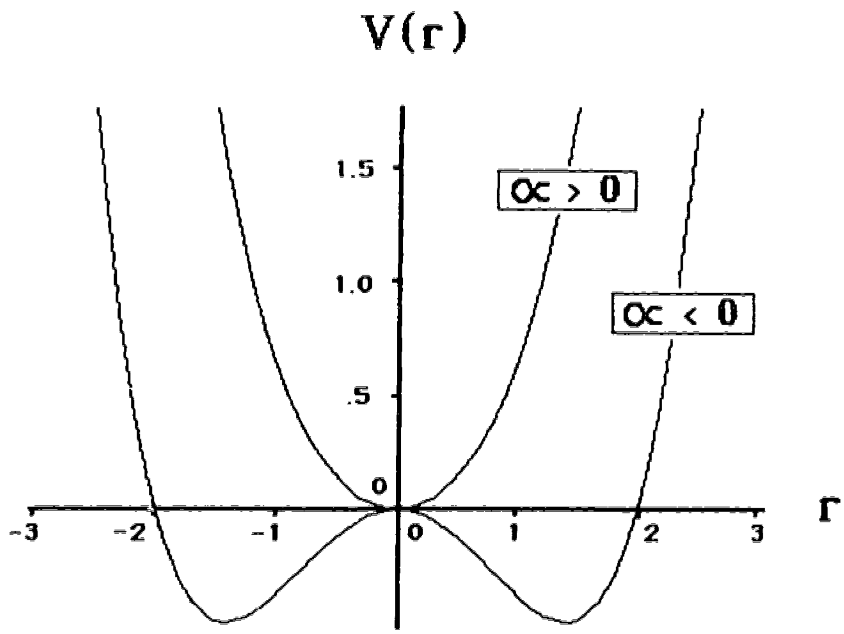


Figure 6. Amplitude potential V as a function of the amplitude, r, for the van der Pol oscillator, when α is less than and greater than zero. Units are arbitrary (see Appendix 1).

Obviously for $\gamma > 0$, the limit cycle of finite amplitude

$$r_0 = \sqrt{|\alpha|/\gamma} \qquad (A1.8)$$

is a stable, stationary solution. A movement with an amplitude close to r_0 relaxes to the limit cycle according to:

$$r(t) = (r(t0) - r_0)e^{-\alpha t} + r_0 \qquad (A1.9)$$

(as can be seen by linearization of (A1.5) around $r = r_0$). Thus this amplitude varies slowly, as long as $|\alpha| << \omega$. This is the above-mentioned self-consistency condition. The time $(1/|\alpha|)$ is called the relaxation time of the amplitude. The equation (A1.6) of the relative phase shows that phase is marginally stable, i.e., does not return to an initial value if perturbed. This can be tested in phase resetting experiments as explained in the General Discussion.

Appendix 2

Here we perform a dimensional analysis to compare different contributions to the oscillator dynamics. To that end we estimate the different forces in the equation of motion (4.5) by their amplitudes when the system is on the limit cycle. The linear restoring force behaves as:

$$\omega^2 x \approx \omega^2 r_o \tag{A2.1}$$

where r_o is the radius of the limit cycle. The linear (negative) damping is:

$$\alpha\dot{x} \approx \alpha\omega r_o \tag{A2.2}$$

The van der Pol nonlinearity is

$$\gamma x^2\dot{x} \approx \gamma\omega r_o^3 \tag{A2.3}$$

while the Rayleigh nonlinearity scales as:

$$\beta\dot{x}^3 \approx \beta\omega^3 r_o^3 \tag{A2.4}$$

Using equation (4.6)

$$r_o = 2\sqrt{|\alpha|/(3\beta\omega^2 + \gamma)} \tag{4.6}$$

as the radius of the hybrid limit cycle, the strength of the nonlinear dissipative terms relative to the linear restoring term is:

$$\frac{\beta\dot{x}^3 + \gamma x^2\dot{x}}{\omega^2 x} \approx \frac{\alpha(\ \beta\omega^2 + \gamma\ )}{\omega(\ 3\beta\omega^2 + \gamma\ )} \tag{A2.5}$$

For either of the simple oscillators this reduces to α/ω.

LANGUAGE MECHANISMS AND READING DISORDER: A MODULAR APPROACH*

Donald Shankweiler† and Stephen Crain†

Abstract. In this paper we consider a complex of language-related problems that research has identified in children with reading disorder and we attempt to understand this complex in relation to proposals about the language processing mechanism. The perspective gained by considering reading problems from the standpoint of language structure and language acquisition allows us to pose specific hypotheses about the causes of reading disorder. The hypotheses are then examined from the standpoint of an analysis of the demands of the reading task and a consideration of the state of the unsuccessful reader in meeting these demands. The remainder of the paper pursues one proposal about the source of reading problems, in which the working memory system plays a central part. This proposal is evaluated in the light of empirical research that has attempted to tease apart structural knowledge and memory capacity both in normal children and in children with notable reading deficiencies.

1. Introduction

There is a growing consensus among researchers on reading that the deficiencies of most children who develop reading problems reflect limitations in the language area, not general cognitive limitations or limitations of visual perception. In this paper we take this for granted.[1] Our concern is with analysis of the language deficiencies that research has identified in poor readers, and with how these deficiencies affect the reading process. Our main goal is to determine whether or not the complex of deficits commonly found in poor readers forms some kind of unity. In order to proceed we will make use of two central ideas. One is the idea of modular organization and the other is the distinction between structure and process. To begin, our conception of reading and its special problems grows out of a biological perspective on language and cognition in which language processes and abilities are taken to be distinct from other cognitive systems. On this perspective, which has long guided research on speech at Haskins Laboratories, the language apparatus forms a biologically-coherent system--in Fodor's terms,

---

*Cognition, in press.
†Also University of Connecticut
Acknowledgment. Portions of this research were supported by NSF Grant BNS 84-18537, and by a Program Project Grant to Haskins Laboratories from the National Institute of Child Health and Human Development (HD-01994). We would like to thank Brian Byrne, Robert Crowder, Alvin Liberman, Virginia Mann, Ignatius Mattingly and three anonymous reviewers for their comments on earlier drafts. The order of the authors' names was decided by a coin toss.

a module (1983)--that is distinguished from other parts of the cognitive apparatus by special brain structures and by other anatomical specializations. An extension of the modularity hypothesis supposes that the language faculty is itself composed of several autonomous subsystems, the phonology, lexicon, syntax, and semantics. These systems, together with a processing system, working memory, constitute the relevant cognitive apparatus. When a person learns to read, this apparatus, which nature created for speech, must be adapted to the requirements of reading.

A modular view of the language mechanism raises the possibility that any number of components of the system might be the source of reading difficulties. At the same time, the fact that these components are related in a hierarchical fashion creates the possibility that a complex of symptoms of reading disorder may derive from a single affected component. Just such a proposal has been offered by M.-L. Kean (1977; 1980) in interpreting the symptom picture in Broca-type aphasia. Kean attributes the agrammatic features in the productions of these aphasics to an underlying deficit at the phonological level. Beyond that, the specific pattern of syntactic errors is predictable from the characteristics of the putative phonological deficit. It is not our intention to defend this particular application of the modularity principle or to assess its empirical adequacy. We mention it as an example of a strategy that can help us to understand the possible connections among the elements of the total symptom picture in poor readers. In later sections, we develop an explanation along similar lines: we interpret the apparent failures of poor readers in syntactic comprehension as manifestations of a low-level deficiency that masquerades as a set of problems extending throughout language. Our account builds also on earlier empirical findings and interpretive discussion of researchers at Haskins Laboratories and on the work of Perfetti and his associates at Pittsburgh.[2]

The second idea that plays an important role in our analysis of reading problems is the distinction between structure and process. By a linguistic structure we mean a stored mental representation of rules and principles corresponding to a formally autonomous level of linguistic knowledge (see Chomsky, 1975). We assume that the language apparatus consists of several structures, hierarchically related, each supported by innately specified brain mechanisms. A processor, crudely put, is a device that brings linguistic input into contact with linguistic structures. The special purpose parsers, which access rules and resolve ambiguities that arise at each structural level of representation--phonologic, syntactic, semantic, lexical--are considered to be linguistic processors. The processor on which much of our discussion focuses is working memory (see Hamburger & Crain, 1984, for related discussion of language processing).

Since reading builds on earlier language acquisition, it is appropriate to begin discussion by considering why the link from the orthography to preexisting language structures and processes should be so difficult for many children to establish. Then we consider the state of the would-be reader who is unsuccessful in meeting the demands of the reading task. The remainder (and largest part) of the paper deals with analysis of poor readers' problems in language comprehension and considers how higher-level problems are related to their difficulties at the level of the word. We review studies that were specifically designed to tease apart deficits in structural knowledge from deficiencies in the working memory system that accesses and manipulates this knowledge. Based on the research findings, we reach the tentative conclusion that a major source of reading difficulties is in working memory processing and in the metalinguistic abilities required to interface the orthography with the existing language subsystems, not a deficit in basic language structures. Throughout this discussion, we emphasize the formative stages of reading, because it is here that the difficulties are most pronounced.

2. Reading Acquisition: Demands of the Task

At first cut, we can roughly identify two levels of processing in reading: (i) deciphering the individual words of the text from their orthographic representations and (ii) processing sentences and other higher-level units of the text. Corresponding to the two levels are two critical kinds of language abilities. The first have to do with forming strategies for identifying the printed word. These may vary in kind with the specific demands posed by different languages and orthographies. Alphabetic orthographies place especially heavy demands on the beginning reader. To gain mastery, the reader must discover how to analyze the internal structure of the printed word and the internal structure of the spoken word, and must discover how the two sets of representations are related. For successful reading in an alphabetic system, the phonemic segmentation of words must become accessible to conscious manipulation, engaging a level of structure of which the listener, qua listener, need never be aware. Explicit conscious awareness of phonemic structure depends on metalinguistic abilities that do not come free with the acquisition of language (Bradley & Bryant, 1983; Liberman, Shankweiler, Fischer, & Carter, 1974; Mattingly, 1972; 1984; Morais, Cary, Alegria, & Bertelson, 1979). The speech processing routines give automatic rapid access to many lexical entries. During the course of learning to read, the orthographic representations of words also become capable of activating this lexical knowledge. But mastery of the orthographic route to the lexicon ordinarily requires a great deal of instruction and practice.

A second set of abilities relate to the syntactic and semantic components of the language apparatus. These abilities take the would-be reader beyond the individual words to get at the meanings of sentences and the larger structures of text. Since reading is compositional, there is an obvious need for some kind of memory in which to integrate spans of words with preceding and succeeding material. The need applies to all languages and orthographies (Liberman, Liberman, Mattingly, & Shankweiler, 1980). Although this is a requirement that reading shares with the perception of spoken sentences, we will argue that reading may make especially severe demands on working memory. Research reviewed in the next section makes it clear that beginning readers are often unable to meet these demands.

3. The State of the Poor Reader

This section draws upon research based on children who have encountered more than the average degree of difficulty in learning to read. Further, since not all of the possible causes of reading failure concern us here (for example, reading problems caused by sensory loss or severe retardation), we have generally required average IQ and a disparity (at least six months for a second-grade child) between the child's measured reading level and the expected level based on test norms. We do not assume that by such means we obtain a tightly homogeneous group. But use of an IQ cutoff and a disparity measure serve to distinguish the child with a relatively specific problem from the child who is generally backward in school subjects, including reading.

The research to which we refer has observed these criteria in selecting the affected subjects. For convenience, we will call them simply "poor readers." Research of the past two decades has identified the following areas of performance in which poor readers characteristically fail or perform at a lower level than appropriately matched good readers.

1. Poor conscious access to sublexical segmentation and poorly developed metalinguistic abilities for manipulation of segments. Beginning readers and older people who have never learned to read do not readily penetrate the internal structure of the word to recover its phonemic structure. Research from several laboratories has shown that weakness or absence of phonemic segmentation ability is characteristic of poor readers and illiterates of all ages (for reviews, see Liberman & Shankweiler, 1985; Morais et al., 1979; Stanovich, 1982; Treiman & Baron, 1981).

2. Difficulties in naming objects. Poor readers frequently have difficulties finding the most appropriate names for objects in speaking (Denckla & Rudel, 1976; Wolf, 1981). They are less accurate than good readers and, under some conditions, also slower. By testing subjects' recognition of the object when the name is given, and by questioning them about the objects they misname, it has been discovered that when the poor reader misnames an object, the problem is less often a semantic confusion than a problem with the name itself. Thus the failure seems to involve the phonological level in some way (Katz, 1986).

3. Special limitations in phonetic perception. Although poor readers usually pass for normal in ordinary perception of spoken language, tests of phonetic perception under difficult listening conditions find them to be less accurate than good readers. For example, it has been found that poor readers were significantly worse than good readers at identifying speech stimuli degraded by noise (Brady, Shankweiler, & Mann, 1983). Since the investigation also found that the poor readers did as well as the good readers in perceiving environmental sounds masked by noise, it is unlikely that a general auditory defect can account for the findings with degraded speech.

4. Deficiencies in verbal working memory. Evidence from several laboratories indicates that children who are poor readers have limitations in verbal working memory that extend beyond the normal constraints (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Mann & Liberman, 1984; Olson, Davidson, Kliegl, & Davies, 1984; Perfetti & Goldman, 1976; Vellutino, 1979). It should be emphasized that these deficiencies are to a large extent limited to the language domain. Other kinds of materials, such as nonsense designs and faces, can often be retained without deficit by poor readers (Katz, Shankweiler, & Liberman, 1981; Liberman, Mann, Shankweiler, & Werfelman, 1982).

Research of the past 20 years offers much evidence that the verbal working memory system exploits phonological structures. It has been shown many times, for example, that the recall performance of normal subjects is adversely affected by making all the items in each set rhyme with one another (Baddeley, 1966; Conrad, 1964, 1972). The strength of the rhyme effect is one indication of the importance of phonological codes for working memory. This prompted members of the reading group at Haskins Laboratories to study children who were good and poor readers on memory tasks while manipulating the phonetic similarity (i.e., confusability) of the stimulus materials (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Mann, Liberman, & Shankweiler, 1980; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). This research has had two major outcomes: first, regardless of whether the stimulus items were presented in printed form or in spoken form, poor readers are consistently worse than good readers in recall of nonconfusable (nonrhyming) items. Second, performance of good readers, like normal adults, is strongly and adversely affected by rhyme; poor readers, on the other hand, typically display only a small relative decrement on the rhyme condition of the recall test.

5. Difficulties in understanding spoken sentences. Failure to comprehend sentences in print that could readily be grasped in spoken form is diagnostic of specific reading disability. Recently, however, it has been found that, under some circumstances, poor readers are less able than good readers even to understand spoken sentences. Special tests employing complex structures are required to bring the difficulties to light (Byrne, 1981; Mann, Shankweiler, & Smith, 1984; Stein, Cairns, & Zurif, 1984; Vogel, 1975). Poor readers have been found to make errors on several syntactic constructions including relative clauses and sentences like John is easy to please, which were contrasted with sentences like John is eager to please (see Section 6).

Having briefly surveyed the performance characteristics of poor readers, we see that their problems are dispersed throughout language. However, it is important to appreciate that the five problem areas are not independent. Although not every one may be demonstrable in all poor readers, the deficits clearly tend to co-occur. There is much evidence, moreover, that difficulties at the level of the word are a common denominator; word recognition measures of reading account for a large portion of the variance in comprehension-related measures of reading (Perfetti & Hogaboam, 1975; Shankweiler & Liberman, 1972). Thus the problems at higher levels would appear to be associated with problems at lower levels.

Researchers at Haskins Laboratories have argued that underlying this diversity in symptoms may be a common problem at the level of the phonology. It is clear that problems (1)-(3) can be seen as manifestations of poor readers' failure to use phonologic structures properly. On the face of it, a different kind of explanation might seem to be required for problems in working memory (4) and in understanding complex spoken sentences (5). However, it has also long been supposed that the verbal working memory system, which is deficient in poor readers, is a faculty that is phonologically-grounded (Conrad, 1964, 1972). Moreover, it has been suggested, in keeping with this view, that poor readers' problems in sentence processing may reflect working memory limitations, and, by extension, phonological limitations (Liberman & Shankweiler, 1985; Mann et al., 1980, 1984).

In what follows we pursue the possibility that all the "symptoms" noted in the preceding section are reflections of a unitary underlying deficit. Our goal is to explain why poor readers sometimes fail to comprehend even spoken language as well as good readers, by asking to what extent problems at the sentence level may be related to problems at the level of the word. It should be emphasized that failure to comprehend a sentence correctly does not necessarily indicate an absence of critical syntactic structures. Understanding a sentence is a complex task in which both structures and processors are engaged. Examples of their interdependence can be found in recent research findings in language acquisition in which young children failed to comprehend complex sentences in some tests, yet were shown (under favorable test conditions) to have the necessary structures. Thus, errors that on the surface might appear to be syntactic have been found, on a closer analysis, to be a result of processing limitations. Later, we discuss some of this research and we show that the same problems of interpretation arise when we encounter failures of sentence understanding in older children who are poor readers.

4. Two Hypotheses About the Source of Reading Difficulties

In order to bring the research on poor readers into sharper focus, we distinguish what we take to be the major alternative positions concerning the relationships between language acquisition and reading. Broadly, two positions can be distinguished: one hypothesis proposes delays in the availability of critical structures; the alternative hypothesis emphasizes processing limitations. Since both are idealized positions, they are not intended to represent fully the views of any individual. We adopt this device because it allows us to draw out differences in the research literature that we believe are fundamental, but that often go unrecognized.

4.1 The Structural Lag Hypothesis

In its most general form, the first hypothesis supposes that reading demands more linguistic competence than many beginning readers command. Although learning to speak and learning to read are continuous processes, some researchers have supposed that reading requires more complex linguistic structures than early speech development. On this view, at the age at which children begin to learn to read, some are still lacking part of the necessary structural knowledge. It is assumed that the inherent complexity of certain structures makes them unavailable until the would-be reader has had sufficient experience with sentences that contain these structures. Thus, this hypothesis about the sources of reading difficulty rests on two assumptions about language acquisition: 1) that linguistic materials are ordered in complexity, and 2) that language acquisition proceeds in a stepwise fashion, beginning with the simplest linguistic structures and culminating when the most complex structures have been mastered.

An advocate of this view might point to evidence of late maturation of the spoken-language competence of poor readers, including late-maturing structures that are required for interpreting complex sentences (see e.g., Byrne, 1981; Fletcher, Satz, & Scholes, 1981; Stein et al., 1984; Vogel, 1975). One might also propose that reading engages linguistic structures or rules that require special experience for their unfolding. The earliest developments in language acquisition require only immersion in a speaking environment; instruction is unnecessary, even irrelevant. In contrast, the later development of language, as well as the early stages of reading, may require more finely-tuned experience.

Since this hypothesis turns out to be more appropriate for some levels of linguistic knowledge than for others, we consider two variants, one at the level of syntax and the other at the level of phonology.

4.1.1 The Syntactic Lag Hypothesis

We ask first what consequences a syntactic delay would have for beginning readers. Let us suppose, for example, that children who are at the age at which reading instruction normally begins have not yet mastered the syntactic rules needed for generating restrictive relative clauses (e.g., "who threw the game" in The referee who threw the game...). It is clear that these children would be unable to learn to read sentences containing relative clauses. A deficiency at this level, then, would establish a ceiling on the abilities of poor readers to comprehend text. Further, the impact of a lag in syntactic knowledge would presumably show up in processing spoken sentences; it could hardly be limited to reading. However, a syntactic deficiency could not explain why poor readers have problems at lower levels of language processing, such as deficits in phonologic analysis and orthographic decoding.

It is apparent then that this hypothesis, by itself, cannot explain why some children have special problems learning to read. If poor readers do in fact have structural deficits at the syntactic level, their reading problems are in no way special. One possibility is that they are a manifestation of a general deficit that depresses all language functions. Another possibility is that poor readers have specific deficits at more than one level of language. In that event, the sentence processing problems of poor readers would simply be unrelated to their deficiencies in orthographic decoding. But if, on the contrary, both the lower-level (orthographic-phonologic) and the higher-level (sentence understanding) problems have a common source in poor readers, then the latter problems could be derivative.

In succeeding sections we make a case for a derivational view by appealing to experimental studies that assess factors influencing the understanding of complex syntactic structures by preschool children and by school-age children who are good or poor readers. First, however, we must consider another variant of the structural lag hypothesis: the view that reading problems are derived from delay in the appearance of needed phonological structures.

4.1.2 The Phonological Lag Hypothesis

The Phonological Lag Hypothesis draws support from empirical correlations between measures of reading skill derived from reading isolated, unconnected words and those derived from reading text with comprehension. There is abundant evidence, as we noted, that word recognition measures account for a large portion of the variance in comprehension-related measures of reading. Since, in addition, there is also evidence pointing to a close link between phonological segmentation abilities and ability to decode words orthographically, the hypothesis that the root problem for many poor readers is a structural deficiency at the phonological level has much to recommend it. It provides a theoretically coherent and empirically testable framework for research and it is consistent with many empirical findings on successful and unsuccessful readers.

There are strong grounds, then, for supposing that orthographic decoding abilities and the phonological knowledge on which they rest are necessary for reading mastery. But are they sufficient? Are orthographic decoding skills the only new thing a would-be reader must acquire in order to read with understanding up to the limit set by spoken-language comprehension? To suppose so would assume that the other abilities needed for understanding printed text are already in place and have long been in use in understanding spoken language. But such an assumption would appear to ignore the other two components of the symptom picture in poor readers: deficiencies in temporary verbal memory and failures in understanding complex spoken sentences. Therefore, at this juncture, we take another direction, and examine the alternative hypothesis that all the problems of poor readers are reflections of a deficiency in processing, rather than a deficiency in linguistic knowledge.

4.2 The Processing Limitation Hypothesis

The Processing Limitation Hypothesis maintains that all the necessary linguistic structures are mastered before the child begins to learn to read, and therefore that the source of reading difficulty lies outside of the phonological and syntactic components of children's internal grammars. This hypothesis acknowledges that decoding skills, and the metaphonological analytic abilities that support them, are necessary for reading mastery in an alphabetic orthography (the individual who lacks them has no means of identifying words newly encountered in print). On this view, however, these are not the only necessary abilities. The Processing Limitation Hypothesis asserts that an additional skill is required by the internal language apparatus in order to interface an alphabetic orthography with preexisting phonological and morphological representations: the efficient management of working memory. This is needed for sentence understanding, both in reading and in spoken language, to bring about integration of the component segments for assembly of higher-level linguistic structures of syntax and semantics.

On this hypothesis, learning to process language in the orthographic mode places extra burdens on working memory with the result that, until the reader is quite proficient, comprehension of text is more limited than comprehension of spoken sentences. It is assumed that speech processing is usually automatic in the beginning reader. One consequence of automaticity is that processing spoken sentences, including even many complex syntactic structures, is conserving of working memory resources. Reading, on the other hand, is extremely costly of these resources until the reader has sufficient mastery of orthographic decoding skills. Moreover, the existence of working memory impairment adds another dimension to the picture of the poor reader. Given sentences that pose unusual memory demands, a poor reader with this impairment can be expected to manifest language deficits that extend beyond reading, involving comprehension of spoken language. In Section 6 we discuss the possibility that the structures that have been found to be stumbling blocks for poor readers in previous research are in fact structures that tax working memory resources.

In contrast to the Structural Lag Hypothesis, the Processing Limitation Hypothesis can, in principle, account for all the basic facts about reading acquisition. Therefore, in the following sections, we adopt this standpoint and we draw out its implications.

5. The Language Processing Mechanism

Since the Processing Limitation Hypothesis assigns an essential role to linguistic memory, it will be useful to sketch our conception of temporary verbal memory. Then we turn to consider the language processing system, and the place of verbal memory in it.

5.1 Short-term Memory versus Working Memory

First, we emphasize that we do not equate "short-term memory" and "working memory," although the former is partly subsumed by the latter. Verbal short-term memory is commonly seen as a passive storage bin for information, whereas working memory is seen as an active processing system, although it has a storage component. Short-term memory is commonly understood as a static system for accumulating and holding segments of speech (or orthographic segments) as they arrive during continuous listening to speech or during reading. This form of memory is verbatim, but highly transient. Presented items are retained in the order of arrival, but are quickly lost unless the material is maintained by continuous rehearsal. Material in short-term memory can also be saved if it can be restructured into some more compact representation (replacing the verbatim record). Put another way, the system is limited in capacity, but the limits are rendered somewhat elastic if opportunities exist for grouping its contents. Finally, it has long been recognized that a phonetic code is important for maintaining material in short-term memory.

In place of the storage bin conception, some workers (Baddeley, 1979; Baddeley & Hitch, 1974; Daneman & Carpenter, 1980; Perfetti & Lesgold, 1977) have argued for a more dynamic notion, endowing this form of memory with processing and not merely storage functions. This conception of working memory makes it an active part of the language processing system. Working memory is seen to play an indispensable role in comprehension both of spoken discourse and printed text (Liberman, Mattingly, & Turvey, 1972).

On the simplest analysis, working memory has only two working parts, although it has access to several linguistic structures. One component is a storage buffer where rehearsal of phonetically coded material can take place. The buffer has the properties commonly attributed to short-term memory. Its phonological store can hold unorganized linguistic information only briefly, perhaps for only one or two seconds. Given this limitation, working memory cannot efficiently store unorganized strings of segments.

The second component of working memory plays an "executive" role (Baddeley & Hitch, 1974). This component has received comparatively little attention, so its exact functions are still opaque. Pursuing an analogy with the compiling of programming languages, we view it as a control mechanism that is capable of fitting together "statements" from the phonological, syntactic, and semantic parsers. As we conceive of it, the control structure integrates written or spoken units of processing with preceding and succeeding material. It facilitates the organization of the products of lower-level processing by relaying information that has undergone analysis at one level to the next-higher level. The first duty of the control mechanism is to transfer phonologically analyzed material out of the buffer and push it upwards through the higher level parsers, thus freeing the buffer for succeeding material. In reading, it is this transfer of information that is constrained by the level of orthographic decoding skill, according to the Processing Limitation Hypothesis.[3]

5.2 Working Memory and the Language Processing Mechanism

The thesis of modular organization of the language system leads us to expect a specific memory component for linguistic material. The question of domain-specific systems of memory has been the subject of considerable research. A good case can be made for the existence of a memory system that is specialized for verbal material. It has been found, in this regard, that verbal retention is selectively impaired by damage to critical regions of the left dominant cerebral hemisphere; damage to corresponding portions of the right nondominant hemisphere results in selective impairments of nonverbal material, such as abstract designs and faces (Corsi, 1972; Milner, 1974). The finding of dissociated memory deficits fits neatly with evidence discussed above, that the memory limitation in poor readers is restricted to linguistic materials.

Although the neuropsychologic evidence clearly points to the existence of a specific verbal memory system, we must ask, nevertheless, whether this system is a part of the language module. On Fodor's (1983) view, the language module as a whole is an "input system": its operations are fast; they are mandatory; they are largely sealed off from conscious inspection; they are also insulated from cognitive inferencing mechanisms external to language. Working memory, as we understand it, does not conform to all of these criteria. Some of its operations consume appreciable time, and some are open to conscious inspection, as in the rehearsal and reanalysis of linguistic material. Nevertheless, it seems to us that working memory belongs in the language module by reason of its intimate association with the parsers that assign phonological, syntactic, and semantic structure to linguistic input. In so far as the working memory system is understood to be a part of the language module, albeit as an "output system," we are forced to differ with Fodor's characterization of the language processing mechanism. For purposes of further discussion, though, we will assume that working memory is part of the language module.

In addition to its storage and rehearsal functions, working memory, as we have characterized it, controls the unidirectional flow of linguistic information through the series of parsers from lower levels to higher levels in the system. Each parser is taken to be a processor that accesses rules and principles corresponding to its level of representation. Each is, roughly, a function from input of the appropriate type to structural descriptions at the given level of representation. We maintain that each of the parsers meets Fodor's criteria for an "input system." Before leaving these architectural matters, we would append a disclaimer: we do not assume that higher-level processors beyond semantic parsing are accessed by the working memory system. Reasoning, planning actions, inference, and metalinguistic operations are not taken to be parts of the language module, though they operate on its contents. We emphasize, therefore, that we are using the term "semantics" in a highly restricted sense, to describe the rule system that determines coreference between linguistic constituents, and 'filler-gap' dependencies (see section 6). Crucially, the term is not being used here to refer to real-world knowledge or beliefs.

5.3 Working Memory in Spoken-language Understanding and Reading

It is pertinent to consider how the components of the language module may interact. (We consider spoken language first, and address remarks specific to reading at the end of the section.) It seems reasonable to suppose that both the operations of the fixed-resource parsing mechanisms as well as the operations of the control mechanism of working memory are subject to the constraints of the limited buffer space. Limited space means that the parsers have a narrow window of input data available to them at any one time. On the one hand, understanding sentences clearly requires working memory, because syntactic and semantic structures are composed over sequences of several words. On the other hand, the assignment even of complex higher-level structures is ordinarily conserving of this limited resource; parsing does not ordinarily impose severe demands on memory in understanding speech. The combinatorial properties of the parsing systems are evidently so rapid that they minimize the role of memory in speech understanding.

Under some circumstances, however, working memory constraints apparently do produce problems in syntactic processing, especially in reading. Memory limitations may impair syntactic processing in two ways, corresponding to the two components of the working memory system. Here we build on the insight of Perfetti and Lesgold (1977), who proposed that if the limitations on the working memory are exceeded, for whatever reason, in the service of low-level processing, higher-level processing may be curtailed. This would apply, first, to poor readers who have inherent limitations in buffer capacity (Mann et al., 1980). They would have insufficient capacity to allow higher-level processing to occur uninhibited, although it may not be brought to a complete halt. We should caution, however, that variation among individuals in buffer capacity is not the most important factor in reading, because, in general, tests of rote recall account for only 10% - 25% of the total variance in the measures of reading (Daneman & Carpenter, 1980; Mann et al., 1984). It was this fact that led us to consider the other component of working memory.

A second way that working memory dysfunction can inhibit syntactic processing is by poor control of the flow of information through the system of parsers. The control structure must efficiently regulate the flow of linguistic material from lower- to higher-levels of representation in keeping with the inherent limitation in the buffer space. From the dual structure of working memory, it may be inferred, as Daneman and Carpenter (1980) and Perfetti and Lesgold (1977) have noted, that studies of retention and rote recall of unorganized materials may provide an incomplete and possibly misleading picture of the active processing capabilities of working memory. In relying exclusively on these measures as indices of working memory capacity, researchers may have overlooked a possibly more important source of variation among readers: in our terms this is the problem of regulating the flow of information between the phonological buffer and the higher-level parsers.

Whichever component of the system is most responsible for the functional limitation on working memory, it should be noted that only those sentence processing tasks that impose unusually severe memory demands are expected to offer significant problems for poor readers in spoken language comprehension. On syntactic tasks that are less taxing of this resource, we would expect them to perform as well as good readers. (This prediction is borne out in two studies reviewed in the next section.)

It remains to compare the involvement of working memory in spoken language and in reading. Since reading and speech tap so many of the same linguistic abilities, it is easy to overlook the possibility that reading may pose more difficulties than speech for some of the language apparatus. In reading, the chores of working memory include the on-line regulation of syntactic and semantic analyses, after orthographic decoding and phonologic compiling have begun. Until the reader is proficient in decoding printed words, we contend that reading is more taxing of working memory resources than speech. We are aware, however, of a contrary claim: it is sometimes argued that the permanence of print, in contrast to the transience of speech, should have exactly the opposite effect, with the result that, other things equal, the demands on working memory in processing print should be <u>less</u>. The advantage of print would obtain because the reader can look back, whereas the listener who needs to reanalyze is forced to rely on the fast-decaying memory trace.

In evaluating this argument, we maintain that other things are <u>not</u> equal, and in the case of the beginning reader and the unskilled reader, the inequality favors speech over reading. In either case, what must be considered is the effect of rate of information flow through the short-term memory buffer. If the rate is too fast, as by rapid presentation in the laboratory, information will be lost; if it is too slow, integration will be impaired. An optimal rate of transmission of linguistic information is achieved so often in speech communications because the language mechanisms for producing and receiving speech are biologically matched (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985). As a consequence, speech processing up to the level of meaning is extremely fast (Marslen-Wilson & Tyler, 1980). Perhaps it must be, given the constraints on the memory buffer.

Reading, on the other hand, is fast only in the skilled reader. It is reasonable to suppose, then, that only the skilled reader can take advantage of the opportunity afforded by print, to reanalyze or to verify the initial analysis of a word string. The unskilled reader cannot make efficient use of working memory because of difficulties in orthographic decoding. But until the reader is practiced enough to become proficient, there is no advantage in being able to look back. For these reasons, we would make the prediction that unskilled readers will be less able than good readers to recover from structural ambiguities that induce a wrong analysis (this so-called "garden path" effect is discussed further in the next section). This would hardly be surprising in reading tasks, but since the normal limitations on verbal working memory are magnified in many poor readers, we would expect them to be less able to recover from wrong syntactic analyses even in spoken language.

6. The Role of Working Memory in Failures of Sentence Comprehension

As sketched above, the Structural Lag Hypothesis supposes that linguistic structures are acquired in order of complexity, so that late emergence of a structure reflects its greater inherent complexity. Poor readers, on this view, are language delayed, and would be expected to make significant errors on tasks that involve comprehension of sentences that have complex syntactic structure. However, as we have emphasized, failure on a comprehension task does not necessarily indicate a lack of the correct structure for the sentences that are misunderstood; inefficient or abnormally limited working memory can also interfere with understanding on some sentence comprehension tasks, as claimed by the Processing Limitation Hypothesis.

In order to pursue the causes of poor readers' failures in comprehension, we first discuss experimental tasks that have been devised to test the contrasting predictions of these hypotheses as they have been applied in the investigation of the linguistic abilities of young children. Following this, two studies are presented in which the spoken language abilities of both good and poor readers were compared, and alternative interpretations of the findings are considered.

6.1 Assessing Linguistic Competence in Young Children

We sketch two experiments that were specifically devised to disentangle structural factors and working memory in the sentence comprehension of normal children. In each case we find that the children's comprehension improves dramatically when the processing demands on memory are reduced.

The first experiment makes use of the contrast between two structural phenomena, coordination and subordination. It is widely held that structures involving subordination are more complex than ones involving coordination. Researchers in language acquisition have appealed to this difference to explain why children typically make more errors in understanding sentences bearing relative clauses (as in 1) than sentences containing conjoined clauses (as in 2), when comprehension is assessed by a figure manipulation ('do-what-I-say') task.

(1) The dog pushed the sheep that jumped over the fence.
(2) The dog pushed the sheep and jumped over the fence.

The usual finding, that (1) is more difficult for children than (2), has been interpreted as revealing the relatively late emergence of the rules for subordinate syntax in language development (e.g., Tavakolian, 1981).

However, it was shown by Hamburger and Crain (1982) that the source of children's performance errors on this task was not a lack of knowledge of the syntactic rules underlying relative clauses. By constructing appropriate pragmatic contexts, they were able to elicit utterances containing relative clauses reliably from children as young as three. In addition, when the pragmatic "felicity conditions" on the use of restrictive relative clauses were satisfied, they found very few residual errors even in the 'do-what-I-say' comprehension task. These findings suggest that nonsyntactic demands of this task had been masking children's competence with this construction in previous studies.

One of the nonsyntactic impediments to successful performance involves working memory (for others, see Hamburger & Crain, 1982, 1984). To clarify this, we would note that even children's correct responses to sentences containing relative clauses can be seen to display the effects of working memory. In the Hamburger and Crain study (1982), it was observed that many children who performed the correct actions associated with sentences like (1) often failed, nevertheless, to act out these events in the same way as adults. Most 3-year-olds and many 4-year-olds would act out this sentence by making the dog push the sheep first, and then making the sheep jump over the fence. Older children and adults act out these events in the opposite order, the relative clause before the main clause. Intuitively, acting out the second mentioned clause first seems conceptually more correct because "the sheep that jumped over the fence" is what the dog pushed. It is reasonable to suppose that this kind of conflict between the order of mention and conceptual order stresses working memory because both clauses must be available long enough to plan the response that represents the conceptual order. We propose that the differing responses of children and adults reflect the more severe limitations in children's working memory. Young children are presumably unable to compile the plan and so must interpret and act out the clauses in the order of mention (see Hamburger & Crain, 1984, for more detailed discussion of plans and planning).

Studies of temporal adverbial clauses have also yielded data that support the twofold claim that processing factors mask children's knowledge of complex structures and that working memory is specifically implicated. Temporal terms like before, after and while dictate the conceptual order of events, and they too may present conflicts between conceptual-order and order-of-mention, as (3) illustrates.

(3) Luke flew the plane after Han flew the helicopter.

In this example, the order in which events are mentioned is opposite the order in which they took place. Several researchers have found that 5-year-olds frequently act out sentences like (3) in an order-of-mention fashion (Clark, 1970; Johnson, 1975). As with relative clause sentences, it is likely that this response reflects an inability to hold both clauses in memory long enough to formulate a plan for acting them out in the correct conceptual order.

There is direct evidence that processing demands created by the requirements of plan formation, and not lack of syntactic or semantic competence, were responsible for children's errors in comprehending sentences bearing temporal terms. The evidence is this: once the demands on working memory were reduced by satisfying the presuppositions associated with this construction, most 4- and 5-year-old children usually give the correct response to sentences like (4).

(4) Push the plane to me after you push the helicopter.

To satisy the presupposition, Crain (1982) had children formulate part of the plan associated with sentences such as (4) in advance, by having them select one of the toys to play with before each trial. For the child who had indicated the intent to push the helicopter on the next trial, (4) could be used. Given this contextual support, children displayed unprecedented success in comprehending the temporal terms before and after.

This brief review shows how the apparent late emergence of a linguistic structure can result from the failure of verbal working memory to function efficiently. The methodological innovations that resulted in these demonstrations of early mastery of complex syntax have been extended to other constructions, including Wh-movement, pronouns, and prenominal adjectives (Crain & Fodor, 1984; Crain & McKee, 1985; Hamburger & Crain, 1984). Although the possibility must be left open that some linguistic structures are problematic for children reaching the age at which reading instruction normally begins, this line of research emphasizes how much syntax has already been mastered by these children. The findings make it clear that the evidence cited above (section 3) that poor readers have difficulty comprehending complex syntactic constructions is compatible with the Processing Limitation

Hypothesis. The proper interpretation of such findings is complicated by the existence of confounding factors. Unfortunately, the techniques discussed above have rarely been applied in reading research. But fortunately, other methods of teasing apart structural and processing factors have been applied, as we now show.

6.2 Assessing Spoken Language Comprehension of Good and Poor Readers

In Section 3, we noted evidence that poor readers have problems in comprehending some kinds of sentences, not only when these are presented to them in printed form, as would be expected, but also when the sentences are processed by ear. We have seen, however, that these findings would receive a different interpretation on each of the two hypotheses advanced in Section 4. The question can be put to the test by comparing the success of good and poor readers on structurally complex sentences. We can infer a processing limitation, and rule out a structural deficit, whenever the following four conditions are met: (i) there is a decrement in correct responses by poor readers but, (ii) they reveal a similar pattern of errors as good readers, (iii) they manifest a high rate of correct responses on some subset of sentences exhibiting the structure in question, and (iv) they show appreciable improvement in performance on problem cases in contexts that lessen the processing demands imposed on working memory.

It is germaine to consider two recent studies that have addressed the question of whether poor readers have a structural or a processing limitation, one by Mann et al. (1984), and the other by Fowler (1985). The study by Mann and her associates asked first whether good and poor readers in the third grade could be distinguished on a speech comprehension task involving sentences with relative clauses. Having found an affirmative answer, these researchers went on to ask whether malformation or absence of syntactic structures accounted for the differences in performance between the good and poor readers.

In the experiment on temporal terms discussed in the previous section, syntax was held constant and aspects of the task were manipulated in order to vary processing load. The experiment of Mann et al. adopted another approach, holding sentence length constant while varying the syntactic structure. Four types of sentences with relative clauses were presented, using a figure manipulation task. As (5) illustrates, each set of sentences contained exactly the same ten words, to control for vocabulary and sentence length.

(5) a) The sheep pushed the cat that jumped over the cow.
b) The sheep that pushed the cat jumped over the cow.
c) The sheep pushed the cat that the cow jumped over.
d) The sheep that the cat pushed jumped over the cow.

It was found that the type of relative clause structure had a large effect on comprehensibility. Sentences of type a) and d) evoked the most errors. These are structures that earlier research on younger children also identified as the most difficult (Tavakolian, 1981).

Good and poor readers did not fare equally well, however. The study confirmed the earlier claims that poor readers can have considerable difficulties in understanding complex sentences even when these are presented in spoken form. But, given our criteria for distinguishing structural deficits from processing limitations, the findings of this study invite the inference that poor readers' problems with these sentences reflect a deficit in processing. First of all, the poor readers were worse than the good readers in comprehension of each of the four types of relative clause structure that were tested. But the poor readers did not appear to lack any type of relative clause structure entirely. In fact, their pattern of errors closely mirrored that of the good readers; they simply did less well on each sentence type. Thus, there was no statistical interaction of group by sentence type. Another reason to think that the source of the poor readers' difficulties is attributable to working memory is that they were also inferior to the good readers in immediate recall of these sentences and on other tests of short-term recall.

A further attempt to disentangle structural knowledge and processing capabilities in beginning readers was carried out by Fowler (1985). Two new experimental tasks were administered to second graders: a grammaticality judgment task, and a sentence correction task (in addition to other tests previously used at Haskins Laboratories to assess short-term recall and metaphonological abilities). The grammaticality judgment task was used to establish a baseline on the structural knowledge of the subjects, for comparison with the correction task. This expectation is motivated, in part, by recent research on aphasia showing that agrammatic aphasic patients with severe memory limitations were able judge the grammaticality of sentences of considerable length and syntactic complexity (Crain, Shankweiler, & Tuller, 1984; Linebarger, Schwartz, & Saffran, 1983; Saffran, 1985). The findings on aphasics suggest that this task taps directly the syntactic analysis that is assigned. The correction task, on the other hand, is expected to stress working memory to a greater extent, because the sentence has to be retained long enough for reanalysis and revision.

As predicted, reading ability was significantly correlated with success on the correction task, but not with success on the judgment task. This is further support for the view that processing complexity, and not structural complexity, is a better diagnostic of reading disability. Two additional findings bear on the competing hypotheses about the causes of reading failure. First, the level of achievement on grammaticality judgments was well above chance for both good and poor readers, even on complex syntactic structures (e.g., Wh-movement and tag questions). Second, results on the test of short-term recall (with IQ partialed out) were more strongly correlated with success on the sentence correction task than with success on the judgment task.

The poor readers in both of the foregoing studies appear to have had the syntactic competence to compute complex structures (see also Shankweiler, Smith & Mann, 1984; Smith, Mann & Shankweiler, in press). We infer, however, from the studies of preschool children reviewed earlier, that some children may display comprehension of certain structures only when contextual supports are available, or where memory demands are minimized. Thus, when reading is put in the perspective of recent data on language acquisition, it is apparent that an explanation that appeals to processing limitations can account for the data. There is no need to impute to the poor reader, in addition, gaps in structural knowledge.

6.3 Other Points of View

The contention that a deficit in working memory is responsible for errors in sentence understanding by poor readers has not gone unchallenged. Here we take up two challenges. First, it has been argued by Byrne (1981) that some differences in comprehension between good and poor readers cannot be attributed to verbal working memory. Comprehension data are presented from an object manipulation study in which good and poor readers responded to sentences containing adjectives like easy and eager. An appeal is then made to earlier findings by C. Chomsky (1969) that children master the syntactic properties of adjectives like easy later than those like eager.

Byrne's poor readers performed less accurately than age-matched good readers on sentences like (6) than sentences like (7). He argues that failures on sentences containing easy reflect the inherent syntactic complexity of this adjective, not its contributions to processing difficulty.

(6) John is easy to please.
(7) John is eager to please.

An explanation invoking the verbal memory system could not explain the difference between easy and eager, according to Byrne, because the two forms "load phonetic memory equally (having identical surface forms)" and, being short, impose relatively modest demands on memory (p. 203).

Results such as these can be accommodated within the Processing Limitation perspective, by attributing them to limitations in working memory function. As pointed out by Mann et al. (1980), short-term memory demands are not just a matter of sentence length or surface form. Despite their simple surface form and brevity, the inherent structural complexity of sentences with adjectives like easy may require additional computation and so may intensify the demands on working memory, as compared to sentences with adjectives like eager. The schematic diagrams below can be used to motivate an explanation invoking working memory to account for the greater difficulty poor readers have in acting out sentences with easy.

(8) The bear is easy (__ to reach __).
(9) The bear is eager (__ to jump).

As the diagram in (8) illustrates, the transitive verb reach has a superficially empty direct object position. In the terms of transformational grammar, the direct object has been "moved." In contrast, the subject position of the infinitival complement is empty in diagram (9), in this case by deletion. Comparing the two diagrams, it is apparent that the distance between the "gap" in the infinitival complement and the lexical NP that is interpreted as its "filler" is greater in (8) than in (9). Another relevant difference is that although both infinitival complements have missing subjects, the referent for the gap in subject position in (8) cannot be found anywhere in the sentence; it must be mentally filled by the listener.

It is widely assumed that holding onto a "filler" (or retrieving one for semantic interpretation) is a process that stresses working memory (see e.g., Wanner & Maratsos, 1978). This would explain why constructions with object gaps are more difficult to process than subject-gap constructions for normal children and adults. It would also explain why other populations with deficits in short-term memory are especially sensitive to this difference (Grodzinsky, 1984, for example, found the asymmetry with Broca-type aphasics). Given these considerations, poor readers also would be expected to perform with less success than good readers in response to structures like (8) even if they have attained an equivalent level of linguistic competence. In order to establish the level of competence of selected poor readers, we are currently investigating several constructions using tasks that minimize demands on working memory. The pursuit of optimal conditions for assessing linguistic competence was discussed in section 7.1. The same methodological prescription has been followed in other areas of cognitive development, with considerable success (for a review, see Gelman, 1978).

The importance of working memory for sentence understanding has been challenged from another standpoint by Crowder (1982). This criticism is based on evidence that the syntactic parsing mechanism is fast. It is argued that claims for the centrality of working memory in language processing are weakened by evidence that the parsing mechanism extracts higher level structure "on line" (Frazier & Fodor, 1978; Frazier & Rayner, 1982). If there is little or no delay in attachment of successive lexical items into the structural analysis being computed, then there is no need, this argument goes, for the memory buffer to store more than a few items at a time.

Findings that indicate that higher-level processing is accomplished within very short stretches of text or discourse do not, in our view, undercut the position that sentence processing imposes burdens of major proportions on short-term memory. On the contrary, high-speed parsing mechanisms are exactly what one would expect to find in a system that has severely limited memory processing capacity. High-speed parsing routines may have evolved precisely to circumvent the intrinsic limitations.

Sentence parsing strategies, on one prominent view (Frazier & Fodor, 1978), are not learned maneuvers. Instead, they reflect the architecture of the language processor, which has several functions to perform and limited time and space for their compilation and execution. One parsing strategy that may have evolved to meet these exigencies encourages listeners or readers to connect incoming material with preceding material as locally as possible (the strategy called "right association" by Kimball, 1973, and "late closure" by Frazier, 1978). For example, the adverb yesterday is interpreted as related to the last mentioned event in (8); though at first reading this strategy may cause a momentary misanalysis, as in (9).

(8) Sam said he got his pay, yest day.
(9) Sam said he will get paid, yesterday.

Although parsing strategies may enable the parser to function more efficiently in many cases, the existence of "garden path" sentences like (9) shows that these strategies are not powerful enough to overcome the liability of a tightly constrained working memory. Garden path phenomena make it clear that the need for working memory is not totally obviated by on-line sentence processing. Again, we should emphasize that some sentences will tax working memory heavily in certain experimental tasks, and those will be problem sentences for poor readers. It is worth noting, also, that there is evidence that children are even more dependent on these strategies than adults, presumably because children's working memories are more severely limited (see Crain & Fodor, 1984). As we have seen already, a clear prediction of the

Processing Limitation Hypothesis is that poor readers will be less able to recover from garden path sentences than good readers, even in spoken language tasks.

7. The Hypotheses Reconsidered

In earlier sections, we attempted to identify the reasons poor readers fail to comprehend complex sentences as well as good readers. In this final section, we return to the hypotheses raised at the outset, and to the question of a unitary underlying deficit that generates the symptom picture of the poor reader (as sketched in Section 3).

The fact that poor readers sometimes have difficulties in understanding spoken sentences raised the possibility that they have a structural deficit at the syntactic level (as the Syntactic Lag Hypothesis claims). The existence of a deficit at this level would jeopardize a unified theory, because if poor readers' problems in sentence understanding are at least in part attributable to missing syntactic structures, then at least two basic deficits must be invoked to account for the total symptom picture. But, as we noted, comprehension difficulties could have another explanation: the problems could be caused by a limitation of a processor, namely, working memory, which is necessary for gaining access to syntactic structures and for their successful manipulation. In reviewing the evidence, we argued that the empirical data, such as they are, can better be accounted for by supposing that the syntactic structures are in place. Poor readers' failures in comprehension are only <u>apparently</u> syntactic: they occur on just those sentences that stress working memory.

An argument against a lag in the development of phonological structures is more difficult to make. We have pointed to the evidence that poor readers lack the necessary metaphonologic skills needed for partitioning words into their phonologic segments and mentally manipulating these segments. These deficits, and others in the phonologic domain to which we have referred (e.g., Brady et al., 1983; Katz, 1986), could reflect delay in the establishment of some aspects of phonologic structure.[4] However, in the absence of any decisive evidence, we would seek to explain them as instead reflecting limitations on <u>use</u> of phonologic structures. Thus, whereas we believe the empirical evidence is sufficient to locate the problem underlying the syndrome of the poor reader at the phonological level, there is no need to suppose that any structures are missing. We recognize that the arguments against a structural deficit in poor readers cannot be conclusive without considerably more data. In the absence of such data we must leave the question open. However, the Processing Limitation Hypothesis has an advantage: by invoking the concept of working memory it can tie together the diverse strands in the symptom complex of the poor reader.

Two properties of the working memory system play an essential role in explaining the language-related problems of poor readers: (i) limitations in either component of the working memory system supporting the analysis of input both in speech and reading, and (ii) the dependence of higher-level (syntactic and semantic) processing on preceding lower-level (orthographic and phonological) analysis of the contents of the buffer. From this combination of properties the possibility arises that unless the resources of working memory are managed efficiently in pursuing the phonological analysis of letter strings, higher-level analysis will be hobbled or inhibited altogether. The poor reader (and indeed any beginning reader) will fail to understand sentences in print that could easily be understood in spoken language. But, in addition, we know that poor readers often have special working memory limitations over and above the normal limitations. Therefore they have a double handicap: poor decoding abilities and unusually constrained immediate memory. The handicap would be expected to show up even in processing spoken language when sentences are costly of memory resources.

It is worth pointing out similarities between our hypothesis about the constraining factors in comprehension and the ideas of Perfetti and his associates. Perfetti and Lesgold (1977) advanced the idea nearly 10 years ago that slow decoding interferes with integration and inhibits reading comprehension in poor readers. The combined result of poor decoding skills and working memory limitations creates a "bottleneck." Like us, these researchers see inefficient low-level processing as a limiting factor in poor readers' reading comprehension, and they maintain, as we do, that poor readers' problems in comprehension are not confined to reading (see Perfetti, 1985, for a comprehensive summary). Perfetti and Lesgold even suggest that there may be a single deficit underlying the bottleneck, but they stop short of identifying the deficit. We have pursued the possibility that a unified explanation can be given of the problems that give rise to the bottleneck. Researchers at Haskins Laboratories have sought an explicit connection between working memory problems and orthographic decoding problems. The bridge currently being investigated is that both orthographic decoding and working memory access phonological structures (Liberman & Shankweiler, 1985; but see also Alegria, Pignot, & Morais, 1982).

There is, in fact, much evidence that what we are calling verbal working memory (one component of which is verbal short-term memory, as traditionally conceived) uses a phonologic output code. Earlier, we noted the empirical basis for this belief: 1) in recalling linguistic material, verbatim retention of the phonologic units of the input is possible within narrow constraints of quantity and time, 2) interference with rehearsal causes errors in recall, 3) the error rate is increased when the items are phonetically similar (as when they rhyme with one another). The buffer component of working memory is surely phonologic in the sense that it incorporates these characteristics. The finding that poor readers show reduced confusability effects in comparison to good readers is evidence that a phonological deficiency may underlie their extra limitations in buffer storage capacity.

Poor readers' working memory problems have not heretofore been related explicitly to the other component of working memory, the control component. The primary job of the control mechanism as it relates to reading is to transfer the contents of the buffer from the phonological level to higher levels. Because we assume that reading is a bottom-up process, a disruption in flow of phonologic information to the other parsers would inevitably result in impaired reading performance. Of course it is possible that other control properties of this mechanism are also deficient. Such deficiencies would set a ceiling on reading, but would not give rise specifically to reading difficulties.

The problem of learning to read is largely to adapt the control component to accept orthographic input and to assign a phonologic analysis. As we have seen, the phonologic analysis of the speech signal is executed entirely within the speech module, whereas phonologic analysis of orthographic input demands the construction of algorithms for relating orthographic structure to phonologic structure. To construct this interface is an intellectual task, which requires overt attention and metalinguistic knowledge that doesn't come free with language acquisition. Until an entire set of analytic metaphonologic strategies are practiced enough to become largely automatic, higher-level processing will be curtailed because working memory is overloaded.

The idea of a computational bottleneck enables us to understand how constriction of the working memory system in handling phonologic information can inhibit higher-level processing of text. Clarification of the peculiar demands of orthographic decoding, together with the properties of working memory, enables us to explain why the poor reader is far less able to understand complex sentences in print than in speech, and it explains difficulties with spoken language that would otherwise appear mysterious. It is our conclusion, then, that deficits that implicate lower-level (phonological) components in the structural hierarchy have repercussions on higher levels. The hypothesis that language-related problems at different levels arise from a common source is the foremost reason, in our view, for adhering to the Processing Limitation Hypothesis. It represents the strongest empirical hypothesis. The explanatory strength and further empirical consequences of this hypothesis are discussed in Crain and Shankweiler (in press).

References

Alegria, J., Pignot, E., & Morais, J. (1982). Phonetic analysis of speech and memory codes in beginning readers. Memory & Cognition, 10, 451-456.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic and formal similarity. Quarterly Journal of Experimental Psychology, 18, 362-365.

Baddeley, A. D. (1979). Working memory and reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), The proceedings of the conference on the processing of visible language (Vol. 1). New York: Plenum.

Baddeley, A. D., & Hitch, G. B. (1974). Working memory. In G. H. Bower (Ed.), The psychology of learning and activation (Vol. 4). New York: Academic Press.

Brady, S., Shankweiler, D., & Mann, V. A. (1983). Speech perception and memory coding in relation to reading ability. Journal of Experimental Child Psychology, 35, 345-367.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read -- a causal connection. Nature, 301, 419-421.

Byrne, B. (1981). Deficient syntactic control in poor readers: Is a weak phonetic memory code responsible? Applied Psycholinguistics, 2, 201-212.

Chomsky, C. (1969). The acquisition of syntax in children from 5 to 10. Cambridge, MA: MIT Press.

Chomsky, N. (1975). Reflections on language. New York: Pantheon Books.

Clark, E. V. (1970). How young children describe events in time. In G. B. Flores d'Areais & W. J. M. Levelt (Eds.), Advances in psycholinguistics. Amsterdam: North Holland.

Conrad, R. (1964). Acoustic confusions in immediate memory. British Journal of Psychology, 3, 75-84.

Conrad, R. (1972). Speech and reading. In J. Kavanagh & I. Mattingly (Eds.), Language by ear and by eye: The relationships between speech and reading. Cambridge, MA: MIT Press.

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. Unpublished doctoral thesis, McGill University.

Crain, S. (1982). Temporal terms: Mastery by age five. In Papers and Reports on Child Language Development Vol. 21, Stanford University.

Crain, S., & Fodor, J. (1984). On the innateness of subjacency. Proceedings of the Eastern States Conference on Linguistics, Vol. 1. Columbus, OH: Ohio State University.

Crain, S. & McKee, C. (1985). Acquisition of structural constraints on anaphora. Proceedings of the North Eastern Linguistics Society, 16. Amherst, MA: University of Massachusetts.

Crain, S., & Shankweiler, D. (in press). Reading acquisition and language acquisition. In A. Davison, G. Green, & G. Herman (Eds.), Critical approaches to readability: Theoretical bases of linguistic complexity. Hillsdale, NJ: Erlbaum.

Crain, S., Shankweiler, D., & Tuller B. (1984). Preservation of sensitivity to closed-class items in agrammatism. Los Angeles, CA: Academy of Aphasia.

Crowder, R. G. (1982). The psycholcgy of reading. New York: Oxford University Press.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, 450-466.

Denkla, M. B., & Rudel, R. G. (1976). Naming of object-drawings by dyslexic and other learning disabled children. Brain and Language, 3, 1-15.

Fletcher, J. M., Satz, P., & Scholes, R. (1981). Developmental changes in the linguistic performance correlates of reading achievement. Brain and Language 13, 78-90.

Fodor, J. A. (1983). The modularity of mind. Cambridge, MA: MIT Press.

Fowler, A. (1985). Do poor readers have a basic syntactic deficit: Evidence from a grammaticality judgment task in second graders. Paper presented at New England Psychological Association; New Haven, CT.

Frazier, L. (1978). On comprehending sentences: Syntactic parsing strategies. Unpublished doctoral dissertation, University of Connecticut.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. Cognition, 6, 291-325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. Cognitive Psychology, 14, 178-210.

Gelman, R. (1978). Cognitive development. Annual Review of Psychology, 29, 297-332.

Gough, P. B., & Hillinger, M. L. (1980). Learning to read: An unnatural act. Bulletin of the Orton Society, 30, 179-196.

Grodzinsky, Y. (1984). Language deficits and linguistic theory. Unpublished doctoral dissertation, Brandeis University.

Hamburger, H., & Crain, S. (1982). Relative acquisition. In S. Kuczaj, II (Ed.), Language development. Volume 1: Syntax and semantics. Hillsdale, NJ: Erlbaum.

Hamburger, H., & Crain, S. (1984). Acquisition of cognitive compiling. Cognition, 17, 85-136.

Johnson, M. L. (1975). The meaning of before and after for preschool children. Journal of Experimental Child Psychology, 19, 88-99.

Katz, R. B. (1986). Phonological deficiencies in children with reading disability: Evidence from an object-naming task. Cognition, 22, 225-257.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. Journal of Experimental Child Psychology, 32, 474-484.

Keil, F. (1980). Development of the ability to perceive ambiguities: evidence for the task specificity of a linguistic skill. Journal of Psycholinguistic Research, 9, 219-230.

Kean, M.-L. (1977). The linguistic interpretation of aphasic syndromes. Cognition, 5, 9-46.

Kean, M.-L. (1980). Grammatical representations and the description of language processing. In D. Caplan (Ed.), Biological studies of mental processes. Cambridge, MA: MIT Press.

Kimball, J. P. (1973). Seven principles of surface structure parsing. Cognition, 2, 15-47.

Kohn, B., & Dennis, M. (1974). Selective impairments of visiospatial abilities in infantile hemiplegics after right cerebral hemidecortication. Neuropsychologia, 12, 505-512.

Liberman, A. M., Cooper, F. S., Shankweiler, D. & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revisited. Cognition, 21, 1-37.

Liberman, A. M., Mattingly, I.G., & Turvey, M. (1972). Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), Coding processes and human memory. Washington, D.C.: Winston and Sons.

Liberman, I. Y. (1983). A language-oriented view of reading and its disabilities. In H. Myklebust (Ed.), Progress in learning disabilities (Vol. 5). New York: Grune & Stratton.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading, and dyslexia. Baltimore, MD: University Park Press.

Liberman, I. Y., Mann, V. A., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and non-linguistic material in relation to reading ability. Cortex, 18, 367-375.

Liberman I. Y., & Shankweiler, D. (1985). Phonology and the problems of learning to read and write. Remedial and Special Education, 6, 8-17.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. Journal of Experimental Child Psychology, 18, 201-212.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), Toward a psychology of reading: The proceedings of the CUNY Conferences. Hillsdale, NJ: Erlbaum.

Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. Cognition, 13, 361-392.

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory: Can they presage early reading problems? Journal of Learning Disabilities, 17, 592-599.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. Memory & Cognition, 8, 329-335.

Mann, V. A., Shankweiler, D., & Smith S. T. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. Journal of Child Language, 11, 627-643.

Marslen-Wilson, W., & Tyler L. (1980). The temporal structure of spoken language understanding. Cognition, 8, 1-71.

Mattingly, I. G. (1972). Reading the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye. Cambridge, MA: MIT Press.

Mattingly, I. G. (1984). Reading, linguistic awareness, and language acquisition. In J. Downing & R. Valtin (Eds.), Languge awareness and learning to read. New York: Springer-Verlag.

Milner, B. (1974). Hemispheric specialization: Scope and limits. In F. O. Schmitt & F. G. Worden (Eds.), The Neurosciences: Third study program. Cambridge, MA: MIT Press.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a seuence of phonemes arise spontaneously? Cognition, 1, 323-331.

Netley, C., & Rovet, J. (1983). Relationships among brain organization, maturation rate, and the development of verbal and nonverbal ability. In S. Segalowitz (Ed.), Language functions and brain organization. New York: Academic Press, pp. 245-265.

Olson, R. K., Davidson, B. J., Kliegl, R., & Davies, S. E. (1984). Development of phonetic memory in disabled and normal readers. Journal of Experimental Child Psychology, 37, 187-206.

Perfetti, C. A. (1985). Reading ability. New York: Oxford University Press.

Perfetti, C. A., & Goldman, S. R. (1976). Discourse memory and reading comprehension skill. Journal of Verbal Learning and Verbal Behavior, 14, 33-42.

Perfetti, C. A., & Hogaboam, T. (1975). The relationship between single word decoding and reading comprehension skill. Journal of Educational Psychology 67, 461-469.

Perfetti, C. A. & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, NJ: Erlbaum.

Pylyshyn, Z. W. (1984). Computation and cognition: Toward a foundation for cognitive science. Cambridge, MA: Bradford.

Saffran, E. M. (1985). Short-term memory and sentence processing: Evidence from a case study. Paper presented at Academy of Aphasia; Pittsburgh, Pa.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A. & Fischer, F. W. (1979). The speech code and learning to read. Journal of Experimental Psychology: Human Learning and Memory, 5, 531-545.

Shankweiler, D., & Liberman, I. Y. (1972). Misreading: A search for causes. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye: The relationships between speech and reading. Cambridge, MA: MIT Press.

Shankweiler, D., Smith, S. T., & Mann, V. A. (1984). Repetition and comprehension of spoken sentences by reading-disabled children. Brain & Language, 23, 241-257.

Smith, S. T., Mann, V. A., & Shankweiler, D. (in press). Good and poor readers' comprehension of spoken sentences: A study with the Token Test. Cortex.

Stanovich, K. E. (1982). Individual differences in the cognitive processes of reading: 1. Word decoding. Journal of Learning Disabilities, 15, 449-512.

Tavakolian, S. L. (1981). The conjoined-clause analysis of relative clauses. In S. Tavakolian (Ed.), Language acquisition and linguistic theory. Cambridge: MIT Press.

Treiman, R., & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. In G. E. MacKinnon & T. G. Waller (Eds.), Reading research: Advances in theory and practice (Vol. 3). New York: Academic Press.

Vellutino, F. R. (1979). Dyslexia: Theory and research. Cambridge, MA: MIT Press.

Vogel, S. A. (1975). Syntactic abilities in normal and dyslexic children. Baltimore, MD: University Park Press

Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), Linguistic theory and psychological reality. Cambridge, MA: MIT Press.

Wolf, M. (1981). The word-retrieval process and reading in children and aphasics. In K. Nelson (Ed.), Children's language (Vol. 3). New York: Gardner Press.

Footnotes

[1]Some of the evidence for this position is sketched in succeeding pages, but space does not allow us to make the complete case here. The interested reader should consult: Gough and Hillinger, 1980; Liberman, 1983; Perfetti, 1985; Vellutino, 1979.

[2]References to the work of investigators at Haskins Laboratories and at Pittsburgh are made throughout the paper. We should also note similarities between the position we have developed on reading disorder and the conclusions of studies of children's cognitive development that indicate a dissociation of language-based skills and nonlinguistic abilities (see, for example, Keil, 1980; Kohn & Dennis, 1974; Netley & Rovet, 1983).

[3]For an insightful general discussion relating computer architecture and models of cognitive processing, see Pylyshyn, 1984. See Hamburger and Crain, 1984, for detailed discussion of the role of "cognitive compiling" in children's language processing.

[4]Although it is easy in principle to draw a distinction between a deficiency in setting up phonological representations and an inefficiency in processing the representations, in practice the distinction is difficult to maintain. Recent work by investigators at Haskins Laboratories clearly points to poor reader's phonological deficiencies in identifying spoken words in degraded contexts (Brady et al., 1983) and in object naming and in judging metalinguistic properties of the retrieved names (Katz, 1986). However, neither study resolves the issue of defective representation versus defective processing.

SYNTACTIC COMPLEXITY AND READING ACQUISITION*

Stephen Crain† and Donald Shankweiler†

1. Introduction

Learning to read is difficult for most people and complete mastery usually requires years of practice. In this paper we explore how the difficulties are related to linguistic structure. We will focus primarily on one component of the language apparatus, the syntactic component, and consider the role of syntactic complexity in the problems of reading. These problems are most transparent at the early stages of learning, and therefore, it should prove most revealing to compare beginning readers who are progressing at the expected rate with those who are failing to make normal progress.

The approach we will develop assumes that the language faculty is composed of several autonomous subsystems, or modules. The modules are autonomous in the sense that they develop and function according to operating principles that are specific to them, that is, not shared by other subsystems of language or other cognitive systems. Although these subsystems are intertwined in normal language use, experiments can be devised to disentangle them. The importance of this step has not always been recognized, however. We will argue that failure to take account of the modular organization of language has led to many apparently conflicting findings concerning the syntactic competence of young children. We will show, moreover, that the concept of language as a modular system has important implications for understanding how reading is acquired and for interpreting the difficulties that so often arise.

A modular view of the language apparatus raises the possibility that a single component may be the source of reading difficulty. We assume that levels of language processing are organized in a hierarchical fashion and that the flow of information is unidirectional and vertical ("bottom up") such that lower levels serve as input to higher levels and not the reverse. This means that if a lower-level component is implicated in reading difficulty, manifestations may appear at higher levels. A lower-level deficit may, therefore, masquerade as a complex of lower-level and higher-level deficits.

---

*In A. Davison, G. Green, & G. Hermon (Eds.), Critical approaches to readability: Theoretical bases of linguistic complexity. Hillsdale, NJ: Erlbaum, in press.

†Also University of Connecticut

Acknowledgment. Portions of this research were supported by NSF Grant BNS 84-18537, and by a Program Project Grant to Haskins Laboratories from the National Institute of Child Health and Human Development (HD-01994). We would like to thank Alice Davison and Ignatius Mattingly for their comments on earlier drafts.

We will argue that this is what often happens in cases of childhood reading disability: the verbal short-term memory system, hereafter called working memory, which briefly retains a phonological record of the input, is largely responsible for difficulties in processing complex syntactic structures. In developing a modular approach to reading difficulties, we were influenced by the work of M. L. Kean (1977) on the analysis of language deficits in aphasia. By seeking a unified account of language problems associated with reading difficulties we may be able to move toward an explanation of what would otherwise look like an aggregate of individual differences between good and poor readers.

There are many unanswered questions about how reading exploits the language apparatus. In order to identify the questions and examine them it is important to say what we mean by the term "language apparatus." We use it to cover both linguistic structures and the processing systems that access and manipulate these structures. The structures include the language user's stored knowledge of rules of phonology, morphology, syntax, semantics, and pragmatics. The processing systems that invoke these structures include the verbal working memory system, the syntactic parsing mechanism, and the semantic and pragmatic processors.

Since our concern is not exclusively with the reading process but more generally with the question of what makes a sentence complex, we have found it appropriate, indeed necessary, to consider the problems associated with reading from the standpoint of language acquisition. For the most part these two aspects of cognitive development have been studied independently, but we have found compelling reasons to bring them together.

Broadly speaking, there are two ways to view the relationship between children's acquisition of language and the subsequent development of reading abilities. Each view of the relationship offers an explanation of the important facts about reading; namely, why it is hard to learn to read, and why reading, unlike speech, is not universal. The differences between the views are fundamental. Each conceives of syntactic complexity in a different way and each has a different conception of language acquisition. One view is that reading demands more syntactic competence than beginning readers have at their disposal. This view is based on the assumption that some aspects of syntax that are necessary for reading are not yet in place in the beginning reader. Since reading problems are seen as a result of missing structures, we shall call this position the Structural Deficit Hypothesis (SDH).

The second view locates the problem elsewhere. It supposes that most syntactic structures are mastered well before the child begins to learn to read, and therefore that the source of reading difficulty lies in the subsidiary mechanisms that are used in language processing, mechanisms that may require modification in order to accommodate print. This position will be called the Processing Deficit Hypothesis (PDH).

These hypotheses are somewhat idealized, but they provide a framework from which to direct the search for causes of the difficulties encountered in mastery of reading, and each offers a distinctive perspective on the nature of syntactic complexity. In the later sections we consider how each hypothesis squares with research on language acquisition (section 3.A), with emphasis on one syntactic construction, the restrictive relative clause (3.B). We then focus on the plight of the poor reader; Section 3.C gives an account of an experiment designed to determine which hypothesis can best explain failures to comprehend sentences containing relative clauses. Section 3.D explores the implications of empirical findings showing that poor readers have problems with lower-level language operations. We raise there the possibility that these difficulties may, in turn, have ramifications for processing language structures at higher levels. We argue, moreover, that written language places special demands on the subsidiary language processors such that reading comprehension is often more limited than comprehension of spoken sentences.

On the empirical side, our conclusions will be tentative; much research remains to be done. On the theoretical side, we will offer a new perspective on reading and its problems--one that ties reading research more securely to current linguistic and psycholinguistic research.

2. Two Hypotheses about Reading Acquisition

In this section we fully sketch the two hypotheses that were briefly introduced above. First we examine their different conceptions of the sources of syntactic complexity. From these conceptions are derived different explanations about what makes reading hard to learn. Ultimately our concern is with the different empirical predictions of the two hypotheses, since, in our view, one of the principal tasks of the psycholinguistics of reading is to discover which hypothesis comes closer to the truth.

A. The Structural Deficit Hypothesis

The first proposal is based on the premise that some syntactic structures are inherently more complex than others. The supposition that linguistic materials are ordered in complexity invites an inference about the course of language acquisition; namely, that language acquisition proceeds in a stepwise fashion, beginning with the simplest structures and culminating only when the most complex structures have been mastered. This view of the course of language acquisition provides a foundation for hypotheses about learning to read and about the factors that distinguish good and poor readers. In this way, the SDH is intimately linked with a particular viewpoint on language development.

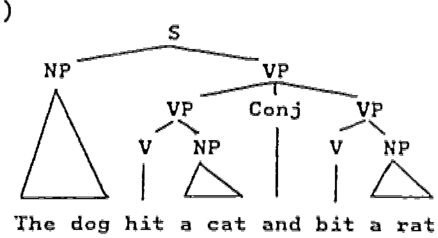
The SDH maintains that, at the time reading instruction begins, children are only partway through the course of language acquisition. If true, this hypothesis of gradually unfolding competence could explain why reading is delayed in most children until they are five to seven years of age. Moreover, the difference between successful and unsuccessful readers could be attributed to further lags in primary language abilities in some children or to deficient instruction and/or experience with written language. This view may also contain implications for the role of experience. Although the early development of language requires only immersion in a speaking environment, the later development of language, as well as the early stages of reading, may require both graded inputs and extensive experience.

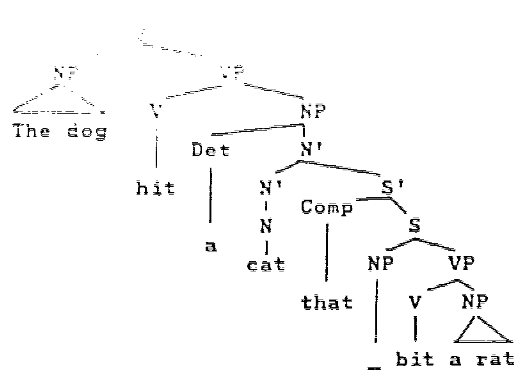
To develop this hypothesis further, we consider first the claim that syntactic structures differ in inherent complexity. As a case in point, it has been claimed that a sentence containing both a main clause and a subordinate clause, such as (2), is more complex than a coordinate structure, as in (1) (see section 3.B).

(1) The dog hit a cat and ...
(2) The dog hit a cat that ...

Syntactic differences between ... a cursory examination of the following, ...

(1*)





One difference is in the number of syntactic constituents in (1*) and (2*). Notice that there is a higher ratio of phrasal categories to words in (2*). Another difference is that (2*) but not (1*) contains a "missing" noun phrase, indicating that a constituent has been "moved" by transformational rule. It is an empirical question whether or not these structural differences contribute to difficulties in processing either in speech or in reading (see Fodor & Garrett, 1967; Kimball, 1973). This possibility could be tested by measuring reaction-time latencies to sentences like (1) and (2) on some reading task that is sensitive to ease of processing. But in the research discussed here the indicator of the relative complexity of syntactic structures is the following: one structure is simpler than another if children can speak and comprehend it first. Returning to our examples, if sentences like (2) take longer to master than sentences like (1), this would be attributable to the relative complexity of (2*) as compared to (1*).

As we noted, the SDH makes an explicit prediction about reading acquisition: the structures that beginning readers and poor older readers find most difficult are just those that appear last in the course of language acquisition. Advocates of the SDH, then, would point to data on the late acquisition of specific structures in poor readers, particularly those structures underlying complex sentences (e.g., Byrne, 1981; Fletcher, Satz, & Scholes, 1981; Vogel, 1975). The SDH regards learning to speak and learning to read as continuous processes that tap the same cognitive abilities, but it is argued that reading is difficult largely because many of the primary linguistic abilities that support it are acquired late.

B. <u>The Processing Deficit Hypothesis</u>

We now introduce an alternative account of the fundamental facts of reading acquisition. Based on a different conception of linguistic complexity, this hypothesis supposes children have already acquired a great deal if not all of the primary linguistic apparatus by the time they begin to learn to read. But in addition to this, reading demands a number of secondary processing mechanisms to interface spoken language and an orthographic system of representation. These subsidiary mechanisms include verbal working memory, routines for identification of printed words, and the syntactic, semantic, and pragmatic processors.

Since many of the same structures are used in reading and speech, it is easy to overlook the possibility that reading may make special demands on the language processing systems beyond those required for speech. In speech processing, word identification, syntactic parsing, and semantic composition of word meanings are all highly automatic from the earliest stages of language acquisition. In reading, these processes must be reshaped to interface with a new input source. At the lowest level, a system for gaining access to the mental lexicon from print must be mastered to the point that it is both rapid and accurate. Until this is accomplished, higher-level processes such as syntactic parsing and semantic composition may be inhibited, reduced to a level far below the level at which they function in speech.

To make this discussion more concrete, suppose that working memory resources are exhausted by the task of identifying words from their orthographic representations. In that case, higher-level syntactic and semantic processing may be preempted. Much evidence exists that word recognition difficulties persist for a long time in early readers and that good and poor readers are sharply distinguished in orthographic ("decoding") skills (Gough & Hillinger, 1980; Perfetti & Hogaboam, 1975; Shankweiler & Liberman, 1972) If it could be shown further that when the pressures on working memory were reduced, beginning readers could comprehend structures that were otherwise problematic, this would provide confirmation for the PDH.

To develop this account, and to explain that the PDH offers a different view of syntactic complexity, we must consider further the implications of the early acquisition of syntax, a tenet we take to be central to this hypothesis. To this end, we will draw upon the modularity hypothesis introduced earlier, which can be contrasted with the view that knowledge of language is a composite of more general cognitive faculties (for a recent statement, see J. A. Fodor, 1983). One tenet of the modularity thesis is the innate specification of language structures. Neurological evidence for the innateness of the language faculty is extensive. Among the facts that should be mentioned are the existence of special brain mechanisms present from birth, and evidence of dissociation between patterns of sparing and loss in language and other cognitive abilities in cases of brain damage (Dennis, 1980; Milner, 1974; Whitaker, 1976).

It is difficult to find psycholinguistic evidence that a particular linguistic structure, such as syntax, constitutes a submodule of the language component. Even the apparent innateness of some ability does not guarantee modular organization. An ability might, in principle, be innate and also multifactorial in composition. There are, however, some general guidelines for detecting modular organization, and tests for innateness are certainly among them. In the best case, an innate system could be expected to unfold rapidly, with much latitude regarding input from the environment, and with minimal interaction with concurrently developing systems (in Fodor's terms, "informationally encapsulated").

The acquisition of syntax adheres closely to these guidelines for innateness and, by extension, seems to conform to the modularity hypothesis. If the recent findings of early mastery of complex structures can be generalized (see section 3A), this would constitute strong empirical support for one tenet of linguistic theory, namely the hypothesis that there is an innately-specified "Universal Grammar." The theory of Universal Grammar maintains that the language module develops into a rich and intricate system of rules much more rapidly than many other cognitive structures because of its innately-specified content. Children seem to know too much too soon and they take too few wrong turns for the acquisition of language to be explained without supposing that it is both guided and constrained by innate principles (for further discussion, see Chomsky, 1971; 1975; 1981; Hamburger & Crain, 1984; and Lasnik & Crain, 1985).

Our specific concern here is with syntactic structure. If syntactic structure is largely built into the blueprint for development, then it makes little sense to ask if some syntactic constructions are harder to learn. Each construction simply develops in its own time, according to a predetermined schedule, regardless of its specific properties. In this way, the PDH calls into question the notion of linguistic complexity advanced by the SDH.

One possible advantage of modular organization, then, is that extreme structural complexity (by pretheoretic standards) can come "prewired." And what is not prewired may nonetheless be rapidly acquired, since the modular character of the linguistic system may endow it with heavy internal constraints on the types of hypotheses that a child can entertain. One way that children's grammar formation is believed to be constrained is in the structure-dependent nature of rules. A structure-dependent rule is one that is based on an abstract schema that partitions sequences of words into constituent structure. By contrast, a structure-independent rule, such as a simple counting rule, is applied directly to sequences of words themselves, without partitioning them into abstract functional units.

The theory of Universal Grammar maintains that children invariantly adopt structure-dependent rules in the course of grammar formation, eschewing structure-independent rules even when much of the available data is consistent with hypotheses of either type (Chomsky, 1971; 1975). Moreover, children are predicted to opt for structure-dependent rules even if structure-independent rules are computationally <u>less complex</u>. In the next section we will present evidence of the children's acquisition of an apparently complex rule at a time when a simpler rule would suffice.

To summarize, the two views we have presented make different predictions because they locate the source of reading difficulties in different components of the language apparatus. In essence, the views turn on the distinction between structure and process. On the first view there is a structural deficit, that is, a deficit in stored knowledge. On the second view the problem is one of process, that is, access and use of this stored knowledge. What is common to these hypotheses is that each attempts to locate the <u>causes</u> of reading difficulties. In this way they go beyond description and move towards explanation.

Each hypothesis attempts to account for the same basic facts about reading, but ultimately they diverge. Both predict that beginning readers will have difficulty reading some linguistic material, but on the SDH they should have trouble understanding complex linguistic structures even when these are presented in the speech mode. This hypothesis maintains that the late emergence of some structures places an upper bound on both the reading skills and the spoken language skills of the young reader. On the PDH, beginning readers will have achieved a high level of mastery of the grammatical operations that are required for speaking and understanding spoken sentences. The strongest version of the PDH would hold that all of the primary language apparatus is in place before formal instruction in reading begins. But even in this strong version, reading and writing will be acquired gradually, with some difficulty and with uncertain results, precisely because they tap abilities that may appear to be peripheral to the language module, though closely associated with it (Liberman, Shankweiler, Fischer, & Carter, 1974; Mattingly, 1972; Rozin & Gleitman, 1977; Shankweiler & Liberman, 1976). The PDH predicts that most beginning readers may be competent to deal with complex linguistic constructions in spoken language, whatever the attained level of reading skill, within the constraints imposed by their limitations in processing capacity.

It is important to point out, in this connection, that we are discussing performance here, and not competence. Poor readers' performance on complex sentences may often be faulty. But, according to the PDH, the failures in comprehension should be ascribed to secondary processing limitations, such as limitations on working memory, and not to lack of syntactic competence *per se*. Beginning readers and those with persisting difficulties may not be able to make use of their underlying grammatical competence because lower-level processing may preempt higher-level processing. Only by experimental means can we assess underlying competence when performance is faulty: the prediction of the PDH is that syntactic competence should be revealed in contexts that reduce the processing demands on the secondary language apparatus. In the next section we will discuss how primary and secondary linguistic abilities may be successfully teased apart in studies of language acquisition.

3. Implications of Language Acquisition for Reading

This section will review aspects of language acquisition that are relevant to the two hypotheses about the sources of reading difficulty. The SDH distinctively predicts, as we noted, that relatively more complex linguistic structures emerge only at the later stages of language development. By contrast, the PDH predicts rapid acquisition of complex syntactic structures. As a test of this difference, the following experiment addresses the claim of Universal Grammar that children adopt only structure-dependent rules even if there exist viable alternative rules that appear to be considerably simpler. Following this, we will shift our attention to the acquisition of another syntactic construction, the restrictive relative clause. We consider first its course of acquisition in normal development; then we present a study of the comprehension of this construction by good and poor readers.

A. Structure-Dependence in Language Acquisition

It is Chomsky's hypothesis that children unerringly adopt structure-dependent rules. To test this hypothesis Crain and Nakayama (1986) developed an experimental task, in the form of a game, to elicit yes/no questions that are amenable in principle either to structure-independent or structure-dependent analyses. For yes/no questions, the structure-independent strategy might be as follows:

Move the *first* "is" (or "can," "will" etc.) to the front of the sentence.

Notice that this principle gives the correct question forms for many simple sentences, as in (3).

(3) John is tall. Is John tall?
Mary can sing very well. Can Mary sing very well?

Since the structure-independent strategy produces the correct forms in simple cases, and since it appears to be computationally simpler than the structure-dependent operation, we might expect some children to adopt it were it not precluded by Universal Grammar. However, the structure-independent rule produces incorrect question forms for more complex cases, as examples (4) and (5) illustrate.

(4) The man who is running is bald.
(5) *Is the man who _ running is bald?
(6) Is the man who is running _ bald?

Applying the structure-independent strategy to sentence (4) results in the ungrammatical question (5). The correct form (6) comes from the application of a rule that treats "the man who is running" as a constituent. It is the auxiliary verb following this constituent, the entire subject noun phrase, that must be fronted.

To discover whether children could be induced to give structure-independent responses such as (5), sentences like (7) were used.

(7) Ask Jabba if the man who is running is bald.

Sentences like (7) evoked corresponding yes/no questions from thirty 3- to 5-year-old children. These children were enjoined by one experimenter to pose questions about a set of pictures to Jabba the Hutt, a figure from "Star Wars," that was concurrently manipulated by a second experimenter. Following each question, Jabba would be made to look at the picture and give an appropriate response. This game was used to determine whether structure-independent questions such as (5) would be produced, as opposed to correct question forms like (6).

Crain and Nakayama found that the children never produced structure-independent utterances. Thus, the structure-independent strategy was not adopted in spite of its simplicity and in spite of the fact that it produces the correct question forms in many instances. Crain and Nakayama also provide evidence that even children as young as three base their rule for forming yes/no questions on the syntactic properties of sentences; they do not restrict its application to referential NPs, as claimed by Stemmer (1982), who advocates a semantic account of the acquisition of this construction. In this connection, Crain and Nakayama's subjects proved to be totally insensitive to the semantic properties of the noun phrases they encountered, which included abstract NPs (e.g., running, love) and expletives (e.g., it, there) in addition to referential NPs (e.g., the boy). Thus, yes/no question formation proved to be an instance of the developmental autonomy of syntax.

This experiment on structure-dependence serves to sustain the modularity hypothesis. Notice that each of the criteria of a modular system is met in this aspect of language development: early acquisition of complex structures, system-internal constraints on hypothesis testing, as illustrated by the formation of yes/no questions, and informational encapsulation, in the form of the developmental autonomy of syntax and semantics. It is worth emphasizing the importance of universal constraints on grammar formation, such as structure-dependence, for language learnability. By forestalling wrong turns that might otherwise be taken, these constraints obviate the need for "negative data," which are presumably unavailable. The findings of Crain and Nakayama, then, provide striking support for the biological efficacy of Universal Grammar.

The concept of language as a modular system has implications both for the acquisition of syntax and for reading. If the language faculty is truly modular, then the primary language abilities of both good and poor readers should be in place before reading instruction begins. It is surprising that research addressing the comprehension of syntax by good and poor readers is so sparse. In the following section, we present the results of recent studies conducted by one of us on the acquisition of relative clauses by young children, and in section 3.C we present a study, by the other author, that suggests that poor readers have these structures, though their processing of them is to some extent impaired.

B. The Acquisition of Relative Clauses

Full syntactic competence is revealed by performance with complex linguistic constructions such as the restrictive relative clause. This construction is complex in its syntactic, semantic, and pragmatic properties. For instance, because it is the product of a movement transformation, it contains a superficially empty noun phrase as one of its constituents. This empty constituent must be assigned an interpretation based on some overt noun phrase elsewhere in the sentence. Difficulties of interpretation may be encountered at sites like these where movement leaves a gap (indicated by "_" in (8)). At these positions principles of semantic interpretation must be applied. For instance, in sentence (8) the relative clause, "who we visited _ in Amherst," depends on the preceding noun phrase "the man" for its interpretation.

(8) The man who we visited _ in Amherst listens to WFCR.

Often, the head noun phrase of a restrictive relative clause refers to a set of entities in the surrounding context. Thus, a sentence like (8) would normally be used when more than one man has been introduced into the discourse. The set referred to by the general term "man" is then restricted in scope by the content of the clause; in the present example, reference is restricted to just the man who was visited in Amherst. Both of these properties of sentences containing relative clauses may contribute to processing complexity, and indeed, such sentences are frequently misinterpreted, especially by people with language impairment, like mentally retarded people (Crain & Crain, in preparation) and aphasics (Caramazza & Zurif, 1976).

The examples in (9) display four types of relative clauses, the characteristics of which are indicated by the preceding code letters. The first letter refers to the grammatical role of the noun phrase that bears the relative clause. In the first two examples the subject of the main clause is modified by a relative clause, whereas, in the last two examples, the relative clause is attached to the direct object. The second code letter refers to the grammatical role of the missing noun phrase in the relative clause. In the first and third examples, the relative clause has a missing subject. The direct object is superficially empty in the second and fourth. These varieties of relative clauses have received the greatest amount of attention in the literature (but also see deVilliers, Tager-Flusberg, Hakuta, & Cohen, 1979).

(9) SS The dog that _ chased the sheep stood on the turtle.
SO The dog that the sheep chased _ stood on the turtle.
OS The dog stood on the turtle that _ chased the sheep.
OO The dog stood on the turtle that the sheep chased _.

It is commonly believed that children even beyond the fifth year frequently misinterpret sentences with relative clauses, especially OS and SO relatives. Both Sheldon (1974) and Tavakolian (1981) found that many children would act out an OS relative, like the example above, by having the (toy) dog stand on the turtle and then chase the sheep. Tavakolian observed that this action sequence is a correct response to a sentence in which the two clauses are conjoined, as in (10).

(10) The dog stood on the turtle and chased the sheep.

This kind of misinterpretation led Tavakolian to suggest that children younger than six have not yet developed the grammatical competence needed to comprehend syntactic structures as complex as relative clauses. She argued that the "conjoined-clause" response reflects a stage of acquisition at which children have not yet attained full competence with the hierarchical constituent structure of relative clauses. She points out further that children are already productively using conjoined clauses at the age at which they misinterpret relative clauses (cf. Brown, 1973; Limber, 1973). It was concluded, therefore, that they tend to adopt a less differentiated conjoined-clause analysis when confronted with sentences with relative clauses, until some later stage of acquisition.

Although Tavakolian's conjoined-clause hypothesis is still widely accepted, several researchers have found that children can be diverted from the conjoined-clause response to relatives by careful selection of test sentences. Solan and Roeper (1978) found that sentences containing relative clauses evoke very different error rates depending on their semantic content. Their subjects produced more errors with sentences like (11) than with sentences like (12), which contain a relative clause that can be interpreted more naturally as modifying the object of the matrix sentence rather than its subject. In addition, Goodluck (1978) found that children made fewer incorrect responses when the number of <u>animate</u> noun phrases was reduced, as in (13).

(11) The dog kicked the sheep that jumped over the pig.
(12) The girl petted the sheep that licked the cow.
(13) The dog kicked the sheep that jumped over the fence.

In accord with the PDH, these findings favor a performance account, rather than a competence account, of children's errors. Given that children misinterpret only a subset of sentences bearing the same structure, a non-structural explanation of their errors seems to be required.

A direct test of the conjoined-clause hypothesis was conducted using a picture verification paradigm (Crain, Epstein, & Long, in preparation). In this study, three- to five-year-old children heard sentences containing relative clauses like (14). Then they were asked to select one of two pictures, which depicted the events expressed in sentences (14) and (15). According to the conjoined-clause hypothesis, children should have preferred the picture corresponding to (15).

(14) A cat is holding hands with a man that is holding nands with a woman.
(15) A cat is holding hands with a man and is holding hands with a woman.

Conjoined-clause responses were evoked only 10% of the time in this task. That is, children matched sentences containing relative clauses with the appropriate pictures and not with pictures representing a conjoined clause interpretation of the sentence. This finding suggests that children's misinterpretations of OS relatives in earlier studies should not be viewed as a reflection of incomplete syntactic development. Instead, misinterpretations in these studies were probably attributable to task complexity. By contrast, the picture verification technique appears to be a simple and direct test of comprehension. Sentences like (14), tested in this way, proved to be well within the capacity of three-year-old children.

Additional evidence that children have mastered the relative clause comes from an elicited production study by Hamburger and Crain (1982) who found that four-year-old children consistently produced and understood restrictive relative clauses in contexts that were appropriate for them but inappropriate for conjoined clauses. These authors argue that previous research ignored what they called the "felicity conditions" on the use of relative clauses. One felicity condition is that the events depicted by the relative clause are presupposed to be true. For example, an utterance of sentence (16) is normally felicitous only if it is already known to both speaker and hearer that a particular cow has previously jumped over some contextually salient fence.

(16) The sheep pushed the cow that jumped over the fence.

A second pragmatic constraint, noted above, requires that there be a set of objects corresponding to the head noun of the relative clause. In the present example, there should be at least one other cow from whom the fence-jumper needs to be distinguished. The relative clause serves to restrict the set, in this case to the cow that jumped the fence. If this constraint is not met, that is, if only a single cow is present, the sentence without the relative clause (i.e. "The sheep pushed the cow") would convey as much information. In the experiments cited above (that evoked high error rates), sentences like (16) were used with only one cow present in the experimental workspace. This fact alone may have resulted in poor performance by children except, perhaps, when other processing demands were sufficiently reduced. As noted, poor performance has sometimes been attributed to children's ignorance of the syntactic rules for relative clause construction. Suppose, however, that a child had mastered not only the syntax of relative clauses, but also the presuppositions associated with their use. Such a child might still be unable to relate sentences with relative clauses to the (inappropriate) circumstances provided by the experiment. Hamburger and Crain propose that the failure to satisfy presuppositions renders sentences quite unnatural in the experimental context, encouraging subjects to think of the task as unrelated to normal contextually-sensitive language use. If so, their responses would not be indicative of their grammatical knowledge.

This brief review shows that different tasks and procedures lead to different conclusions about the acquisition of complex syntax. Resolution of these conflicting results is important for reaching a decision on whether the SDH or the PDH gives a better account of the source of reading difficulty. We would appeal to the competence-performance distinction as an aid to resolve the conflict. Since performance and not competence is what is directly observed, negative findings are not necessarily indicative of children's incompetence. Though elusive, syntactic competence can be revealed in contexts that minimize semantic and pragmatic processing complexities. By eliciting successful performance in these controlled contexts, we can be confident that competence exists.

These observations underscore the need to disentangle aspects of structure and process. We have just seen that if a test sentence contains presuppositions that go unheeded in an experimental task, it cannot validly assess a subject's knowledge of syntax. The fact that syntax, semantics, inference, and so forth, are normally interwoven in discourse makes it difficult to isolate any one of these, even by experimental design. Although these methodological problems may seem obvious when pointed out, a large proportion of the existing research both on normal and language-impaired populations has paid them little heed. As a result, the research literature may give a misleading picture of the linguistic competence of young children, portraying them as ignorant of complex structures until well after the age at which reading instruction begins. Thus, much of the research appears to support the SDH. However, a reinterpretation of the empirical findings on the acquisition of syntax leads to a different conclusion. Several recent studies, which have respected the methodological problems we have been discussing, seem to show that even three-year-old children have acquired the complex syntax denied by earlier investigators. These findings, then, support the PDH.

C. Comprehension of Complex Syntax by Good and Poor Readers

Until now we have not discussed the problems of the poor reader directly. We have presented several issues in the assessment of syntactic competence in young children, and attempted to show how these issues bear on the two hypotheses about the nature of the obstacles that lie in the way of becoming a good reader. We are now ready to apply the findings on language acquisition to the problems of learning to read with comprehension.

The literature we have reviewed on the acquisition of the restrictive relative clause has shown that very young children sometimes produce and comprehend complex syntactic structures of this sort. We know, however, from other work, including the findings presented in this section, that even much older (school-age) children who are poor readers have difficulties understanding complex spoken sentences, including those containing restrictive relative clauses. Our task in this section is to explain how the difficulties in understanding these structures might have arisen. To that end, we will present the results of a recent study designed to locate the source of comprehension failures in poor readers, using a variety of sentences containing the restrictive relative clause. These studies underscore many of the theoretical and methodological problems that concerned us in the preceding discussion.

In the light of the foregoing findings on young children, it is to be expected that relative clause structures should already be well established in the internalized grammars of eight- or nine-year-old children. It is conceivable, however, that even by this age some children (i.e., poor readers) may have attained only partial mastery of these structures. It is important to find out whether certain forms of relative clause structure are missing from their grammars, because, as we have argued, if poor readers were absolutely unable to comprehend some types of restrictive relative clauses, this would be strong support for the SDH.

According to the PDH the difference between good and poor readers should be one of degree. The PDH, too, would predict that poor readers would have difficulties understanding complex structures such as relative clauses, but crucially, they should not fail to comprehend them altogether. If they give the same pattern of responses as good readers, but do not achieve as high a rate of success, this would support the PDH. In this event, we would have to go on to ask what secondary processing mechanisms must be invoked to explain their difficulties.

We now discuss in some detail the results of an experiment that attempts to test directly the possibility that a certain processing deficit is responsible for poor readers' difficulties in understanding complex sentences not only in reading but also in spoken language. As we will see, the answer turns on the role of working memory in processing connected discourse. In spoken language comprehension, only structures that severely stress working memory will be expected to cause notable difficulties. We maintain that comprehension difficulties that are manifested in spoken language will be magnified in reading because reading places greater demands than speech processing on limited working memory resources. Until orthographic decoding skills are mastered and highly practiced, a reader cannot be expected to perform with print up to the ceiling set by performance in spoken language. The comparison between speech and reading is treated in the next section (3.D), and at greater depth in Shankweiler and Crain (1986). See also Perfetti (1985) and Perfetti and Lesgold (1977).

Comprehension and recall of complex sentences containing four relative clause structures (as in sample sentences (9) above) were studied by Mann, Shankweiler, and Smith (1984). The children's comprehension was tested first, using a toy manipulation paradigm; on a later day, the taped sentences were presented again and rote recall was tested. Both tests were administered to the same groups of good and poor readers in the third grade.

The experiment was designed to hold certain processing demands constant while varying the type of relative clause structure. Each of the test sentences mentioned three (animate) objects. As the examples in (9) illustrate, each set of test sentences mentioned the same objects, and each set contained the same ten words. Therefore, any differences in their meanings were carried by syntactic structure. The importance of controlling sentence length in a test of this kind is well recognized. Indeed, readability formulas assume that this is the most important variable in determining ease of understanding (Dawkins, 1975). But, as we will see, structure has large effects on comprehensibility that are independent of length.

The good and poor readers in this study were compared both with respect to the kinds of errors that occurred and the way these errors were distributed between the groups. As to the kinds of errors, it was expected that a conjoined-clause response might more often be made by poor readers than by good readers. This could mean that poor readers are heavily influenced by non-syntactic processing factors, just as younger normal children are. Alternatively, these responses could imply, as the SDH would predict, that the grammars of poor readers are less differentiated than those of normal adults and more mature children of the same age.

The way the errors are distributed is also relevant to the two hypotheses. If there exists a specific syntactic deficiency over and above the difficulties of processing, we would expect, other things being equal, to find a different pattern of accuracy between groups on the four sentence types. Figure 1 displays the mean errors for each of the four sentence types, separately for good and poor readers. As expected, the types were not equal in difficulty. The poor readers made more errors than the good readers on each. But when the four types were ranked in order of difficulty for good and poor readers separately, the ordering was the same for both groups. The lack of statistical interaction means that the poor readers were generally worse than the good readers in comprehension of relative clause sentences, but within this broad class, they were affected by syntactic variations in the same way as the good readers. The results give no evidence, then, that the poor readers in this study were deficient on any facet of the grammar pertaining to the interpretation of these relative clause sentences. The competence they displayed was essentially like that of the good readers.

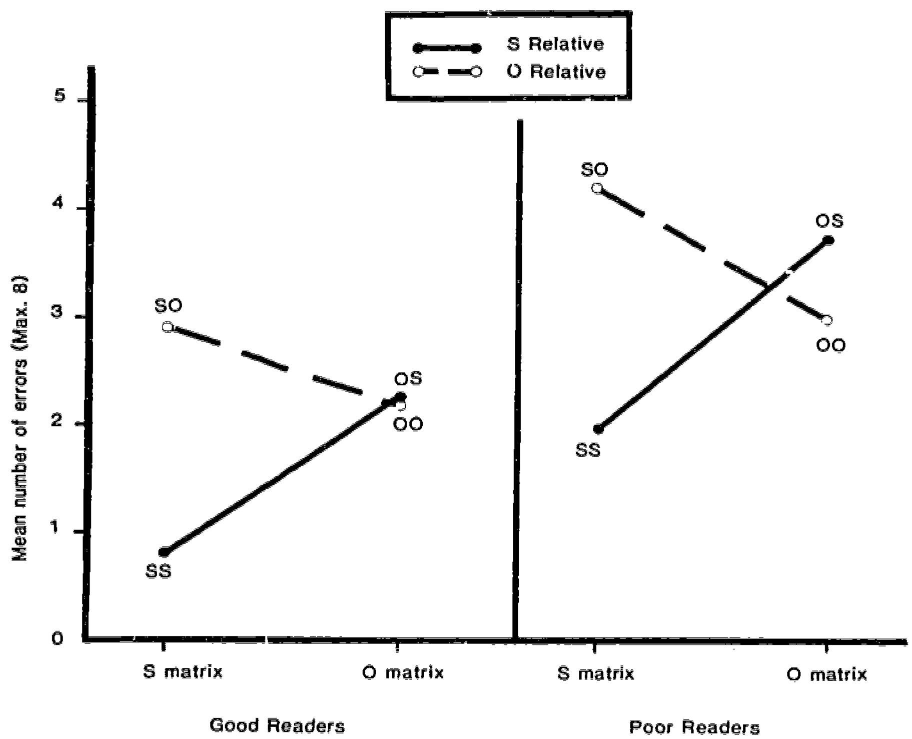


Figure 1. Mean errors of good and poor readers in the third grade on four types of relative clause constructions (from Mann, Shankweiler, & Smith, 1984).

We must nevertheless account for the fact that the poor readers made somewhat more errors than the good readers on the comprehension of each type of relative clause sentence. A likely explanation is found by comparing the groups on the test of rote recall of the sentences. As we noted earlier, the taped sentences were presented to the children a second time on another day and immediate recall was tested. In working memory for the sentences, as in the previous test of comprehension, the poor readers made significantly more errors than the good readers, and, again, the differences between the groups did not favor one type of sentence more than another. These results fit well with much earlier work that indicates that poor readers do consistently less well than good readers on a variety of tests of verbal working memory (see Jorm, 1979; Mann, Liberman, & Shankweiler, 1980; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979).

In keeping with the modularity hypothesis, it is important to appreciate that the memory deficits of poor readers are largely limited to verbal material. Tests of working memory for nonverbal material, such as unfamiliar faces and nonsense designs, do not distinguish good and poor readers (Katz, Shankweiler & Liberman, 1981; Liberman, Mann, Shankweiler, & Werfelman, 1982). Thus the failure of the poor readers to do as well as the good readers on the test of sentence comprehension is probably largely a reflection of specifically-linguistic working memory limitations on the part of the poor readers. But it is a limitation on efficiency of linguistic processing and not a limitation of structural competence. To make a further test of this possibility, it will be important to find out if poor readers have a higher success rate when the same structures are placed in contexts that minimize, not just control for, processing demands (such as presuppositions and parsing) that are otherwise confounded with syntactic complexity.

Having discussed the basis of poor readers' difficulties in sentence understanding in speech, we now turn to the consequences of these problems for reading. We have cited the evidence that poor readers have special limitations in use of the verbal working memory system that supports on-line language processing. We can now guess how handicapping such a limitation must be for reading, since the poor reader is also generally slow in decoding the individual words of the text. If the individual words are read too slowly, comprehension suffers, even if all the words are read correctly, because the integrative processes are disturbed by the slow rate of input. Perfetti and his colleagues have suggested that working memory limitations create a "bottleneck" that restricts the utilization of the higher level language processing systems, preventing proper comprehension of what is read (see, e.g., Perfetti & Lesgold, 1977).

The bottleneck hypothesis takes us some distance toward an explanation of the high correlation that has repeatedly been noted between 1) the speed and accuracy of identifying words and pseudowords in isolation and 2) various measures of reading comprehension (Calfee, Venezky, & Chapman, 1969; Perfetti & Hogaboam, 1975; Shankweiler & Liberman, 1972). We view this correlation as a particularly strong indication that a low-level deficit can give rise to apparent deficits at higher levels. Because syntactic structure and propositional content are conveyed by sequences of words, it is generally supposed that working memory is needed for sentence comprehension, whether by speech or by reading. Since the verbatim record of incoming speech or printed text is extremely fleeting, the input to the working memory system is lost unless it is rapidly converted into a more durable form (Sachs, 1967).

Because the working memory representation is so brief in duration and so limited in span, it has been proposed that the sentence parsing mechanism works rapidly on small chunks of text to decode linguistic information into more durable memory representations (Frazier & Fodor, 1978; Liberman, Mattingly, & Turvey, 1972).

We conclude this section with some remarks on the role of context in determining whether or not a sentence will be understood. Consider first the role context plays in spoken language comprehension. It was seen that children who are poor readers sometimes fail to comprehend spoken sentences that impose heavy processing demands on working memory. It was not easy, however, to demonstrate that poor readers are not as adept as good readers in sentence processing. The problems of the poor reader are ordinarily well masked; they are revealed only under rather stringent conditions of testing, without contextual supports, in the "null context" (see Crain & Steedman, 1985). The difficulty in bringing these problems to light should not surprise us. Under ordinary conditions, listeners do have contextual support. It is only when we artificially deprive poor readers of this support that they are apt to fail. When support is available, ten-year-old poor readers display clear ability to benefit from it (Perfetti, Goldman, & Hogaboam, 1979). This too is not surprising. We have shown that even three- and four-year-old children are able to understand complex sentences in appropriate contexts.

In reading, the situation is complicated by the demands of orthographic decoding. It is obvious that young poor readers have a problem in comprehending complex sentences that are set down in print. But why can't they use context here as effectively as they do in perception of spoken sentences? Our response is that a working memory limitation has a more profound effect on reading comprehension than on comprehension of speech. As noted earlier, the beginning reader is required to develop a whole new apparatus for word recognition, incorporating a set of rules for getting from the orthography to preexisting lexical entries. Until the rules and the strategies for invoking their use are automatized, the would-be reader cannot use syntactic and pragmatic context effectively, because nearly the whole of the processing capacity is consumed by lower-level functions. This, we assume, is the point of the bottleneck hypothesis of Perfetti and his associates. The remainder of this section will be concerned with working out the detailed implications of poor readers' lower-level deficits for performance on sentence processing tasks.

D. Consequences of a Low-level Deficit for Higher-level Processing

The preceding section gives a rough sketch of the source of comprehension difficulties that plague the beginning reader and many others who, though no longer beginners, are still struggling to gain mastery. If our analysis is on the right track, we have now moved beyond the stage of identifying correlates of reading difficulties. To the extent that we now have the beginnings of a theory, we stand in a position to make fairly detailed predictions about what will be difficult for children to read and to offer tentative suggestions about how these difficulties might be circumvented.

Since our concern here is with sentence understanding, our predictions involve syntactic structures and the mechanism that invokes them. We have no reason to suppose that different mechanisms perform this function in reading than in speech. But the bottleneck hypothesis anticipates that the syntactic parsing mechanism will be less efficient in reading at the early stages, when the reader is preoccupied with the identification of words in print. The poor beginning reader, as we saw, labors under a double handicap, since he or she has less than normal working memory capacity to begin with. In this section we will discuss two ways that a working memory deficit may affect syntactic parsing: limitations in the use of syntactic parsing strategies, and the consequent over-reliance on nonsyntactic parsing strategies.

The syntactic parser is a processor that tends to favor certain structures where more than one grammatical possibility exists partway through a sentence. Demonstrated parsing preferences have been used as an indicator of the relative complexity of syntactic structures. The subject's resolution of structural ambiguities is accomplished by decision-making strategies. Parsing strategies are used on line for ambiguity resolution, but they do not always result in the adoption of the correct structural analysis. When a listener or a reader is led to expect one particular syntactic organization by the first part of the sentence but is later required to reinterpret the structure, one might say that the perceiver has been led down a "garden path." As a consequence of limited working-memory storage we would expect poor readers to show greater susceptibility to garden path effects.

Eye movements in reading can reveal these garden path effects, in the view of Frazier and Rayner (1982). These investigators measured eye fixation times during sentences that would demand restructuring if the syntactic parsing strategy "Minimal Attachment" was being used (Frazier & Fodor, 1978). This is a parsing strategy that induces the reader to resolve local ambiguities most economically, by using the fewest possible nonterminal nodes in the constituent structure being assigned to the fragment of the sentence currently under analysis. Minimal Attachment predicts that a garden path will be pursued in example (17).

(17) John believed the big burly policeman was lying.

The minimal analysis of the noun phrase beginning "the big..." would assign it the grammatical role of Direct Object of "believe". But "believe" also permits a Sentential Complement, and, in this example, the phrase "the big burly policeman was lying" serves this grammatical role. Since sentence parsing strategies are applied on-line, according to Frazier and Fodor, the Direct Object analysis should be pursued first, producing a garden path effect when the word "was" is encountered, since it is this word that indicates the necessity for reanalysis.

Investigation of eye movements in reading sentences like (17) revealed that eye fixations are prolonged on the word that was predicted to initiate reorganization, indicating that the Minimal Attachment analysis had been adopted (Frazier & Rayner, 1982). Measurement of eye movements is useful not only in evaluating models of the sentence processing mechanism by which structural ambiguities are resolved, but it can also potentially inform us about differences in the use of this mechanism by good and poor readers. One testable hypothesis, using the eye-fixation tracking technique, is that poor readers are less likely than good readers to recover from garden paths because of their working memory limitations (see Shankweiler & Crain, 1986, for further discussion of this hypothesis).

The need for working memory in sentence processing might seem to be vitiated by parsing strategies, such as Minimal Attachment, that have the parser operate on small segments of speech or text. In our view, however, the existence of on-line strategies strengthens, not weakens, the argument that working memory plays an essential role in language processing. As Frazier and Fodor (1978) point out, the fact that verbal working memory decays rapidly and has limited capacity requires parsing decisions to be made quickly. Since for many poor readers, working memory limitations are even greater than normal, we would expect them to be more dependent on these on-line strategies for ambiguity resolution.

Over-reliance on nonsyntactic processing strategies is another expected manifestation of a working memory limitation. For example, upon encountering a pronoun in extended text, the reader must initiate a search for a referent. Although there are syntactic constraints on which noun phrases can serve as legitimate antecedents (Lasnik, 1976), we expect working memory limitations to lead poor readers to adopt nonsyntactic strategies based on proximity rather than hierarchical structure. In a recent study of this problem, it was found that poor readers tend to rely on a minimal distance strategy more often than good readers in determining the reference of reflexive pronouns, although the difference did not reach significance statistically (Shankweiler, Smith, & Mann, 1984).

It is worth emphasizing again that even rigid adherence to a structure-independent strategy by poor readers would not necessarily be indicative of syntactic incompetence, since there are so many other factors besides syntax involved in sentence understanding. Parsing preferences must be neutralized or factored out when the objective is assessment of active mastery of a particular syntactic structure. It is crucial that a subject's proclivity to use one structure at the expense of another must not be taken uncritically to indicate an incapacity to use the latter (Crain & McKee, 1985; Hamburger & Crain, 1984; Lasnik & Crain, 1985).

Summary and Conclusions

Previous research extending across languages and cultures indicates that the abilities that distinguish successful and unsuccessful readers are primarily in the language domain and not in the general cognitive domain, or in visual processing (Katz, Shankweiler, & Liberman, 1981; Liberman et al., 1982; Liberman & Shankweiler, 1985). Our focus, within this domain, has been on the relevance of syntactic complexity to reading acquisition and difficulties in comprehending text. We argued that in order to understand the special problems of comprehension in reading, we must address the problems of sentence understanding more broadly, by considering comprehension of speech as well. In pursuing these questions about the nature of syntactic complexity, we appealed to the distinction between structure and process, a distinction that enabled us to identify two possible sources of linguistic complexity in understanding spoken and written sentences. On one view, linguistic structures are taken to be ordered in complexity; on the other, it is not the structures themselves that make comprehension difficult, but the demands these structures make on the subsidiary processing mechanisms, especially verbal working memory.

Distinct predictions about the course of language acquisition arose from the different views of linguistic complexity. On the one hand, by adopting the thesis that language is a self-enclosed system, a module, the PDH predicts rapid acquisition of complex structures. On the other hand, a premise of the SDH is that some structures are inherently more complex than others. This would lead one to predict gradual, staged acquisition. These different conceptions of the course of language acquisition, in turn, yield different ways of viewing the problems of the beginning reader and the older unsuccessful reader. The SDH holds that these groups may not have acquired some of the language structures needed for learning to read successfully. The alternative is that the beginning reader has the language structures but has not yet managed to construct an efficient interface between these preexisting structures and the orthography, nor is he or she able to integrate the words of the text into higher order structures because of limitations on working memory. Each hypothesis can account for most of the basic facts about reading, and indeed, each often makes the same predictions. However, they identify different causes for failure to comprehend complex sentences, and these difference are amenable to empirical test.

Having developed the predictions, the next step was to examine the relevant empirical findings. First, it was shown that complex structures such as restrictive relative clauses and yes/no questions could be elicited successfully from children as young as three. These studies supported the rapid-acquisition scenario that the modularity hypothesis predicts and offered no support for the alternative staged-acquisition view. This led us to the second step in our argument. We asked whether subsidiary language mechanisms and not the language structures themselves might be the source of observed difficulties in the comprehension of complex syntax in reading. We expected the early stages of learning to read to be the most revealing. Accordingly, we sought an answer to this question by examining good and poor readers in the early grades. Studies of good and poor readers were presented that confirmed earlier claims that poor readers have difficulties in understanding complex sentences even when presented in spoken form. But these studies went on to suggest that the source of these difficulties was not a syntactic deficit as such. Instead, we found that good and poor readers were distinguished in efficiency of working memory, a subsidiary processing mechanism, rather than in syntactic competence. It is not clear whether the limitation is in the capacity of working memory <u>per se</u>, or whether it is in the "executive" or control component (Baddeley & Hitch, 1974). In Shankweiler and Crain (1986) we speculate that the control component of verbal memory is the site of the primary problem. In all events, the memory constraint would be expected to show up beyond sentence boundaries, for example, in relating pronouns to their antecedents.

In the preceding section, we examined the implications of working memory limitations of the poor reader for the reading process itself. Building on the bottleneck hypothesis of Perfetti and his associates, we explained how a working memory limitation could be expected to inhibit higher-level processing of text, by slowing word decoding and making it less accurate. This perspective tells us why poor readers are far less able to understand complex sentences in print than in speech, and it also explains their difficulties with spoken language. Finally, this hypothesis yields fairly specific predictions about the strategies for syntactic parsing on which beginning readers and poor readers should be expected to rely (although the research to test these predictions has not yet been done).

It follows from the bottleneck hypothesis that if our goal is to increase reading comprehension in beginning readers and unsuccessful readers, the first priority is to improve skills in recognizing printed words. It was argued that deficits implicating lower-level components in the structural hierarchy may have important repercussions at higher levels. In this connection, we would add that there is evidence that the abilities that underlie word decoding can be successfully taught at any age (see Liberman & Shankweiler, 1985; Liberman, Shankweiler, Blachman, Camp, & Werfelman, 1980). If we are correct in our other conclusion that the syntactic structures needed for sentence interpretation are already in place long before children actually encounter these structures in print, then the main thrust of efforts to improve reading should be directed to the inculcation of those lower level skills that pertain to use of the orthography. Only then can the working memory system be used effectively to gain access to the higher level syntactic, semantic, and pragmatic structures.

The position we have developed has definite implications, we believe, for the design and evaluation of appropriate text materials for beginning readers. It has long been appreciated that the beginning reader has special needs, but what these needs are has often been misunderstood. If the acquisition of the relative clause is indicative of the syntactic capacities of beginning readers, we should suppose that text designed for beginners need not simplify sentence structure. Since, in fact, the child of five or six is producing complex sentences in appropriate contexts, the avoidance of these complex structures in the text would likely be perceived as unnatural.

The findings we presented on early acquisition of complex structures suggest a caution, however. Complex syntactic structures, when used in reading materials, should appear in contexts that satisfy the presuppositions on their use, if good comprehension is to be achieved. We have seen that children as old as ten may have difficulties comprehending some sentences containing relative clauses, when these presuppositions are not met. One can expect, then, that without contextual supports, young children will often fail to display successful comprehension, but with these supports even texts containing complex syntactic structures may be read with understanding.

References

Baddeley, A. D., & Hitch, G. B. (1974). Working memory. In G. H. Bower (Ed.), The psychology of learning and activation (Vol. 4). New York: Academic Press.

Byrne, B. (1981). Deficient syntactic control in poor readers: Is a weak phonetic memory code responsible? Applied Psycholinguistics, 2, 201-212.

Brown, R. (1973). A first language. Cambridge, MA: Harvard University Press.

Calfee, R. C., Venezky, R. & Chapman, R. (1969). Pronunciation of synthetic words with predictable and unpredictable letter-sound correspondences (Tech. Rep. No. 71). Madison: Wisconsin Research and Development Center.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. Brain and Language, 3, 572-582.

Chomsky, N. (1971). Problems of knowledge and freedom. New York: Pantheon Books.

Chomsky, N. (1975). Reflections on language. New York, Pantheon Books.

Chomsky, N. (1981) Lectures on government and binding: The Pisa lectures. Dordrecht, Holland: Foris Publications.

Crain, S., & Crain, W. M. (in preparation). Restrictions on the comprehension or relative clauses by mentally retarded adults. Unpublished manuscript, University of Connecticut. University of Massachusetts, Amherst, MA.

Crain, S., Epstein, S., & Long, Y. (in preparation). Syntactic theory as a theory of language acquisition. Unpublished manuscript, University of Connecticut.

Crain, S., & McKee, C. (1985). Acquisition of structural constraints on anaphora. Proceedings of the North Eastern Linguistics Society 16, Amherst, MA: University of Massachusetts.

Crain, S., & Nakayama M. (1986). Structure-dependence in grammar formation. Unpublished manuscript, University of Connecticut.

Crain, S., & Steedman M. (1985). On not being led up the garden path: The use of context by the syntactic processor. In D. R. Dowty, L. Karttunen, & A. Zwicky (Eds.), Natural language parsing: Psychological, computational, and theoretical perspectives. Cambridge: Cambridge University Press.

Dawkins, J. (1975). Syntax and readability. Newark, DE: International Reading Association.

Dennis, M. (1980). Capacity and strategy for syntactic comprehension after left or right hemidecortication. Brain and Language, 10, 287-317.

de Villiers, J. G., Tager-Flusberg, H. B. T., Hakuta, K., & Cohen, M. (1979). Children's comprehension of relative clauses. Journal of Psycholinguistic Research, 8, 499-518.

Fletcher, J. M., Satz, P., & Scholes, R. (1981) Developmental changes in the linguistic performance correlates of reading achievement. Brain and Language, 13, 78-90.

Fodor, J. A. (1983). The modularity of mind. Cambridge, MA: MIT Press.

Fodor, J. A., & Garrett, M. (1967). Some syntactic determinants of sentential complexity. Perception & Psychophysics, 2, 289-296.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. Cognition, 6, 291-325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. Cognitive Psychology, 14, 178-210.

Goodluck, H. (1978). Linguistic principles in children's grammar of complement subject interpretation. Unpublished doctoral dissertation, University of Massachusetts.

Gough, P. B., & Hillinger, M. L. (1980). Learning to read: An unnatural act. Bulletin of the Orton Society, 30, 179-196.

Hamburger, H., & Crain, S. (1982). Relative acquisition. In S. Kuczaj (Ed.), Language development: Syntax and semantics (pp. 245-274). Hillsdale, NJ: Erlbaum.

Hamburger, H., & Crain, S. (1984). Acquisition of cognitive compiling. Cognition, 17, 85-136.

Jorm, A. (1979). The cognitive and neurological basis of developmental dyslexia: A theoretical framework and review. Cognition, 7, 19-33.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. Journal of Experimental Child Psychology, 32, 474-484.

Kean, M. L. (1977). The linguistic interpretation of aphasic syndromes: Agrammatism in Broca's aphasia, an example. Cognition, 5, 9-46.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. Cognition, 2, 15-47.

Lasnik, H. (1976). Remarks on coreference. Linguistic Analysis, 2, 1-22.

Lasnik H., & Crain, S. (1985). On the acquisition of pronominal reference. Lingua, 65, 135-154.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading, and dyslexia. Baltimore, MD: University Park Press.

Liberman, I. Y., Mann, V. A., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and non-linguistic material in relation to reading ability. Cortex, 18, 367-375.

Liberman, A. M., Mattingly, I. G., & Turvey, M. (1972). Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), Coding processes and human memory. Washington, DC: Winston and Sons.

Liberman I. Y., & Shankweiler, D. (1985). Phonology and the problems of learning to read and write. Remedial and Special Education, 6, 8-17.

Liberman, I. Y., Shankweiler, D., Blachman, B. A., Camp, L., & Werfelman, M. (1980). Steps toward literacy. In P. Levinson & C. C. Harris Sloan (Eds.), Auditory processing and language: Clinical and research perspectives. New York: Grune and Stratton.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. Journal of Experimental Child Psychology, 18, 201-212.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), Toward a psychology of reading: The proceedings of the CUNY Conferences. Hillsdale, NJ: Lawrence Erlbaum.

Limber, J. (1973). The genesis of complex sentences. In T. E. Moore (Ed.), Cognitive development and the acquisition of language. New York: Academic Press.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. Memory & Cognition, 8, 329-335.

Mann, V. A., Shankweiler, D., & Smith, S. T. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. Journal of Child Language, 11, 627-643.

Mattingly, I. G. (1972). Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye. The relationships between speech and reading. Cambridge, MA: MIT Press.

Mattingly, I. G. (1984). Reading, linguistic awareness, and language acquisition. In J. Downing & R. Valtin (Eds.), Languge awareness and learning to read. New York: Springer-Verlag.

Milner, B. (1974). Hemispheric specialization: Scope and limits. In F. O. Schmitt & F. G. Worden (Eds.), The Neurosciences: Third study program. Cambridge, MA: MIT Press

Perfetti, C. A. (1985). Reading ability. New York: Oxford University Press.

Perfetti, C. A., Goldman, S., & Hogaboam, T. (1979). Reading skill and the identification of words in discourse context. Memory & Cognition, 7, 273-282.

Perfetti, C. A., & Hogaboam, T. (1975). The relationship between single word decoding and reading comprehension skill. Journal of Educational Psychology, 67, 461-469.

Perfetti, C. A., & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, NJ: Erlbaum.

Rozin, P., & Gleitman, L. R. (1977). The structure and acquisition of reading II: The reading process and the acquisition of the alphabetic principle. In A. S. Reber & D. L. Scarborough (Eds.), Toward a psychology of reading: The proceedings of the CUNY Conference. Hillsdale, NJ: Erlbaum.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. Perception & Psychophysics, 2, 437-442.

Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorder, a modular approach. Cognition, to appear in November issue.

Shankweiler, D., Smith, S. T., & Mann, V. A. (1984). Repetition and comprehension of spoken sentences by reading-disabled children. Brain & Language, 23, 241-257.

Shankweiler, D., & Liberman, I. Y. (1976). Exploring the relations between reading and speech. In R. M. Knights & D. J. Bakker (Eds.), The neuropsychology of learning disorders: Theoretical approaches. Baltimore: University Park Press.

Shankweiler, D., & Liberman, I. Y. (1972). Misreading: A search for causes. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye: The relationships between speech and reading. Cambridge, MA: MIT Press.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. Journal of Experimental Psychology: Human Learning and Memory, 5, 531-545.

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. Journal of Verbal Learning and Verbal Behavior, 13, 272-281.

Solan, L., & Roeper, T. W. (1978). Children's use of syntactic structure in interpreting relative clauses. In H. Goodluck & L. Solan (Eds.), Papers in the structure on development of child language (pp. 105-126). University of Massachusetts Occasional Papers in Linguistics, Vol. 4. 105-126.

Stemmer, N. (1982). A note on empiricism and structure dependence. Journal of Child Learning 8, 629-633.

Tavakolian, S. L. (1981). The conjoined-clause analysis of relative clauses. In S. Tavakolian (Ed.), Language acquisition and linguistic theory. Cambridge, MA: MIT Press.

Vogel, S. A. (1975). Syntactic abilities in normal and dyslexic children. Baltimore, MD: University Park Press.

Whitaker, H. (1976). A case of isolation of the language function. In H. Whitaker H. A. Whitaker (Eds.), Studies in Neurolinguistics (Vol. 2). New York: Academic Press.

PHONOLOGICAL CODING IN WORD READING: EVIDENCE FROM HEARING AND DEAF READERS*

Vicki L. Hanson and Carol A. Fowler†

Abstract. The ability of prelingually, profoundly deaf readers to access phonological information during reading was investigated in three experiments. The experiments employed a task, developed by Meyer, Schvaneveldt, and Ruddy (1974), in which lexical decision response times to orthographically similar rhyming (e.g., WAVE-SAVE) and nonrhyming (e.g., HAVE-CAVE) word pairs were compared against response times to orthographically and phonologically dissimilar control word pairs. The subjects of the study were deaf college students and hearing college students. In the first two experiments, in which the nonwords were pronounceable, the deaf subjects, like the hearing subjects, were facilitated in their RTs to rhyming pairs, but not to nonrhyming pairs. In the third experiment, in which the nonwords were consonant strings, both deaf and hearing subjects were facilitated in their RTs to both rhyming and nonrhyming pairs, with the facilitation being significantly greater for the rhyming pairs. These results indicate that access to phonological information is possible despite prelingual and profound hearing impairment. As such, they run counter to claims that deaf individuals are limited to the use of visual strategies in reading. Given the impoverished auditory experience of such readers, these results suggest that the use of phonological information need not be tied to the auditory modality.

There is evidence that under some experimental conditions skilled readers with normal hearing access phonological information about the words they read. One such set of experimental conditions has been described by Meyer, Schvaneveldt, and Ruddy (1974). In their procedure, subjects are shown pairs of letter strings to which they respond "yes" if both letter strings are words and "no" if one or both are nonwords. There are four types of word pairs.

*Memory & Cognition, in press.
†Also Dartmouth College
Acknowledgment. This research was supported by Grant NS-18010 from the National Institute of Neurological and Communicative Disorders and Stroke and by Grant HD-01994 from the National Institute of Child Health and Human Development. We are grateful to individuals at Gallaudet College who made it possible for us to conduct the research. In particular we wish to thank Drs. Horace Reynolds, Donald Moores, and Pat Cox for their cooperation. We would also like to thank John Richards, Ignatius Mattingly, Rena Krakow, Alvin Liberman, Carol Padden and Nancy McGarr for their valuable discussions regard this research, and Nancy Fishbein, Debbie Kuglitsch, and Beth Schwenzfeier for their help in testing subjects.

Type 1 words rhyme and are spelled alike except for the first letter (for example, BRIBE-TRIBE). Type 2 words are neither orthographically nor phonologically similar; they are repairings of words of the first type and serve as control pairs for them. Type 3 word pairs consist of words that are spelled alike except for the first letter, but do not rhyme (for example, FLOWN-CLOWN). The fourth type of word pair consists of control words for these nonrhyming pairs.

Meyer et al. argued that if word reading were done on a completely visual basis, then the following equation should hold for response times:

$$\text{Type 2} - \text{Type 1} = \text{Type 4} - \text{Type 3}.$$

If, however, there was a phonological influence, then:

$$\text{Type 2} - \text{Type 1} \neq \text{Type 4} - \text{Type 3}.$$

The inequality was upheld in their study. Meyer et al. found a small facilitation effect for rhyming words (Type 1) as compared to control items of Type 2. They found a large interference effect for nonrhyming, orthographically similar pairs (Type 3) as compared with control items of Type 4. Because the rhyming and nonrhyming test pairs were equally similar orthographically, the differential outcome on the rhyming and nonrhyming pairs could be ascribed unambiguously to the differences in the phonological relationship between members of the two pair types.

Research subsequent to that of Meyer et al. has revealed that this pattern of facilitation and interference is dependent on task variables (Evett & Humphreys, 1981; Shulman, Hornak, & Sanders, 1978). For example, the pattern has been found to be related to the nonword distractors used in the task: When the nonwords are pronounceable nonwords (i.e., "pseudowords"), the pattern obtained by Meyer et al. (1974) is apparent, but when the nonwords are unpronounceable, there is facilitation for orthographically similar word pairs, whether rhyming or nonrhyming (Shulman et al., 1978). These latter findings have been used to argue against the notion of an obligatory phonological mediation in lexical access. However, the interpretation that the response time difference obtained with the procedure of Meyer et al. (1974) is caused by the discrepant phonological representations of the nonrhyming pairs of words remains unquestioned.

Our interest in the procedures of Meyer et al. derives from the information they may provide about word reading by deaf individuals. We ask here whether skilled deaf readers are able to access phonological information about a word under conditions in which skilled hearing readers do so. Therefore, any bias that the procedures may introduce toward accessing phonological information will be to our advantage.

There are at least two ways in which a prelingually, profoundly deaf reader might acquire information about the phonological forms of words. First, the alphabetic orthography itself provides phonological information. According to some theorists (Chomsky & Halle, 1968; Gleitman & Rozin, 1977; see also Crowder, 1982), the English orthography maps onto the phonological representations of words most directly at the level of the "systematic phoneme," which, putatively, is the level of phonological representation specified in the lexical entries of mature users of the language (but see

Linell, 1979; Steinberg, 1973). Although deaf readers might be able to acquire information about the systematic phonological forms of words from the orthography, any information that the orthography may thus provide will not distinguish the rhyming from the nonrhyming orthographically similar word pairs in the Meyer et al. paradigm. That is, there is nothing in the written forms of SAVE and WAVE on the one hand and HAVE and CAVE on the other, for example, that could reveal to a reader otherwise ignorant of the phonological forms of these words that the first pair of words is rhyming and the second pair nonrhyming. A second way in which a deaf reader might acquire information about the phonological forms of words is by learning to speak and/or lipread the language. This would enable acquisition of a phonetic or classical phonemic representation.

In the first two experiments that we describe below, we use the task of Meyer et al. (1974) to ask whether deaf readers access phonological information in a form that leads to facilitation when orthographically similar words rhyme and to interference when they do not. The deaf subjects of our study were college students, and thus, presumably, represent the more successful of deaf readers. To provide baseline data for interpreting the performance of the deaf subjects, a group of hearing college students was also tested.

Experiment 1

Method

Stimuli and design. The Word/Word pairs were the same pairs of words used by Meyer et al. (see Meyer et al. for a more complete discussion of the selection procedures for these pairs). These pairs were of four types. Type 1 (rhyming) word pairs were orthographically and phonologically similar (e.g., MARK-DARK, LOAD-TOAD). Type 2 pairs were control pairs that were both orthographically and phonologically dissimilar. These control pairs were constructed by interchanging the first and second members of the Type 1 pairs (e.g., MARK-TOAD, LOAD-DARK). Type 3 (nonrhyming) pairs were orthographically similar although phonologically dissimilar (e.g., GONE-BONE, PAID-SAID). Type 4 pairs were control pairs for the Type 3 pairs. These Type 4 pairs were both orthographically and phonologically dissimilar and were constructed by interchanging the two members of the Type 3 pairs (e.g., GONE-SAID, PAID-BONE). There were 48 pairs of each of the four types.

In addition to these 192 Word/Word pairs, 192 Word/Nonword pairs were constructed by pairing each word of the Type 1 and Type 3 Word/Word pairs with a pseudoword (pronounceable nonword). The pseudowords were formed by replacing the first letter of each word with a letter that made the string a pseudoword. Thus, as with the Word/Word pairs, half of the Word/Nonword pairs were orthographically similar (e.g., MARK-WARK, NAID-PAID) and half were orthographically dissimilar (e.g., ROWN-TOAD, PAID-TOST).

Using these Word/Word pairs and Word/Nonword pairs, two stimulus sets of 192 pairs each were constructed. Each set had half of each type of Word/Word pair and half of the Word/Nonword pairs. The two sets were constructed so that the words appearing in the Type 1 pairs in one set appeared in the Type 2 pairs of the other set. Similarly, the words that appeared in the Type 3 pairs in one set appeared in the Type 4 pairs of the other set. Thus, no word appeared twice in a Word/Word pair within either set. The Word/Nonword pairs of each set contained one member from each of the Word/Word pairs in the set. For half of the Word/Nonword pairs the word appeared first (on top); for the other half of these pairs, the nonword appeared first. For each stimulus set, a random order of pair presentations was generated, and this list was divided into six blocks of 32 trials each.

Two practice blocks of 34 trials each were generated. The stimulus pairs in these practice blocks were constructed in a manner consistent with the experimental blocks.

Procedure. The start of each trial was signaled by a 250 ms fixation point (a "+") presented in the center of a CRT display. Following this there was a 250 ms blank interval prior to stimulus presentation. The two letter strings for a trial were then presented, the first string centered one line above where the fixation cross had occurred and the second string centered one line below this point. The strings remained in view until either the subject pressed a response key or until 5 s had elapsed.

Feedback was given on each trial. The feedback consisted of the subject's response time (RT) for that trial (in ms), which, if the subject had been in error, was preceded by a minus sign. If the subject had failed to respond within the 5 s time limit, the words "TOO SLOW" appeared as feedback. The feedback, displayed for 250 ms, was centered six lines below the fixation cross. There was approximately a 2.5 s interval before the start of the next trial.

Subjects were instructed that on each trial they would be presented with two letter strings and that their task was to decide, as quickly and as accurately as possible, whether or not both letter strings were English words. The instructions were written, and the experimenter answered any questions that the subjects may have had about either the task or the feedback. For the deaf subjects, the experimenter was a deaf recent graduate of Gallaudet College who communicated with the subjects by signing. For the hearing subjects, the experimenter was a hearing person.

The subjects were shown the two response keys, one labeled "YES" and the other labeled "NO." If both letter strings on a trial were English words, subjects were to press the YES key; if both were not English words, they were to press the NO key. Subjects were instructed to keep their index fingers resting one on each key to achieve fastest response times (RTs).

All subjects were presented the two practice blocks, followed by testing with one of the two experimental stimulus sets. In addition, the assignment of the YES/NO keys for the two hands was counterbalanced across subject group and stimulus set.

Following this lexical decision task, subjects were presented with a rhyme judgment task. This was given to determine whether deaf readers could distinguish rhyming from nonrhyming pairs of words. The rhyme task was a paper and pencil test in which subjects were required to indicate whether or not the two words of each pair rhymed. Word pairs of Types 1-4 were typed (in lowercase letters), followed by a blank line. The written instructions informed subjects that they were to write YES on the blank line if the two words of a pair rhymed, and to write NO on the line if the two words did not. Two forms of the test were constructed. One form used the Word/Word pairs from Set 1 and their order of presentation from the lexical decision task; the other form used the Word/Word pairs of Set 2 in their previous order of presentation. Subjects received the form corresponding to the stimulus set they had received in the lexical decision task.

Deaf subjects. Deaf participants were 16 students from Gallaudet College, a liberal arts college specifically for deaf students. Measures of hearing loss and speech intelligibility were obtained from records at the college. As criteria for inclusion in the experiment, deaf subjects had to be prelingually deaf and have a profound hearing loss. Three of the participants failed to meet these criteria due to postlingual deafness (ages 3 or older) and were dropped from the study. In addition, the data of one deaf subject were excluded owing to an mean error rate more than 2.5 standard deviations greater than that of the group average. This resulted in 12 deaf subjects; Eleven were congenitally deaf, and the other was adventitiously deafened before the age of two years. Six of these subjects had deaf parents. All had a hearing loss of 90 dB or greater, better ear average. Half of the subjects were tested with one stimulus set; the other half with the second set.

The speech intelligibility ratings of the deaf subjects were based on a scale of 1 to 5, in which 1 is readily intelligible speech and 5 is unintelligible speech. Of the 12 deaf subjects in this experiment, two had speech that was rated '3,' meaning that the general public has some difficulty understanding the speech initially, but can understand it after repeated exposure to it; four of the subjects had speech that was rated '4,' meaning that the speech is very difficult for the public to understand; and six of the subjects had speech that was rated '5,' meaning that it cannot be understood.

The reading level of the deaf subjects was assessed by means of the comprehension subtest of the Gates-MacGinitie Reading Test (1969, Survey F, Form 2), which was administered following the rhyme judgment task. Survey F of the test is designed for hearing students in grades 10 through 12. On this comprehension test, a percentile score was determined for each subject based on grade level 10.1. The percentiles ranged from 97 to 7 (N=12, median = 22.5).

Hearing subjects. Hearing subjects were 16 students from Yale University who reported no history of hearing impairment. Eight of these subjects were tested with each experimental set.

Results and Discussion

For analysis purposes, RTs in the lexical decision task were stabilized by eliminating RTs in each condition that differed from the cell mean by more than two standard deviations. Table 1 provides the mean correct RTs (in ms) and mean percentage errors for each group and condition.

A difference score was obtained for each subject for phonological similarity (Type 2 minus Type 1) and phonological dissimilarity (Type 4 minus Type 3). Table 1 also provides the mean difference scores for the two subject groups. The Table shows that the hearing subjects exhibited the response pattern found by Meyer et al.; namely, a small facilitation effect on rhyming word pairs and a larger interference effect on nonrhyming pairs. The deaf subjects also responded differentially to the rhyming and nonrhyming pairs, but exhibited a somewhat different response pattern. These subjects showed relatively large facilitation on rhyming words, but neither facilitation nor interference on nonrhyming words.

Table 1

Mean RTs (in ms) in the lexical decision task of Experiment 1. Mean percentage errors are given in parentheses.

	Hearing	Deaf
Word/Word Pairs		
Phonologically similar	775 (7.5)	602 (13.9)
Control	800 (4.9)	657 (11.9)
Difference score	25 (-2.6)	55 (-2.0)
Phonologically dissimilar	845 (12.7)	631 (9.8)
Control	793 (6.1)	633 (11.3)
Difference score	-52 (-6.6)	2 (1.5)
Pseudoword/Word Pairs	931 (19.6)	732 (40.9)

Note: A positive number for the difference scores indicates facilitation and a negative indicates interference.

Using the difference scores, an analysis of variance was performed on the within-subjects factor of phonological relation (similar, dissimilar), and the between-subjects factors of group (deaf, hearing) and stimulus set (set 1, set 2). Stimuli were treated as a fixed effect owing to the constraints imposed upon stimulus selection in this experiment, and in Experiments 2 and 3 (see also Evett & Humphreys, 1981; Shulman et al., 1978). The factors of interest here are phonological relation and any interaction that may involve subject group.

The analysis yielded a significant main effect of phonological relation, $F(1,24) = 26.17$, $MSe = 2169.65$, $p < .001$, with difference scores in the phonologically similar condition tending to be positive (reflecting facilitation for rhyming word pairs) and difference scores for the phonologically dissimilar condition tending to be negative (reflecting interference). This main effect did not interact with either subject group or stimulus set, both $Fs < 1$. Thus, there was no significant difference in the magnitude of the effect of phonological relation for the hearing and deaf subjects. The higher order interaction involving these variables also was not significant, $F < 1$. This pattern of RTs is inconsistent with the hypothesis that graphemic information alone is utilized in this task by either hearing or deaf subjects. A main effect of group, $F(1,24) = 5.24$, $MSe = 4675.98$,

$p < .05$, reflected the fact that for the hearing subjects the mean of the difference scores was negative (reflecting a large interference effect and a smaller facilitation effect), while for the deaf subjects the mean of the difference scores was positive (reflecting only a facilitation effect).

An analysis of variance on the difference scores for the error data revealed no significant main effects or interactions (all $ps > .05$).

In the rhyme judgment task, deaf subjects made many errors, particularly on the orthographically similar but phonologically dissimilar (Type 3) word pairs. A similar error pattern was obtained for the hearing subjects, although their error rate was much lower. The mean percentage of errors for each word type for hearing and deaf subjects is shown in Table 2. This pattern of responding suggests that subjects' responses in this task were influenced by the orthographic similarity of the stimulus pairs. In fact, one deaf subject exemplified this strategy perfectly, by not making any errors on the Type 1 (orthographically and phonologically similar) word pairs on the rhyme judgment task but making an error on each of the 24 Type 3 pairs. One hearing subject showed much the same pattern by not making any errors on the Type 1 pairs and making errors on 17 of the 24 Type 3 pairs in this task.

Table 2

Mean percentage errors for deaf and hearing subjects in the rhyme judgment task.

	Type 1	Type 2	Type 3	Type 4
Deaf	28.1	5.6	70.8	3.5
Hearing	2.3	.8	11.2	.8

For the deaf subjects, correlations were computed between their speech intelligibility rating, reading achievement, accuracy on the rhyme judgment task, and RTs on the lexical decision task. The only correlation to reach significance was the correlation between speech intelligibility and errors on Type 3 word pairs on the rhyme task, $r(10) = -.81$, $p < .01$, two-tailed, which indicated that the more intelligible the speech the greater the accuracy on these pairs. Other correlations with speech intelligibility, although not significant (all $ps > .10$, two-tailed), were in the expected direction: The better the rated speech intelligibility, the greater the overall accuracy on the rhyme judgment task, $r = -.32$, and the larger the RT effect of phonological relation in the lexical decision task, $r = -.47$.

Experiment 2

Experiment 2 was similar to Experiment 1 with differences between the experiments in the stimulus sets, instructions to subjects, and in the form of rhyme judgment task.

The change in stimuli was motivated by the desire to replicate the findings of Experiment 1 on a new set of stimuli, an approach proposed by Wike and Church (1976) for showing generalization over stimuli. In our new stimulus set, we attempted to control for possible differences in the size of the orthographic neighborhoods of rhyming and nonrhyming words by selecting pairs of rhyming and nonrhyming words from a common neighborhood. For example, one rhyming pair of words in Experiment 2 was DONE-NONE; the corresponding nonrhyming pair was BONE-GONE.

The change in instructions was motivated by the high error rate among the deaf subjects in the first experiment; in this second experiment, we requested that subjects try to maintain a level of accuracy at or better than 90%.

The change in the rhyme judgment task was designed to force deaf subjects to try to make their judgments based on phonological information rather than orthographic similarity, if they could. As noted in Experiment 1, the deaf subjects and one hearing subject identified most of the Type 3 words as rhyming. We thought that this manner of responding might have been promoted by the fact that just one-fourth (rather than one-half) of the word pairs in the test rhymed. In Experiment 2, therefore, only pairs of Types 1 and 3 were included in the rhyming test. In addition, words were presented in matched pairs (for example, DONE-NONE was presented with BONE-GONE) and subjects had to select which of the two matched word pairs rhymed.

Method

Stimuli and design. The Word/Word pairs were chosen so that for each rhyming pair (Type 1) there was a nonrhyming pair (Type 3) that was orthographically similar, e.g, SAVE-WAVE and HAVE-CAVE; DONE-NONE and BONE-GONE. There were 32 such matched pairs. These stimuli are given in Appendix A.

In all other respects, the design of this experiment followed that of Experiment 1. Rhyme controls (Type 2 words) were generated by repairing the Type 1 words. Nonrhyme controls (Type 4 words) were generated by repairing the Type 3 words. Pseudowords were formed by replacing the first letter of each word with a letter that made the string a pronounceable nonword. In total, there were 128 Word/Word pairs and 128 pairs in which one of the items was a Pseudoword.

Two stimulus sets were constructed. Assignment of pairs to a list was made as in Experiment 1 with the one additional constraint that the matched orthographically similar pairs never both occur in the same list. Half of the Type 1 and Type 3 word pairs occurred in one set; the remaining word pairs in the other set. For each stimulus set, a random order of pair presentations was generated and presented as four blocks of 32 trials each.

Two practice blocks of 32 trials each were constructed in a manner consistent with list construction in the experimental blocks.

Procedure. The procedure of the lexical decision task was identical to that of Experiment 1 except that instructions to subjects stressed accuracy. Subjects in both groups were told to try to be at least 90% accurate. All subjects pressed the YES response key with their right hand and the NO response key with their left hand.

Following the lexical decision task, subjects were asked to complete a rhyme judgment task. This rhyming task was a paper and pencil test that consisted of 32 trials using pairs of words from the lexical decision task. On each trial, two word pairs were presented, one pair to the right of the other. The two words of each pair were orthographically similar, but words in one pair on each trial rhymed (a Type 1 pair) and the words in the other did not (a Type 3 pair). The two pairs on each trial were always the ones matched for orthographic similarity. Thus, for example, subjects would have to indicate whether it was the pair SAVE-WAVE or the pair HAVE-CAVE that rhymed. Each pair had a short blank line preceding it. Subjects were told that on each trial one of the two pairs rhymed. They were to indicate which of the two pairs rhymed by making a check on the line in front of the rhyming pair. For each subject, one of the two pairs on each trial had been tested in the lexical decision task; half of these previously seen stimuli were rhymes (Type 1 pairs) and half were nonrhymes (Type 3 pairs).

Deaf subjects. The deaf subjects were 16 students from Gallaudet College of whom 4 were subsequntly eliminated from the study. Two deaf subjects were ineligible due to postlingual deafness (age 3), and one due to a reported hearing loss less than the criterion of 85 dB. The data of a fourth deaf subject were eliminated due to excessive error rate (more than 2.5 standard deviations greater errors than the mean for the deaf subject group). Eleven of the remaining 12 subjects were congenitally deaf, and the other was adventitiously deafened before the age of one year. Four of these subjects had deaf parents. Five were tested on Set 1 and seven on Set 2.

Speech intelligibility ratings were available for eleven of the twelve subjects in this experiment. One of these subjects had speech that was rated a '2,' three had speech that was rated a '3,' four had speech that was rated a '4,' and three had speech that was rated a '5.'

Following the lexical decision and rhyme judgment tasks, the reading level of the deaf subjects was assessed by means of the comprehension subtest of the Gates-MacGinitie Reading Tests (1978, Level F, Form 2), designed for hearing students of grades 10-12. The percentile scores for the subjects ranged from 97 to 10 (N = 12, median = 48).

Hearing subjects. The hearing subjects were 14 students from Yale University. Seven subjects were tested with each stimulus set.

Results and Discussion

Consistent with the analyses in Experiment 1, RTs that differed by more than two standard deviations from a subject's mean in each cell were discarded. Shown in Table 3 are the means (in ms) for the correct RTs and the mean percentage errors for the two subject groups in each condition. Also shown are the difference scores for phonological similarity and dissimilarity. As can be seen from the table, the performance of the hearing subjects in Experiment 2 was remarkably similar to that of the hearing subjects in Experiment 1. The deaf subjects in Experiment 2 were slower and more accurate than those in Experiment 1, presumably because of our change in instructions emphasizing accuracy. Despite this change in position along the speed-accuracy continuum, the deaf subjects showed a pattern similar to that exhibited by the deaf subjects in Experiment 1; namely, a large facilitation on the rhyming pairs but little interference on the nonrhyming pairs.

Table 3

Mean RTs (in ms) in the lexical decision task of Experiment 2. Mean percentage errors are given in parentheses.

	Hearing	Deaf
Word/Word Pairs		
Phonologically similar	778 (7.5)	972 (8.7)
Control	804 (11.6)	1026 (10.1)
Difference score	26 (4.1)	54 (1.4)
Phonologically dissimilar	848 (15.2)	1003 (9.9)
Control	801 (10.3)	986 (9.7)
Difference score	-47 (-4.9)	-17 (- .2)
Pseudoword/Word Pairs	804 (14.8)	1078 (16.0)

Note: A positive number for the difference scores indicates facilitation and a negative number indicates interference.

The analysis of the RT difference scores indicated a main effect of phonological relation, $F(1,22) = 6.20$, $MSe = 9742.00$, $p < .05$, that did not significantly interact with group, $F < 1$. No other main effects or interactions were significant (all $ps > .25$). The main effect of phonological similarity reflected the fact that for rhyming pairs there was a response time facilitation, while for nonrhyming pairs there was a response time interference. This result indicated that both hearing and deaf subjects were influenced by the phonological similarity of the word pairs. The magnitude of the phonological similarity effect was not significantly different for the deaf and hearing subjects.

The analysis of the error data also indicated a main effect of phonological relation, $F(1,22) = 7.09$, $MSe = 38.98$, $p < .05$, and interactions of this variable with stimulus set, $F(1,22) = 12.63$, $p < .01$, and subject group, $F(1,22) = 6.24$, $p < .05$. The main effect resulted from fewer errors on the rhyming items than on the control, and from more errors on the nonrhyming items than on the control. The interaction with set reflected a larger influence of phonological relationship for one stimulus set than for the other. The interaction with subject group reflected a larger influence of phonological relationship for hearing subjects than for deaf subjects.

The lack of an interference effect among the deaf subjects in Experiment 1, and the relatively small interference effect among these subjects in Experiment 2, may have one of two origins. First, it may be that the deaf subjects individually, as well as collectively, showed no interference. Alternatively, some deaf subjects may have shown interference while others, failing to distinguish rhyming words from orthographically similar nonrhyming words, showed facilitation on both sets of words. To distinguish between these two possibilities, we looked at individual performances in the rhyming and nonrhyming conditions in Experiments 1 and 2. As a comparison, we also looked at individual hearing subjects.

The results of this classification are more in line with the the second alternative. As shown in Table 4, the individual responses revealed that for both hearing and deaf subjects, roughly half (slightly fewer) of the subjects showed facilitation on the phonologically similar word pairs and interference on the phonologically dissimilar word pairs. The magnitude of these facilitation and interference effects, as shown in Figure 1, was similar for the hearing and the deaf subjects (with the possible exception of the deaf subjects in Experiment 2 who actually showed a larger interference effect than the hearing subjects in that experiment). Inspection of the individual responses further revealed that the differences in pattern in the group data resulted from the fact that more deaf than hearing subjects exhibited facilitation on both of these word types, while more hearing than deaf subjects exhibited interference on both the phonologically similar and dissimilar pairs.

The results of the rhyme judgment task indicated that the hearing subjects more accurately discriminated between the rhyming and nonrhyming orthographically similar pairs than did the deaf subjects; the mean percentage correct responses were 99.6% and 64.1% for hearing and deaf subjects, respectively. Despite the fact that the deaf subjects thus made a considerable number of errors, their performance was significantly better than chance in this two-choice task, $t(11) = 4.05$, $p < .002$.

For the deaf subjects, further analyses yielded no significant correlations between individual subject characteristics (speech intelligibility and reading achievement) and accuracy on the rhyme judgment task or RT on the lexical decision task. The correlation between speech intelligibility and accuracy on the rhyme task was in the same direction as that in Experiment 1, $r(10) = -.56$, $.05 < p < .10$, two-tailed. The correlation between speech intelligibility and phonological relation in the lexical decision task was in the same direction as Experiment 1, although small, $r = -.13$.

Experiment 3

In Experiments 1 and 2, both the hearing and the deaf subjects responded differentially to rhyming and nonrhyming orthographically similar word pairs in the lexical decision task. This pattern of differential facilitation as a function of phonological similarity shown by both deaf and hearing subjects is consistent with the notion that subjects in both of these groups were accessing phonological information.[1] This outcome for hearing subjects is not remarkable, but it is surprising that prelingually, profoundly deaf subjects showed evidence of access to phonological information.

Table 4

Results of the analysis of individual subjects' data in Experiments 1 and 2. Shown are the mean percentages of hearing and deaf subjects whose response times revealed facilitation or interference as a function of whether the word pairs were rhyming (phonologically similar) or nonrhyming (phonologically dissimilar).

	Phonologically Dissimilar			
	Interference		Facilitation	
	Experiment		Experiment	
Phonologically Similar	1	2	1	2
Facilitation				
Hearing	50	43	25	14
Deaf	42	33	50	50
Interference				
Hearing	25	36	0	7
Deaf	8	0	0	17

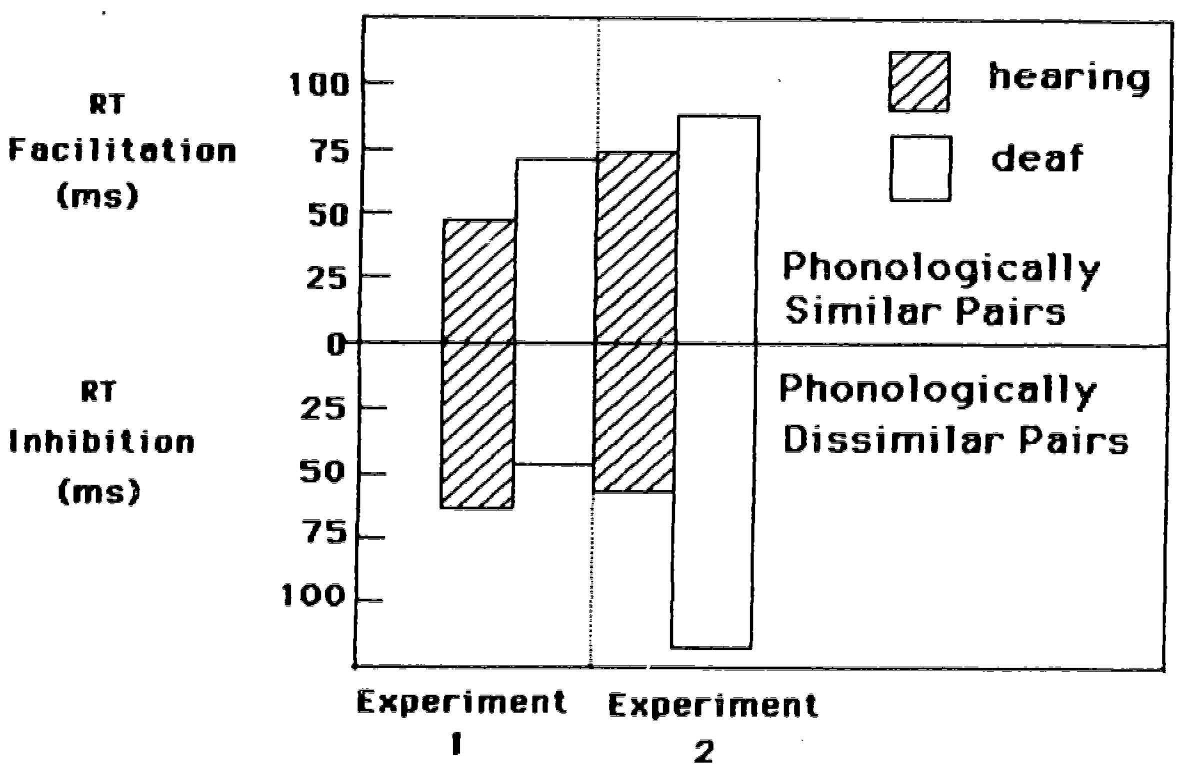


Figure 1. Mean response time difference scores (in ms) for the deaf and hearing subjects in Experiments 1 and 2 who showed both facilitation on rhyming (Type 1) word pairs and interference on nonrhyming (Type 3) word pairs.

To substantiate our conclusion that the response pattern of deaf subjects indeed reflects access to phonological information, we next used a manipulation that does not change the Word/Word pairs, but has been reported to change hearing subjects' pattern of performance in the paradigm of Meyer et al. (1974). This manipulation, performed by Shulman et al. (1978), uses consonant strings as nonword distractors. With hearing subjects, Shulman et al. found that this manipulation facilitated responding on orthographically similar nonrhyming (Type 3) as well as rhyming (Type 1) word pairs. The result was taken as evidence that the use of phonological information was reduced. The finding of semantic priming by Shulman et al. with orthographically and phonologically irregular nonwords indicated that subjects were accessing the lexicon in their task, not simply truncating the decision process after determining the regularity of the letter string.

If the effects obtained for the deaf subjects in our previous experiments could possibly be attributed to unidentified nonphonological factors (e.g., visual similarity or sign similarity differences in the rhyming and nonrhyming word pairs) then changing only the nonword distractors in the experiment should not influence their pattern of responding; that is, they should continue to show facilitation on the rhyming but not the nonrhyming pairs. If the pattern of results obtained for the deaf subjects in our previous two experiments were attributable to the phonological relationships between the two words of a pair, then when the distractor items are orthographically and phonologically irregular the deaf subjects, like the hearing subjects in the study of Shulman et al., should show facilitation on both the rhyming (Type 1) and nonrhyming (Type 3) word pairs.

In Experiment 3, therefore, we used the Word/Word pairs of Experiment 1, but altered the pseudowords of the experiment so that they were orthographically and phonologically irregular strings. Thus, the only difference between the stimuli of Experiments 1 and 3 was in the distractor items. Any difference in responding in the two experiments could therefore be attributed to this change.

Method

Stimuli and design. The two stimulus sets (and the two practice blocks) of Experiment 1 were used, with the items presented in the same order as in that previous experiment. The Word/Word pairs were identical to those of Experiment 1. The pseudowords of that experiment were changed to consonant strings by replacing each vowel in a pseudoword with a consonant.

Procedure. The procedure in the lexical decision task was identical to that of Experiment 2, with the exception that no specific mention was made of accuracy.

Following the lexical decision task, subjects were again given a rhyme judgment task. This task was similar to that of Experiment 2 in that on each trial subjects had to indicate which of two orthographically similar pairs rhymed. All of the rhyming (Type 1) and nonrhyming (Type 3) pairs from the two stimulus sets were used, resulting in 48 pairs. On each trial, both pairs were from the same set. Two forms of the test were constructed, using the same word pairs, but re-pairing the rhyming and nonrhyming pairs. These two forms of the test were given to different subjects.

Deaf subjects. The deaf subjects were 15 students from Gallaudet College. As in the first two experiments, the criteria for inclusion in the study were that subjects be prelingually, profoundly deaf. Two subjects were excluded due to postlingual deafness (ages 3 or older) and one was eliminated owing to a hearing loss less than the criterion of 85 dB. Twelve experimental subjects remained. Eleven of these subjects were congenitally deaf, and the other had been deafened before the age of two years. Two of these subjects had deaf parents. Five subjects were tested in experimental Set 1, and seven in Set 2. Due to experimenter error, three of the twelve subjects were not given the rhyme judgment task.

The speech intelligibility ratings were available from Gallaudet College for all but two of the subjects. The ratings for the remaining subjects were as follows: one subject had a rating of '2,' two subjects had a rating of '3,' five subjects had a rating of '4,' and two subjects had a rating of '5.'

Due to experimenter error, reading tests were given to only nine of the subjects in this experiment. Of those nine, four were given the comprehension test used in Experiment 1 (*Gates-MacGinitie Reading Tests*, 1969, Survey F, Form 2), and five were given the comprehension test of the more recent version of the test used in Experiment 2 (*Gates-MacGinitie Reading Tests*, 1978, Level F, Form 2). The percentile scores for these subjects in relation to grade 10.1 ranged from 79 - 9 (N=9, median=49).

Hearing subjects. The hearing subjects were 15 students from the University of Connecticut. Seven were tested with stimulus Set 1, and eight with Set 2.

Results and Discussion

As in Experiments 1 and 2, RTs that differed by more than two standard deviations from a subject's mean in each cell were eliminated from analysis. Shown in Table 5 are the means (in ms) for the correct RTs and the mean percentage errors for the two subject groups in each condition. The difference scores for phonological similarity and dissimilarity are also shown in Table 5. The difference scores for each of the two subject groups indicated facilitation on both the phonologically similar (Type 1) and phonologically dissimilar (Type 3) words. As can be seen from Tables 1, 3, and 5, changing the pseudowords to consonant strings resulted in an increased facilitation on rhyming pairs for both hearing and deaf subjects. Moreover, it resulted in facilitation of the nonrhyming, but orthographically similar, pairs as well.

The RT difference scores were entered into an analysis of variance on the factors of phonological relation, stimulus set, and subject group. The analysis indicated a main effect of phonological relation, $\underline{F}(1,23) = 6.28$, $\underline{MSe} = 2955.98$, $\underline{p} < .02$. No other main effects or interactions were significant (all $\underline{p}$s > .50). The main effect of phonological relation indicated that although there was facilitation for both the rhyming and nonrhyming words, the facilitation was greater for the rhyming pairs.

The analysis of difference scores for errors indicated no significant effects (all $\underline{p}$s > .25).

Table 5

Mean RTs (in ms) in the lexical decision task of Experiment 3. Mean percentage errors are given in parentheses.

	Hearing	Deaf
Word/Word Pairs		
Phonologically similar	592 (2.1)	542 (5.1)
Control	707 (6.1)	645 (12.6)
Difference score	115 (4.0)	103 (7.5)
Phonologically dissimilar	605 (2.4)	537 (1.8)
Control	681 (5.2)	603 (11.1)
Difference score	76 (2.8)	66 (9.3)
Nonword/Word Pairs	633 (8.3)	558 (14.8)

Note: A positive number for the difference score indicates facilitation.

The pattern of response times in Experiment 3 differed from that of Experiments 1 and 2 in that orthographic similarity facilitated responding, regardless of whether the words of a pair were phonologically similar or dissimilar. This was Shulman et al.'s finding, interpreted as evidence that access to phonological information is eliminated with consonant strings as nonwords. One aspect of our outcome leads us to treat this interpretation with caution. In Experiment 3, the effect of phonological similarity, for both hearing and deaf subjects, was still significant, albeit somewhat smaller, numerically, than in Experiments 1 and 2. Evidently, the procedures of Shulman et al. (1978) did not eliminate the influence of phonological information. This same pattern was obtained by Shulman et al. (1978). In their experiments, too, there was greater facilitation for the orthographically similar rhyming than nonrhyming pairs when irregular nonwords were used. Although the difference was not statistically significant in their study, the differences they obtained with irregular nonwords were consistent in direction and magnitude with the differences obtained here; the facilitation was greater for the rhyming pairs by 37 ms in their Experiment 1, by 31 ms in their Experiment 2, and by 24 ms in their Experiment 3. In the present experiment, the facilitation was greater for the rhyming pairs by an average of 38 ms (39 ms for the hearing subjects and 37 ms for the deaf subjects).

Rather than eliminating access to phonological information, the inclusion of the consonant strings as nonwords appears to have increased reliance on orthographic information in responding. This increased reliance on orthographic information can be seen as a criterion shift in Experiment 3, leading to fast rejection and somewhat quick acceptances. Comparison of RTs in Tables 1, 3, and 5 shows faster RTs in this third experiment, particularly on nonwords. This faster responding with orthographically and phonologically illegal nonwords was also obtained by Shulman et al. (1978).

The results of the rhyme judgment task were similar to those of Experiment 2. The deaf subjects were considerably less accurate than the hearing subjects, but again were better than chance. The mean percentage correct responses were 99.4% and 60.2% for hearing and deaf subjects, respectively. Despite the fact that the deaf subjects made a considerable number of errors, their performance was significantly better than chance in this two-choice task, $\underline{t}(8) = 3.21$, $\underline{p} = .02$.

For the deaf subjects, correlations were small and nonsignificant between individual subject characteristics (speech intelligibility and reading achievement) and accuracy on the rhyme task or RT on the lexical decision task. The correlation between speech intelligibility and accuracy on the rhyme task was in the same direction as those in Experiments 1 and 2, $\underline{r} = -.19$. The correlation between speech intelligibility and phonological relation in the lexical decision task was essentially zero, $\underline{r} = -.03$. The failure to obtain significant correlations consistently across the three experiments may be due to the restricted range of speech intelligibility scores and the relatively small numbers of subjects in the experiments.

General Discussion

The evidence from these studies suggests that deaf readers have access to phonological information in word reading (see Footnote 1). In the lexical decision tasks of all three experiments, the responses of both hearing and deaf subjects were affected by the phonological relationship between the orthographically similar pairs. This result was obtained using two different sets of Word/Word pairs (Experiments 1 and 2) and even when consonant strings were used as nonwords (Experiment 3).

The results obtained here argue against the possibility that the deaf subjects' differential responding to rhyming and nonrhyming pairs could have been due to differences in the visual similarity or the sign similarity of these pairs. The first argument against these interpretations is that Experiments 1 and 3 used the same Word/Word pairs, only the nonwords differed for the two experiments. This manipulation, while obviously not altering the visual or sign similarity of the word pairs, did alter the deaf (and hearing) subjects' pattern of responding. A second argument against a visual similarity interpretation is that the same pattern of results was obtained with two different sets of Word/Word pairs, with the visual similarity of the rhyming and nonrhyming pairs tightly controlled in Experiment 2. A second argument against a sign similarity interpretation is that there is no correspondence between American Sign Language signs and English phonology. There is no reason to expect, therefore, that the rhyming (Type 1) pairs of the experiments should be signed similarly while the nonrhyming (Type 3) pairs would not. Indeed, inspection of the word pairs used in these experiments showed that only one rhyming pair (and no nonrhyming pairs) could be considered to have similar signs.[2]

The major difference in the performance of the two groups was that the deaf subjects in Experiments 1 and 2 overall showed more facilitation on the rhyming pairs and less interference on the nonrhyming pairs than did the hearing subjects. Inspection of the individual patterns of performance in these two experiments showed, however, that some deaf subjects did exhibit both facilitation and interference comparable to that of the hearing subjects. The difference between the deaf and hearing subjects in the group data can be accounted for, primarily, by the tendency of some deaf subjects to show facilitation on both the rhyming and nonrhyming pairs, and, secondarily, by the tendency of a few hearing subjects to show interference on both types of word pairs. Thus, there were many subjects, in both the hearing and deaf groups, whose pattern of facilitation and interference gave evidence for the use of phonological information; there were also some subjects in both groups whose response pattern indicated that they failed to distinguish the rhyming from nonrhyming pairs. There was some suggestion that the pattern of facilitation and interference for the deaf subjects was related to rated speech intelligibility, with those subjects having the better rated speech showing the larger effects of phonological relation.

An outcome of the present study that requires further consideration is the deaf readers' performance on the rhyme task. In Experiment 1, their response pattern indicated a strong tendency to rely on orthographic similarity in making their rhyme judgments. This finding is consistent with other work on deaf individuals' explicit judgments of rhyme (e.g., Blanton, Nunnally, & Odom, 1967). However, in the rhyming tasks of Experiments 2 and 3, in which subjects were forced to make a rhyming judgment without relying on orthographic information, the deaf subjects demonstrated that they could make these judgments with better than chance accuracy.

Two features of the present study are particularly striking. The first is that not only were the deaf subjects accessing phonological information, but that they were doing so in a speeded task. It might be supposed that deaf readers would be confined to accessing phonological information in situations in which they have time to laboriously recover learned pronunciations. In the present study, however, they were found to access phonological information quite rapidly, suggesting that such accessing is a fundamental property of reading.

The second striking feature of this study is that the deaf subjects were not from predominantly oral backgrounds. All had received speech instruction in school, but considered sign to be their primary language. It is noteworthy that in their reading of English they utilized their phonological abilities. In this, the present results converge with evidence from short-term memory studies in which deaf readers, most notably the better ones, are sensitive to phonological similarity manipulations (Conrad, 1979; Hanson, 1982; Hanson, Liberman, & Shankweiler, 1984; Lichtenstein, in press).

We cannot determine from our research the nature of the deaf readers' phonological representations of words. We can conclude only that their representation of words must include phonological as well as orthographic information. Our findings are compatible with any hypothesized type of phonological representation as long as it captures the phonological similarity of our rhyming pairs and the dissimilarity of the nonrhyming pairs.

The representation could correspond closely to the detailed articulatory form of the word or it could be more abstract. An articulatory representation would not be incompatible with our findings that phonological information is accessed even by those deaf subjects whose speech is only poorly intelligible. It may well be the case that deaf individuals' ability to use some form of speech-based representation when reading is not well reflected in the intelligibility ratings of their speech. These intelligibility ratings are based on listeners' ability to understand the deaf speakers' utterances, not on the deaf individuals' ability to utilize speech in reading. Further research will be required to make the discrimination as to the type of phonological representation used.

In summary, the present study indicates access to phonological information by deaf readers. As such, the results run counter to claims that deaf individuals are limited to the use of visual strategies in reading. In interpreting these results, however, it is necessary to bear in mind that the deaf subjects in this study were college students, thus being some of the best educated of deaf individuals. Therefore, these results do not necessarily indicate that the use of phonological information is typical in the reading of deaf individuals. Rather, they indicate that access to this information is possible despite prelingual and profound hearing impairment. Given the impoverished auditory experience of such readers, these results suggest that the use of phonological information need not be tied to the auditory modality.

References

Blanton, R. L., Nunnally, J. C., & Odom, P. B. (1967). Graphemic, phonetic, and associative factors in the verbal behavior of deaf and hearing subjects. Journal of Speech and Hearing Research, 10, 225-231.

Chomsky, N., & Halle, M. (1968). The sound pattern of English. New York: Harper & Row.

Conrad, R. (1979). The deaf school child. London: Harper & Row.

Crowder, R. G. (1982). The psychology of reading: An introduction. New York: Oxford University Press.

Evett, L. J., & Humphreys, G. W. (1981). The use of abstract graphemic information in lexical access. Quarterly Journal of Experimental Psychology, 33A, 325-350.

Gates-MacGinitie Reading Tests. (1969). New York: Columbia Teachers' College Press.

Gates-MacGinitie Reading Tests, 2nd Edition. (1978). Boston: Houghton Mifflin Company.

Gleitman, L. R., & Rozin, P. (1977). The structure and acquisition of reading: I. Relations between orthographies and the structure of language. In A. S. Reber & D. L. Scarborough (Eds.), Toward a psychology of reading. Hillsdale, NJ: Erlbaum.

Hanson, V. L. (1982). Short-term recall by deaf signers of American Sign Language: Implications for order recall. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 572-583.

Hanson, V. L., Liberman, I. Y, & Shankweiler, D. (1984). Linguistic coding by deaf children in relation to beginning reading success. Journal of Experimental Child Psychology, 37, 378-393.

Lichtenstein, E. H. (in press). The relationships between reading processes and English skills of deaf college students: Parts I and II. _Applied Psycholinguistics_.

Linell, P. (1979). _Psychological reality in phonology_. Cambridge: Cambridge University Press.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual word-recognition. _Memory & Cognition_, _2_, 309-321.

Shulman, H. G., Hornak, R., & Sanders, E. (1978). The effects of graphemic, phonetic and semantic relationships on access to lexical structures. _Memory & Cognition_, _6_, 115-123.

Steinberg, D. (1973). Phonology, reading and Chomsky's optimal orthography. _Journal of Psycholinguistic Research_, _2_, 239-258.

Wike, E. L., & Church, J. D. (1976). Comments on Clark's "The language-as-fixed-effect fallacy." _Journal of Verbal Learning and Verbal Behavior_, _15_, 249-255.

Footnotes

[1]We do not mean to imply by this that deaf readers are using _only_ phonological information when they read. We focus on their use of phonological information only because it is so remarkable, given orthographic presentation of items to deaf individuals with poor speech intelligibility.

[2]The signs for the rhyming pair TOUGH-ROUGH in Experiment 2 are similarly produced. They are made with similar movement and location, but differ in handshape.

Appendix

Word pairs of Experiment 2

Type 1	Type 3
SAVE-WAVE	HAVE-CAVE
DONE-NONE	BONE-GONE
RUSH-GUSH	HUSH-BUSH
GOOD-WOOD	FOOD-HOOD
CARD-HARD	WARD-LARD
YARN-BARN	EARN-DARN
LIGHT-MIGHT	EIGHT-FIGHT
TON-WON	CON-SON
GULL-LULL	DULL-PULL
LORD-FORD	WORD-CORD
MATCH-PATCH	CATCH-WATCH
KID-BID	AID-RID
ROSE-HOSE	NOSE-LOSE
NEAR-REAR	DEAR-WEAR
HINT-TINT	MINT-PINT
MAID-RAID	PAID-SAID
SO-NO	GO-DO
DOVE-LOVE	MOVE-COVE
PUNT-HUNT	AUNT-RUNT
TOUGH-ROUGH	COUGH-DOUGH
TAR-FAR	BAR-WAR
FIVE-DIVE	HIVE-GIVE
HOST-POST	LOST-MOST
COW-VOW	NOW-LOW
RASH-DASH	CASH-WASH
CUT-BUT	PUT-NUT
HAND-LAND	WAND-SAND
TOMB-WOMB	BOMB-COMB
FEW-PEW	SEW-NEW
BAT-HAT	CAT-OAT
DOWN-GOWN	MOWN-TOWN
FAST-PAST	EAST-LAST

STRATEGIES FOR VISUAL WORD RECOGNITION AND ORTHOGRAPHIC DEPTH: A MULTI-LINGUAL COMPARISON*

Ram Frost,† Leonard Katz,†† and Shlomo Bentin†††

Abstract. The psychological reality of the concept of orthographic depth and its influence on visual word recognition were investigated by examining naming performance in Hebrew, English, and Serbo-Croatian. Three experiments were conducted using native speakers and identical experimental methods in each language. Experiment 1 revealed that the lexical status of the stimulus (high-frequency words, low-frequency words, and nonwords) significantly affected naming in Hebrew (the deepest of the three orthographies). This effect was only moderate in English, and nonsignificant in Serbo-Croatian (the shallowest of the three orthographies). Moreover, lexical status had similar effects on naming and lexical decision performance only in Hebrew. Experiment 2 revealed that semantic priming effects in naming were larger in Hebrew than in English and were completely missing in Serbo-Croatian. Experiment 3 revealed that a large proportion of nonlexical tokens (nonwords) in the stimulus list affects naming words in Hebrew and in English, but not in Serbo-Croatian. These results were interpreted as strong support for the orthographic depth hypothesis, and suggest that, in general, phonology in shallow orthographies is generated directly from print, whereas phonology in deep orthographies is derived from the internal lexicon.

Recognition of a word presented in the visual modality is ultimately based upon a match between a printed string of letters and a lexical representation. This match can be mediated by two types of codes: One that is based on some abstract representation of the orthography, and one that refers to phonemic information that is represented by the graphemic structure.

---

*Journal of Experimental Psychology: Human Perception and Performance, in press.

†Department of Psychology, Hebrew University
††Also Department of Psychology, University of Connecticut
†††Aranne Laboratory of Human Psychophysiology, Hadassah Hospital and The Institute of Advanced Studies, Hebrew University

Acknowledgment. This work was supported in part by National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories. The study is based on a doctoral dissertation presented by the first author to the Hebrew University. The authors gratefully acknowledge the very generous help provided by Georgije Lukatela, Predrag Ognjenović, Aleksandar Kostić, and all the other members of the Psychology Laboratory at the University of Belgrade. Without their support, this study would not have been completed.

There is some agreement that both code types are automatically activated during the process of word recognition, and act in parallel (but asynchronously) to mediate lexical access (but see Humphreys & Evett, 1985, for a critical review). The relative use of the orthographic and phonemic codes is determined by factors such as the subject's reading ability, the complexity of the stimuli, and task demands. For example, orthographic codes gain priority when the subjects are fluent readers, when the stimuli are very familiar, or phonemically irregular, and when the task emphasizes the graphemic aspects of the printed words. In contrast, phonological codes are employed relatively more by inexperienced readers, when the stimuli are more complex, and when the phonemic aspects of the material are emphasized by the task (for a review, see McCusker, Hillinger, & Bias, 1981).

The data on which the above suggestions have been based were provided primarily by studies conducted in English. Research outside of the English language suggested that in addition to these three factors, a bias toward one or the other code type may be tied to the depth of the language's orthography (Lukatela, Popadić, Ognjenović, & Turvey, 1980). Alphabetic orthographies can be classified according to the complexity of their letter-to-sound correspondences. In a shallow orthography, the phonemic and the orthographic codes are isomorphic; the phonemes of the spoken word are represented by the graphemes in a direct and unequivocal manner. In contrast, in a deep orthography, the relation of spelling to sound is more opaque. The same letter may represent different phonemes in different contexts; moreover, different letters may represent the same phoneme. Comparison of the English and Serbo-Croatian orthographies exemplifies the above distinction. The Serbo-Croatian writing system directly represents the phonology of the word; each grapheme unequivocally represents a single phoneme and each phoneme is represented by only one grapheme. Therefore, it is considered shallower than the English spelling system, which simultaneously represents both the phonology and morphology and mixes these representations inconsistently from word to word (Gleitman & Rozin, 1977). Consequently, generation of phonemic codes from print should be easier in Serbo-Croatian than in English. Several studies revealed that, indeed, lexical access in English is mediated by both orthographic and phonemic codes, while native readers of Serbo-Croatian are biased towards using phonemic codes in word recognition (Lukatela et al., 1980; Feldman, 1980).

The influence of orthographic depth on word recognition has been suggested in several studies that compared lexical decision and naming performance. It has been argued that in a shallow orthography, the extensive use of grapheme-to-phoneme translation[1] for word recognition might efficiently provide the articulatory codes used for pronunciation and, therefore, would minimize involvement of the lexicon in naming printed words. On the other hand, if the grapheme-to-phoneme translation is complex, the translation may be excessively costly in terms of time, and naming may be mediated by a lexical representation of the word. In this case, lexical access would have been achieved by means of an orthographic code, which then affords the word's stored pronunciation. Consequently, lexical processes in naming printed words should be more conspicuous in English than in Serbo-Croatian. Recently, Katz and Feldman (1983) compared pronunciation and lexical decision in English and Serbo-Croatian. In both languages naming was faster than lexical decision, but the difference was smaller in English. Furthermore, semantic priming facilitated lexical decision but not pronunciation performance in Serbo-Croatian, whereas in English, semantic priming was effective in both tasks. These results are in perfect accordance with the English language's putatively greater dependence on the lexicon for pronunciation.

The influence of orthographic depth on word recognition processes was apparently confirmed by the comparisons between English vs. Serbo-Croatian, but this conclusion is not without criticism. Orthographic depth is not the only dimension along which these two languages differ. English and Serbo-Croatian have different grammatical structures and possibly different lexical organizations (Lukatela, Gligorijević, Kostić, & Turvey, 1980). Since it is not known how those other factors may affect word recognition in English and Serbo-Croatian, attribution of differences in performance only to orthographic depth might be incorrect. Moreover, the effect of orthographic depth on printed word processing is not unanimously accepted. Recently, the claim has been made that the manner in which an orthography encodes phonology has little effect on skilled word recognition (Seidenberg & Vidanović, 1985).

One way to test the validity and psychological reality of the concept of orthographic depth is to find a third language that, although different from either of the other two in many aspects, would represent a third point along the continuum of orthographic depth. Assuming that orthographic depth is indeed the relevant factor and that there is no other relevant dimension on which the three languages may be aligned along a continuum, the effects found in a two-language comparison should be found to extend in a systematic manner to the three-language comparison. An appropriate ordering of the three languages on a given measure would corroborate the psychological reality of the orthographic depth factor more strongly because the predicted ordering would be one out of six possibilities of order (for the three languages), instead of only one out of two possibilities (for the two languages).

The Hebrew language provides a natural third point on the continuum of orthographic depth. In Hebrew, consonantal information is represented by letters, and vowels are mainly conveyed by small diacritical marks added to the consonants. These vowel marks, however, are omitted from regular reading material such as literature (except poetry), newspapers, advertisments, street signs, etc. Although the full writing system (consonants and vowel marks) is taught in the first two grades of elementary school, the adult reader is exposed almost exclusively to unvowelized print. Therefore, the Hebrew orthography is an extreme example of ambiguity. Because several words may share an identical consonant structure, many consonant strings can be pronounced in several ways, each producing a different legal Hebrew word. This is in complete contrast to Serbo-Croatian and is essentially different from English as well. (English has only a few heterophonic homographs: bow, wind, read, etc. An English reader can get some feeling for the adult Hebrew orthography by imagining an English orthography in which the vowels are omitted: The string "bttr" would stand for "batter," "better," "bitter," and "butter" and, of course, a large number of nonwords, e.g., "bottir," etc.). Clearly, in Hebrew orthography, the full phonemic code of the word is less transparent than it is in either English or Serbo-Croatian, Hebrew representing the third, deepest, point along the continuum of orthographic depth.[2]

Previous studies have already suggested that, in Hebrew, orthographic codes play a more important role in the process of word recognition than do phonemic codes, especially in comparison with the roles played in other languages. For example, it was found that rejection of nonwords in a lexical decision task is slower if they are orthographically similar than if they are phonemically similar to real words (Bentin, Bargai, & Katz, 1984). Moreover, addition of the missing phonemic information (vowel marks) did not facilitate or even delay lexical decision (Bentin & Frost, in press; Koriat, 1984).

None of the studies mentioned, however, made a controlled comparison between languages, and therefore the relative importance of phonemic and orthographic codes in different languages was not directly examined. The present study sought to fill this gap. We hoped to improve the validity of the interlanguage comparison (1) by using identical methodology and apparatus for all three languages and (2) by studying all three in their native environments to insure that the language a subject was tested on was, in fact, the actual language environment of the subject. Thus, we hoped to provide clearer evidence for or against the notion that the directness with which an orthography represents its language's phonology determines the relative use of orthographic vs. phonological codes in printed word perception.

General Methods

Because this study was conducted in three different countries, special care was taken to standardize procedures, materials, and apparatus. The same experimenter ran identical experiments in Israel, Yugoslavia, and the United States.

Stimulus selection and word-frequency evaluation. All stimuli in each languge were two-syllable nouns that had a stop consonant as their first letter. Native speakers constructed the word data base in each language, but, wherever possible, literal translations were used. Homographs and homophones were not used. In Hebrew all stimuli were undotted (i.e., without vowel marks), but could be pronounced as only one real word. Because there are no reliable sources of standard objective word frequency in Hebrew and in Serbo-Croatian, we devised a procedure for estimating subjective frequencies. Recently, subjective and standard objective word frequencies were found to be highly correlated (Gordon, 1985). The same procedure of frequency estimation was used in all three languages: Two hundred words that conformed to the above criteria were printed on two pages (in Hebrew, without the vowel marks). Fifty undergraduates were asked to rate each word on a five-point scale ranging from least frequent (1) to most frequent (5). Estimated frequency for each word was calculated by averaging the ratings across all fifty judges. Based on these ratings, groups of high- and of low-frequency words were selected.

Experiment 1

This experiment was designed to assess the effect of lexical factors on naming of words and of nonwords across Serbo-Croatian, English, and Hebrew and to relate naming to lexical decision performance in the three languages. This technique was employed in order to assess the hypothesis that the deeper the orthography, the more the reader will depend on lexical information for naming.

Previous studies in English suggested that naming and lexical decision performance are significantly correlated; this correlation was interpreted as evidence that naming in English is usually lexically mediated (Forster & Chambers, 1973; see also Forster, 1979; Theios & Muise, 1977; and West & Stanovitch, 1982). More recently, Katz and Feldman (1983) used the same technique to assess the extent of lexical mediation for naming in Serbo-Croatian and to compare it to English. In that study, they also used semantic priming as a manipulation that was assumed to affect only lexically mediated processes. Semantic priming effects were found for lexical decision in both languages, but for naming, only in English. Moreover, in English, but not in Serbo-Croatian, they found significant lexical decision-naming correlations, regardless of the semantic relationship between target words and previously presented primes. Based on these results, the authors concluded that naming is less mediated by lexical information in the orthographically shallow Serbo-Croatian; naming in that language was apparently dependent on prelexical phonological coding. However, attributing the differences between the two languages specifically to orthographic depth is problematic because they differ in other ways as well (as discussed above).

Assessment and interpretation of the effects that orthographic depth might have on word recognition is further complicated by the results of two recent studies: Hudson and Bergman (1985) revealed that if only words are to be named (in a blocked condition), lexical involvement (as reflected by word frequency effects) may be found even in Dutch, which has a shallow orthography. In addition, phonological manipulations had a similar effect on naming words in the deep Chinese logography and in the English alphabetic orthography. In both languages only very infrequent words were affected (Seidenberg, 1985).

The present study addresses these controversies; we attempted to assess the validity of the orthographic depth hypothesis (1) by using a three-language comparison and (2) by using the lexical decision results for each language only as a reference point against which its naming results could be interpreted. Thus, Experiment 1 investigates how factors that are generally agreed to involve lexical processing affect naming in each language; we consider the effects of the same factors on lexical decision only as a point of reference.

The most obvious lexical factor is the difference between words and nonwords. Although some authors have suggested that nonwords might be pronounced by referring to related lexical entries for words (Glushko, 1979), few would suggest that nonwords are represented in the lexicon. Therefore, one can reasonably assume that, in most cases, pronounciation of nonwords is mediated by a process of grapheme-to-phoneme translation performed outside the lexicon. However, if the grapheme-to-phoneme translation is the route chosen for naming both words and nonwords, the lexical status of the stimulus should have only a small effect on performance. On the other hand, if the lexical route is the strategy usually chosen for naming words, naming nonwords should be delayed by the lack of a lexical entry.

A similar argument can be made about the effects of word frequency. It can be expected that for processes that depend on lexical search, word frequency should affect performance more than for processes that do not involve lexical mediation. Although the frequency of the word may confound prelexical and lexical factors, there is little doubt that both levels influence lexical access. To the extent that naming depends primarily on prelexical (i.e., phonologic) information, word frequency should affect word pronunciation less.

Assuming that orthographic depth indeed affects word recognition, we predicted that: (1) Lexical factors would influence naming in Hebrew more than in English, and in English more than in Serbo-Croatian, and (2) because naming and lexical decision would share more commonality in deep than in shallow orthographies, the influence of lexical factors on the two tasks should be more similar in Hebrew than in English, and more similar in English than in Serbo-Croatian.

Methods

Subjects. The subjects were all under-graduates who participated as part of the requirements of psychology courses. There were 48 students from the Hebrew University, 48 from the University of Connecticut, and 48 from the University of Belgrade. They were all native speakers of Hebrew, English, and Serbo-Croatian, respectively. A different set of 24 subjects in each language were employed in the naming and in the lexical decision tasks.

Stimuli and apparatus. The same list of 48 words and 48 nonwords was used for lexical decision and for naming. All stimuli were 3 to 7 letters long. Because vowels are omitted in Hebrew, the average number of letters per word was smaller than in either English or Serbo-Croatian, which did not differ between themselves. Note, however, that the range of phonemes per word was similar in the three languages (4 to 6 phonemes), and the means did not differ significantly. The word stimuli were composed of 24 high-frequency and 24 low-frequency words selected from those that were rated above 4.0 or below 2.0 respectively. The mean ratings of the high-frequency groups were 4.42, 4.40, and 4.30, and of the low-frequency groups were 1.72, 1.71, and 1.68 in Hebrew, English, and Serbo-Croatian, respectively. Each nonword was produced by replacing one letter of a real word. All letters were normal characters generated by a computer on the center of a CRT screen. On the average, a stimulus subtended a visual angle of approximately 2.5 degrees.

Lexical decisions were communicated by pressing either a "Yes" or a "No" button. The dominant hand was always used for the "Yes" (i.e., Word) responses and the other hand for the "No" (i.e., Nonword) responses. In the naming task, subjects' verbal responses were recorded by a Mura-DX 118 microphone connected to a voice key. Reaction times were measured in milliseconds from stimulus onset.

Procedure. Subjects were randomly assigned to either the lexical decision or the naming task. They were tested individually in a semi-darkened room. The instructions were to respond as quickly and as accurately as possible by pressing one button (in the lexical decision task) or by pronouncing the word (in the naming task). Following the instructions, 15 practice trials were presented. A trial began by presenting a stimulus, which was removed by the subject's response. Following the practice trials, the 96 test trials were presented in one block at a 3-sec inter-trial interval. In the naming task, incorrect pronunciations were recorded by the experimenter and scored as errors.

Results

Each subject's distribution of reaction times was normalized by excluding RTs that were above or below two standard deviations from the subject's own mean. The percent of outliers was similar across all word conditions, less than 2.5%. This procedure was followed for all subsequent experiments reported in this paper. An analysis of variance assessed the effects of Language (Hebrew, English, Serbo-Croatian), Task (Lexical Decision, Naming), and Stimulus Group (High-Frequency Words, Low-Frequency Words, Nonwords). The mean reaction times for conditions are presented in Figure 1.

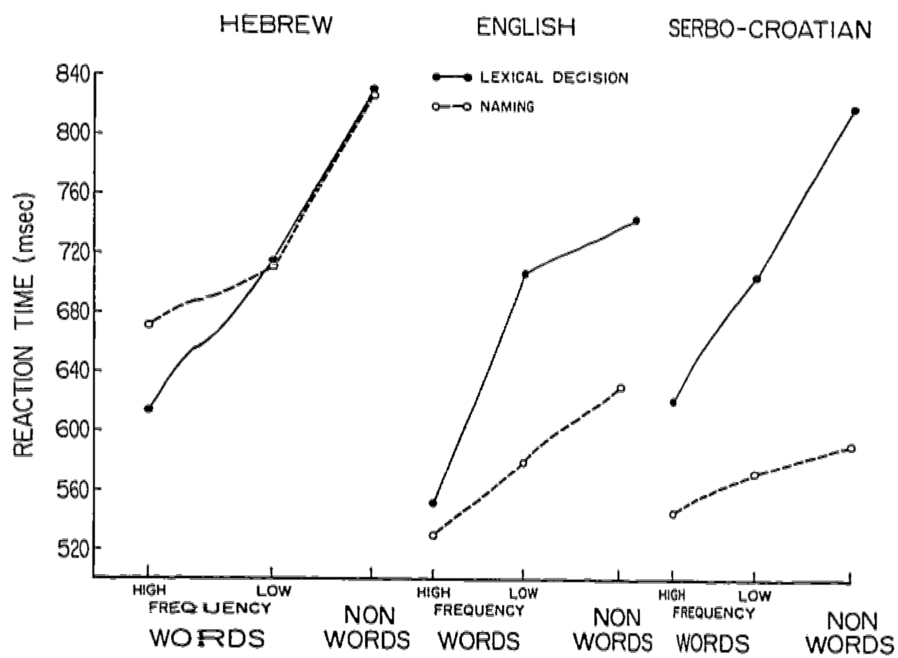


Figure 1. Average response time to high-frequency words, low-frequency words and nonwords in the naming task (dashed lines), and in the lexical decision task (solid line), in Hebrew, English, and Serbo—Croatian.

All main effects and two-way interactions were significant; however, the most important result was the three-way interaction, which was significant both for the stimulus analysis, $\underline{F}(4,207)=9.07$, MSe=1482, and for the subject analysis, $\underline{F}(4,276)=6.30$, MSe=2051, with minF'(4,46)=3.72, $\underline{p}<.02$. This interaction demonstrates that each task was affected differently by the stimulus group manipulation in each language. In the naming task, the reaction time difference between nonwords and high-frequency words systematically decreased from 157 ms in Hebrew, to 101 ms in English, and to 56 ms in Serbo-Croatian. In contrast, in the lexical decision task, these differences were similar across languages (217 ms, 192 ms, and 198 ms for Hebrew, English, and Serbo-Croatian, respectively).

The influence of the language on naming was more conspicuous when comparing words and nonwords than when comparing high- and low-frequency words. This observation was tested by analyzing separately those two effects on naming. First, a Frequency X Language ANOVA (with the exclusion of nonwords) revealed that Hebrew words were named significantly slower than English and Serbo-Croatian, with no difference between the latter two, $\underline{F}(2,69)=21.24$, MSe=14314. High-frequency words were named faster than low-frequency words, $F(1,69)=137.99$, MSe=478. However, the interaction between word frequency and language was not significant, $F(2,69)=1.57$, MSe=478. The effect of lexicality was assesed by a second ANOVA that compared nonword performance with the mean of high- and low-frequency words. It revealed that both main effects were significant: $F(2,69)=26.36$, MSe=1954 for language, and $F(1,69)=159.16$, MSe=1566 for stimulus type. More importantly, the interaction between the stimulus type (words/nonwords) and language was significant, $F(2,69)=20.0$, MSe=1566, suggesting that lexicality influenced naming differently in each language.

A second important result is the Task by Language interaction. Across stimulus groups, naming was 149 ms faster than lexical decision in Serbo-Croatian, 88 ms faster in English, but it was 17 ms slower in Hebrew. Note, however, that the difference between naming and lexical decision in Hebrew was significant only for the high-frequency words.

Table 1 presents the mean percentage of errors in each condition.

Table 1

Percent of Errors in the Naming and in the Lexical Decision Tasks in Hebrew, English, and Serbo-Croatian.

	LEXICAL DECISION			NAMING		
	Hfreq	Lfreq	Nonwords	Hfreq	Lfreq	Nonwords
Hebrew	1.0	8.2	3.1	0.3	7.3	7.5
English	0.1	12.5	5.0	0.2	2.5	2.6
Serbo-Croatian	0.5	7.8	3.7	0.3	0.5	5.0

Because the number of errors was small, their distribution did not permit a three-factor analysis. However, some trends were observed. The pattern of Stimulus Group effects on lexical decision was similar across languages. In contrast, in the naming task the pattern of the effects was different in each language. This difference is most conspicuous when high- and low-frequency words are compared. In Hebrew, the difference between the number of low- and high-frequency words that were pronounced incorrectly is practically the same as the difference between the number of incorrect lexical decisions made with low- and with high-frequency words. In English this difference is considerably reduced in the naming task relative to the lexical decision task, whereas in Serbo-Croatian an equal percentage of errors was found with high- and low-frequency words.

Discussion

The results of this experiment substantiated the hypothesis that the deeper the orthography, the more lexical mediation occurs. In Hebrew, naming was affected by lexicality and word frequency, variables that are believed to affect processing in the lexicon. Consistent with the orthographic depth hypothesis, the effects of these factors were smaller in English, and even smaller in Serbo-Croatian. Furthermore, similarity--or the lack of similarity--between tasks in their sensitivity to word frequency and lexicality was revealing. In Hebrew, naming and lexical decision performance were similarly affected by the lexical nature of the stimulus and except for the high-frequency words, the reaction times in the two tasks were practically identical. In contrast, in Serbo—Croatian, which represents the other end of the orthographic depth continuum, lexical factors had only a slight influence on naming even though their effect on lexical decision was almost as strong as in Hebrew. In English, as might be expected by its place on the continuum (deeper than Serbo-Croatian but more shallow than Hebrew), lexical factors affected both lexical decision and naming.

Note that the lexicality effect (the difference between nonwords and the mean of high- and low-frequency words) discriminated between languages more than the word frequency effect. In fact, in agreement with Seidenberg and Vidanović (1985), word frequency (high vs. low) did not have a significantly different effect on naming English and Serbo-Croatian words. However, when we consider nonwords, we see a slightly sharper diffence between Serbo-Croatian and English and, when we extend our view to Hebrew, that difference becomes very marked, indeed. If only Serbo-Croatian and English had been studied, and only word frequency effects considered, the results may have led to the incorrect conclusion that orthographic depth has no influence on word recognition (see also Fredriksen & Kroll, 1976).

Further analysis supports this view. A common finding in English and Serbo-Croatian is that naming is faster than lexical decision (cf. Forster & Chambers, 1973; Fredriksen & Kroll, 1976; Katz & Feldman, 1983); this finding is replicated in the present study. In contrast, there is no such ordering for Hebrew; indeed, the opposite appears to be true, at least for high-frequency words. Apparently in Hebrew, naming cannot be accomplished before lexical decision. This result is consistent with the suggestion that naming is Hebrew is lexically mediated and replicates our previous observations (Bentin et al., 1984; Bentin & Frost, in press). Presumably, naming depends on lexical information in Hebrew because the print provides only partial phonemic information.

With regard to the two other languages, the comparison of English and Serbo-Croatian reveals that the difference between lexical decision and naming is smaller for English. This greater similarity between the two tasks for English suggests that naming in English shares more of its processing with lexical decision than is true for Serbo-Croatian. Nevertheless, it can be seen that the relation between naming and lexical decision time is similar in Serbo-Croatian and English for low—frequency words, while the difference is most conspicuous for high-frequency words: In Serbo-Croatian, naming was 86 ms faster than lexical decision, in English, 22 ms, while in Hebrew, the effect was reversed and naming was 43 ms slower than lexical decision. This pattern partially supports the hypothesis that there is a tendency to recognize high-frequency words on a graphemic basis (Seidenberg, 1985).

However, the results of our experiment suggest the qualification that the use of graphemic/orthographic codes depends additionally on the orthography being read. In a very shallow orthography, even recognition of high-frequency words may be mediated by grapheme-to-phoneme translation done outside the lexicon.

Experiment 2

In Experiment 1 we assessed the extent of lexical mediation for naming by manipulating the lexical status of the stimulus (high or low frequency, word or nonword). Although suggestive, these manipulations cannot be unequivocally interpreted as influencing lexical processing and Experiment 2 was designed to offer converging evidence. High-frequency words, low-frequency words, and nonwords differ not only on lexical dimensions but also on orthographic familiarity, a factor that might influence prelexical processes of word recognition (cf., Mason, 1975). Furthermore, recent evidence suggests that lexical access is not the only process that is influenced by word frequency in pronunciation tasks; postlexical processes are also affected (Balota & Chumbley, 1985). Therefore, the link between orthographic depth and the degree of lexical information used in naming needs to be subjected to further evaluation. Experiment 2 was designed to assess lexical involvement in naming Hebrew, English, and Serbo-Croatian words by using the semantic priming technique.

Since the first Meyer and Schvaneveldt (1971) report, numerous studies have shown that words are recognized faster if they are presented simultaneously with, or immediately following, a semantic associate than if they are paired with an unrelated word. Most of these studies used a lexical decision task. Several studies, however, suggested that semantic priming might also be effective in naming (Becker & Killion, 1977; Meyer, Schvaneveldt, & Ruddy, 1975). In comparison to lexical decision, semantic priming effects in naming are usually smaller, and can be obtained only with strongly associated word pairs (Forster, 1981; Lupker, 1984). One possible interpretation of the weaker effect of semantic priming on naming than on lexical decision is that in English the use of lexical information for naming may be limited to a only a subset of words (such as the exception words), while others are named by means of at least some phonological coding directly from print. This interpretation is consistent with the orthographic depth hypothesis; it implies that the size of the semantic priming effect on naming should correlate positively with orthographic depth.

Several authors have suggested that the magnitude of semantic priming can be influenced by the "depth" at which words are analyzed (Henik, Friedrich, & Kellogg, 1983; Smith, Theodor, & Franklin, 1983; In an analogous way, the rationale of Experiment 2 was based on the assumption that semantic information used to prime a target word in the lexicon should facilitate pronunciation of the target only if the naming process is able to utilize lexically activated information. More generally, we would expect the effect of semantic priming to be greatest for an orthography that putatively depends most strongly on lexical information for pronunciation (e.g., Hebrew) and smallest for the orthography that depends least on lexical mediation (e.g, Serbo-Croatian). The effect on English should be intermediate to the others.

Previous inter-language comparisons of semantic priming effects on naming contrasted only English and Serbo-Croatian, and were contradictory. One study reported that in English, semantic priming equally facilitated lexical decision and naming performance, while in Serbo-Croatian naming was not facilitated at all (Katz & Feldman, 1983). Different results, however, have been recently reported (Seidenberg & Vidanovic, 1985). In that study, naming printed words was equally facilitated by semantic priming in both English and Serbo-Croatian. A major methodological difference between the studies of Katz and Feldman (1983) and Seidenberg and Vidanović (1985) is the selection of subjects. The first study was conducted in Yugoslavia where the subjects were undergraduate students; the subjects in the second study were mainly less well-educated Yugoslavian workers in Montreal. A second difference concerned the kind of relation between prime and target. For Katz and Feldman, this was a superset-subset relation (e.g., music - jazz), while for Seidenberg and Vidanovic, an associative relation was used. Nevertheless, in both studies, the same stimuli produced semantic facilitation in a lexical decision task. Thus, although the different results may reflect methodological differences between the two studies, Seidenberg and Vindanovic's report raises doubts about the generality of the orthographic depth effect.

In the present experiment, we hoped that the careful matching of methodology and stimuli in all three languages would allow an unequivocal cross-language comparison. If there is an orthographic depth effect on word recognition, semantically related primes should faciliate naming performance in Hebrew more than in English, and in English more than in Serbo-Croatian.

Methods

Subjects. The subjects were undergraduates studying in Jerusalem, St.orrs, and Belgrade who participated in the experiment as part of the requirements of their respective courses in psychology. There were 48 native speakers in each language.

Stimuli and design. The critical stimuli were 32 target words, none of which had been used in Experiment 1. The mean word frequency rating was approximately the same for the three languages: 3.32, 3.24, 3.20, for Hebrew, English, and Serbo-Croatian, respectively. Each target word was paired with a semantically related prime. A target and a prime were different examples of one semantic category (e.g., lion-tiger; rifle-canon). Semantic categories were used only once. Whenever possible straightforward translations from language to language were made; otherwise, two other examples from the same category were usually selected. In addition to the critical targets, 16 words (mean frequency 3.30, 3.31, and 3.30 for Hebrew, English, and Serbo-Croatian, respectively) were paired with nonwords. The 48 stimulus pairs were compiled into two stimulus lists. In each list, only 16 out of the 32 critical targets were presented in conjunction with their related primes. The remaining 16 primes were redistributed between the other 16 targets such that no obvious semantic relationship could be found between a prime and a target. Targets presented with semantically related primes in one list were unrelated in the other list, and vice-versa. The nonword-word pairs were the same in both lists.

Half of the subjects were randomly assigned to each list. Each subject was presented with 16 semantically related, 16 semantically unrelated, and 16 nonword-word pairs; each critical target was semanticaly related to its prime for 24 subjects and unrelated for the other 24.

Procedure. An experimental session consisted of 15 practice trials, followed by one block of 48 test trials. Each trial contained three events: a warning signal, and two consecutive test stimuli, the prime and the target. Subjects were instructed to make a lexical decision for the prime (by pressing a "Yes" or a "No" button as in Experiment 1), and to read the subsequent target aloud as soon as they could. The ISI between the warning stimulus and the prime was 1000 ms. The exposure of the prime was terminated by the subject's manual response. The target's onset was 500 ms from the prime offset, and it was removed from the screen by the subject's vocal response. The inter-trial interval was 3 seconds.

All stimuli were presented at the center of a CRT. The physical characteristics of the stimuli and the apparatus were identical to those in Experiment 1.

Results

Table 2 presents the average reaction times for semantically primed and unprimed conditions in each language. In Hebrew, the target words were named 21 ms faster when they were semantically related to the prime. In English, the priming effect was reduced to 16 ms and in Serbo-Croatian it was nonexistent.

These data were analyzed by a Language (Hebrew, English, Serbo-Croatian) X Semantic Relationship (Related, Unrelated) mixed model ANOVA, with repeated mesures. Both main effects were significant for both the stimulus and the subject analysis. However, the most important result was the interaction between the two factors. This interaction was significant for the subject analysis, $F(2,141)=4.54$, MSe=645, $p<.013$, but not for the stimulus analysis, $F(2,93)=2.40$, MSe=731, $p<.097$, probably because the priming effects in Hebrew and English were not significantly different. Nevertheless, planned t-tests revealed that the priming effect was significant in Hebrew, $t(47)=3.94$, $p<.0001$, and $t(31)=4.1$, $p<.0001$ for the subjects and stimulus analysis, respectively), and in English, $t(47)=6.86$, $p<.0001$, and $t(31)=2.08$, $p<.046$ for the subjects and stimulus analysis, respectively. In contrast, for Serbo-Croatian, the differences in naming time in the related and unrelated conditions were insignificant.

Table 2

Naming Time in Ms (and SEMs) for Semantically Primed and Unprimed Words in Hebrew, English, and Serbo-Croatian.

	HEBREW	ENGLISH	SERBO-CROATIAN
Unprimed	619 (15.3)	499 (11.4)	565 (15.7)
Primed	598 (17.3)	483 (10.6)	565 (17.9)
Facilitation	21	16	0

Some additional insight was provided by comparing words that were preceded by nonwords with words that were preceded by unrelated words. In Hebrew, words that were preceded by nonwords were named significantly more slowly than words that were preceded by unrelated words (650 ms and 619 ms, respectively). A small difference, but in the same direction, was found for English (509 ms and 499 ms, respectively) but not for Serbo-Croatian (564 ms and 565 ms). A Language X Stimulus Group mixed model ANOVA and Tukey post-hoc analysis revealed that the interaction was significant, $F(2,141)=14.79$, $MSe=526$, $p<.0001$, and that the Stimulus Group effect was significant only in Hebrew.

Discussion

The results of Experiment 2 revealed that semantic priming had the strongest priming effect in Hebrew, a slightly smaller effect in English, and no effect in Serbo-Croatian. This pattern replicates Katz and Feldman's (1983) suggestion that naming in Serbo-Croatian is not strongly influenced by lexical processes. Most importantly, the results of this study corroborate the hypothesis that the Serbo-Croatian, English, and Hebrew orthographies are on different points of a dimension that influences the amount of lexical involvement in naming. The reasonable conclusion is that orthographic depth is this dimension.

The discrepancy between our findings in Serbo-Croatian and the findings reported by Seidenberg and Vidanović (1985) is puzzling. One possible explanation is that the discrepancy was caused by differences in the stimuli or subjects employed. Seidenberg's subjects were native Serbo-Croatian speakers, but they were residents in Montreal. It was reported that they spoke little French or English and therefore it was presumed that the environment in which they were tested had little if any influence on their performance. Even if that was the case, there was a second difference to consider. Seidenberg and Vidanović's subjects were less educated than our subjects, who were university students. This difference in the level of education might have decreased the subjective word frequency of the stimuli for these subjects. Since the size of the semantic priming facilitation is larger for low- than for high-frequency words (Becker, 1979), the difference in the subjective frequency of the stimuli in the present study and the former results might explain the difference of the results. Nevertheless, the results of Seidenberg and Vidanović could suggest that in Serbo-Croatian, as in other languages, lexical involvement in naming can be manipulated.

In Hebrew, words that followed nonwords were named significantly slower than words that followed words. As discussed in Experiment 1, the only way one can pronounce a nonword is by some process that includes grapheme-to-phoneme translation. Thus, naming words after processing nonwords might have been slowed down by a change from a naming strategy that involves grapheme-to-phoneme translation (for nonwords), to one that is primarily lexically mediated (for words). Note that in Serbo-Croatian, where we have assumed that the same strategy (grapheme-to-phoneme translation) is in effect for naming both words and nonwords, the lexicality of a stimulus has no effect on naming the subsequent stimulus. Therefore, results showing a difference between the naming time for words preceded by nonwords and words preceded by unrelated words might also be determined by orthographic depth. Experiment 3 examined this assumption.

Experiment 3

In Experiment 2 we observed that, in Hebrew, words that followed nonwords were named slower than words that followed words, whereas in Serbo-Croatian this factor had no effect whatsoever on naming time. It is possible that this effect is characteristic only of deep orthographies because only in deep orthographies does switching from naming nonwords to naming words involve a strategic change of the naming mechanism. In shallow orthographies we assume that the same mechanism (grapheme-to-phoneme translation) is employed for naming both words and nonwords, and therefore, naming them in alternation is without cost. In Experiment 3 we attempted to examine this hypothesis by using a technique that discouraged subjects from giving priority to a lexical strategy for naming strings of consonants.

Previous studies in English revealed that word recognition strategies (at least in lexical decision tasks) can be influenced by task demand characteristics and/or by the nature of the stimuli employed. For example, it has been reported that lexical decision for high-frequency words is faster if the list includes only high-frequency words than if high- and low-frequency words are mixed in the list (Glanzer & Ehrenreich, 1979). More relevant to our study are findings suggesting that the use of phonemic encoding of printed words is discouraged if the stimulus list includes a large proportion of homophones (Hawkins, Reicher, Rogers, & Peterson, 1976), whereas the use of visual codes in word recognition is discouraged by backward masking the stimuli (Spoehr, 1978).

We attempted to influence naming strategies by manipulating the proportion of nonwords in the stimulus list. Previous studies reported that frequency effects in naming were more conspicuous when the list contained only words than when words and nonwords were intermixed (Frederiksen & Kroll, 1976; Hudson & Bergman, 1985). One possible interpretation of these results is that naming is less likely to involve lexical mediation when there are many extralexical tokens in the stimulus list. A high proportion of nonwords in the list may have discouraged the subject from using lexical mediation because this route was inefficient most of the time. Consequently, if the grapheme-to-phoneme translation is the natural naming strategy in a shallow orthography and the lexical route is usually employed in a deep orthography, a high proportion of nonwords should impair word naming performance (i.e., increase percentage of errors and RTs in the latter but not in the former orthography).

Methods

Subjects. The subjects were 48 undergraduates from the Hebrew University, 48 from the University of Connecticut, and 48 from the University of Belgrade. None of the subjects employed in this experiment had participated in Experiments 1 or 2, but they were part of the same population of students.

Stimuli and Design. Two lists of 160 stimuli each were assembled in Hebrew, English, and Serbo-Croatian. List 80%-NW consisted of 128 nonwords and 32 words (80% nonwords), and list 20%-NW consisted of 128 words and 32 nonwords (20% nonwords). The target stimuli were 20 words (identical between the lists) that were the last words in the list, dispersed without disrupting the nonword/word ratio in either of the two lists. Both high- and low-frequency words were included among the targets; the mean frequency rating was 2.97, 2.95, and 2.94 in Hebrew, English, and Serbo-Croatian, respectively.

Subjects in each language group were randomly assigned to the two lists, half to each list. The apparatus and experimental conditons were identical to Experiment 2.

Procedure. The procedure was similar to that used in the previous experiments. Subjects were instructed to name words and nonwords that were presented on the screen as quickly and as accurately as possible. In Hebrew, the stimuli were presented without the vowel marks, and the subjects were told to assign the nonwords any vowel combination they prefered. All 160 stimuli were presented in one uninterrupted block. During performance, the experimenter recorded errors verbatim for subsequent qualitative analysis.

Results

The number of mispronounced targets in each list were compared in each language separately. In Hebrew, the number of errors in list 80%-NW (the list with the high proportion of nonwords), was significantly higher than in list 20%-NW, $t(46)=5.44$, $p<.0001$. The same tendency was found in English; however, the difference between the two conditions was marginally significant, $t(46)=1.90$, $p<.063$. In Serbo-Croatian, there were no errors in list 80%-NW and only a nonsignificant number of errors in list 20%-NW (see Table 3).

Table 3

Naming Time in Ms and Percent of Errors in the 80%-NW and the 20%-NW Lists, in Hebrew, English, and Serbo-Croatian.

LIST	HEBREW	ENGLISH	SERBO-CROATIAN
80%-NW			
RT (SEM)	557 (15.3)	565 (8.0)	578 (15.3)
ERRORS	12.6%	5.0%	0%
20%-NW			
RT (SEM)	627 (14.3)	501 (8.5)	558 (11.0)
ERRORS	4.1%	2.7%	0.6%

The average naming time to the 20 target words in each list was analyzed by analyses of variance on subjects and stimuli. The interaction between the Language and List factors was significant for both the subject and the stimulus analyses (subject analysis, $F(2,138)=14.087$, $MSe=4019$; stimulus analysis, $F(2,57)=130.10$, $MSe=351$; $\min F'(2,58)=12.66$, $p<0.001$). In Serbo-Croatian, the target words were named slightly faster in List 20%-NW than in List 80%-NW. However, Tukey-A post-hoc comparison revealed that the difference between the two lists was not significant $p>.05$). In English and Hebrew, Tukey-A comparisons revealed that the differences in RTs between the two lists were significant $p<.01$), but in different directions. In English, the target words were named 61 ms faster in List 20%-NW than in List 80%-NW. In contrast, words in Hebrew, were named 72 ms faster in List 80%-NW than in List 20%-NW.

Discussion

Changing the proportion of nonwords in the stimulus list influenced naming performance differently in each language. Consider first the error data. In Hebrew, the 80%-NW list yielded 8.5% more errors on the target words than the 20%-NW list. In English, the difference was in the same direction but smaller (2.3%). In Serbo-Croatian, there were practically no errors. This systematic pattern is congruent with the hypothesis of orthographic depth. If we assume that naming words in Hebrew is normally mediated by the internal lexicon, subjects must change this strategy for naming nonwords (which have no lexical representation). When the stimulus list contains a high proportion of nonwords, the nonlexical naming strategy is the more efficient one. One consequence of this strategy is that many strings that would have made a word in Hebrew if given the appropriate vowelization, were in fact assigned with incorrect vowels using the nonlexical naming strategy. Thus, the subject pronounced these as a nonword instead of a word. Indeed, in Hebrew, all erroneously read target words were pronounced as nonwords. In contrast to Hebrew, we hypothesized that in Serbo-Croatian naming does not strongly involve the lexicon. Therefore, there is no obvious reason why different strategies should be used for naming words and nonwords. All stimuli can be named via the same mechanism that is based on grapheme-to-phoneme translation. Consequently, the proportion of nonwords in the list should have little influence on word naming performance, as indeed was revealed in this experiment. English, more than the other two languages, combines both lexical and nonlexical routes for naming (cf. Coltheart, 1980). It is conceivable that the nonlexical strategy is used for nonwords and for a subset of words (e.g., low frequency, phonemically regular words), while the lexical strategy is used for naming other words. Consequently, the reinforcement of a nonlexical strategy by the high proportion of nonwords in the list influenced reading of only a part of the words, which explains why the effect was smaller than in Hebrew.

Observations of the nature of the errors in Hebrew and in English provided further support of the hypothesis. If a nonlexical strategy is applied to name words that are usually named via the lexicon, the errors should primarily consist of naming words as nonwords. On the other hand, lexical substitution should be more prevalent when the lexicon is involved in naming. Indeed, in English, the nature of the errors found in the 80%-NW list was as in Hebrew, primarily reading words as nonwords. In contrast, in the 20%-NW list, the errors were mainly substitutions of words by other words (for example "degree" instead of "decree").

The influence of the proportion of nonwords in the list on naming time was less systematic. Although the results in Serbo-Croatian and English appear straightforward, those for Hebrew are not. In Serbo-Croatian, naming time in the two nonword conditions was similar, thereby supporting the hypothesis that word naming strategy was not influenced by the proportion of nonwords in the list. In English, words were named significantly slower in the 80%-nonwords condition than in the 20%-nonwords condition. Because the increase in naming time was followed by an increase in the percent of errors, it is clear that naming performance was interfered with when the stimulus list contained a high percent of nonwords. Although other interpretations might be possible, the explanation provided by the orthographic depth hypothesis is simple and straightforward. In contrast to Serbo-Croatian, grapheme-to-phoneme translation in English is not only constrained by phonemic rules, but also by morphophonemic factors (Chomsky & Halle, 1968). Therefore, in the absence of lexical information, the grapheme-to-phoneme translation is sometimes a complex and painful process. No wonder it takes more time to complete.

The interpretation of the naming data in Hebrew is less straightforward. According to the simple rationale elaborated above, we should have observed a delay in naming words in the 80%-NW condition, which should have been even larger than in English. In contrast, naming the word targets in this condition was significantly faster than in the 20%-NW condition. There is, however, one interpretation that, although admittedly post-hoc, can account for these results. This explanation is based on the insight that in contrast to English, Hebrew phonology has very few constraints on naming nonwords; the subject is free to choose almost any arbitrary set of vowels in order to make the consonantal structure pronounceable. Therefore, when the list contains both words and nonwords, the limiting factor operating on naming time is determined by some (as yet unspecified) competition between lexical search and grapheme-to-phoneme translation for consonants coupled with arbitrary addition of vowels. Before applying a nonlexical strategy, the subject must make sure, with some degree of certainty, that the presented consonant string is not a word. Therefore, naming is slow not only for words but also for nonwords. However, when subjects do not expect words, as in the 80%-NW list, they may be more inclined to change their strategy, and use an idiosyncratic arbitrary selection of vowels for both words and nonwords. This change should result in different speed-accuracy trade-off strategies in the 20%-NW and the 80%-NW lists. In the latter list, subjects might have been inclined to spend less time analyzing the stimulus in reference to lexical information. In many cases the fast analysis was sufficient to generate a correct response, but in some cases incorrect (nonword) responses were made. In the 20%-NW list, since subjects expected words to appear, they referred to the lexicon to find the correct pronounciation of the stimulus. This procedure increased naming time but decreased the probability that a word would be read as a nonword.

This interpretation is certainly not the only one possible but it is supported by several observations. Recall that the average naming time was calculated based on all 20 targets, that is, both errors and correct responses were included. Comparison of the naming time in the 80%-NW list revealed that the same words were read slower by the subjects who named them correctly (559 ms), than by the subjects who read them as nonwords (531 ms). Therefore, some portion of the difference between the naming time in the two lists can be explained by the errors that may have been read without lexical mediation. However, even considering only correctly read words, the difference in naming time between the two lists remains considerably large (561 ms in 80%-NW vs. 625 ms in the 20%-NW list). Therefore, naming difference between the two lists should be accounted for by a factor that affects naming of all stimuli in the list, words and nonwords. As previously mentioned, a change in the speed-accuracy trade-off strategy is a reasonable explanation. There is additional evidence in support of our interpretation. If naming strategy was changed in the way we suggest, naming nonwords should have also been accelerated when words are not expected. We verified this hypothesis by comparing naming time to 32 nonwords presented in List 20%-NW with naming time to the same nonwords presented in List 80%-NW. In agreement with our prediction, naming the same nonwords was faster if words were not highly expected than if they were (635 ms vs. 505 ms, respectively).

In conclusion, we suggest that the error data are in complete agreement with the orthographic depth hypothsis, and the RT data can be reasonably explained without a necessity to change it. Therefore, we consider the results of Experiment 3 as additional support for the validity of the orthographic depth concept.

General Discussion

This study was designed to test the psychological validity of the orthographic depth hypothesis. According to this hypothesis, in a shallow orthography lexical word recognition is mediated primarily by phonemic information generated outside the lexicon by grapheme-to-phoneme translation. In contrast, in a deep orthography, lexical access for word recognition relies strongly on orthographic cues, while phonology is derived from the internal lexicon. One implication of this hypothesis is that in a shallow orthography, the normal strategy for naming is to generate the major phonological information needed for word pronunciation prelexically by means of grapheme-to-phoneme translation. In contrast, in a deep orthography, such prelexical information for naming is either absent or too complex to be used efficiently. Therefore, pronunciation is based on information stored in the lexicon.

We tested the hypothesis by investigating naming performance in Hebrew, English, and Serbo-Croation. These languages are located, respectively, at deep, average, and shallow points on the orthographic depth continuum. Our rationale was to examine the effects on naming performance of factors that were assumed to influence lexical processing; comparisons were made among the three languages, that is, as a function of orthographic depth. Thus, we were primarily interested in the interactions of effects of the lexical manipulations with the language factor, rather than in main effects that could reflect a multitude of factors.

In Experiment 1, we showed that the lexical status of the stimulus (i.e., being a high-frequency word, a low-frequency word, or a nonword) affected the speed of naming in Hebrew more than in English, and in English more than in Serbo-Croatian. Furthermore, only in Hebrew were the effects on naming very similar to the effects on lexical decision. In Experiment 2, the results suggested that semantic priming (a factor that presumably operates on the lexicon) facilitates naming in Hebrew, has a smaller effect in English, while in Serbo-Croatian it has no effect at all. Finally, in Experiment 3, the results indicated that presenting a large proportion of nonlexical items (nonwords) in a stimulus list encouraged the use of a nonlexical strategy that, in Hebrew, speeded naming at the expense of treating many words as nonwords. This manipulation had a similar, but smaller, effect on the reader of English, while in Serbo-Croatian the proportion of nonwords in the list had no effect on naming. We interpreted this to mean that Hebrew readers normally use an orthographic code to access the lexicon for naming but may abandon it when it becomes intractable (as when they must name many nonwords, which have no lexical representation). Note that in each experiment there were six different permutations possible for ordering the three languages in terms of a given effect, but only one order predicted by the orthographic depth hypothesis. Most importantly, in all the three experiments, different lexical factors affected naming systematically in perfect agreement with the order predicted by the orthographic depth hypothesis. Therefore, we suggest that the concept of orthographic depth is psychologically real and that it influences word recognition. In what follows, we will elaborate some of the implications of this conclusion and incorporate it into current thinking on the process of word recognition.

Many reports suggest that the reader of English uses both orthographic and phonemic cues in word recognition (see the review by McCusker at al., 1981). Even when performance requires the generation of phonological codes for output, as in naming, grapheme-to-phoneme translation is the only available route (Coltheart, 1980; Forster & Chambers, 1973; Fredriksen & Kroll, 1976; Is this strategic flexibility limited to orthographies, like English, located in the middle of the orthographic depth continuum? The results of this study suggest that this is probably not the case. The change in strategies observed in Experiment 3 suggests that the nonlexical route, although not accurate, can predominate even in Hebrew when it appears to be more efficient. Note, however, that there is no evidence in our data to imply that words were named correctly without previous lexical access. The extent of using the lexical route in Serbo-Croatian has not been tested. However, other studies support this possibility (Seidenberg, 1985; Seidenberg & Vidanović, 1985). Considering the converging evidence, it seems plausible that both orthographic and phonological information are available prelexically in all languages and probably interact during the process. Therefore, the relevant question should not be what are the codes used in each specific situation but what is the nature of this interaction and how orthographic depth might influence it.

One attempt to disclose the nature of the interaction between orthographic and phonemic codes in word recognition is the version of McClelland and Rumelhart's (1981) parallel interactive model suggested by Seidenberg and his associates (Seidenberg, Waters, & Barnes, 1984; Seidenberg, 1985). Their version of the model emphasizes the relative time course of the phonological and orthographic code activation suggesting that prelexical generation of the two code types is mandatory, and that since the phonological code depends on prior orthographic analysis, it usually lags behind. Consequently, they suggest that orthographic information accumulates faster than phonemic information and, for many (perhaps most) words, lexical access occurs before a recognizable prelexical phonological code is generated. Recognition on the basis of an orthographic code, however, automatically provides the lexical representation of the phonological code, which can then be used in overt naming. We adopt this model as a working hypothesis, but suggest some extensions to explain how orthographic depth may affect naming strategies. Most authors have so far emphasized the importance of the time course of the orthographic code for word recognition, implicitly assuming that it is coding speed that determines whether word naming is based on nonlexical processes or is mediated by the lexicon. However, this implicit assumption may be incorrect. Instead, one could plausibly assume that the time lag between generation of the graphemic and the phonemic codes may begin with the prior onset of the orthographic analysis but then is increased or decreased by the time course of orthographic code generation and by the time course of the phonological code. The results of the present study suggest that the time course of phonological code generation is affected mainly by the simplicity of the rules governing the spelling-sound correspondance. Thus, it is possible that at the extreme shallow end of the orthographic depth continuum, a sufficient portion of the phonolological code can accumulate before the orthographic analysis can help word recognition.

The time it takes to extract sufficient graphemic information for lexical access is determined primarily by the familiarity of the stimulus (Balota & Chumbley, 1984; Mason, 1975). It is possible, however, that prelexical generation of phonemic codes is performed by a parallel interactive process, analogous to the process of orthographic code generation (see also Seidenberg et al., 1984). In this case, nodes would consist of phonemic rather than graphemic features, but their activation would be governed by the same rules as proposed by McClelland and Rumelhart (1981). This hypothesis implies that stimulus familiarity should affect the time course of the phonemic code generation much in the same way as it affects the generation of the orthographic code, and probably for the same reasons. Thus, familiarity of the stimulus should not change the relationship between the time course of the two code types. However, in addition to familiarity, prelexical generation of phonemic codes is also affected by orthographic depth. In a shallow language, the process might be based on idiosyncratic application of grapheme-to-phoneme translation rules. Such a process might get very fast access to the word's phonology and, in parallel, provide the articulatory mechanism with the necessary phonological information. In contrast, in a deeper orthography, simple grapheme-to-phoneme translation is difficult and may frequently lead to incorrect responses. Therefore, generation of the phonemic code is more complex, is frequently dependent on units larger than the single grapheme, and is, therefore, slower. In extreme cases like unvoweled Hebrew a full phonemic code cannot be generated before information about the whole word has accumulated and some lexical decisions about its meaning are made. Therefore, we suggest that the major factor that determines the origin of the phonetic codes in naming is not the speed of orthographic code generation, but rather the ease of the generation of the phonemic codes.

This study concentrated on naming performance. Nevertheless, we suggest that orthographic depth affects lexical access in a similar way. Indeed, in Experiment 1, we observed that the lexical status of the stimulus had a similar pattern of influence on lexical decision performance in each language. However, we agree with Balota and Chumbley (1985) that the lexical decision task is probably not a very good way to examine this hypothesis. A better approach would be to examine, in each language, how factors that are related to the phonology and the orthography influence word recognition performance in semantic tasks. To this end, we believe that the data of the present study strongly support the validity of the orthographic depth factor in word recognition.

References

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical accesss? The role of word frequency in the neglected decision stage. Journal of Experimental Psychology: Human Perception and Performance, 10, 340-357.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effect in the pronunciation task: Lexical access and/or production? Journal of Memory and Language, 24, 89-106.

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. Journal of Experimental Psychology: Human Perception and Performance, 5, 252-259.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. Journal of Experimental Psychology: Human Perception and Performance, 3, 389-401.

Bentin, S., Bargai, N., & Katz, L. (1984). Orthographic and phonemic coding for lexical access: Evidence from Hebrew. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 353-368.

Bentin, S., & Frost, R. (in press). Processing lexical ambiguity and visual word recognition in a deep orthography. Memory & Cognition.

Chomsky, N., & Halle, M. (1968). The sound pattern of English. New York: Harper & Row.

Coltheart, M. (1980). Reading, phonological recording, and deep dyslexia. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), Deep dyslexia. London: Routledge & Kegan Paul.

Feldman, L. B. (1980). Visual word recognition in Serbo-Croatian is primarily phonological. Unpublished doctoral dissertation, University of Connecticut.

Feldman, L. B., & Turvey, M. T. (1983). Word recognition in Serbo-Croatian is phonologically analytic. Journal of Experimental Psychology: Human Perception and Performance, 9, 288-298.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.), Sentence processing: Psycholinguistic studies presented to Merrill Garret. Hillsdale, NJ: Erlbaum.

Forster, K. I. (1981). Frequency blocking and lexical access: One lexicon or two? Journal of Verbal Learning and Verbal Behavior, 20, 190-203.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. Journal of Verbal Learnimg and Verbal Behavior, 12, 627-635.

Fredriksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. Journal of Experimental Psychology: Human Perception and Performance, 2, 361-379.

Glanzer, M., & Ehrenreich, S. L. (1979). Structure and search of the internal lexicon. Journal of Verbal Learning and Verbal Behavior, 18, 381-398.

Gleitman, L. R., & Rozin, P. (1977). The structure and acquisition of reading: Relation between orthography and the structure of language. In A. S. Reber & D. L. Scarborough (Eds.), Toward a psychology of reading: The Proceedings of the CUNY Conferences, 1-5. Hillsdale, NJ: Erlbaum.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. Journal of Experimental Psychology: Human Perception and Performance, 9, 674-691.

Gordon, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. Journal of Memory and Language, 24, 631-654

Hawkins, H. L., Reicher, G. M., Rogers, M., & Peterson, L. (1976). Flexible coding in word recognition. Journal of Experimental Psychology: Human Perception and Performance, 2, 380-385.

Henik, A., Friedrich, F. J., & Kellogg, W. A. (1983). The dependence of semantic relatedness effects upon prime processing. Memory & Cognition, 11, 366-373.

Hudson, P. T. W., & Bergman M. W. (1985). Lexical knowledge in word recognition: Word length and word frequency in naming and lexical decision tasks. Journal of Memory and Language, 24, 46-58.

Humphreys, G. W., & Evett, L. J. (1985). Are there independent lexical and nonlexical routes in word processing? An evaluation of the dual-route theory of reading. The Behavioral and Brain Sciences, 8, 689-740.

Katz, L., & Feldman, L. B. (1983). Relation between pronounciation and recognition of printed words in deep and shallow orthographies. Journal of Experimental Psychology: Learning, Memory and Cognition, 9, 157-166.

Koriat, A. (1984). Reading without vowels: Lexical access in Hebrew. In H. Bouma, & D. G. Bouwhuis (Eds.), Attention and performance X: Control of language processes. Hillsdale, NJ: Erlbaum.

Lukatela, G., Gligorijević, B., Kostić, A., & Turvey, M. T. (1980). Representation of inflected nouns in the internal lexicon. Memory & Cognition, 8, 415-423.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. (1980). Lexical decision in a phonologically shallow orthography. Memory & Cognition, 8, 124-132.

Lupker, S. J. (1984). Semantic priming without association: A second look. Journal of Verbal Learning and Verbal Behavior, 23, 709-733.

Mason, M. (1975). Reading ability and letter search time: Effects of orthographic structure defined by single letter positional frequency. Journal of Experimental Psychology: General, 104, 146-166.

McClelland, J., & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychological Review, 88, 375-407.

McCusker, L. X., Hillinger, M. L., & Bias, R. G. (1981). Phonological recording and reading. Psychological Bulletin, 89, 375-407

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology, 90, 227-234.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word-recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), Attention and performance V. New York: Academic Press.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. Cognition, 19, 1-30.

Seidenberg, M. S., & Vidanovic, S. (1985). Word recognition in Serbo-Croatian and English: Do they differ? Paper presented at the XXVI Annual Meeting of the Psychonomic Society, Boston.

Seidenberg, M. S., Waters, G. S., & Barnes, M. A. (1984). When does irregular spelling or pronunciation influence word recognition? Journal of Verbal Learning and Verbal Behavior, 23, 383-404.

Smith, M. C., Theodor, L., & Franklin, P. E. (1983). On the relationship between contextual faciliatation and depth of processing. Journal of Experimental Psychology: Learning, Memory, & Cognition, 4, 697-712.

Spoehr, K. T. (1978). Phonological recoding in visual word recognition. Journal of Verbal Learning and Verbal Behavior, 17, 127-141.

Theios, J., & Muise, J. G. (1977). The word identification process in reading. In N. J. Castellan, Jr., D. B. Pisoni, & G. R. Potts (Eds.), Cognitive theory (Vol 2). Hillsdale, NJ: Erlbaum.

West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. Journal of Experimental Psychology: Learning Memory and Cognition, 8, 385-399.

Footnotes

[1]Throughout this paper we will use the term grapheme-to-phoneme translation with the understanding that the process often involves units larger than single letters.

[2]The ambiguity in English pronunciation and the ambiguity in Hebrew pronunciation are different in kind. Nevertheless, it seems reasonable to assume that it is more difficult to get to the correct pronunciation in Hebrew words than in English.

THE INFLECTED NOUN SYSTEM IN SERBO-CROATIAN: LEXICAL REPRESENTATION OF MORPHOLOGICAL STRUCTURE*

Laurie B. Feldman† and Carol A. Fowler††

Abstract. Repetition priming is examined for alternating and nonalternating morphologically-related inflected nouns. In Experiments 1 and 2, latencies to targets in nominative and dative/locative cases, respectively, were invariant over case of prime. In Experiment 3, latencies to nominative-case nouns were the same whether they were primed by forms in which the spelling and pronunciation of the common stem were shared ("nonalternating") or not ("alternating") with nominative form. Results are interpreted as reflecting lexical organization among the members of a noun system. In Experiments 1 and 2, the pattern of latencies to primes suggests a satellite organization in which nominative forms are more strongly linked to oblique forms than are oblique forms to each other. In Experiment 3, atypical cases of alternating forms showed a different pattern of prime latencies suggesting that the organization within a noun system may differ for alternating and nonalternating forms.

The research we describe examines the role of morphology in the reading lexicon of speakers of Serbo-Croatian, the dominant language of Yugoslavia. The morphology of Serbo-Croatian is particularly interesting to study because it is substantially richer than that of English. Generally, in Serbo-Croatian inflectional affixes are appended to nouns and adjectives with the particular termination varying according to case, gender, and number. Analogously, for verbs, the inflectional suffixes and sometimes the infixes may vary with tense, aspect, person, number, and sometimes gender of the subject. The formation of diminutives, agentives, and other derivations--which are characteristic of Slavic languages--is similarly complex. Consequently, each Serbo-Croatian base word has many variants, yielding extensive families of morphologically-related words.

---

*Memory &Cognition, in press.
†Also University of Delaware
††Also Dartmouth College
Acknowledgment. We wish to thank the following students for collecting data: Jasmina Cesić, Sanda Parezanović, Dara Andelković, and Teodora Vujin. Experiment 3 was suggested by Suzanne Boyce and Louis Goldstein. In addition, we thank Vicki Hanson and Jasmina Moskovljević for many helpful comments on the manuscript. This research was supported by funds from the National Academy of Sciences and the Serbian Academy of Sciences to Laurie B. Feldman; by NICHD Grant HD-01994 to Haskins Laboratories, and by NICHD Grant HD-08495 to the University of Belgrade. Portions of this paper were presented to the meeting of the Psychonomic Society in San Diego, CA in 1983 and in San Antonio, TX in 1984.

The present series of experiments explores in particular how the singular case inflected forms of a word are related in the internal lexicon of adult readers who are native speakers of Serbo-Croatian. The experiments represent an extension of earlier work by Lukatela and his colleagues (Lukatela, Gligorijević, Kostić, & Turvey, 1980; Lukatela, Mandić, Gligorijević, Kostić, Savić, & Turvey, 1978) that investigated how individual inflected forms are recognized.

There are seven cases of inflected noun forms in Serbo-Croatian and they differ in their frequency of occurrence in printed text (Kostić, 1965). When singular inflected cases were presented in a lexical decision task, decision times for the nominative singular form of a noun were less than decision times for the same noun in (a) dative/locative and instrumental singular cases (Lukatela, et al., 1978) and (b) genitive and instrumental cases (Lukatela, et al., 1980). The decision times for all non-nominative (oblique) forms were equivalent. Lukatela et al. (1978, 1980) proposed that in the lexicon, the singular cases of a noun comprise a satellite-like system where the nominative singular of the noun or base form has a special status in that it provides a nucleus around which the oblique cases cluster in a uniform fashion. This organization applies for inflected forms of both familiar and less familiar base words. That is, frequency of the nominative base word but not frequency of inflectional case governs reaction time.

The satellite-entries model reflects a position on a debated issue in the literature on how morphological structure may influence word recognition (see Caramazza, Miceli, Silveri, & Laudanna, 1985). In that literature, the lexical entries are considered to consist of stem morphemes or, alternatively, of whole words. In the former case, polymorphemic words are decomposed into stem and affix prior to lexical access (Taft, 1979; Taft & Forster, 1975); in the latter, they are not. Instead, the lexicon may comprise a morphological principle of organization so that morphologically-related words are near-neighbors (Stanners, Neiser, Hernon, & Hall, 1979) but lexical entries are accessed from whole words. In the studies by Lukatela et al. (1978; 1980), the same general pattern of decision latencies was obtained for masculine and feminine nouns in nominative and oblique singular cases despite differences in the number of morphological transformations between nominative and oblique cases. Specifically, in masculine words, the nominative singular is uninflected and therefore serves as the base morpheme for other inflected forms. In feminine words, the nominative singular is inflected. It includes an "A" affix, which is replaced to form other inflected forms. This finding suggested to Lukatela et al. that entries for each case in a noun system are represented completely--that is, they are not "decomposed" into a shared base morpheme plus an affix. It should be pointed out that no direct comparisons of gender were reported in that work, although failure to find evidence of a case-by-gender interaction is critical to support the non-decomposition characterization of satellite entries.

The results of a second study also suggest that morpheme bases do not constitute the units of access to the noun entries in a Serbo-Croatian lexicon. In an experiment designed to evaluate BOSS structure (Taft, 1979) as a unit of lexical access in Serbo-Croatian (Feldman, Kostić, Lukatela, & Turvey, 1983), BOSS units (which included the first unprefixed syllable as well as the longest sequence of consonants that can legally occur in syllable-final position) and base morphemes were fully redundant. This was due, in part, to general constraints on orthographic structure for Serbo-Croatian and in part to the criteria for selecting stimulus materials.

The outcome of the experiment was that where two different phonological interpretations of a letter string were equally possible such that letter strings were "bivalent," latencies in lexical decision were retarded as long as the entire word, that is, the base morpheme, and the inflectional affix were bivalent. When only the base morpheme was bivalent, decision latencies were not changed relative to unequivocal controls. Feldman and her colleagues (1983) argued that most varieties of models that entail decomposition to a base morpheme as the unit for lexical access in Serbo-Croatian would predict that all words that included a bivalent base morpheme should be affected.

These outcomes have served as the basis of arguments against decomposing isolated inflected nouns to a base morpheme in order to access their lexical entry. It should be noted, however, that an interpretation of some of these outcomes as evidence for or against a morphemic representation for access may be inconclusive, in part because a distinction between morphological processes that arise prior or subsequent to lexical access may not be possible in a lexical decision task (Burani, Salmaso, & Caramazza, 1984; Henderson, Wallis, & Knight, 1984; Seidenberg & Tanenhaus, 1986).

The present series of experiments extends the satellite-entries account along two lines of inquiry. 1) We ask whether decision latencies to inflected forms of a noun correlate strongly. If members of a noun system are associated in the lexicon, then, nonlexical factors being equal, decision latencies to inflected forms of a word will tend to be correlated. 2) We ask whether the nominative singular can prime and be primed by its oblique-case satellites as effectively as can an oblique case by other oblique cases or by a nominative. Reductions in decision latency to words in appropriate contexts or facilitation by priming is sometimes explained in terms of activation among entries in the lexicon and is assumed to reflect, at least in part, lexical organization (e.g., Seidenberg & Tanenhaus, in press). Magnitude of facilitation then can provide an index of the cohesion among lexical entries in a noun system. A variation of the lexical decision procedure, repetition priming, permits extensive investigation of the organization among regular and alternating inflected forms in the Serbo-Croatian lexicon.

In the repetition priming procedure (Forbach, Stanners, & Hochhaus, 1974; Scarborough, Cortese, & Scarborough, 1977; Stanners et al., 1979) each word and pseudoword is presented twice (with a lag of intervening items) for a lexical decision judgment and the facilitation to decision latency or priming due to repetition is measured. (The first presentation of the item is the "prime." The second presentation is the "target.") With English materials, it is not necessary that the identical word be repeated as prime and target for facilitation to occur. Generally, morphologically-related words including inflections and derivations also reduce target decision latency--sometimes as fully as an identical repetition (Fowler, Napps, & Feldman, 1985; Stanners et al., 1979). For example, both the inflected form "manages" and the derived form "management" can facilitate a subsequent presentation of "manage." Sometimes, the effect is equivalent to an identical presentation of "manage." (When the facilitation with morphological relatives as primes is statistically equivalent to the facilitation with an identical repetition [following Fowler et al., 1985], the outcome is "full" repetition priming. Priming that is significant, but significantly less than with an identical prime, is "partial.")

Repetition priming does not occur among orthographically-similar but morphologically-unrelated words, e.g., "ribbon" and "rib" (Hanson & Wilkenfeld, 1985; Murrell & Morton, 1974; Napps & Fowler, 1986) but it does occur when morphologically related primes and targets have discrepant pronunciations and/or spellings, e.g., "health" and "heal" (Hanson & Wilkenfeld, 1985; Fowler et al., 1985; Napps & Fowler, 1983). Results such as these support an interpretation of repetition priming effects as primarily lexical in origin (Fowler et al., 1985; Stanners et al., 1979) although there may also be a nonlexical or episodic component (Feustel, Shiffrin, & Salasoo, 1983). Episodic contributions to repetition priming based on an examination of derivational forms in Serbo-Croatian are considered elsewhere (Feldman, 1984; Feldman, in press; Feldman & Moskovljević, 1986). Currently, it appears that facilitation due to presentation of morphological relatives reflects lexical organization, but the difference between numerically full and partial priming may be at least in part episodic (Fowler et al., 1985). The longevity of the effect with morphologically-related words has been offered as evidence that repetition priming may be distinct from semantic or associative priming (Dannenbring & Briand, 1982; Henderson et al., 1984; Napps, 1985). One way to capture this distinction is by proposing that morphological relatives activate the same lexical entry whereas semantically associated words activate different entries.

Recent research has also identified a strategic contribution to the repetition priming effect (Forster & Davis, 1984; Oliphant, 1983). As anticipated by Fowler et al. (1985), the large proportion of affixed primes followed after a lag by their base forms may have permitted subjects to predict future targets from the prime. However, they found priming at long lags between prime and target (48 items). Moreover Napps (1985) has demonstrated significant facilitation by morphological relatives even when only a very small proportion of morphemes is repeated. In light of these findings, the facilitation evidenced in repetition priming cannot be predominantly strategic in origin. Nevertheless, the experimental design introduced in that study as well as in the present one does not prevent adoption of such a strategy by the subject, especially when base words serve as targets and inflections and derivations serve as primes.

The present series of experiments employs the repetition priming paradigm to investigate the lexical organization of Serbo-Croatian inflected noun systems in adults and proceeds as follows: In Experiment 1, nominative case words served as targets and we asked whether, for real words, repetition priming was full such that primes morphologically related to their targets were as effective as identity primes. As a byproduct, this procedure permitted a replication of the original study on the satellite-entries account; specifically, it allowed an examination of the pattern of decision latencies for nominative and non-nominative forms of many words as it reflects the structure of the noun system. In addition, word gender was treated as a variable to ascertain that it did not interact with other effects as a decomposition account might predict. Finally, the pattern of correlations among pairs of satellite entries was examined. As discussed above, according to the satellites-entries account (Lukatela et al., 1978, 1980), the nominative singular case of both masculine and feminine words enjoys a privileged status in the satellite configuration. Taken in isolation, therefore, the outcome of Experiment 1 is ambiguous. Plausibly, it reflects the coherence of the noun system. Alternatively, it reflects the special role of the nominative. In Experiment 2 the pattern of facilitation for an oblique (viz., dative/locative) case target was investigated. Once again, we examined the pattern of facilitation by various primes to ask about the lexical organization of satellite-entries and specifically about whether the nominative singular case has a special status relative to oblique cases. In Experiment 3, the lexical organization for nouns that undergo sound and spelling changes in at least one of their inflected case forms was investigated. Accordingly, the similarity of form between prime and target was reduced. Generally, decision latencies to primes and the pattern of intercorrelations were interpreted with respect to the structure of the satellite system and the pattern of facilitation in repetition priming was interpreted to reflect the coherence or organization within the noun system. Together, Experiments 1, 2, and 3 provide an elaborated account of the structure and coherence of the noun system of the mature reader of Serbo-Croatian, thereby characterizing the skilled reader's sensitivity to aspects of morphological structure.

Experiment 1

The first experiment examined priming of nominative case nouns by identical and morphologically-related words. It addressed three questions: 1) Does the presentation of an inflected form of a noun facilitate lexical decision to a subsequently presented nominative form of the same noun? Positive evidence suggests that the skilled reader of Serbo-Croatian is sensitive to morphological relatedness among words in that accessing one form necessarily accesses its morphological relatives. 2) Do decision latencies for prime presentations of masculine and feminine words pattern in different ways? If not, then in replication of Lukatela and his colleagues (1978; 1980), inflected nouns do not appear to be accessed from a base morpheme and then transformed or checked (in a fashion that affects reaction time) for the appropriateness of its affix. 3) Do decision latencies for inflected forms of a noun correlate? A positive correlation in conjunction with significant facilitation due to repetition suggests that all inflected forms of a noun access the same lexical entry.

Method

Subjects. Forty-two students from the Department of Psychology at the University of Belgrade participated in the experiment. All were native speakers of Serbo-Croatian and all had vision that was normal or corrected-to-normal. They participated in the study in partial fulfillment of course requirements.

Stimulus materials. Twenty-four Serbo-Croatian words and twenty-four pseudowords were included in the experiment. Words contained four or five letters in their nominative form and all were judged by four independent raters to be very familiar. Half were feminine and half were masculine in gender; words in the two genders were matched on length. No words were included that contained sequences of more than two consonants. Pseudowords were generated by changing one or two letters (vowel with vowel or consonant with consonant) in other real words with the same orthographic structure as the real words in the experiment. All materials were printed in Roman characters.

Each word appeared in three different singular cases: nominative, dative/locative, and instrumental singular. Each pseudoword also appeared with affixes for masculine or feminine words in the same inflectional cases. Words were chosen so that inflectional suffixation did not alter the spelling of the base form. Examples of regular masculine and feminine words in their seven inflected-case forms appear in Table 1.

Table 1

Examples of Regular Masculine and Feminine Singular Inflected Nouns and their Frequencies

CASE	MASCULINE	(FREQ.)	FEMININE	(FREQ.)
Nominative (N)	DINAR	13	RUPA	9
Genitive (G)	DINARA	9	RUPE	8
Dative (D)	DINARU	1	RUPI	<1
Accusative (A)	DINAR	6	RUPU	6
Instrumental (I)	DINAROM	2	RUPOM	2
Locative (L)	DINARU	4	RUPI	2
Vocative (V)	DINARE	<1	RUPO	<1

Procedure. Subjects performed a lexical decision task. As each letter string appeared, they hit a telegraph key with both hands to indicate whether or not it was a word. They hit the farther key (with index fingers) to signal "yes" and the closer key (with thumbs) to signal "no." All letter strings were typed in Roman script, then photographed and mounted as slides. Stimuli were projected from a carousel projector equipped with a modified camera lens as a shutter and displayed on a screen until after subjects responded (approximately 750 ms). Subjects viewed the screen from a distance of 1 m and letterstrings subtended a visual angle between 2.6° and 3.9°. A dark field immediately preceded and followed the display. The interval between experimental trials was controlled by the experimenter and lasted about 2000 ms. Reaction times were measured from the onset of the stimulus display and subjects were tested individually.

Design. Three test orders were created. Each one included three priming conditions distinguished by the inflectional case of the prime, that is, nominative singular, dative/locative singular, or instrumental singular. (Case of prime was indicated as N1, D1, or I1, respectively.) All targets were in nominative case. Half were masculine gender and half were feminine. (The conditions of nominative targets preceded by nominative, dative/locative, and instrumental singular primes were indicated as NN, DN, and IN, respectively.) Words appeared in the same serial position across all test orders although the inflectional form of the prime varied. For example, the word RUPA (meaning "hole") was presented in its nominative form as the target in the same serial position in all three test orders but it was preceded in the same position by either RUPA, RUPI, or RUPOM as a prime.

Each subject viewed one test order. Therefore, subjects saw each morpheme twice, once in a prime and once in a target. The average lag between the presentation of the prime and the target was ten items and lags ranged from seven to thirteen. Filler items were introduced to maintain appropriate lags and a practice list of ten items preceded the test list.

To summarize the experimental design, across test orders each target word in nominative case was preceded by its prime in nominative, dative/locative, and instrumental form. Within each order, a base morpheme occurred once in a target and once in a prime, and case of prime varied with item. Stated alternatively, all subjects viewed the three cases of prime on different target items and across test orders, and each word was preceded by each case of prime.

Results

Errors and extreme reaction times (greater than 1200 ms or less than 350 ms) were excluded from all analyses. This procedure eliminated fewer than 4% of all responses. In addition, when a subject responded incorrectly to one member of a prime-target pair, both responses were excluded from subsequent analyses. The error pairing procedure eliminated an additional 3% of all responses.

Mean reaction time for correct responses to nominative forms (Conditions N1, NN, DN, IN) of masculine and feminine words were calculated and subjected to analyses of variance. Each comparison included an analysis for subjects averaging over items (F_1) and for items, averaging over subjects (F_2; items' analysis, reported in parentheses). Means for Experiment 1 are summarized in Table 2.

For words, the effect of condition (N1, NN, DN, IN) was significant $F_1(3,114) = 26.53$, $MSe = 1759$, $p < .001$ ($F_2(3,66) = 11.41$, $MSe = 1244$, $p < .001$). The effect of gender was not significant although the interaction of condition by gender approached significance in a subjects' analysis but not in the items' analysis, $F_1(3,144) = 2.38$, $MSe = 1651$, $p < .07$ ($F_2(3,66) = .26$, $MSe = 1058$, $p < .85$). A second analysis including only nominative targets (NN, DN, IN) revealed no significant differences among targets as a function of case of prime, and no interaction involving gender. Therefore, the significant effect of condition in the earlier analyses is due to the difference between the N1 condition on the one hand and the three target conditions on the other; thus priming was full.

For pseudowords, neither the effect of condition nor gender was significant although their interaction was significant by a subject's analysis only, $F_1(3,114) = 3.98$, $MSe = 2032$, $p < .01$ ($F_2(3,66) = 1.33$, $MSe = 1864$, $p < .27$). Inspection of pseudoword means indicated that familiarity with pseudoword targets slowed rejection latencies in the case of pseudo-masculine noun forms and speeded rejection latencies in the case of pseudo-feminine noun forms. Because the effect of condition on pseudowords was not significant, no analysis combining words and pseudowords is included.

An analysis of variance on mean reaction times for correct responses to word primes (Conditions N1, D1, I1) revealed a significant effect of case, $F_1(2,76) = 40.22$, MSe = 4269, $p < .001$ ($F_2(2,44) = 25.95$, MSe = 2036, $p < .001$). There was no effect of gender and importantly for the satellite interpretation, there was no interaction of case by gender $F_1(2,76) = 1.89$, MSe = 2750, $p < .16$ ($F_2(2,44) = .78$, MSe = 2036, $p < .47$). Inspection of word means showed that for both masculine and feminine words, the nominative case was recognized faster than the oblique cases and oblique cases did not differ between themselves.

Table 2

Mean Reaction Times (ms) to Nominative Targets (NN, DN, IN) and their Respective Nominative (N1), Dative/Locative (D1), and Instrumental (I1) Case Primes in Experiment 1

WORDS

MASCULINE				FEMININE				COMBINED			
PRIME		TARGET		PRIME		TARGET		PRIME		TARGET	
N1	600	NN	533	N1	576	NN	536	N1	588	NN	534
D1	665	DN	539	D1	672	DN	544	D1	668	DN	541
I1	680	IN	543	I1	661	IN	548	I1	670	IN	545

PSEUDOWORDS

MASCULINE				FEMININE				COMBINED			
PRIME		TARGET		PRIME		TARGET		PRIME		TARGET	
N1	682	NN	704	N1	729	NN	671	N1	705	NN	687
D1	723	DN	695	D1	721	DN	683	D1	722	DN	688
I1	768	IN	700	I1	773	IN	708	I1	770	IN	703

An analogous analysis on pseudoword primes showed a significant effect of case, $F_1(2,76) = 20.76$, MSe = 1300, $p < .001$ ($F_2(2,44) = 3.92$, MSe = 7006, $p < .03$), and an interaction of case by gender that was significant by the subjects analysis only, $F_1(2,76) = 4.90$, MSe = 2800, $p < .01$ ($F_2(2,44) = .60$, MSe = 7006 $p < .56$). The pattern of pseudoword means revealed longer rejection latencies for instrumental forms than for nominative forms. For pseudo-feminines, dative/locative latencies were similar to nominatives. For pseudo-masculines, however, dative/locative latencies were intermediate between nominative and instrumental latencies and significantly different from each. All contrasts were significant at $p < .01$.

No analyses were performed on the error data because some subjects made no errors and all subjects were very accurate. Out of 8 possible errors per condition, the mean number of errors in conditions N1, D1, I1 for words and pseudowords respectively were .47, .49, .18 and .52, .69, .64. The mean number of errors on targets in each condition (NN, DN, IN) computed independently of the error pairing procedure was less than .20 for both words and pseudowords.

Finally, mean reaction times for each prime word in its Nominative (N1), Dative/Locative (D1), and Instrumental (I1) form were computed and inflected forms of each word were correlated. To the extent that the various members of a noun system share a lexical entry or are equivalent on factors that contribute to reaction time in a lexical decision task (Balota & Chumbley, 1984), correlations between latencies for any pair of inflected forms will be significant and all pair-wise correlations will be equal. The correlations of nominative with dative (N1, D1), nominative with instrumental (N1, I1) and dative with instrumental (D1, I1) were $r = .57$, $r = .49$, and $r = .67$, respectively. (For correlations based on 24 items where df = 22, values of r greater than |.40| are significant at the .05 level.) Analogous correlations computed on pseudo-nominative, pseudo-dative, and pseudo-instrumental latencies did not approach significance.

Discussion

Significant priming of nominative targets occurred when real words were presented for lexical decision in a repetition priming procedure. The effect was obtained with both identity primes (NN) and inflected relatives (i.e., morphological primes DN, IN). The means of the three target conditions did not differ significantly and their numerical values differed overall only by 10 ms. This outcome, namely, statistically full priming with small numerical differences between means replicated results reported previously with English materials (Fowler et al., 1985). One account provided by Fowler et al. is that the small numerical differences in priming may reflect an episodic component that augments the lexical effects of repetition priming by selectively inflating the identity prime condition. However, as argued elsewhere, this effect cannot be visual in nature (Feldman, 1984; Feldman & Moskovljević, in press) because the magnitude of facilitation is as large when prime and target are printed in different alphabets as when they are printed in the same alphabet. Evidently, in the present experiment, presentation of related inflected-case forms of a word facilitated subsequent lexical decision about that word in nominative case and both identical and morphological forms primed fully. This outcome can be captured in terms of a full spreading of activation among individual inflected forms of a noun system (i.e., satellite entries) and its nominative nucleus.

The suggestion of an interaction of condition by gender for words indicated that the magnitude of the facilitation due to repetition was larger for masculine nouns than feminine nouns. However, inspection of means revealed that the effect was carried by a difference between masculine and feminine nominative primes (N1) rather than by targets (NN, DN, IN) and the outcome of an analysis restricted to target latencies supported this interpretation. In summary, decision latencies to masculine and feminine target words were equally fast when an identical or morphologically-related prime preceded it.

Among pseudowords, evidence of a condition-by-gender interaction made the absence of any overall facilitation with repetition equivocal. Inspection of means suggested that masculine-gender targets were slowed by a previous presentation of the identical prime whereas feminine-gender targets were facilitated (collapsing over gender, therefore, gave no evidence of facilitation with repetition). This effect is curious because neither gender nor the interaction of condition by gender was significant for real word targets and because the only difference between masculine and feminine nominative case pseudowords was the addition of an "A" suffix on feminine forms. In all other respects, the assignment of gender and consequently, inflectional affixes to the two groups of pseudowords was essentially arbitrary. At this point, we can suggest no explanation as to why repetition sometimes facilitated and sometimes impeded decision latencies for pseudowords.

The primary outcome of the present experiment, based on the pattern of facilitation using the repetition priming procedure, was that both nominative and oblique case forms can prime a nominative target. Identity primes and morphologically-related primes both exhibited statistically-full priming with nominative targets. Following Stanners et al. (1979) and Fowler et al. (1985), we interpret repetition priming as an index of the interrelation among forms of a noun in the internal lexicon. By this convention, all oblique-case forms were tightly linked to their nominative nucleus. The facilitation evidenced in the repetition priming procedure with inflected nouns of Serbo-Croatian can be conceptualized to mean that once a satellite entry is accessed, the nominative nucleus of the noun system is also activated.

The latency data for word primes provided a replication of previous results on inflected forms in Serbo-Croatian (Lukatela et al., 1978; 1980). Nominatives were recognized faster than other cases and the oblique cases did not distinguish among themselves. This outcome suggested that nominative forms are most accessible in the internal lexicon. Importantly, there was no interaction with gender. Masculine and feminine words displayed the same pattern of latencies among inflected forms despite differences in the complexity of deriving inflected forms from a nominative form. Equally strong correlations between mean latencies of two oblique cases of a word (D1, I1) or of a nominative and one oblique case (N1, D1; N1, I1) support this interpretation.

In conclusion, the outcome of the present experiment buttresses the interpretation of Lukatela et al. (1980), in that it provided no evidence that the morphological relatedness among inflected forms of a noun was represented in the lexicon by a shared base morpheme and a set of transformations whose complexity governs recognition latency. It appears that masculine nouns, where the nominative singular and base morpheme are isomorphic, and feminine nouns, where the nominative singular includes an "A" affixed to a base morpheme, are represented lexically in the same manner.

In the pseudoword prime data, decision latencies varied with number of letters. For pseudo-feminine items, nominative and dative/locative forms had the same number of letters and they had similar reaction times. Both differed from instrumental forms, which are one letter longer. For pseudo-masculine items, by contrast, nominative forms, which have the fewest letters, were recognized significantly faster than dative/locatives, which are one letter longer than nominatives. Both of these were faster than instrumentals, which are two letters longer than nominatives. Length effects for orthographically regular but meaningless letter strings in lexical decision have been reported previously in English and in other languages (e.g., Feldman & Turvey, 1983; Hudson & Bergman, 1985).

As reviewed above, the satellite entries account posits a separate and complete entry for each affixed word and grants a special status to the nominative case. Because nominative case forms served as targets in Experiment 1, the outcome of the experiment (viz., full priming with nominative targets) is inconclusive with respect to lexical organization within the noun system. The present outcome may reflect the alleged privileged position of the nominative case in the satellite configuration. Alternatively, the same result could also arise if the nominative singular case of a noun did not possess a special status within the noun system, that is, if the principle of organization were uniform among all inflected forms. According to the homogeneous interpretation, however, the same pattern of full priming effects would emerge with any oblique-case target. In Experiment 2, we continue to explore the characteristics of the noun system. We use the pattern of facilitation in repetition priming to look for inhomogeneities in organization among entries. As above, it was our intention to ascertain how the principle of morphological relatedness operates within the noun system, specifically, whether as predicted by the satellite-entries account, there exist some inflected-case forms that retain a privileged status when the oblique form of a noun must be activated.

Experiment 2

In Experiment 2, we asked whether primes that are morphologically related to their targets facilitate recognition of oblique case targets as effectively as they facilitate recognition of nominative targets. Priming of dative/locative case targets by nominative, dative/locative, and instrumental cases was examined. As in the first experiment, an identity prime condition served as the criterion for determining full repetition priming. If inflected forms are defined only relative to the nominative singular, as posited in the satellite-entries account, then the instrumental singular case of a noun may facilitate lexical decision on the dative/locative singular case of a noun less than the dative/locative case itself. That is, the priming of dative/locative target by instrumental case forms may be partial. Alternatively, if the organization among cases of a noun is homogeneous, then priming for oblique targets should be comparable to priming with nominative targets.

Method

Subjects. Thirty-nine first-year students from the Department of Psychology at the University of Belgrade participated in Experiment 2. None had participated in Experiment 1. All were native speakers of Serbo-Croatian, had normal or corrected-to-normal vision, and had never participated previously in a psycholinguistic experiment.

Stimulus materials. The same words and pseudowords presented in Experiment 1 were used in Experiment 2. Moreover, the original order of presentation was preserved with one exception. In the test list for Experiment 2, the dative/locative form rather than the nominative form appeared as the target.

In Experiment 2, as in the first experiment, all letter strings were printed in Roman characters.

Procedure. The procedure in Experiment 2 was identical to that of the previous experiment.

Results

Errors and extreme response times were eliminated from the present analyses according to the same criteria as in Experiment 1. Fewer than 4% of all responses were eliminated according to these criteria. An additional 2% of all responses were eliminated by the error pairing procedure. Table 3 summarizes the mean recognition times for dative/locative target words and pseudowords in Experiment 2.

Table 3

Mean Reaction Times (ms) to Dative/Locative Targets (ND, DD, ID) and their Nominative, Dative/Locative, and Instrumental Case Primes (N1, D1, I1) in Experiment 2

WORDS

MASCULINE				FEMININE				COMBINED			
PRIME		TARGET		PRIME		TARGET		PRIME		TARGET	
N1	614	ND	576	N1	593	ND	551	N1	603	ND	563
D1	636	DD	563	D1	649	DD	542	D1	642	DD	552
I1	675	ID	580	I1	655	ID	566	I1	665	ID	573

PSEUDOWORDS

MASCULINE				FEMININE				COMBINED			
PRIME		TARGET		PRIME		TARGET		PRIME		TARGET	
N1	712	ND	691	N1	715	ND	686	N1	714	ND	688
D1	722	DD	688	D1	710	DD	679	D1	716	DD	684
I1	782	ID	710	I1	739	ID	699	I1	761	IN	705

Analyses of variance with condition (D1, ND, DD, ID) and gender as independent variables were performed using subjects and items (in parentheses) as random variables. Consistent with the outcome of Experiment 1, the effect of condition was significant for real words, $F_1(3,114) = 59.48$, MSe = 2158, $p < .001$ ($F_2(3,66) = 27.54$, MSe = 1435, $p < .001$). The effect of gender and the interaction of condition by gender were significant in the subjects' analysis but not in the items' analysis, $F_1(1,38) = 6.27$, MSe = 1728, $p < .02$ (F_2

(1,22) = .93, MSe =3589, $p < .35$) and $F_1(3,114) = 2.98$, Mse = 1913, $p < .04$ (F_2 (3,66) = 1.20, MSe = 1435, $p < .32$), respectively.

A subsequent set of analyses including only dative/locative target latencies (conditions ND, DD, ID) revealed a significant effect of prime condition, $F_1(2,76) = 4.02$, MSe = 2028, $p < .02$ ($F_2(2,44) = 3.17$, MSe = 790, $p < .05$) such that identity primes were more effective than instrumental primes. There was also a significant effect of gender by the subjects' analysis, $F_1(1,38) = 20.77$, MSe = 1125, $p < .001$ but not by the items' analysis (F_2 (1,22) = 3.07, MSe = 2340, $p < .09$). The interaction of condition by gender was not significant.

An analogous analysis of pseudoword latencies showed a significant effect of prime condition, $F_1(3,114) = 6.77$, MSe = 2582, $p < .001$ ($F_2(3,66) = 2.75$, MSe = 1952, $p < .05$); there was no effect of gender and no interaction of condition by gender. A subsequent analysis of pseudoword targets indicated a significant effect of condition such that instrumental case primes facilitated less than did dative/locative or nominative case primes, $F_1(2,76) = 3.37$, MSe = 2848, $p < .04$. This effect was not significant in the stimulus analysis, however, F_2 (2,44) = 2.06, MSe = 1430, $p < .14$. When word and pseudoword latencies (D1, ND, DD, ID) were entered into one analysis, the interaction of condition by lexicality was significant, $F_1(3,114) = 15.72$, MSe = 1938, $p < .001$ ($F_2(3,132) = 4.81$, MSe = 1950, $p < .003$). Words were facilitated more by repetition than were pseudowords.

An analysis of word primes revealed a significant effect of case, $F_1(2,76) = 27.49$, MSe = 2762, $p < .09$ ($F_2(2,44) = 22.14$, MSe = 1055, $p < .001$). Neither the effect of gender nor the interaction of case by gender approached significance. An analogous analysis of pseudoword prime latencies revealed a significant effect of case, F_1 (2,76) = 19.32, MSe = 2826, $p < .001$ ($F_2(2,44) = 7.02$, MSe = 2393, $p < .002$) and a significant effect of gender, $F_1(1,38) = 9.18$, MSe = 1913, $p < .01$ ($F_2(1,22) = .75$, MSe = 7156, $p < .40$). The interaction of case by gender was significant by the subjects analysis only, F_1(2,76 = 3.68), MSe = 3046, $p < .03$ ($F_2(2,44) = 1.44$, MSe = 2393, $p < .25$).

No analysis could be performed on the error data. Out of 8 possible errors per condition, the mean number of errors on conditions N1, D1, I1 for words and pseudowords respectively were .27, .29, .42 and .34, .24, .51. The mean number of errors on targets in each condition (ND, DD, ID) computed independently of the error pairing procedure was less than .30 for both words and pseudowords.

Finally, mean recognition latencies for prime words in their nominative, dative/locative, and instrumental forms were computed and correlated for each word pair. For nominative with dative (N1, D1), nominative with instrumental (N1, I1), and dative with instrumental (D1, I1), the correlations were $r = .69$, $r = .66$, and $r = .71$, respectively. These correlations with df = 22, are all significant at the $p < .05$ level. No pseudoword correlations were significant.

Discussion

Overall, decision latencies were prolonged in the second experiment relative to the first. In light of the claim by Forster and Davis (1984) that magnitude of facilitation varies with word frequency (and hence reaction time) in unmasked presentations, no comparisons across experiments are offered. Inspection of decision latencies for word and pseudoword primes revealed a deviation from the characteristic satellite entries outcome. For words, dative/locative case primes were faster than instrumental case primes. Moreover, for pseudowords of both genders, dative/locative and nominative case primes were nearly equivalent. It appears that the preponderance of dative/locative case target words and pseudowords may have facilitated all dative/locative forms. This finding does not invalidate the analysis of repetition priming, however, because all comparisons are on dative/locative case targets.

The strategy for interpreting repetition priming effects adopted in the present study has been to compare identity prime and morpheme prime conditions and to define "full" facilitation as effects that are not different from the identity prime condition. Consistent with the first experiment, the second experiment showed that lexical decision to nouns in the dative/locative case was facilitated by prior presentation of a morphologically-related form. In contrast to the first experiment, the second experiment showed that the instrumental singular primes produced only partial facilitation of dative/locative targets. Assuming that degree of facilitation indexes closeness of relation or extent of activation spread among morphological relatives, it appears that connections within a noun system are not uniform. In Experiment 2, oblique cases were primed more fully by themselves than by other oblique cases. This effect was demonstrated both for masculine nouns whose (base morpheme and) nominative were fully repeated in all oblique forms and for feminine nouns whose nominative was not completely reiterated in any oblique form. Evidently, the lexical organization for a system of inflected nouns includes connections that vary in strength. Moreover, appreciation of morphological relatedness does not depend on a full overlap of the letters that constitute the nominative case form.

A comparison of the pattern of full and partial priming effects in Experiments 1 and 2 revealed some asymmetries in organization for inflected forms that argue against a homogeneous organization of morphological relatives. By the satellite-entries alternative, however, asymmetries are easily accommodated because the nominative form functions as the nucleus of an inflected-noun system. Specifically, the relationship between nominative and oblique cases was as strong as the relationship between oblique and nominative cases in that neither was significantly different from the identity prime condition. In that instrumental primes were significantly different from identity primes, the relationship between two different oblique cases appears to be relatively attenuated. If inflected cases of a noun formed a homogeneous structure--either as fully represented but independent lexical entries or as entries sharing a base morpheme, a claim sometimes made for English (e.g., Kempley & Morton, 1982), then priming should have been equal among all inflected forms. Counter to the claim of a homogeneous representation, identity primes and morphologically-related primes were not equally effective for all targets. In summary, the pattern of partial facilitation obtained in Experiment 2 argues against a uniformly coherent noun system. Moreover, the observed asymmetry in the facilitation among entries of an inflected noun system can be interpreted to support the alleged special status of the nominative singular case proposed by the satellite-entries account.

The effect of presenting a morphologically-related word prior to the presentation of a target word was significantly greater than the analogous manipulation on pseudowords. However, the small but, nevertheless, significant effect of repetition on inflected pseudowords in Experiment 2 implicates a non-lexical contribution to facilitation in the repetition priming paradigm. The nature of inflectional processes in Serbo-Croatian guarantees that members of a satellite system generally will be both orthographically and phonologically very similar. Consequently, all morphologically-related prime-target pairs were visually and phonologically similar in their initial portion. The third and final experiment was designed to examine appreciation of morphological relatedness in word pairs, with diminished orthographic and phonological similarity.

Experiment 3

Experiment 3 asked whether nouns that include sound and spelling changes in some of their inflected forms are represented in the lexicon by a satellite constellation. The present experiment included nouns with two types of sound and spelling changes: (1) feminine words with palatalization in their dative/locative forms and (2) masculine words with changed nominative/accusative forms that include either a) a moveable "A" or b) an "O" that elsewhere appears as "L." We will refer to morphemes that occur in more than one form as "alternating." It is important to note that by linguistic accounts, these alternations are regular and can be described by rules, although they are no longer productive. The repetition priming paradigm was again used with nominative targets preceded by an identical prime and by two morphological primes. For half of the items presented (viz., masculine alternating nouns such as PETAK), both morphological primes differed in spelling and pronunciation from the target forms (PETKU, PETKOM). For the other half of the items (viz., feminine alternating nouns such as NOGA), half of the morphological primes differed in spelling and pronunciation from the target (i.e., NOZI) and half were identical in spelling and pronunciation of the stem morpheme to that of the target (i.e., NOGOM). As in previous experiments, decision latencies to targets as a function of type of prime addresses the issue of cohesion among inflected members of a noun system and the pattern of decision latencies (and correlations) among primes hints at the structure of the noun system.

Method

Subjects. Forty-two first-year students from the Department of Psychology at the University of Belgrade participated in Experiment 3. All had participated in either Experiment 1 or Experiment 2 approximately 6-8 weeks earlier.

Stimulus materials. Twenty-one alternating masculine words and twenty-one alternating feminine words were included in Experiment 3. All of the masculine words had changed spellings in the nominative/accusative singular case and this constituted the atypical form. For most masculine items, the alternation took the form of the addition of a vowel before the last consonant of the base form, thus eliminating certain consonant sequences in word-final position that occurred as a consequence of the disappearance of a weak semivowel in word final position (e.g., PETAK vs. PETKU [nominative singular] vs. [dative singular]). For other masculine forms, it comprised the deletion of "l" and its replacement by "o" in syllable- and word-final position (e.g., PETAO vs. PETLU [nominative/accusative vs. dative/locative singular]); This process occurred in 14th century Serbo-Croatian and again, it was related to the disappearance of a weak semivowel following syllable-final "l" (Belić, 1976). In each case, nominative/accusative and dative/locative forms contained the same number of letters.

All of the feminine words had changed spellings in the dative/locative form where the alternation entailed palatalization of velar consonants (viz., the consonants k, g, h change to c, z, s when followed by "i" derived from "o" or the letter jot (second palatalization)" (Belić, 1976). By comparison, the instrumental singular forms for both masculine and feminine words were typical in construction. One consequence of the locus of the changed case form was that for masculine words the dative/locative and instrumental forms shared spelling and pronunciation, whereas for feminine words the nominative and instrumental forms were similar. Masculine and feminine pseudowords were constructed to include the same style of spelling and sound changes that occurred in words. Examples of alternating masculine and feminine words in their inflected case forms are presented in Table 4.

Table 4

Examples of Alternating Masculine and Feminine Singular Inflected Nouns

CASE	MASCULINE		FEMININE
Nominative (N)	PETAK*	PETAO*	NOGA
Genitive (G)	PETKA	PETLA	NOGE
Dative (D)	PETKU	PETLU	NOZI*
Accusative (A)	PETAK*	PETAO*	NOGU
Instrumental (I)	PETKOM	PETLOM	NOGOM
Locative (L)	PETKU	PETLU	NOZI*
Vocative (V)	PETCE	PETLE	NOGO

*indicates atypical form

The test order and composition of the list(s) were analogous with those of Experiment 1. In the present experiment target words were presented in nominative case and all items were printed in Roman script. As in previous experiments, lags between target and prime averaged ten items with a range of seven to thirteen. With the exception of the number of words in a test order, the testing procedure was identical with that described above.

Results

Errors and extreme response times were eliminated from the present analyses according to the same criteria applied in previous experiments. Fewer than 3% of all responses were eliminated according to these criteria. An additional 2% of all responses were eliminated by the error pairing procedure. Table 5 summarizes the mean recognition times for nominative targets of alternating words and pseudowords.

Table 5

Mean Reaction Times (ms) to Nominative Targets (NN, DN, IN) with Sound and Spelling Alternations and to their Nominative, Dative/Locative, and Instrumental Primes (N1, D1, I1) in Experiment 3

WORDS

MASCULINE				FEMININE				COMBINED			
PRIME		TARGET		PRIME		TARGET		PRIME		TARGET	
N1	664*	NN	631	N1	706	NN	622	N1	685	NN	627
D1	728	DN	641	D1	785*	DN	634	D1	757	DN	638
I1	726	IN	633	I1	739	IN	631	I1	732	IN	632

PSEUDOWORDS

MASCULINE				FEMININE				COMBINED			
PRIME		TARGET		PRIME		TARGET		PRIME		TARGET	
N1	771*	NN	744	N1	773	NN	765	N1	772	NN	754
D1	758	DN	741	D1	796*	DN	777	D1	777	DN	759
I1	806	IN	750	I1	816	IN	757	I1	811	IN	753

*indicates form that undergoes sound and spelling change

Analyses of variance with prime condition (N1, NN, DN, IN) and gender as independent variables were performed on real word latencies using subjects and items as random variables. Consistent with the outcome for repetition priming of nominative targets for regular words, there was a significant effect of prime condition, $F_1(3,123) = 37.30$, MSe = 1630, $p < .001$ ($F_2(3,120) = 19.57$, MSe = 1553, $p < .001$). The interaction of gender by prime condition was also significant, $F_1(3,123) = 10.38$, MSe = 1191, $p < .001$ ($F_2(3,120) = 3.98$, MSe = 1553, $p < .01$). All feminine targets showed more facilitation relative to unprimed nominatives (N1) than did masculine targets. In subanalyses including only target word latencies (viz., NN, DN, IN), neither the effect of gender nor the effect of prime condition approached significance.

An analogous analysis of pseudoword latencies indicated a significant effect of prime condition, $F_1(3,123) = 3.44$, MSe = 1775, $p < .02$; a significant effect of gender, $F_1(1,41) = 8.98$, MSe = 2481, $p < .005$, and an interaction of condition by gender, $F_1(3,123) = 3.74$, MSe = 1338, $p < .01$. None of these was significant by the items analysis, however: ($F_2(3,120) = 1.76$, MSe = 1739, $p < .16$; $F_2(1,40) = 1.50$, MSe = 7425, $p < .23$; and $F_2(3,120) = 1.44$, MSe = 1739, $p < .23$, respectively).

Inspection of the latency data for word primes suggested an interesting deviation from the familiar equivalence among oblique-case latencies predicted by the satellite-entries account. Results of analyses of variance indicated a significant effect of case, $F_1(2,82) = 41.76$, MSe = 2654, $p < .001$ ($F_2(2,80) = 17.98$, MSe = 3082, $p < .001$); of gender, $F_1(1,41) = 55.60$, MSe = 1602, $p < .001$ ($F_2(1,40) = 2.77$, MSe = 16051, $p < .01$); and an interaction of case by gender, $F_1(2,82) = 4.97$, MSe = 2079, $p < .009$, that was not significant by stimulus analysis ($F_2(2,80) = 1.68$, MSe = 3082, $p < .19$). For both genders, nominative forms were recognized most quickly. For masculine forms, oblique cases (neither of which had changed spellings) were equivalent. By contrast, for feminine forms, instrumentals whose stem morphemes were identical in sound and spelling to those of their nominative form were significantly faster than dative/locative forms in which the stem morpheme was not identical, $t(41) = 4.57$, $p < .01$.

An analogous analysis of alternating pseudoword primes indicated that the effect of case was significant, $F_1(2,82) = 17.36$, MSe = 2147, $p < .001$ ($F_2(2,80) = 5.45$, MSe = 3418, $p < .01$) as was the effect of gender, $F_1(1,41) = 11.57$, MSe = 1489, $p < .002$ ($F_2(1,41) = 11.57$, MSe = 1489, $p < .002$). The interaction of case by gender was also significant, $F_1(2,82) = 3.12$, Mse = 2400, $p < .05$ ($F_2(2,82) = 3.12$, MSe = 2400, $p < .05$).

No analyses were performed on the error data because some subjects made no errors and all subjects tended to be extremely accurate. Out of 14 possible errors per condition, the mean number of errors in conditions N1, D1, I1 for words and pseudowords respectively were .63, 1.04 .85 and .65, .42, .69. The mean number of errors on targets in each condition (NN, DN, IN), computed independently of the error pairing procedure, was less than .20 for both words and pseudowords.

Finally, means for prime words in each inflected case were computed and correlated treating Nominative-Dative, Nominative-Instrumental, and Dative-Instrumental latencies as pairs. In consideration of the case-by-gender interaction among primes, separate correlations were made for feminine and masculine word pairs. The correlations of N1, D1; N1, I1, and D1, I1, were $r = .38$, $r = .77$, and $r = .22$ for feminine words and $r = .69$, $r = .77$, and $r = .82$ for masculine words, respectively. No analogous pseudoword correlations approached significance. With 21 words (and 19 degrees of freedom) correlations of $r = |.44|$ are significant at the .05 level. In summary, with the exception of correlations involving feminine dative/locative case, correlations among all inflected forms of a noun were significant.

Discussion

Differences between prime and target in spelling and pronunciation of the shared morpheme did not eliminate repetition priming. Facilitation with repetition obtained both when target and prime maintained a common spelling and pronunciation and when they did not. This outcome is consistent with that obtained by Fowler et al. (1985), which showed statistically full priming for alternating English words, and also with many of the results reported by Stanners et al. (1979). It is not the same, however, as the outcome of an experiment by Kempley and Morton (1982) in which irregular morphologically-related words were presented auditorily for recognition in noise and no priming obtained between irregular and regular forms. Evidently, the outcome of the present study indicates that regular alternations in sound and spelling do not mask morphological relationships. Relative to the identity prime condition (NN), there was no significant reduction in facilitation due to repetition when morphological primes differed from targets in spelling and pronunciation (viz., dative and instrumental masculine primes and dative feminine primes). Statistically, priming was full in all instances. Secondarily, and as described above, Fowler et al. (1985) have reported that a nonsignificant numerical loss in priming typically occurs when affixes of prime and target are not identical. An analysis of target latencies alone in the present experiment replicates the outcome of Fowler and her colleagues in English. There is a tendency for prime target pairs with nonidentical affixes to show very small and nonsignificant reductions in the magnitude of facilitation. Based on these data, overlap in sound and spelling between target and prime (interpreted as a nonlexical or an episodic contribution) did not systematically modify the facilitation that occurs in the repetition priming task. The coherence among satellite entries of an alternating noun system appears not to differ from that of nonalternating nouns.

Among pseudowords, inspection of means suggested that the magnitude of facilitation averaged over gender was 18 ms when prime and target differed (Experiment 3) and was 58 ms in one condition when prime and target remained the same (viz., feminine pseudowords in Experiment 1). In Experiment 3, the analyses of variance were significant only by the subject's analysis and in Experiment 1, there was no facilitation with repetition for masculine pseudowords. Nevertheless, it is important to point out that the differences among latencies to pseudoword targets cannot readily be ascribed to overlap of surface characteristics with their prime. Inspection of means suggested that, irrespective of case of prime and in contrast to the outcome of Experiment 1, alternating masculine pseudoword targets were primed more consistently than were alternating feminine pseudoword targets. However, morphological primes were consistently less similar to their targets for masculine pseudowords (whose nominative/accusative was different from all oblique forms) than for feminine pseudowords (whose nominative overlapped formally with instrumental morphological primes but not with dative morphological primes). In summary, the magnitude of facilitation was significantly reduced in alternating pseudoword targets relative to regular pseudoword targets but similarities of surface characteristics do not account satisfactorily for the pattern.

For alternating primes, the interaction of case by gender and the pattern of correlations among recognition latencies indicated that the structure of the noun system for masculine and feminine nouns contrasts. Latencies for masculine nouns supported the usual primacy for the nominative and the equivalence among oblique cases described by a satellite-entries account, whereas latencies for feminine nouns suggested that recognition of the dative/locative was impeded because its spelling and pronunciation were changed relative to its nominative and to other oblique cases. This outcome suggests that, at least for feminine alternating nouns, the structure of the noun system may differ from the typical satellite configuration. Pair-wise correlations between mean latencies for each word in its nominative, dative/locative, and instrumental forms supported this interpretation. For masculine nouns, all cases were strongly correlated whereas for feminine nouns, the changed dative/locative did not correlate significantly with its more regular forms, although the regular cases did correlate with each other.

In summary, deviations in spelling and pronunciation affect the structure of the inflected noun system as evidenced by latencies for changed dative/locative forms of feminine alternating nouns that served as primes. The failure to demonstrate an analogous effect in masculine nouns was ambiguous, however. It might reflect a qualitative difference in the irregular spellings. The phonetic environment for the application of the moveable "a" rule or the "o" to "l" alternation is perhaps less simply described than the environment for palatalization. Alternatively, it may provide further evidence for the primacy of the nominative case. If typicality within a satellite system is defined relative to the nominative form, then changed nominative forms of alternating masculine nouns may not, in effect, be deviant. The pattern of correlations supports the latter interpretation.

In conclusion, the latency data for changed primes suggested that deviation in spelling and pronunciation alter initial accessibility of inflected forms and the structure of the noun system, whereas the repetition priming data on target words suggested that once an entry has been activated, the nominative nucleus of its noun system is activated as well. Deviations in spelling and pronunciation may affect the structure of the noun system; it appears, however, that once the satellite entry of either a regular or an alternating noun system has been accessed, the entire noun system is activated.

General Discussion

In the first experiment, nouns in the nominative case were primed by identical or morphologically-related forms, namely, dative/locative and instrumental cases. The outcome was statistically full facilitation by repetition in all prime conditions. This outcome is consistent with the claim that inflected-noun forms in Serbo-Croatian are strongly cohesive in the lexicon. The pattern of latencies for the primes replicated the pattern from which the satellite-entries account originated (Lukatela et al., 1978; 1980). Moreover, the latencies of primes were significantly correlated. A critical characteristic of the satellite-entries account is that the nominative singular case has a special status in the lexical organization. One consequence of its privileged position might be that the nominative can prime and be primed more fully by non-nominative cases than can any oblique case. The outcomes of Experiments 1 and 2 support this interpretation. In Experiment 1, we found full facilitation of nominative targets by both identical and morphological primes. In Experiment 2, lexical decision latency to nouns in the dative/locative case was facilitated by a prior presentation of a morphologically-related inflected form. However, instrumental singular primes produced only partial facilitation of dative/locative targets. The statistically significant pattern of full and partial priming was interpreted as evidence that the lexical organization among inflected cases of a noun is not homogeneous; that is, connections among inflected nouns are not uniformly represented in the lexicon. In particular, the connection between two satellites of an entry appears to be weak relative to the connection between a satellite and the nucleus. Insofar as inhomogeneities in organization are evident, it is difficult to conceive of a representation in which all inflected forms of a noun either share a base morpheme or are fully independent lexical entries.

In the third and final experiment, nouns that undergo regular sound and spelling changes in at least one of their inflected-case forms were presented as targets in the nominative case. Decision latency was equally facilitated by a prior presentation of all morphologically-related primes. Thus, the pattern of facilitation observed does not depend on maintaining phonological and orthographic similarity between prime and target: The same outcome obtained with pairs including a sound and spelling change and pairs including no change. Likewise, the pattern of facilitation with pseudowords could not be accounted for entirely by sound and spelling overlap. Collectively, the results suggest that the representation that underlies repetition priming must be sufficiently abstract to accommodate changes in the base morpheme of morphologically-related words.

The effect of repetition priming was consistently more robust with words than with pseudoword targets and this outcome is interpreted as implicating, at least in part, lexical processes. Insofar as facilitation reflects activation among lexical entries, results indicate that in addition to capturing inflectional rules that are productive, these representations also encompass alternations among forms that are probably no longer productive.

Finally, the pattern of decision latencies for regular noun primes and the correlation among forms indicates that inflected forms of a noun are associated. This outcome is interpreted as reflecting the structure of the noun system. For feminine alternating nouns, however, latencies were associated only when both words had identical base morphemes. Failure to observe a significant correlation between atypical and typical forms of alternating nouns lends support to the assumption that the pattern of correlations reflects, at least in part, lexical factors. The pattern of decision latencies and correlations for masculine alternating nouns that had a changed nominative/accusative case indicated that they were handled like regular nouns: All cases were associated. This outcome permits two interpretations: Either the nominative case is special such that alternation is defined relative to the nominative or alternatively, the particular sound and spelling changes that appear in the present set of masculine words are different from the changes that occur in feminine words. Discussion of the specifics by which alternating inflectional forms are represented and their role in defining the satellite organization among entries should not be allowed to obscure the basic result. The outcome of the present series of experiments is consistent with the claim that inflected cases of a noun are represented fully but not independently and that morphological relatedness provides a principle of organization in the lexicon. In this respect, the present experiments conducted in the highly inflected language of Serbo-Croatian are consistent with results of repetition priming studies conducted with English materials (Fowler et al., 1985).

In summary the present study extends the satellite-entries account of Lukatela et al. (1978; 1980) in the following ways: The equivalence of decision latencies for all oblique forms observed with nonalternating nouns was not observed with feminine alternating nouns. These data, in conjunction with the correlations between latencies for inflected-case forms, support the claim that alternating nouns do not configure in the typical satellite fashion. In the present study, the pattern of full and partial facilitation in repetition priming was deployed to probe the organization among satellite entries as a further extension of Lukatela's work. Among regular noun systems, the facilitation was always full for nominative targets, whereas facilitation was significantly diminished when an oblique-case target was preceded by a different oblique-case prime. Interpreting magnitude of facilitation as an index of the organization within the inflected noun system, these results reveal inhomogeneities in the coherence of the satellite system. Specifically, the connections between two satellite entries that represent different inflected-case forms are weaker than the connection between an entry and its nucleus. In contrast, the connections between the nominative nucleus and all of its inflected-case satellites are equally strong. The latter outcome can be interpreted as further evidence for the primacy of the nominative. Finally, when typical and atypical forms of alternating nouns were presented as primes, decision latencies to nominative targets revealed a pattern of facilitation that was comparable to that reported with nonalternating nouns. This outcome, namely full facilitation, suggests that once a satellite entry is activated, all components of its noun system are accessed and that this is true both for alternating and nonalternating nouns. In conclusion, although the noun system of alternating and nonalternating may differ, once access to an entry occurs it necessarily entails the activation of its entire noun system.

References

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. Journal of Experimental Psychology: Human Perception and Performance, 10, 1-201.

Belić, A. (1976). Fonetika. University of Belgrade: Naučna Knijiga.

Burani, C., Salmaso, D., & Caramazza, A. (1984). Morphological structure and lexical access. Visible Language, 18, 348-358.

Caramazza, A., Miceli, G., Silveri, M. C., & Laudanna, A. (1985). Reading mechanisms and the organization of the lexicon: Evidence from acquired dyslexia. Cognitive Neuropsychology, 2, 81-114.

Dannenbring, G. L., & Briand, K. (1982). Semantic priming and the word repetition effect in a lexical decision task. Canadian Journal of Psychology, 36, 435-444.

Feldman, L. B. (1984). Lexical organization of morphological relatives. Paper presented at the Psychonomic Society Meeting, San Antonio, TX.

Feldman, L. B. (in press). Phonological and morphological analysis by skilled readers of Serbo-Croatian. In A. Allport, D. Mackay, W. Prinz, & E. Scheerer (Eds.), Language perception and production. London: Academic Press.

Feldman, L. B., Kostić, A., Lukatela, G., & Turvey, M. T. (1983). An evaluation of the "Basic Orthographic Syllabic Structure" in a phonologically shallow orthography. Psychological Research, 45, 55-72.

Feldman, L. B., & Moskovljević, J. (1986). Repetition priming is not purely episodic in origin. Journal of Experimental Psychology: Learning, Memory and Cognition.

Feldman, L. B., & Turvey, M. T. (1983). Morphological processes in word recognition. Paper presented at the Psychonomic Society Meeting, San Diego, CA.

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. Journal of Experimental Psychology: General, 112, 309-346.

Forbach, G. B., Stanners, R. F., & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. Memory & Cognition, 2, 337-339.

Forster, K., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 680-698.

Fowler, C. A., Napps, S. E., & Feldman, L. B. (1985). Lexical entries are shared by regular, irregular, and morphologically-related words. Memory & Cognition, 13, 241-255.

Hanson, V. L., & Wilkenfeld, D. (1985). Morphophonology and lexical organization in deaf readers. Language and Speech, 28, 269-280.

Henderson, L., Wallis, J., & Knight, D. (1984). Morphemic structure and lexical access. In H. Bouma & D. Bouhuis (Eds.), Attention and performance X. London: Erlbaum.

Hudson, P. T. W., & Bergman, M. W. (1985). Lexical knowledge in word recognition: Word length and word frequency in naming and lexical decision tasks. Journal of Memory and Language, 24, 46-58.

Kempley, S. T., & Morton, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. British Journal of Psychology, 73, 441-454.

Kostić, D. (1965). Frequency of occurrence of words in Serbo-Croatian. Unpublished manuscript, Institute of Experimental Phonetics and Speech Pathology, University of Belgrade.

Lukatela, G., Gligorijević, B., Kostić, A., & Turvey, M. T. (1980). Representation of inflected nouns in the internal lexicon. Memory & Cognition, 8, 415-423.

Lukatela, G., Mandić, Z., Gligorijević, B., Kostić, A., Savić, M., & Turvey, M. T. (1978). Lexical decision for inflected nouns. Language and Speech, 21, 166-173.

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. Journal of Experimental Psychology, 102, 963-968.

Napps, S. (1985). Morphological, semantic, and formal relations among words and the organization of the mental lexicon. Unpublished doctoral dissertation, Dartmouth College.

Napps, S., & Fowler, C. A. (1983). Orthographic organization of lexical forms. Paper presented at the Eastern Psychological Association Meeting.

Napps, S., & Fowler, C. A. (in press). The effect of orthography on the organization of the mental lexicon.

Oliphant, G. (1983). Repetition and recency effects in lexical memory. Australian Journal of Psychology, 35, 393-403.

Scarborough, D. L., Cortese, C., & Scarborough, H. (1977). Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance, 3, 1-17.

Seidenberg, M. S., & Tanenhaus, M. K. (1986). Modularity and lexical access. In I. Gopnik (Ed.), Papers from the McGill Cognitive Science Workshops. Norwood, NJ: Ablex.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. Journal of Verbal Learning and Verbal Behavior, 18, 399-412.

Taft, M. (1979). Lexical access via an orthographic code: The Basic Orthographic Syllable Structure (BOSS). Journal of Verbal Learning and Verbal Behavior, 18, 21-40.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. Journal of Verbal Learning and Verbal Behavior, 14, 638-647.

REPETITION PRIMING IS NOT PURELY EPISODIC IN ORIGIN*

Laurie B. Feldman† and Jasmina Moskovljević††

Abstract. The sufficiency of similarity among surface attributes of word pairs to account for the pattern of facilitation obtained in the repetition priming paradigm (Stanners, Neiser, Hernon & Hall, 1979) was evaluated. In this variation of the lexical decision task, target words are preceded earlier in the list by identical and morphologically related prime words. Target latency as a function of type of prime is examined and reduction in decision latency relative to an unprimed presentation is measured. In Experiment 1, morphological relatives were singular inflected case forms of Serbo-Croatian words and visual similarity of prime and target was manipulated by alternating the two alphabets in which the Serbo-Croatian language is written. Results indicated that the magnitude of facilitation in the alphabetically alternating condition was not reduced relative to the nonalternating condition (RUPI-RUPI vs. РУПИ-RUPI), which suggested that visual similarity is not a necessary condition for facilitation in the present task. In Experiment 2, related pairs included (a) base forms with diminutives, a class of highly productive and semantically related derivations marked in Serbo-Croatian by suffixes such as ČIĆ and ICA (STAN-STANČIĆ) and (b) base words with morphologically unrelated monomorphemic words whose orthographic pattern encompassed the target in initial position and a sequence of letters in final position that elsewhere functions as a diminutive suffix (STAN-STANICA). Results of the second experiment showed no facilitation of word targets by orthographically similar but morphologically unrelated primes although there was a tendency toward facilitation among structurally similar pseudowords. Collectively, the experiments suggested that structural similarity of prime and target is not a sufficient condition for facilitation in the repetition priming paradigm.

---

*Journal of Experimental Psychology: Learning, Memory, and Cognition, in press.

†Also University of Delaware

††University of Belgrade

Acknowledgment. This research was supported by funds from the National Academy of Sciences and the Serbian Academy of Sciences to the first author and by NICHD Grants HD-01994 to Haskins Laboratories and HD-08495 to the University of Belgrade. Special thanks to Georgije Lukatela and the staff of the Laboratory of Experimental Psychology at the University of Belgrade.

[HASKINS LABORATORIES: Status Report on Speech Research SR-86/87 (1986)]

Introduction

Serbo-Croatian, the dominant language of Yugoslavia, possesses some special properties that create ideal conditions under which to investigate how morphological structure of words is captured in the internal lexicon of the adult reader. First, numerous inflected and derived variants are used productively in Serbo-Croatian and their formation is complex in that there is no simple relation between form of affix and function. Typically, a word comprises a base morpheme to which may be affixed one or more derivational suffixes that modify the meaning of the base (and sometimes change its word class) as well as an inflectional suffix that serves a syntactic function. Words whose constituent structure includes a common base morpheme are morphologically related. Second, in Serbo-Croatian a simple mapping between grapheme and classical phoneme is always preserved, and many predictable alternations are represented in the orthographic form of a word. As a result, morphological relatives may have discrepant spellings if the shared morpheme undergoes a phonological change in some, but not all variants. Finally, one language is transcribed in two different alphabets, Roman and Cyrillic, and most characters are unique to their respective alphabets. Educational policy mandates that competence in both alphabets be demonstrated by elementary school children and the prevalence of printed material in each alphabet guarantees that this competence is maintained by adults. As a result, most words in Serbo-Croatian have two printed forms that are visually quite distinct and equally familiar to the skilled reader.

One methodology for exploring the role of morphological structure in lexical organization employs a variation of the lexical decision task known as repetition priming. Accordingly, the pattern of facilitation among related forms is interpreted to reflect, at least in part, how those forms are organized in the lexicon of the user. Critics of this approach have claimed that the effect is largely episodic in origin, reflecting the formation of a memory trace for the test materials (Feustel, Shiffrin, & Salasoo, 1983) or alternatively, coding fluency due to prior presentation of the same visual configuration (Jacoby & Brooks, 1984). Episodic memory and perceptual fluency accounts are based on the perceptual analysis of a particular visual pattern or the memory trace thereof and collectively they can be contrasted with a lexical alternative that claims that the representations that underlie facilitation in the repetition priming task necessarily include linguistic knowledge about the meaning and constituent morpheme structure of a letter string if it is a word. In the present study, the sufficiency of the episodic perceptual fluency account of repetition priming is examined for morphologically-related words in Serbo-Croatian. Before describing individual experiments, the paradigm and its various interpretations are summarized.

In the repetition priming procedure (Forbach, Stanners & Hochhaus, 1974; Scarborough, Cortese, & Scarborough, 1977; Stanners, Neiser, Hernon, & Hall, 1979), each word and pseudoword is presented twice (with a lag of intervening items) for a lexical decision judgment. The reduction in decision latency relative to a first presentation or facilitation due to repetition is measured. (The first presentation of the item is the "prime." The second presentation is the "target.") For facilitation to occur with English materials, it is not necessary that the identical word be repeated as prime and target. Generally, morphologically-related words including inflections and derivations also reduced target decision latency--sometimes as fully as an identical repetition. For example, the inflected form "manages" and the derived form "management" both facilitated a subsequent presentation of "manage" and decision latencies to the target when preceded by inflectionally or derivationally related words were equal to an identical presentation of "manage" (Fowler, Napps, & Feldman, 1985; cf. Stanners et al., 1979). (The magnitude of facilitation with morphological relatives as primes is defined relative to the facilitation with an identical repetition [following Fowler et al., 1985]. The outcome is "full" repetition priming when identity and morpheme primes produce equivalent results. Priming that is significant, but numerically less than full, is "partial.") In addition, full repetition priming occurred when target and prime had slightly discrepant pronunciations and/or spellings, e.g., "health" and "heal" (Fowler et al., 1985; Hanson & Wilkenfeld, 1986). By contrast, it did not occur among morphologically-unrelated words whose initial letters overlapped, e.g., "ribbon" and "rib" (Hanson & Wilkenfeld, 1986; Napps, 1985; Napps & Fowler, submitted). Results such as these support an interpretation of repetition priming effects as at least partially lexical in origin (Fowler et al., 1985; Monsell, 1985; Stanners et al., 1979), although they do not explicitly assess the nature and extent of a nonlexical or episodic component (Feustel et al., 1983) or alternatively, of a task-dependent strategic contribution (Forster & Davis, 1984; Oliphant, 1983; Ratcliff, Hockley, & McKoon, 1985). Finally, it has been proposed that facilitation in repetition priming may comprise several factors, including a transitory component that is evident only at prime target lags of two items or less, as well as a more stable lexical component that is evident at longer lags (Monsell, 1985; Ratcliff et al., 1985). The present discussion of repetition priming in lexical decision is restricted to studies that incorporated lags that purportedly exceed the duration of any short-term component. In summary, it appears that facilitation due to presentation of morphological relatives reflects the influence of a lexical factor and that the difference between full and partial priming may reflect an episodic increment to the lexical effect when priming is full that is absent when priming is partial (Fowler et al., 1985). The relationship between the extent to which the surface characteristics are retained and the magnitude of (partial) facilitation at lags greater than four items or 16 sec has been explored systematically for words (but not pseudowords) by Kirsner and colleagues (Kirsner & Dunn, 1985) and is consistent with this characterization.

The present series of experiments was designed to probe the organization of morphologically-related words in the internal lexicon of readers of Serbo-Croatian, a language with a complex morphology. The repetition priming procedure of Stanners et al. (1979) was used to investigate the lexical organization of inflected and derived forms. In light of the confounding of lexical effects by episodic effects or perceptual analysis that may be inherent in the repetition priming paradigm, special consideration is given to the nature of these factors. As discussed above, one interpretation of the repetition priming procedure is that the priming is principally an effect based on the retrieval of information from specific prior episodes or perceptual identification of the same pattern in a similar format and context (Feustel et al., 1983). In Experiment 1, repetition priming among inflected forms of a noun was investigated. Visual contributions to episodic effects were eliminated by presenting the first and second occurrence of an item in different alphabets. We asked whether the effect of repetition priming was unchanged when the similarity of the surface characteristics of prime and target was eliminated. Stated alternatively, we asked whether the basis for facilitation by primes is sufficiently abstract to tolerate changes in alphabet without producing a reduction in the magnitude of facilitation due to repetition. In Experiment 2, the role of orthographic (and phonological) similarity between prime and target in repetition priming was investigated by comparing real derivatives (namely, diminutives) with an unrelated monomorphemic word whose initial portion was orthographically (and phonologically) similar to the target and whose final portion inappropriately suggested that it was a derived form. Taken together, these experiments attempt to find evidence that two nonlexical sources of facilitation govern the effects in repetition priming. To anticipate, effects defined neither by repetition of the same alphabetic characters (visual pattern) nor by repetition of the same "abstract" orthographic (structural) pattern can account adequately for the pattern of facilitation obtained with words in the repetition priming paradigm.

Experiment 1

Morphologically-related primes reduce decision latencies to their targets, an outcome that has been interpreted as an index of lexical organization (Fowler et al., 1985; Kempley & Morton, 1982; Stanners et al., 1979). An alternative to the lexical interpretation, derived from a slightly different paradigm and a recognition measure, emphasizes the formation of an episodic trace (Feustel et al., 1983) or of fluency of perceptual identification (Johnston, Dark, & Jacoby, 1985). Allegedly, it is the "visual characteristics of the display and the configuration of letters in the item that are probably preserved between successive presentations of a letter string" (Feustel et al., 1983, p. 344). Admittedly, in most studies that purport to explore morphological relatedness as a principle of lexical organization, related pairs of words are visually similar as well as morphologically related. One exception is a study by Morton (1979) in which the similarity of surface characteristics of words in the study and test phases was manipulated by alternating handwritten and typed presentations. Whereas the outcome of that study revealed a numerically small and statistically nonsignificant reduction in identification levels when writing style alternated relative to when it was maintained, it could be argued that the critical attributes of the handwritten and typed formats of a word are more similar than different. A second technique used to reduce visual similarity has been to examine repetition priming for words that undergo changes of sound and spelling, including suppletions (e.g., sleep-slept; go-went) (Feldman & Fowler, in press; Fowler et al., 1985; Kempley & Morton, 1982; Napps, 1985). Facilitation is still observed under these conditions but the magnitude of facilitation is attenuated relative to other experiments with words that do not undergo change. Insofar as the magnitude of facilitation is reduced when prime and target are less similar visually, episodic effects are implicated. Nevertheless, explicit attempts to relate visual similarity defined by extent of letter overlap of prime and target to magnitude of facilitation have not proven successful (Napps, 1985).

In general, it appears that structurally similar primes can augment overall facilitation in the repetition priming paradigm by introducing an episodic contribution, but is a nonlexical component sufficient to provide a full account? Discussion of lexical effects hinges on the assumption that words, but not pseudowords, can benefit from the contribution of lexical factors. According to a strong lexical view, evidence of facilitation with word targets and the absence of an effect with pseudoword targets is generally interpreted as evidence in favor of a lexical interpretation. Supporters of the episodic view have argued, however, that pseudowords as well as words have memory representations and that the outcome with pseudowords in this paradigm is equivocal because the tendency to respond "no" may be offset by the availability of an episodic trace that increases with multiple presentations, whereas the tendency to respond "yes" is enhanced (Feustel et al., 1983). Similarly, the availability of a lexical representation or of item meaningfulness may affect perceptual identification so that the interdependence of performance measures on fluency and recognition tasks is greater for pseudowords than it is for words (Johnston et al., 1985). In summary, to the extent that repetition effects occur with pseudowords--an effect that is taken to be nonlexical in nature--an interpretation of the effect with words as purely lexical in origin is not supported. Nevertheless, lexical information appears to facilitate or alternatively impair the formation or utilization of nonlexical codes in repetition priming and related tasks.

In the first experiment, the contribution of visual similarity between prime and target was investigated in an attempt to identify nonlexical contributions, defined on visual characteristics of the display. Morphologically-related prime-target pairs were inflected case forms of masculine and feminine nouns in Serbo-Croatian. As noted above, the Serbo-Croatian language possesses a special property permitting an experimental manipulation that minimizes the visual overlap of target and prime: Words may be printed in either Roman or Cyrillic alphabet characters where the two forms are generally quite dissimilar in appearance and skilled readers are equally facile with both systems. In Experiment 1a, reported as Experiment 2 in Feldman and Fowler (in press), both targets and primes were printed in Roman script. In Experiment 1b, the same targets were again printed in Roman but the preceding primes were printed in Cyrillic. Replication of an experiment within one alphabet context and across alternating alphabet contexts permitted an evaluation of whether the visual similarity of surface attributes of target and prime necessarily figures in the magnitude of facilitation demonstrated in the repetition priming paradigm.

Method

Subjects. Thirty-nine first-year students from the Department of Psychology at the University of Belgrade participated in Experiment 1a. Forty-eight second year students from the same department participated in Experiment 1b. All were native speakers of Serbo-Croatian and fluent readers of both the Roman and Cyrillic alphabets.[1] All had vision that was normal or corrected to normal. As implied above, no subject participated in more than one experiment, although all had prior experience in other reaction-time studies during their first year of study at the University.

Stimulus materials. Twenty-four Serbo-Croatian words and twenty-four pseudowords were included in each part of Experiment 1. Words were familiar nouns that contained four or five letters in their nominative form. Half were feminine and half were masculine. No words were included that contained sequences of more than two consonants. Pseudowords were generated by changing one or two letters (vowel with vowel or consonant with consonant) in other real words with the same orthographic structure. The same words and pseudowords were used in Experiments 1a and 1b.

Each word appeared in three different inflectional cases: nominative, dative/locative, and instrumental singular. Each pseudoword also appeared with inflectional affixes for masculine or feminine words in the same cases. Words were chosen so that inflectional suffixes did not alter the spelling of the base morpheme. Examples of masculine and feminine words in their Roman and Cyrillic inflected-case forms appear in Table 1.

Table 1

Examples of Regular Masculine and Feminine Singular Inflected Nouns Printed in Roman and Cyrillic

	GENDER			
INFLECTED CASE	MASCULINE		FEMININE	
	ROMAN	CYRILLIC	ROMAN	CYRILLIC
Nominative (N)	DINAR	ДИНАР	RUPA	РУПА
Genitive (G)	DINARA	ДИНАРА	RUPE	РУПЕ
Dative (D)	DINARU	ДИНАРУ	RUPI	РУПИ
Accusative (A)	DINAR	ДИНАР	RUPU	РУПУ
Instrumental (I)	DINAROM	ДИНАРОМ	RUPOM	РУПОМ
Locative (L)	DINARU	ДИНАРУ	RUPI	РУПИ
Vocative (V)	DINARE	ДИНАРЕ	RUPO	РУПО

In Experiment 1a, all letter strings were printed in Roman characters. In Experiment 1b, prime items were printed in Cyrillic characters and target items were printed in Roman. Stimulus items were selected to maximize the visual distinctiveness of Roman and Cyrillic transcriptions by avoiding those words that predominated in phonemes that had a common graphemic form in the two alphabets. For example, the word RUPA-РУПА was included, but JAJE-JAJE was not. (Here, the first transcription of the word is in Roman and the second is in Cyrillic.)

Procedure. Subjects performed a lexical decision task. As each letter string appeared, they had to press a telegraph key with both hands to indicate whether or not it was a word. They pressed the farther key to signal "yes" and the closer key to signal "no." All letter strings were typed, then photographed and mounted as slides. The stimuli were projected from a carousel projector and displayed on a screen until subjects responded (approximately 750 ms). A dark field immediately preceded and followed the display. Reaction times were measured from the onset of the stimulus display. The interval between experimental trials was controlled by the experimenter and averaged about 2000 ms.

Design. In each part of the experiment, three test orders containing 100 items were created. Forty-eight items were primes and forty-eight items were targets. In addition, there were four filler items. Words and pseudowords were equally represented in each category. Test orders included three prime conditions distinguished by the inflectional case of the prime, that is, nominative, dative/locative, or instrumental. (Case of prime was indicated as N1, D1, or I1, respectively.) All targets were in dative/locative case. Half were masculine gender and half were feminine. (The conditions of dative/locative targets preceded by nominative, dative/locative, and instrumental primes were indicated as ND, DD, and ID, respectively.) Words appeared in the same serial position across all test orders although the inflectional case of the prime varied. For example, the word RUPI (meaning "hole") was presented in its dative/locative form as the target in all three test orders but within each test order it was preceded, in the same position, by either RUPA, RUPI, or RUPOM as a prime.

Each subject viewed one test order. Therefore, each subject saw every morpheme twice, once in a prime and once in a target. The average lag between the presentation of the prime and the target was ten items. Lags ranged from seven to thirteen and were binominally distributed around a lag of ten. Filler items were introduced to maintain appropriate lags and a practice list of ten items preceded the test list.

To summarize the experimental design: across test orders each target word or pseudoword in dative/locative case was preceded by its prime in nominative, dative/locative, and instrumental form. Within each order, a base morpheme occurred once in a target and once in a prime, and case of prime varied with item. Stated alternatively, all subjects viewed the three cases of prime on different target items, and across test orders each word was preceded by each case of prime. In Experiment 1a, primes and targets were printed in Roman script. In Experiment 1b, primes were printed in Cyrillic and targets were printed in Roman.

Results

Errors and extreme response times (greater than 1200 ms or less than 350 ms) were eliminated from all analyses. This procedure eliminated fewer than 4% of all responses. In addition, when a subject responded incorrectly to one member of a prime-target pair, both responses were excluded from subsequent analysis. The error-pairing procedure eliminated an additional 3% of all responses. Table 2 summarizes the mean recognition times over subjects for dative/locative target words and pseudowords in Experiments 1a (Nonalternating) and 1b (Alternating). They are discussed in that order. It also includes two measures of facilitation based on a) the difference in reaction time to first and second presentation of dative/locative forms D1-DD b) that difference expressed as a percentage of D1 latency.

Analyses of variance on Roman-Roman pairs with condition (D1, ND, DD, ID) and gender as independent variables were performed using subjects (F_1) and items (F_2 in parentheses) as random variables. The outcome reported previously (Feldman & Fowler, in press; Feldman & Turvey, 1983) showed that the effect of condition was significant for real words, $F_1(3,114) = 59.48$, MSe = 2158, $p < .001$ ($F_2(3,66) = 27.54$, MSe = 1435, $p < .001$). The effect of gender and the interaction of condition by gender were significant in the analysis by subjects but not in the analysis by items, $F_1(1,38) = 6.27$, MSe = 1728, $p < .02$, and $F_1(3,114) = 2.98$, Mse = 1913, $p < .04$, respectively.

Table 2

Mean Reaction Times 9ms) to Roman Alphabet Dative/Locative Targets (ND, DD, ID) and their Alphabetically Alternating and Nonalternating Dative/Locative Prime (D1).

	CONDITION				FACILITATION	
	D1	ND	DD*	ID		
PRIME:						
NONALTERNATING	RUPI	RUPA	RUPI	RUPOM	D1-DD	D1-DD / D1
ALTERNATING	РУПИ	РУПА	РУПИ	РУПОМ		
TARGET:		RUPI	RUPI	RUPI		
		WORDS				
Nonalternating	642	563	552	573	90	14%
Alternating	678	595	588	607	90	13%
		PSEUDOWORDS				
Nonalternating	716	688	684	705	32	4%
Alternating	736	712	701	705	35	5%

*Indicates identity prime condition

A subsequent set of analyses including only dative/locative target latencies (conditions ND, DD, ID) revealed a significant effect of prime condition, $F_1(2,76) = 4.02$, MSe = 2028, $p < .02$ ($F_2(2,44) = 3.17$, MSe = 790, $p < .05$) and inspection of means indicated that identity primes were more effective than instrumental primes. There was also a suggestion that the effect of gender was significant, $F_1(1,38) = 20.77$, $p < .001$ (significant by the subject's analysis but not by the item's analysis). The interaction of condition by gender was not significant.

An analysis of pseudoword latencies (D1, ND, DD, ID) showed a significant effect of prime condition, $F_1(3,114) = 6.77$, MSe = 2582, $p < .001$ ($F_2(3,66) = 2.75$, MSe = 1952, $p < .05$); there was no effect of gender and no interaction of condition by gender. A subsequent analysis of pseudoword targets (ND, DD, ID) suggested a significant effect of condition such that instrumental case primes facilitated less than did dative/locative or nominative case primes, $F_1(2,76) = 3.37$, MSe = 2848, $p < .04$. This effect was not significant in the stimulus analysis, however. When words and pseudoword latencies were entered into one analysis, the interaction of condition by lexicality was significant, $F_1(3,114) = 15.72$, MSe = 1938, $p < .001$ ($F_2(3,132) = 4.81$, MSe = 1950, $p <$

.003). Inspection of means indicated that words were facilitated more by repetition than were pseudowords.

When primes were printed in Cyrillic characters and dative/locative targets (real words) were printed in Roman characters (Experiment 1b), an analysis of target latencies (D1, ND, DD, ID) indicated that the main effect of condition was significant, thus replicating the outcome of Experiment 1a, $\underline{F}_1(3,141) = 54.15$, MSe = 3033, $\underline{p} < .001$ ($\underline{F}_2(3,66) = 36.94$, MSe = 1112, $\underline{p} < .001$) both in magnitude as well as in pattern of the significance of its main effects. The effect of gender and the interaction of gender by prime condition were not significant by either analysis in Experiment 1b. As in Experiment 1a, subanalyses on target words alone (conditions ND, DD, ID) and inspection of means replicated a significant effect of case of prime, $\underline{F}_1(2,94) = 3.88$, MSe = 2241, $\underline{p} < .02$ ($\underline{F}_2(2,44) = 3.14$, MSe = 692, $\underline{p} < .05$) whereby identity primes produced faster recognition times for dative/locative targets than did instrumental primes.

An analysis of pseudoword latencies in Experiment 1b replicated the outcome of Experiment 1a. There was a significant effect of condition, $\underline{F}_1(3,141) = 7.60$, MSe = 3170, $\underline{p} < .001$ ($\underline{F}_2(3,66) = 29.71$, MSe = 1382, $\underline{p} < .001$). No other main effect or interaction approached significance. In contrast to Experiment 1a, however, a subanalysis of pseudoword targets indicated no significant difference among targets as a function of inflectional case of prime. When words and pseudoword latencies were entered into one analysis, the interaction of condition by lexicality was significant, $\underline{F}_1(3,141) = 13.61$, MSe = 2553, $\underline{p} < .001$ ($\underline{F}_2(3,132) = 5.53$, MSe = 1572, $\underline{p} < .001$). Words were facilitated more by repetition that were pseudowords.

Discussion

The major outcome of the present experiment was that the magnitude of facilitation in the repetition priming procedure with inflected forms of words and pseudowords was as large when target and prime were printed in different alphabets as when they were printed repeatedly in one. In fact, the magnitudes of priming in the ID, DD, and ND conditions as assessed by subtracting D1 times from them are remarkably similar.

As noted above, facilitation was assessed by comparing second presentations of dative/locative case nouns printed in Roman characters (ND, DD, IN) to the first presentations of those same items (D1) printed in either Roman (Nonalternating Experiment 1a) or Cyrillic (Alternating Experiment 1b). In asserting the appropriateness of a baseline that varies with respect to alphabet, it is important to note that based on latency measures in a lexical decision task, skilled readers of Serbo-Croatian show no systematic alphabet bias for phonologically unambiguous words. This outcome, namely, equivalent reaction times to the Roman and Cyrillic transcriptions of a letterstring, has been reported both in designs where alphabet is treated as a between-subject (Feldman & Turvey, 1983) and as a within-subject variable (footnote 1).

Allegedly, the visual overlap of target and prime is an essential condition for nonlexical facilitation. If the effects in repetition priming were predominantly episodic in origin in the sense that proponents of the episodic view have claimed, then two appearances of the same orthographic configuration in a repetition priming task should have facilitated recognition more than two presentations in different alphabet transcriptions. In the present experiment, it did not. The identity prime condition (D1-DD) produced 90 ms of facilitation both when the same visual pattern was repeated (by using Roman characters throughout, as in Experiment 1a) and when the visual pattern was not repeated (because primes were in Cyrillic print and targets were in Roman, as in Experiment 1b). Likewise for pseudowords, whether prime and target were alphabetically nonalternating or alternating, the effect of condition (D1, DN, DD, ID) was significant by both the subjects and the items analysis of variance. Moreover, the numerical difference between D1 and DD latencies was comparable in the nonalternating and the alternating alphabet conditions (32 ms vs 35 ms). With respect to both the order of magnitude of facilitation and the reduced facilitation relative to that observed with real word targets, these results with pseudowords are consistent with those reported in other repetition priming studies that introduce a comparable range of lags (Feldman & Fowler, in press). Although it cannot be visual in nature, the pseudoword results implicate a nonlexical source of facilitation.

Subjects in the alphabetically nonalternating condition tended to be faster overall than subjects in the alternating condition. Two plausible accounts are offered. As noted above, first-year students participated in the former and second-year students in the latter condition. Perhaps first-year students were more practiced at reaction time studies than their second-year counterparts because experimental participation is a requirement of the first year curriculum. Alternatively, the mixed alphabet design may have produced an overall slowing in reaction times relative to the pure alphabet design. The discrepancy due to alphabet makes a direct comparison of mean latencies across experiments 1a and 1b difficult to interpret (although contrasts within an experiment are not affected). Importantly, magnitude of facilitation was equal in alphabetically alternating and nonalternating contexts despite the tendency for slower targets to be facilitated more in variations of the present task (Forster & Davis, 1984).[2] Because the range of lags was binominally distributed in the test orders, no analysis of facilitation by lag was attempted. It is important to note that when such analyses have been reported for lexical decision with English materials and prime-target intervals of 0, 1, 3, and 10 items, the effect of lag is not significant (Napps, 1985). Nevertheless, significant facilitation has been demonstrated for alternating language conditions at an interval of 0 but not at intervals of 2 and 32 items (Kirsner, Smith, Lockhart, King, & Jain, 1984). It is conceivable that facilitation between visually discrepant prime-target pairs may vary with a more extensive range of lags.

Under both alternating and nonalternating alphabet conditions the facilitation of word targets by identical primes was significantly greater than by morphological primes whose affixes differed from the target affix (viz. instrumental primes for dative/locative targets). Fowler et al. (1985) have proposed that the full-partial distinction in magnitude of priming reflects the decreased contribution of an episodic factor, defined by letter overlap, in the morphologically-related prime condition relative to the identity prime condition. As long as "letter" is defined abstractly, the present results for words are consistent with that claim. The foregoing result was not significant for pseudowords, however. An alternative possibility, also suggested by Fowler, is that it reflects degree of association among words in the lexicon where morphological relatives are associated less closely than are words to themselves.

Insofar as changes in alphabet did not diminish the magnitude of facilitation, the basis for similarity must be more abstract than visual descriptors defined with respect to letter identity. In this respect, changes of alphabet appear similar to case alternations (Scarborough, Cortese, & Scarborough, 1977) and different from changes of language or modality (Kirsner et al., 1984; Scarborough, Gerard, & Cortese, 1984) in terms of the relevance of surface attributes (Jacoby & Brooks, 1984) or specificity of the representations that underlie facilitation in the present task (Kirsner & Dunn, 1985). In summary, one source of facilitation in repetition priming is necessarily more abstract than the surface characteristics of a visually-presented letter string, and this factor may apply to letter strings with or without a lexical representation. Based on other evidence about reading processes in Serbo-Croatian, it is proposed that this code may be phonological in nature although it must tolerate systematic phonological alternations as well (Feldman & Fowler, in press).

The outcome of Experiment 1 indicated that visual attributes of prime and target could not account for the pattern of facilitation. However, a morphological principle was not directly assessed. As noted above, all prime-target word pairs that were structurally similar necessarily shared a base morpheme. In Experiment 2, the sufficiency of nonlexical effects to account for the pattern of facilitation is investigated by comparing prime-target pairs that share extensive orthographic and phonological similarity with and without morphological relatedness.

Experiment 2

As noted above, the Serbo-Croatian language has a complex morphology that comprises many derived forms including diminutives, augmentatives, and agentives (see Table 3). Typically, derivatives are formed by appending an affix to the base form of a noun. So, for example, the word KORICA, which means thin crust, is derived from the word KORA, which means crust, and the word STANČIĆ, which means little apartment, is derived from the word STAN, which means apartment (these are feminine and masculine examples, respectively). The most common diminutive suffixes are (Č)IĆ, ICA, ENCE, and AK and, it is important to note, they are used productively in Serbo-Croatian. As contrasted with other derived forms that do not always respect the word class of their base word (e.g., agentives such as BAKER, a noun, derived from the verb BAKE), diminutives are more like inflections in that they entail only a slight alteration to the meaning of their base word. Thus, diminutives are classified as derivations of subjective judgment (Stevanović, 1979) and are considered almost as similar semantically to their base word as are inflections.

In the second experiment, nominative case base words were presented as targets in a repetition priming paradigm. Analogously to the first experiment, they were sometimes preceded by the identical word and sometimes by a morphologically related word, the diminutive of that word. In order to assess whether abstract letter or phonological similarity can account for the facilitation obtained in the repetition priming task, unrelated words that were orthographically and structurally similar to the word were also included as primes.

Table 3

Examples of Morphologically-Related Words Formed with the Base Morpheme "STAN"*

EXAMPLE	DERIVATIONAL PREFIX	BASE MORPHEME	DERIVATIONAL SUFFIX	INFLECTIONAL SUFFIX	MEANING
STAN		STAN			apartment
STANOVI		STAN		OVI	apartment (plural)
STANČIĆ		STAN	ČIĆ		small apartment
STANAR		STAN	AR		tenant
POSTANAR	PO	STAN	AR		subtenant
STANARINA		STAN	ARINA		rent

*words are in nominative singular unless otherwise noted

Primes that are orthographically similar but morphologically unrelated to the target have been presented previously in repetition priming studies conducted with English materials (Hanson & Wilkenfeld, 1986; Murrell & Morton, 1974; Napps, 1985). Typically, similarity is defined such that both the prime and target have the initial sequence of letters in common (e.g., RIBBON-RIB), but the extent of orthographic overlap is variable and the final portion of the longer word is essentially unconstrained. In the present study, each orthographically and structurally similar unrelated prime was a monomorphemic word in which the initial portion contained the full stem (base morpheme) and the final portion comprised one of the sequence of letters (viz., (Č)IĆ, ICA, ENCE, AK) that elsewhere forms the diminutive suffix (e.g., KORAK, STANICA). These words are termed pseudodiminutives. In this way, the structural similarity to target words of morphologically related and unrelated primes was maximized. By one account, lexical access time depends on the time to access the base morpheme in the internal lexicon and search is conducted from most to least common. (In this case decision latencies for pseudodiminutives might vary as a function of the frequency of the (inappropriate) base morpheme).

In summary, in Experiment 2 the facilitative effect of diminutive and pseudodiminutive primes on lexical decision latency to base target words was investigated. We asked whether the facilitative effect for words in the repetition priming task can be attributed solely to the orthographic and phonological similarity of prime and target or alternatively, whether it necessarily reflects morphological relatedness as well.

Methods

Subjects. Forty-five students enrolled in the Introductory Psychology course at the University of Belgrade participated in the experiment. They received course credit for their participation and all had prior experience in reaction-time tasks.

Stimulus materials. Twenty-four nouns containing four to six letters in their nominative singular form were selected so that they met two criteria. 1) Each noun permitted a diminutive derivation that included the stem (nominative for masculine words, nominative minus position final "A" for most feminine words) with no changes in segmental structure. 2) There existed a monomorphemic word that was orthographically similar to it in that the initial portion included the entire stem and the final portion included the sequences of letters (viz., (Č)IĆ, ICA, ENCE, AK) that elsewhere forms a diminutive suffix.

Consider the triples KORA, KORICA, and KORAK and STAN, STANČIĆ, STANICA. The first two members of each triple represent a nominative word and its diminutive derivation. They are orthographically and phonologically similar but morphologically unrelated to the last member of each triple, which is a pseudodiminutive. To reiterate, pseudodiminutive words are 1) morphologically unrelated to the target word, 2) monomorphemic in structure but 3) appear (inappropriately) to contain a diminutive affix. KORAK and STANICA mean step and station, respectively. The mean frequency (Lukić, 1970) for base, diminutive, and pseudodiminutive words was 329, 16.5, and 64, respectively.

Nominative pseudowords were constructed according to the criteria described in Experiment 1. Diminutive pseudowords were formed by adding a real diminutive affix (viz., (Č)IĆ, ICA, ENCE, AK) to a pseudoword base. Pseudodiminutive items were formed by adding meaningless affixes (i.e, TRA, IZO, ITRA, AT) to pseudoword bases.

Design. Three tests orders were created according to the constraints adopted in Experiment 1. Each was composed of 24 target words and 24 target pseudowords that were preceded seven to thirteen items earlier in the order by their prime. Equal numbers of (non-derived) nominatives, diminutives, and pseudodiminutives served as primes in each test order. Test orders were distinguished by the form of the primes: base, diminutive, or pseudodiminutive (indicated as B1, D1, or P1, respectively). For both words and pseudowords, items were always in nominative case. (The conditions of nominative case base word preceded by base, diminutive, and pseudodiminutive primes were indicated as BB, DB, and PB, respectively). As above, words appeared in the same serial position across all test orders and the form of the prime varied.

To summarize the experimental design, across test orders each target word in its base form was preceded by its prime in base, diminutive, or pseudodiminutive form. Within each order, a stem occurred once in a target and once in a prime and case of prime varied with item.

Results

Incorrect responses and extreme scores (greater than 1250 ms or less than 350 ms) were eliminated from all analyses. These criteria eliminated 2% of all responses. The error pairing procedure eliminated another 2% of all responses. Results of an analysis of variance on correct responses to target words (B1, BB, DB, PB) indicated a significant effect of condition by both the subjects F_1 and stimuli (F_2 in parentheses) analyses, $F_1(3,132) = 16.38$, MSe = 1377, $p < .001$ ($F_2(3,69) = 7.82$, MSe = 1538, $p < .001$). Inspection of mean latencies by condition, summarized in Table 4, indicated no facilitation for target words preceded by pseudodiminutives and significant facilitation for targets preceded by identical and diminutive primes. A protected t-test (Cohen & Cohen, 1975) indicated that targets preceded by identical primes were faster than by diminutive primes, $t(44) = 2.9$, $p < .01$.

Table 4

Mean Reactions Times (ms) to Base Word Targets (BB, DB, PB) as a Function of their Base (B1), Diminutive (D1) and Pseudodiminutive (P1) Primes

	CONDITION				FACILITATION	
	B1	BB	DB	PB		
PRIME:	STAN	STAN	STANČIĆ	STANICA	B1-BB	(B1-BB)/B1
TARGET:		STAN	STAN	STAN		
LEXICALITY						
WORDS	610	563	585	609	47	8%
PSEUDOWORDS	750	712	723	736	38	5%

Mean reaction time for word primes followed the pattern predicted by their respective frequencies such that base forms were faster than pseudodiminutives, which, in turn, were faster than diminutives. In order to explore the relationship between latencies for the three cases of prime, mean decision latency was computed for each word in its base, diminutive, and pseudodiminutive form and correlations were run on means for each pair of cases. For base-diminutive, base-pseudodiminutive, and diminutive-pseudodiminutive pairs, the correlations were $r = .30$, $r = -.01$, $r = -.08$, respectively. Finally, in order to determine whether frequency of the base form influenced decision latency for derived or pseudoderived forms, diminutives and pseudodiminutives were split dichotomously according to the frequency of their base form. Latencies for diminutives but not for pseudodiminutives followed base form frequency. Diminutive and pseudodiminutive reaction times for high and low frequency base words are reported in Table 5.

Table 5

Mean Reaction Times (ms)
to Base (B1), Diminutive (D1), and Pseudodimunitive
(P1) Primes as a Function of the Frequency of its Nominative Base Form

	CASE		
FREQUENCY OF NOMINATIVE	B1	D1	P1
	STAN	STANČIĆ	STANICA
HIGH	560	723	717
LOW	660	785	719

For pseudoword targets (B1, BB, DB, PB), the effect of condition was significant, $\underline{F}_1(3,132) = 7.41$, MSe = 1619, $\underline{p} < .001$ by the subjects analysis but not by the items analysis, $\underline{F}_2(3,92) = 1.47$, MSe = 4336, $\underline{p} < .23$. For pseudoword primes (B1, D1, P1), the effect of case was significant by the subjects analysis, $\underline{F}_1(2,88) = 8.98$, MSe = 2449, $\underline{p} < .001$ but not by the items analysis. Finally, none of the pseudoword correlations reached significance.

Discussion

The most important outcome of Experiment 2 was that significant facilitation occurred for words in a repetition priming paradigm only when prime and target were morphologically related. Elsewhere, the same pattern has been interpreted as reflecting, at least in part, a morphological principle of organization in the internal lexicon of the skilled reader (Feldman & Fowler, in press; Fowler et al., 1985; Hanson & Wilkenfeld, 1986; Kempley & Morton, 1982). Moreover, structural similarity of the initial and final portion of prime-target pairs was not sufficient to produce even partial facilitation among word pairs. Specifically, pseudodiminutive word primes produced no facilitation (1 ms) for structurally similar but morphologically-unrelated targets. This result is noteworthy because the composition of words that served as pseudodiminutive primes were visually quite similar and because they conceivably could have fostered a special strategy on the part of the subject such that subjects were able to predict upcoming targets. For example, given the constraints on selecting pseudodiminutive primes such as KORAK or STANICA, subjects could have anticipated KORA and STAN as targets and activated these lexical entries accordingly. This did not occur.

Although present numerically, effects of facilitation with pseudoword targets were absent statistically in Experiment 2 due to a failure to reach significance by the stimulus analysis. We chose, therefore, to interpret the results of the second experiment as failing to show facilitation with pseudowords although in the first experiment analogous effects with pseudowords were significant. It is possible, however, that the failure to reach significance reflects variability in the data due to a few atypical pseudowords. Significant facilitation with pseudoword targets has been reported previously (Feldman & Fowler, in press, Experiment 1) and was interpreted as episodic in origin.

It has been claimed that prior to lexical access, a reader tries to parse all potentially polymorphemic words into stem and affixes and that reaction time in a lexical decision task is largely a function of isolating and identifying the appropriate lexical unit (Taft, 1979; Taft & Forster, 1975). Inspection of prime latencies for pseudodiminutive words provided no evidence that subjects inappropriately parsed these forms into stem and diminutive prior to making a lexical judgment. First, latencies for these items followed the pattern predicted by their frequency (B1 <P1 <D1). Second, the data for pseudodiminutive words grouped by the frequency of their base forms had nearly equivalent means. Thus, they provide no evidence of slowing due to a frequency-sensitive search to the inappropriate nominative form in the course of lexical access.

The foregoing results are consistent with work conducted in English in that it is generally quite difficult to demonstrate evidence of an inappropriate morphemic parsing for real words. For this and related reasons Caramazza and Lukatela, among others, have suggested that a reader's appreciation of morphology is represented lexically. Caramazza modeled morphological structure in terms of a shared base morpheme (Caramazza, Miceli, Silveri, & Laudanna, 1985; cf. Burani, Salmaso, & Caramazza, 1984). Alternatively, Lukatela and his colleagues posited morphological relatedness as a principle of lexical organization among complete inflected case forms in the satellite entries model (Lukatela, Gligorijević, Kostić, & Turvey, 1980; Lukatela, Mandić, Gligorijević, Kostić, & Turvey, 1978; see also Feldman, Kostić, Lukatela, & Turvey, 1983). For the present purposes, it suffices to point out that morpheme parsing prior to lexical access is not the only way to capture a reader's appreciation of morphology.

Taken together, the results of the present experiments indicate that visual similarity of prime and target is not necessary to obtain full facilitation of targets in the repetition priming task and this outcome calls into question a simple episodic or perceptual fluency account of facilitation that is based on the preservation over successive presentations of attributes that are visually similar. Evidently, in order to tolerate changes across alphabet, a nonlexical basis of facilitation needs to be defined on an abstract structure. Moreover, the availability of lexical knowledge appears to govern the potential contribution of structural similarity. When lexical information is absent (namely, pseudoword prime-target pairs), structural similarity provides a sufficient condition for facilitation. In contrast, when lexical information is present (namely, real word prime target pairs), visual similarity is neither necessary nor sufficient.

In conclusion, nonlexical effects defined by structural similarity appear to contribute to the pattern of facilitation in the repetition priming task, but the adequacy of this account is contingent on the absence of lexical information. Generalizing over perceptual and memory tasks, we have borrowed the term episodic for this source of nonlexical facilitation although it might be claimed that the results with our alphabet manipulation critically alter the character of the episodic trace. When lexical information is available, however, structural characteristics figure only marginally. In conclusion, episodic effects cannot account for the facilitation of word targets in the repetition priming task.

In the discussion of the role of phonological analysis in lexical access, researchers currently focus on the time course or interaction rather than on the competition between phonological and lexical codes. We believe an analogous characterization applies to the role of episodic and lexical effects in the repetition priming task. Evidently, the reader can consider both nonlexical and lexical sources of similarity, but neither is sufficient in itself to accommodate the accumulated body of data. Ultimately, the key is to come to understand how they work together such that the availability of lexical knowledge may serve to mitigate the utility of other codes that may underlie facilitation in the present task.

References

Burani, C., Salmaso, D., & Caramazza, A. (1984). Morphological structure and lexical access. Visible Language, 18, 348-358.

Caramazza, A., Miceli, G., Silveri, M. C., & Laudanna, A. (1985). Reading mechanisms and the organization of the lexicon: Evidence from acquired dyslexia. Cognitive Neuropsychology, 2, 81-114.

Cohen, J., & Cohen, P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum.

Feldman, L. B., & Fowler, C. A. (in press). The inflected noun system in Serbo-Croatian: Lexical representation of morphological structure. Memory & Cognition.

Feldman, L. B., Kostić, A., Lukatela, G., & Turvey, M. T. (1983). An evaluation of the "Basic Orthographic Syllabic Structure" in a phonologically shallow orthography. Psychological Research, 45, 55-72.

Feldman, L. B., & Turvey, M. T. (1983). Morphological processes in word recognition. Paper presented at the Psychonomics Society Meeting.

Feldman, L. B., & Turvey, M. T. (1983). Visual word recognition in Serbo-Croatian is phonologically analytic. Journal of Experimental Psychology: Human Perception and Performance, 9, 288-298.

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. Journal of Experimental Psychology: General, 112, 309-346.

Forbach, G. B., Stanners, R. F., & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. Memory & Cognition, 2, 337-339.

Forster, K., & Davis, C. (1984). Repetition priming on a frequency attentuation in lexical access. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 680-698.

Fowler, C. A., Napps, S. E., & Feldman, L. B. (1985). Lexical entries are shared by regular, irregular, and morphologically-related words. Memory & Cognition, 13, 241-255.

Hanson, V. L., & Wilkenfeld, D. (1986). Morphophonology and lexical organization in deaf readers. Language and Speech, 28, 269-279.

Henderson, L., Wallis, J., & Knight, D. (1984). Morphemic structure and lexical access. In H. Bouma & D. Bouhuis (Eds.), Attention and performance, X (pp. 211-226). London: Lawrence Erlbaum.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception and concept learning. The Psychology of Learning and Motivation, 18, 1-46.
Johnston, W., Dark, V., & Jacoby, L. (1985). Perceptual fluency and recognition judgments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 3-11.
Kempley, S. T., & Morton, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. British Journal of Psychology, 73, 441-454.
Kirsner, K., & Dunn, J. (1985). The perceptual record: A common factor in repetition priming and attribute retention. In M. I. Posner & O. S. M. Marin (Eds.), Attention and performance XI (pp. 547-565). Hillsdale, NJ: Lawrence Erlbaum.
Kirsner, K., & Smith, M. C. (1974). Modality effects in word identification. Memory & Cognition, 2, 637-640.
Kirsner, K., & Smith, M. C., Lockhart, R. S., King, M. L., & Jain, (1984). The bilingual lexicon: Language-specific units in an integrated network. Journal of Verbal Learning and Verbal Behavior, 23, 519-539.
Kostić, Dj. (1965). Frequency of occurrence of words in Serbo-Croatian. Unpublished manuscript, Institute of Experimental Phonetics and Speech Pathology, University of Belgrade.
Lukatela, G., Gligorijević, B., Kostić, A., & Turvey, M. T. (1980). Representation of inflected nouns in the internal lexicon. Memory & Cognition, 8, 415-423.
Lukatela, G., Mandić, Z., Gligorijević, B., Kostić, A., Savić, M., & Turvey, M. T. (1978). Lexical decision for inflected nouns. Language and Speech, 21, 166-173.
Lukić, V. (1970). Active written vocabulary of pupils at the elementary school age (in Serbo-Croatian). Belgrade: Zavod za Izdavanje Udzbenika SR Srbije.
Monsell, S. (1985). Repetition and the lexicon. In A. W. Ellis (Ed.), Progress in the Psychology of Language (pp. 147-195). London: Lawrence Erlbaum.
Morton, J. (1969). The interaction of information in word recognition. Psychological Review, 76, 165-178.
Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P. A. Kolers, W. E. Wrolstad, & M. Bouma (Eds.), Processing visible language. New York: Plenum.
Morton, J. (1982). Disintegrating the lexicon: An information processing approach. In J. Mehler, E. Walker, & M. Garrett (Eds.), Perspectives on mental representation (pp. 89-109). Hillsdale, NJ: Lawrence Erlbaum.
Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. Journal of Experimental Psychology, 102, 963-968.
Napps, S. (1985). Morphological, semantic, and formal relations among words and the organization of the mental lexicon. Unpublished doctoral dissertation, Dartmouth College.
Napps, S., & Fowler, C. A. (Submitted for publication). The effect of orthography on the organization of the mental lexicon.
Napps, S., & Fowler, C. A. (1983). Orthographic organization of lexical forms. Paper presented at the Eastern Psychological Association Meeting.
Oliphant, G. (1983). Repetition and recency effects in lexical memory. Australian Journal of Psychology, 35, 393-403.
Ratcliff, R., Hockely, W., & McKoon, G. (1985). Components of activation: Repetition and priming effects in lexical decision and recognition. Journal of Experimental Psychology: General, 114, 435-450.

Scarborough, D. L., Cortese, C., & Scarborough, H. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 1-17.

Scarborough, D. L., Gerard, L. & Cortese, C. (1984). Independence of lexical access in bilingual word recognition. *Journal of Verbal Language and Verbal Behavior*, *23*, 84-99.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior*, *18*, 399-412.

Stevanović, M. (1983). *Grammar of the Serbo-Croatian Language* (in Serbo-Croatian). Cetinje: Išro Obod OOUR Izdavačka Delatnost SR Crna Gora.

Taft, M. (1979). Lexical access via an orthographic code: The Basic Orthographic Syllable Structure (BOSS). *Journal of Verbal Learning and Verbal Behavior*, *18*, 21-40.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*, 638-647.

Footnotes

[1]In order to insure that subjects from this population have balanced control of both alphabets, a different sample of 34 first-year students were asked to perform a lexical decision judgment on 24 words and 24 pseudowords selected from but not identical with the test materials presented in Experiment 1a and b. In this experiment, alphabet and lexicality were within-subject variables. That is, each subject saw words and pseudowords printed in both Roman and Cyrillic and across groups of subjects, each word and pseudoword appeared in both alphabetic transcriptions. Mean decision latency to words in their Roman and Cyrillic forms were 646 ms and 644 ms, respectively. For pseudowords, the latencies were 693 ms and 698 ms. The effect of alphabet did not approach significance for words, $F_1(1,33) = .08$, MSe = 792.98, $p < 1.0$ ($F_2(1,23) = .04$, MSe = 2516.14, $p < 1.0$) nor for pseudowords, $F_1(1,33) = .6$, MSe = 558.29, $p < 1.0$ ($F_2(1,23) = .11$, MSe = 3642.44, $p < 1.0$). This outcome supports the claim that our population of skilled readers of Serbo-Croatian are equally facile in Roman and Cyrillic and legitimizes the appropriateness of comparisons across alphabets.

[2]The foregoing claim is contingent on the appropriateness of the D1 baseline under alphabetically alternating conditions, which assumes that skilled readers are equally facile with both alphabets. Under conditions of dominance in either alphabet, an alternative comparison would be required.

PUBLICATIONS

APPENDIX

PUBLICATIONS

Baer, T., Gore, J. C., Boyce, S., & Nye, P. W. (in press). Application of MRI to the analysis of speech production. Magnetic Resonance Imaging.

Brady, S., Mann, V. A., & Schmidt, R. (in press). Errors in short-term memory for good and poor readers. Memory & Cognition.

Crain, S., & Shankweiler, D. (in press). Syntactic complexity and reading acquisition. In A. Davidson, G. Green, & G. Hermon (Eds.), Critical approaches to readability: Theoretical bases of linguistic complexity. Hillsdale, NJ: Erlbaum.

Feldman, L. B., & Fowler, C. A. (in press). The inflected noun system in Serbo-Croatian: Lexical representation of morphological structure. Memory & Cognition.

Feldman, L. B., & Moskovljević, J. (in press). Repetition priming is not purely episodic in origin. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Fowler, C. A. (in press). Perceivers as realists, talkers too: Commentary on papers by Strange, Diehl and Verbrugge, and Rakerd. Journal of Memory and Language.

Frost, R., Katz, L., & Bentin, S. (in press). Strategies for visual word recognition and orthographic depth: A multi-lingual comparison. Journal of Experimental Psychology: Human Perception and Performance.

Goldstein, L., & Browman, C. P. (1986). Representation of voicing contrasts using articulatory gestures. Journal of Phonetics, 14, 339-342.

Horiguchi, S., & Bell-Berti, F. (in press). The velotrace: A device for monitoring velar position. Cleft Palate Journal.

Katz, L., Boyce, S., Goldstein, L., & Lukatela, G. (in press). Grammatical information effects in auditory word recognition. Cognition.

Katz, R. B. (1986). Phonological deficiencies in children with reading disability: Evidence from an object-naming task. Cognition, 22, 225-257.

Kay, B., Kelso, J. A. S., Saltzman, E. L., Schöner, G. (in press). Space-time behavior of single and bi-manual rhythmical movements: Data and model. Journal of Experimental Psychology: Human Perception and Performance.

Kelso, J. A. S. (1986). Pattern formation in speech and limb movements involving many degrees of freedom. In H. Heuer & C. Fromm (Eds.), Generation and modulation of action patterns (Experimental Brain Research Series 15, pp. 105-128). Berlin: Springer-Verlag, 1986.

Kelso, J. A. S., Schöner, G., & Scholz, J. P. (in press). Phase-locked modes, phase transitions, and component oscillators. Physica Scripta.

Krakow, R. A., Beddor, P. J., Goldstein, L. M., & Fowler, C. A. (in press). Coarticulatory influences on the perceived height of nasal vowels. Journal of the Acoustical Society of America.

Lisker, L. (in press). "Voicing" in English: A catalog of acoustic features signaling /b/ versus /p/ in trochees. Language and Speech.

Lukatela, G., Carello, C., Kostić, A., & Turvey, M. T. (in press). Low constraint facilitation in lexical decision with single word contexts. American Journal of Psychology.

Lukatela, G., Carello, C., Savić, M., & Turvey, M. T. (1986). Hemispheric asymmetries in phonological processing. Neuropsychologia, 24, 341-350.

Lukatela, G., Carello, C., & Turvey, M. T. (in press). Lexical representation of regular and irregular inflected nouns. Language and Cognitive Processes.

Lukatela, G., Crain, S., & Shankweiler, D. (in press). Sensitivity to inflectional morphology in agrammatism: Investigation of a free-word-order language. Brain and Language.

Lukatela, G., Kostić, A., Todorović, D., Carello, C., & Turvey, M. T. (in press). Type and number of violations and the grammatical congruency effect in lexical decision. Psychological Research.

Lukatela, G., & Turvey, M. T. (in press). Loci of phonological effects in the lexical access of words written in a shallow orthography. Proceedings of the Inaugural Conference of European Cognitive Psychology (Nijmegen, Netherlands, September 9-12, 1985).

MacNeilage, P. F., Studdert-Kennedy, M., & Lindblom, B. (in press). Primate handedness reconsidered. The Behavioral and Brain Sciences.

Mann, V. A. (1986). Why some children encounter reading problems: The contribution of difficulties with language processing and phonological sophistication to early reading disability. In J. K. Torgeson & B. Y. L. Wong (Eds.), Psychological and educational perspectives on learning disabilities (pp. 133-159). Orlando, FL: Academic Press.

Mann, V. A. (in press). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listener's perception of [l] and [r]. Cognition.

Mattingly, I. G., & Liberman, A. M. (in press). Specialized perceiving systems for speech and other biologically significant sounds. In G. M. Edelman, W. E. Gall & W. M. Cowan (Eds.), Functions of the auditory system. New York: Wiley.

Napps, S. E., & Fowler, C. A. (in press). Formal relationships among words and the organization of the mental lexicon. Journal of Psycholinguistic Research.

Nittrouer, S., & Studdert-Kennedy, M. (1986). The stop-glide distinction: Acoustic analysis and perceptual effect of variation in syllable amplitude envelope for initial /b/ and /w/. Journal of the Acoustical Society of America, 80, 1026-1029.

Remez, R. E., Rubin, P. E., Nygaard, L. C., & Howell, W. A. (in press). Perceptual normalization of vowels produced by sinusoidal voices. Journal of Experimental Psychology: Human Perception and Performance.

Repp, B. H. (in press). The sound of two hands clapping: An exploratory study. Journal of the Acoustical Society of America.

Repp, B. H., & Williams, D. R. (in press). Categorical tendencies in imitating self-produced isolated vowels. Speech Communication.

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. In H. Heuer & C. Fromm (Eds.), Generation and modulation of action patterns (Experimental Brain Research Series 15, pp. 129-144). New York: Springer-Verlag.

Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. Cognitive Psychology, 18, 452-499.

Shankweiler, D., & Crain, S. (in press). Language mechanisms and reading disorder: A modular approach. Cognition.

Smith, S. T., Mann, V. A., & Shankweiler, D. (in press). Spoken sentence comprehension by good and poor readers: A study with the Token Test. Cortex.

Studdert-Kennedy, M. (1986). Some developments in research on language behavior. In N. J. Smelser & D. R. Gerstein (Eds.), Behavioral and social science: Fifty years of discovery (pp. 208-248). Washington: National Academy Press.

Studdert-Kennedy, M. (in press). The phoneme as a perceptuomotor structure. In A. Alpert, D. MacKay, W. Prinz, & E. Scheerer (Eds.), Language perception and production. London: Academic Press.

Studdert-Kennedy, M. (in press). Speech development. In G. Adelman (Ed.), Encyclopedia of neuroscience. Boston: Birkhauser Boston, Inc.

Thelen, E., Kelso, J. A. S., & Fogel, A. (in press). Self-organizing systems and infant motor development. Development Review.

Urošević, Z., Carello, C., Savić, M., Lukatela, G., & Turvey, M. T. (in press). Some word order effects in Serbo-Croat. Language and Speech.

Verbrugge, R. R., & Rakerd, B. (in press). Evidence of talker-independent information for vowels. Language and Speech.

Zinna, D., Liberman, I. Y., & Shankweiler, D. (1986). The development of children's sensitivity to factors influencing vowel reading. Reading Research Quarterly, XXI, 465-480.

APPENDIX

Status Report		DTIC	ERIC
SR-21/22	January - June 1970	AD 719382	ED 044-679
SR-23	July - September 1970	AD 723586	ED 052-654
SR-24	October - December 1970	AD 727616	ED 052-653
SR-25/26	January - June 1971	AD 730013	ED 056-560
SR-27	July - September 1971	AD 749339	ED 071-533
SR-28	October - December 1971	AD 742140	ED 061-837
SR-29/30	January - June 1972	AD 750001	ED 071-484
SR-31/32	July - December 1972	AD 757954	ED 077-285
SR-33	January - March 1973	AD 762373	ED 081-263
SR-34	April - June 1973	AD 766178	ED 081-295
SR-35/36	July - December 1973	AD 774799	ED 094-444
SR-37/38	January - June 1974	AD 783548	ED 094-445
SR-39/40	July - December 1974	AD A007342	ED 102-633
SR-41	January - March 1975	AD A013325	ED 109-722
SR-42/43	April - September 1975	AD A018369	ED 117-770
SR-44	October - December 1975	AD A023059	ED 119-273
SR-45/46	January - June 1976	AD A026196	ED 123-678
SR-47	July - September 1976	AD A031789	ED 128-870
SR-48	October - December 1976	AD A036735	ED 135-028
SR-49	January - March 1977	AD A041460	ED 141-864
SR-50	April - June 1977	AD A044820	ED 144-138
SR-51/52	July - December 1977	AD A049215	ED 147-892
SR-53	January - March 1978	AD A055853	ED 155-760
SR-54	April - June 1978	AD A067070	ED 161-096
SR-55/56	July - December 1978	AD A065575	ED 166-757
SR-57	January - March 1979	AD A083179	ED 170-823
SR-58	April - June 1979	AD A077663	ED 178-967
SR-59/60	July - December 1979	AD A082034	ED 181-525
SR-61	January - March 1980	AD A085320	ED 185-636
SR-62	April - June 1980	AD A095062	ED 196-099
SR-63/64	July - December 1980	AD A095860	ED 197-416
SR-65	January - March 1981	AD A099950	ED 201-022
SR-66	April - June 1981	AD A105090	ED 206-038
SR-67/68	July - December 1981	AD A111385	ED 212-010
SR-69	January - March 1982	AD A120819	ED 214-226
SR-70	April - June 1982	AD A119426	ED 219-834
SR-71/72	July - December 1982	AD A124596	ED 225-212
SR-73	January - March 1983	AD A129713	ED 229-816
SR-74/75	April - September 1983	AD A136416	ED 236-753
SR-76	October - December 1983	AD A140176	ED 241-973
SR-77/78	January - June 1984	AD A145585	ED 247-626
SR-79/80	July - December 1984	AD A151035	ED 252-907
SR-81	January - March 1985	AD A156294	ED 257-159
SR-82/83	April - September 1985	AD A165084	ED 266-508
SR-84	October-December 1985	AD A168819	ED 270-831
SR-85	January-March 1986	AD A173677	ED 274-022

Information on ordering any of these issues may be found on the following page.

**DTIC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
3900 Wheeler Avenue
Alexandria, VA 22304-5110

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Language Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER SR-86/87, 1986	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Haskins Laboratories Status Report on Speech Research		5. TYPE OF REPORT & PERIOD COVERED Continuation Report April-September, 1986
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Staff of Haskins Laboratories: Michael Studdert-Kennedy, President		8. CONTRACT OR GRANT NUMBER(s) HD-01994 NS13870 NO1-HD-5-2910 NS13617 RR-05596 NS18010
9. PERFORMING ORGANIZATION NAME AND ADDRESS Haskins Laboratories 270 Crown Street New Haven, CT 06511-6695		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS National Institutes of Health National Science Foundation		12. REPORT DATE September, 1986
		13. NUMBER OF PAGES 330
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) As above		15. SECURITY CLASS. (of this Report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

UNLIMITED: Contains no information not freely available to the general public. It is distributed primarily for library use.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

As above

18. SUPPLEMENTARY NOTES

N/A

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Speech Perception:
psychophysics, modules, biological significance, voicing acoustic features, trochees, imitation, isolated vowels, coarticulation, Catalan, Spanish, VCV sequences, aeroacoustic, phonation, theory

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report (1 January-31 March) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

- -The role of psychophysics in understanding speech perception
- -Specialized perceiving systems for speech and other biologically significant sounds
- -"Voicing" in English: A catalog of of acoustic features signaling /b/ versus /p/ in trochees
- -Categorical tendencies in imitating self-produced isolated vowels

8. Contract or Grant Numbers (Continued)

BNS-8111470
BNS-8520709
N00014-83-K-0083

19. Key Words (Continued)

Speech Articulation:
coarticulation

Motor Control:
clapping, hands, sound, pattern formation, limbs, articulators, degrees of freedom, rhythm, space-time

Reading:
language, mechanism, modular, syntactic complexity, acquisition, phonological coding, hearing readers, deaf readers, word recognition, orthographic depth, multilingual, strategies, inflection, nouns, Serbo-Croatian, morphological structure, lexical representation, priming

20. Abstract (Continued)

-An acoustic analysis of V-to-C and V-to-V: Coarticulatory effects in Catalan and Spanish VCV sequences
-The sound of two hands clapping: An exploratory study
-An aeroacoustics approach to phonation: Some experimental and theoretical observations
-Pattern formation in speech and limb movements involving many degrees of freedom
-The space-time behavior of single and bimanual rhythmical movements: Data and model
-Language mechanisms and reading disorder: A modular approach
-Syntactic complexity and reading acquisition
-Phonological coding in word reading: Evidence from hearing and deaf readers
-Strategies for visual word recognition and orthographical depth: A multi-lingual comparison
-The inflected noun system in Serbo-Croatian: Lexical representation of morphological structure
-Repetition priming is not purely episodic in origin