DOCUMENT RESUME

ED 266 508	CS 505 201
AUTHOR	Studdert-Kennedy, Michael, Ed.; 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, 1985.
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-82/83(1985)
PUB DATE	85
CONTRACT	NICHHD-N01-HD-5-2910; ONR-N00014-83-K-0083
GRANT	NICHHD-HD-01994; NICHHD-HD-16591; NIHBRS-RR-05596; NINCDS-NS-13617; NINCDS-NS-13870; NINCDS-NS-18010; NSF-BNS-8111470
NOTE	289p.
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/PC12 Plus Postage.
DESCRIPTORS	*Articulation (Speech); Cognitive Processes; *Communication Research; Language Acquisition; *Language Research; Linguistic Theory; *Morphology (Languages); Orthographic Symbols; Perceptual Development; Perceptual Motor Coordination; Phonemes; *Speech Communication; Spelling

ABSTRACT

One of a regular series on the status and progress of research into the nature of speech, instrumentation for its investigation, and practical applications, this report consists of 17 papers dealing with the following topics: (1) vagueness and fictions as cornerstones of a theory of perceiving and acting--a commentary on D.O. Walter; (2) the informational support for upright stance; (3) determining the extent of coarticulation--effects of experimental design; (4) the roles of phoneme frequency, similarity, and availability in the experimental elicitation of speech errors; (5) on learning to speak; (6) the motor theory of speech perception revised; (7) linguistic and acoustic correlates of the perceptual structure found in an individual differences scaling study of vowels; (8) perceptual coherence of speech--stability of silence-cued stop consonants; (9) development of the speech perceptumotor system; (10) dependence of reading on orthography--investigations in Serbo-Croatian; (11) the relationship between knowledge of derivational morphology and spelling ability in fourth, sixth, and eighth graders; (12) relations among regular and irregular, morphologically related words in the lexicon as revealed by repetition priming; (13) grammatical priming of inflected nouns by the gender of possessive adjectives; (14) grammatical priming of inflected nouns by inflected adjectives; (15) deaf signers and serial recall in the visual modality--memory for signs, fingerspelling, and nt; (16) did orthographies evolve? and (17) the development of ldren's sensitivity to factors influencing vowel reading. (HOD)

SR-82/83 (1985)

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 1985

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 numbers of previous Status Reports.)

Michael Studdert-Kennedy, Editor-in-Chief
Nancy O'Brien, Editor
Gail Reynolds, Technical Coordinator

ACKNOWLEDGMENTS

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

National Institute of Child and Health and Human Development
Grant HD-01994
Grant HD-16591

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

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*
Thomas Baer
Patrice S. Beddor+
Fredericka Bell-Berti*
Catherine Best*
Geoffrey Bingham+
Gloria J. Borden*
Susan Brady*
Catherine P. Browman
Franklin S. Cooper*
Stephen Crain*
Robert Crowder*
Laurie B. Feldman*
Anne Fowler+
Carol A. Fowler*
Louis Goldstein*
Vicki L. Hanson
Katherine S. Harris*
Sarah Hawkins++
Satoshi Horiguchi[1]
Leonard Katz*
J. A. Scott Kelso
Gary Kidd+
Ascher Laufer[2]
Andrea G. Levitt*
Alvin M. Liberman*
Isabelle Y. Liberman*
Leigh Lisker*
Virginia Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Richard McGowan
Kevin G. Munhall+++
Susan Nittrouer++
Patrick W. Nye
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Mary Smith
Michael Studdert-Kennedy*
Betty Tuller*
Michael T. Turvey*
Ben C. Watson++
Douglas H. Whalen

Technical/Support

Michael Anstett
Philip Chagnon
Alice Dadourian
Vincent Gulisano
Donald Hailey
Raymond C. Huey*
Sabina D. Koroluk
Bruce Martin
Frank Merewether
Betty J. Myers
Nancy O'Brien
Gail K. Reynolds
William P. Scully
Richard S. Sharkany
Edward R. Wiley

Students*

Joy Armson
Dragana Barac
Sara Basson
Eric Bateson
Suzanne Boyce
Teresa Clifford
Andre Cooper
Margaret Dunn
Jan Edwards
Jo Estill
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
Robin Seider
Suzanne Smith
Katyanee Svastikula
David Williams

---

*Part-time
[1] Visiting from University of Tokyo, Japan
[2] The Hebrew University of Jerusalem
+NIH Research Fellow
++NRSA Training Fellow
+++Natural Sciences and Engineering Research Council of Canada Fellow

Status Report on Speech Research

Haskins Laboratories

CONTENTS

ON VAGUENESS AND FICTIONS AS CORNERSTONES OF A THEORY OF PERCEIVING AND ACTING: A COMMENT ON WALTER (1983)*

Claudia Carello† and M. T. Turvey††

"I don't want realism. I want magic!"

Blanche DuBois, Scene 9, A Streetcar Named Desire

Vagueness or unclarity of thought is considered by Walter (1983) as a worthy and necessary state of (human) mind for modeling. He appeals to quantum mechanics (and, in particular, non-pure states) as, perhaps, the only fruitful model by which to understand such phenomena. The analogy takes the following form: The clarity that indeterminant ideas derive from rumination and discussion parallels the reduction of uncertainty in a parameter of a submicroscopic system that accompanies its quantum measurement. Walter suggests that with an allowance for quantum-like brain states, brains can be classified as physical symbol systems--processors that read, write, store, and compare symbols--of the type described by Newell and Simon (Newell, 1981; Newell & Simon, 1976; Simon, 1981).

As a revealing aside (developed more fully in Walter, 1980), Walter (1983) asserts that both scientists' theorizing about perceiving and animals' perceiving are largely story-telling. His implication seems to be that we invent fictions that may or may not pertain to what is _really_ going on but, at least, help us muddle through our laboratories and our environments. Scientists fashion explanations (in a manner of speaking) in an attempt to sort out reaction times, thresholds, and so on, while perceivers contrive hypotheses to sort out patches of color, horizontal lines, and so on. The story's relation to reality is inconsequential as long as it is useful, where useful seems to be read as leading to the next (preferably consistent) fiction. If a fiction loses its usefulness to scientist or perceiver, it can be replaced with a new one--no more real but, ideally, more useful.

As he rightly points out, Walter's position is in conflict with ecological realism. Beyond that assessment, however, whatever it is that Walter describes as ecological realism bears little resemblance to the framework carved out by Gibson over some 30 years (e.g., Gibson, 1966, 1979, 1982) and elaborated by others (e.g., Michaels & Carello, 1981; Reed & Jones, 1982; Shaw & Turvey, 1982; Shaw, Turvey, & Mace, 1983; Turvey & Carello, 1981; Turvey, Shaw, Reed, & Mace, 1981). In what follows, we shall point out where Walter

---

*Cognition and Brain Theory, 1984, 7, 247-261.
†State University of New York at Binghamton
††Also University of Connecticut
Acknowledgment. The writing of this paper was supported in part by Office of Naval Research contract N0014-83-0083 awarded to Haskins Laboratories.

missteps in his treatment of realism, clarify our conflict with his strategy, and elaborate our own strategy for modeling behavior at the ecological scale of animal-environment systems (see below). In so doing, we shall attempt to show that Walter's posture on realism, while understandable in the beleaguered heroine of Tennessee Williams' play, is less sympathetic in a (reasonably content) scientist.

Alternative or Contradictory Descriptions Do Not Deny Realism

While Walter's discontent with ecological realism includes our neglect of quantum-like brain phenomena, he sees the existence of fictions--be they scientists' oft-changing models of the world or animals' deceptive behavior in times of danger or play--as a more fundamental difficulty because they belie the claim that reality can be apprehended.

> The pervasiveness of fictions, deception, play, and so on, make the whole ideology of "realism" seem rather unlikely to me, as a productive model for mammalian nervous systems. A notion of useful fictions ("useful" perhaps to be defined in neo-Darwinian terms) seems more likely than either ecological, or naive, realism, to yield an adequate description of this most complicated organ system. (p. 233)

Not surprisingly, we do not agree with this evaluation of the ramifications of such phenomena. First, dubbing them "fictions" is inaccurate and misleading. And, second, it is unlikely that fictions, with the suggestion that the attainment of goals is accidental, could ever be reliably useful. Let us elaborate this argument.

The notion that science engages in the fabrication of useful fictions has a parallel in legal practice (Walter, 1980). Just as it is convenient but incorrect to conceive of a corporation as a single person in certain legal circumstances so, too, is it useful but fictitious to conceive of space as Euclidean in some circumstances and curved in others. Walter claims that science would be better served by acknowledging that its models, however useful, are fictions "because the inconsistencies between scientific views of 'reality' in different contexts will be more damaging" (Walter, 1980, p. R366).

But do the seeming contradictions entailed by different characterizations of space, for example, remove all characterizations from the realm of reality (unqualified by quotation marks)? In other words, if a given notion changes relative to changes in the problem of interest, does this relativity preclude a consideration of that notion as objective and real? We have argued elsewhere that it does not and, indeed, that the concept of an absolute reality that would be appropriate for all grains of analysis is untenable (Gibson, 1979; Michaels & Carello, 1981; Shaw, Turvey, & Mace, 1982; cf. Prigogine & Stengers, 1984, chap. 7).

Appropriateness is the key idea here--the level of description of reality must be commensurate with the level of inquiry, that is, with the type of systemic interactions that are of interest (cf. Rosen, 1978). Although Walter (1980) says, "When making human-scale measurements, for example, precision seldom requires us to incorporate either relativistic space curvature or super-spacelike microtopological fluctuations" (p. R367), it is not disembodied "precision" that renders such analyses unnecessary. Rather, those analyses are inappropriate because human activities do not occur at those levels. Human (and animal) behavior occurs with reference to the animal-specific, activity-relevant properties of the environment--what Gibson has termed affordances (1979). Affordances, it is proposed, are the appropriate level of description of reality for the ecological scale. The lengthy, difficult search initiated by Grinnel (1917) and Elton (1927) to find a systematic and evolutionarily consistent way to define the econiche--the related environmental realities supporting a given species' lifestyle--has begun to focus on the view of the econiche as an affordance structure (Alley, in press; Patten, 1982).

Affordances are both relative--they are defined with reference to a particular animal--and objective--they are defined by persisting properties of the environment. As an example, consider a brink in a surface. For an animal of a given size, that brink affords stepping down; for an animal of a given smaller size, that brink affords falling off. The reality of that particular layout of surfaces as a step-down place or a falling-off place is relative to the animal. Yet the nature of those relative realities is determined by the independent character of the surface layout--for example, that it is comprised of vertically separated substantial surfaces rather than liquid ones. This echoes a point made by Lewis (1929):

> Relativity is not incompatible with, but requires, an independent character in what is thus relative. And second, though what is thus relative cannot be known apart from such relation ... all such relative knowledge is true knowledge of that independent character which, together with the other term or terms of this relationship, determines this content of our relative knowledge. (pp. 172-173)

The coexistence of contradictory descriptions of reality (e.g., step-downable vs. not step-downable, curved vs. Euclidean space) does not mean that these descriptions are fictions (cf. Ben-Zeev, in press). It simply means that different problems appeal to different aspects of reality. No one description is universally privileged (cf. Alley, in press; Rosen, 1978). Indeed, contrary to Walter's efforts to marshal quantum phenomena in opposition to realism, the same point has been made for that domain by Prigogine and Stengers (1984):

> The irreducible plurality of perspectives on the same reality expresses the impossibility of a divine point of view from which the whole of reality is visible (p. 224). The real lesson to be learned from the principle of complementarity [italics added] a lesson that can perhaps be transferred to other fields of knowledge, consists in emphasizing the wealth of reality, which overflows any single language, any single logical structure. (p. 225)

Biased by his concern about what scientists do when they theorize about the world, Walter is confused in his attitude toward what animals (including humans) do when they perceive their environments. He claims that the fictions by which scientists think they understand the universe have parallels in those cases where perceivers are duped by deceptions. We have already argued that scientific models of natural phenomena need not be considered fictions, even if models of the same phenomenon at different levels are inconsistent. But surely there are scientific models that are just plain wrong--phlogiston, aether, and spontaneous generation, to name a few. Do these speak to the possibility of perceivers knowing reality? They do not because they involve issues of scientific realism, not perceptual realism (see Blackmore, 1979). That is to say, the question of whether or not scientists can be successful in understanding nature is independent of whether or not perceivers are successful in knowing the environment as it constrains their day-to-day activities. Scientists can flounder for any number of reasons--religious dogma bad experiments, stupidity--but for animals to "move so they can eat, and eat so they can move" (Iberall, 1974) and thereby survive, they must be in contact with the facts of their environments. Animals cannot act effectively with respect to fictions.

What of Walter's contention that the fictions are useful? Doesn't that empower them to guide activity? It is not at all clear how a fiction, unfettered as it is by actual states of affairs, could ever be useful. What guides the construction of a fiction so that it is at least relevant to an intended action--for example, a given layout of surfaces is fictionalized as being in the realm of stepping (on) or falling (off) rather than swimming (in), squeezing, eating, ad infinitum? And by what criterion might a given fiction be deemed useful? There must be some standard of comparison. If the actual state of affairs provides the comparison, realism cannot be avoided.

Deception Presupposes Realism

Walter's example of deceptive animal behavior might seem tailor-made for a fiction framework. A mother bird saves her offspring by feigning injury so that a fox will follow and attack her in the mistaken belief that her broken wing will prevent her escape. She has created a fiction--the predator perceives an injury that does not exist--that is useful in preserving her species. Such circumstances are quite rare in nature, however; not all animals engage in deception, and, for those that do, deception constitutes a small part of their behavioral repertoires. Deception provides a disputable foundation, therefore, upon which to build an account of perceiving. Nonetheless, we would emphasize the lawful basis that allows the mother to enact a successful charade and the fox to act upon it. She must constrain her musculature in just that way that will produce postural and joint adjustments specific to a particular dynamic condition (viz., material structure too weak to support the characteristic wing movement). For his part, the fox must detect the dynamics that underlie the bird's kinematic display. In order to pursue a realist basis for deceptive behavior, we will elaborate this so-called kinematic specification of dynamics (or KSD) principle (Runeson, 1977/1983; Runeson & Frykholm, 1983).

The principle starts with the reasonable assumption that, because the body is composed of certain masses and lengths and types of joints, only certain movements will be biomechanically possible. The biomechanics will also determine what one must do to maintain balance and cope with reactive forces (those "back-generated" by the act of moving). The kinematic properties of an action (its variously directed motions, its accelerations and decelerations) are determined by the dynamic conditions that underlie it--the forces produced intentionally and unintentionally by the animal and those supplied by the surrounding surfaces of support. The KSD principle suggests that a reciprocal relationship also exists: The kinematic properties of acts are transparent to the dynamic properties that caused them. For an observer, this principle reads: The ambient optic array (see Gibson, 1979; Lee, 1974, 1976) is structured by an animal's movements such that macroscopic qualitative properties of the optic array are specific to and, therefore, information about, the forces that produced the movements.

The principle finds support in experimental investigations of human movement perception that use Johansson's (1973) patch-light technique. This methodology entails limiting an observer's view of actors (i.e., people who engage in activities) to small lights that are attached to their major joints. When a person engages in some activity, a transforming pattern of lights is generated. Perceivers find this limited optical structure to be informative about a number of properties, including metrical (length of throw of an invisible thrown object of unknown mass [Runeson & Frykholm, 1983]), biomechanic (gender of a walker [Cutting, Proffitt, & Kozlowski, 1978; Kozlowski & Cutting, 1977; Runeson & Frykholm, 1983]), and kinetic (the weight of a lifted box [Runeson & Frykholm, 1981]). Importantly, Runeson and Frykholm (1983) have shown that perceivers are not easily fooled by actors' efforts to be deceptive. Despite attempts to fake the weight of a lifted box, observers not only perceive the real weight but are aware of the deceptive intention and the intended deception (i.e., what weight is being faked) as well. Similar results are found in attempts to be deceptive about one's gender (through gait and carriage in a variety of actions)--observers are aware of both real gender and faked gender. The point to be underscored is that an actor can structure light in ways that provide information about conditions that do not exist (see Gibson, 1966; Michaels & Carello, 1981; Turvey et al., 1981, for realist accounts of this fact) while simultaneously (and unavoidably) providing information about conditions that do exist, and perceivers can be aware of both.

Runeson and Frykholm draw a parallel with the dual reality of pictures, especially as it has been described by Gibson: There is information about objects represented in the picture and information about the picture itself as an object. "The duality of information in the array is what causes the dual experience" (Gibson, 1979, p. 283). The possibility of dual awareness may speak to the dearth of true deceptions in nature. For very sound physical reasons, situations that lend themselves to single awareness deception are, contrary to what Walter seems to imply, difficult to manufacture and, in consequence, quite rare. Intraspecific threat and play behavior, on the other hand, are found throughout the animal kingdom. But it seems to be a misnomer to label these "deceptions" in the sense of trickery. Baboons who bare their teeth have not fabricated a fearsome weapon. They are suggesting that they would rather not use the ones they have. Chimpanzees who play attack-and-flee are not deluded; they behave differently in true fight-escape circumstances (Loisos, 1969). Play provides an opportunity to learn about one's environment, conspecifics, and one's own behavioral possibilities.

We have argued that characterizing perception as useful fictions is inadequate to explain behavior in natural circumstances. An explanation of effective behavior requires a realist framework with the animal-environment system as the unit of analysis. Walter, however, is skeptical of whether such an analysis is possible. We contend that his objection is based on an overevaluation of what can be distilled from brain state accounts and a misunderstanding of what "animal-environment system" means. We will deal with each of these issues in the next two sections.

Brain States Are An Inadequate Basis For Ascribing Intentional Content

Walter implies that any perspective that does not advert to observations of brain states cannot provide a dynamically useful formulation of behavior. However, he prudently avoids any discussion of how observations of brain states would yield the proposed useful formulation. Presumably, Walter's advocated observations or measurements of the brain--no matter how precise or vague those measurements may be--would provide only extensional descriptions. And, presumably, a physical or biological theory of the brain strictly consistent with such observations could only be extensional. At best, observations of brain states, purely interpreted, would lead to an account roughly of the form: In the context of functional brain organizations P and Q, functional brain organization R has the capacity of inducing functional brain organization S. This would not be a dynamically useful formulation of behavior. No matter how elaborate and detailed such an extensional account becomes, it will never allow Walter to answer apparently straightforward questions about prosaic behaviors. For example, how does an outfielder know to charge in rather than retreat to catch a ball (Todd, 1981)? Why does a child, on seeing a particular surface, initiate crawling rather than walking to traverse the surface (E. Gibson, 1983)? The important ingredient missing from the foregoing brain-state based account of behavior is intentionality.

A dynamically useful formulation of behavior grounded in observations of brain states requires minimally (1) a principled basis for individuating brain states, and (2) a principled basis for ascribing content to individuated brain states. The latter refers to the problem of systematically upgrading the extensional characterizations of brain states to intentional characterizations, ordinarily expressed by intensional statements (Dennet, 1969; Fodor, 1981; but see Searle, 1983). The point is that without identifying the contents (the significances, the meanings, the message functions, the signalling functions, etc.) of brain states, the brain theorist's view of brain function in relation to behavior is empty. The intentional characterization earns for the brain theorist the luxury of addressing the question of what the brain states are <u>about</u>. From what observations and on what grounds would an advocate of the explanatory power of brain states fashion intentional characterizations? Those characterizations arise at and are the sine qua non of the ecological scale of animal-environment systems.

Intentional characterizations should not be interpreted as referring to systemic states that are in addition to or separate from those extensionally characterized. Intentional characterizations usually comprise <u>alternative</u> (discrete, symbolic) descriptions of a system's states, descriptions that complement the extensional (continuous, dynamical) accounts of <u>how</u> a system is doing what it is doing. Pattee (e.g., 1973, 1977) has been foremost in identifying the problem of understanding how these two complementary modes of description of any complex system can be treated in a physically consistent way. The ecological approach to perception and action has been concerned similarly with the complementarity of intentional and extensional characterizations (e.g., Carello, Turvey, Kugler, & Shaw, 1984), but it has been concerned more directly with elaborating the extensional basis for ascribing intentionality to states of the animal-environment system in a principled manner (e.g., Gibson, 1979; Kugler, Kelso, & Turvey, 1980, 1982; Turvey et al., 1981). This strategy has been chosen because the principled ascription of content to the states of a system rests ultimately on the accuracy and specific predictions of the extensional account of the system. As Dennett (1969) puts it:

> The ascription of content is thus always an ex post facto step, and the traffic between the extensional and the intentional levels of explanation is all in one direction. (p. 86)

To the extent that the extensional basis for a system's phenomena is underestimated and/or unknown, the intentional characterization of the system is likely to be ungrounded and fatuous; ordinary systemic states get ascribed near magical functions or powers (section below). And this latter statement identifies, in a nutshell, the danger and inadequacy of seeking an account of behavior, as Walter advocates, in observations limited to brain states.

The Animal-Environment System as the Appropriate Unit of Analysis

Walter focuses his attack on realism on Turvey and Carello (1981). He discusses the position thusly:

> This position claims that the joint situation of an organism and its environment is the only correct fundamental concept for brain/mind modeling...I regard their presumption that a state of the brain-and-environment nexus can be observed as a fatal flaw in ecological realism. In my view, the state of a mammal's brain cannot, in most situations, usefully be observed...without so severely interfering with that state, by your observing...that the state will change in an unpredictable and uncontrollable way.... (p. 231)

Interestingly, the word "brain" never appears in the Turvey and Carello manuscript. Indeed, eschewing brains as the appropriate entities to model for an understanding of psychological phenomena is at the heart of using ecological to modify our brand of realism. We are interested in how organisms (including humans) are able to perceive their propertied environments in a way that will allow them to behave effectively with respect to those environments. A runner--be it human, gnu, or cockroach--does not steer around representations or brain states; it avoids real obstacles and goes through real openings. Couching problems in such terms is not, as Walter claims, simply a "programmatic and descriptive phase" that ecological realism is going through. The "dynamically useful formulation of behavior" that Walter asserts is unavailable from our strategy not only is found in a realist framework but, we would argue, can only be provided by such a perspective. One of Gibson's favorite examples--the problem of controlled collisions in locomotion--will be used to buttress this argument.

As an animal moves through a cluttered surround, it sometimes steers around objects, sometimes contacts them gently, and sometimes collides with them violently. In order to control encounters with the environment, activity-relevant (dynamically useful) information must be available. This includes information specific to what is moving (e.g., the animal or the objects that surround it), direction of locomotion, obstacles and apertures in one's path, time to contact (if it should occur), and force of contact (if it should occur). This information has been demonstrated by a number of investigators (e.g., E. Gibson, 1983; J. Gibson, 1979; Lee, 1976, 1980; Lishman & Lee, 1973; Schiff, 1965) to exist in what might be termed the morphology of the optic flow field (Kugler, 1983; Kugler & Turvey, in press; Solomon, Carello, & Turvey, 1984). We will highlight some of the findings here but for detailed analyses, the reader should refer to the cited works.

Although the problem of distinguishing one's own movement from displacements of the surround has been a long-standing puzzle in orthodox accounts of perceiving, Gibson (1979) provided a simple solution, viz., global, smooth change in the optic array specifies egomotion, local discontinuous change specifies motion of an object in the environment. Moreover, one's direction of locomotion is also specified by the form of the optic flow field: Global optical expansion specifies forward movement (where the focus of expansion specifies the point toward which one is moving) while global optical contraction specifies retreat (where the focus of contraction specifies the point from which one is moving). If the appropriate flow fields are generated, the appropriate actions will be constrained (e.g., in the face of simulated global optical expansion, a person will make postural adjustments backward to compensate for the perceived forward movement [Lishman & Lee, 1973]; when confronted with local optical expansion, a person [or animal] will duck [Schiff, 1965]). The same sort of analysis distinguishes obstacles from apertures: A closed contour is specified as an obstacle when there is a loss of structure outside the contour during approach; it is specified as an opening when there is a gain of structure inside the contour during approach (J. Gibson, 1979). Infants as young as six months will duck from approaching obstacles but try to look inside approaching openings (E. Gibson, 1983).

If an animal wishes to steer around objects, it must move in such a way that optical expansion is centered in openings rather than obstacles. In order to contact objects (and to vary the force with which they are contacted), two more optical flow properties are needed. The inverse of the relative rate of dilation of a topologically closed region of the optical flow field (e.g., that structured by a wall) specifies the time at which a moving animal will contact that region. The derivative of the time-to-contact variable is information about the imminent momentum exchange: If it is greater than a certain critical value, the animal will stop short of contact; if it is equal to that critical value, the contact will be soft; if it is less than that critical value, there will be a momentum exchange and the contact will be hard (Kugler, Turvey, Carello, & Shaw, 1984; Lee, 1976, 1980).

Notice that these properties do not exist in the animal or in the environment but *are only defined for the animal-environment system*. The components of the system are not ruled by the indeterminacy that governs conjugate variables in quantum mechanics. That is to say, an exact description of one component does not mean that the other component cannot be determined. On the contrary, measuring one of the components in isolation not only fails to provide an understanding of the system but gives a misleading picture of the component that is being measured. This is the problem of overdecomposing a partial system from the total system that includes it (Turvey & Shaw, 1979; cf. Ashby, 1963; Humphrey, 1933; Weiss, 1969). Although science requires decomposition to a certain extent in order to make its problems manageable, the parsing of systems cannot be done cavalierly. An unprincipled selection of a system in which a phenomenon is thought to reside may make the phenomenon appear capricious and compel the scientist to attribute magical powers or content to the partial system (Ashby, 1963; Turvey & Shaw, 1979). The appropriate grain of analysis, however, may reveal the law-governed determinacy that is unavailable in the partial system (Weiss, 1969).

For example, if we take a climber-stairway system (Warren, 1984) as an instance of an animal-environment system, several points can be illustrated. First, there is optical information for a category boundary for action--perceivers can see which of a variety of stairways (constructed with risers of varying heights) are climbable in the normal way (i.e., without using hands or knees). Second, there is a perceptual preference for stairways that would be easiest to climb (as determined by measures of energy expenditure during climbing). Third, both of these relationships can be described by a method of intrinsic measurement, in which one part of given system (e.g., on the animal side) acts as a natural standard against which a reciprocal part of the system (e.g., on the environment side) can be measured (Warren, in press; Warren & Shaw, 1981; cf. Bunge, 1973; Gibson, 1979). Thus, the critical riser height/leg ratio, indexing the action boundary, is .89 whereas the optimal ratio, indexing minimum energy expenditure, is .26. These ratios are the same for all climbers, short and tall. Finally, each of these ratios is a measure of animal-environment fit; each is an index of the state of that system. Notice that, unlike Walter's quantum systems, the state does not change by measuring it and predictions are not invalidated by observations. For a given individual, if the ratio of riser height to leg is less than or equal to .89, the stair will be climbable; if the ratio equals .26, that stair will be (relatively) energetically cheap to climb. Those relationships do not change. And nowhere in this analysis is it suggested that brain states can be or ought to be observed.

Brainstates Are Not the Touchstone for Theories of Knowing

Walter would not deny that behaviors like stairclimbing are observable without interference from the observer but he would, no doubt, claim that they are not useful or worthwhile to model.

> I have (Walter, 1980) characterized those aspects of behavior that are predictable from less severely interfering observations, as rather gross and physicalistic (contrasted with "psychodynamic"); they seem to obey a correspondence principle or classical limit. They also tend toward conspiring to give a systematically misleading impression...that they are a closed system, adequate to describe the brain. (pp. 231-232)

Though "gross" may be used pejoratively, perceiving and acting are unabashedly macrophenomena. Walter's implication that the only interesting behavior is a microbehavior will sever him from consideration of a gannet's dive for a fish (Lee & Reddish, 1981), the baseball fielder's catch of a deep fly ball (Solomon, Carello, & Turvey, 1984; Todd, 1981), and his own efforts to avoid destruction on the San Diego Freeway (Gibson & Crooks, 1938). While microphenomena may have their place, that place is not a privileged one. They need not and will not serve all of science. Once again, this attitude is not idiosyncratic to ecological realists. Rosen (1978), for example, in stressing the functional and organizational character of certain physical systems, observed:

> what seemed to be emerging from such considerations was apparently the antithesis of the reductionist program: instead of a single ultimate set of analytic units sufficient for the resolution of any problem, we find that distinct kinds of interactions between systems determine new classes of analytic units, or subsystems, that are ap-

> propriate to the study of that interaction. (p. xvi) [These] families of analytic units, all of which are equally "real" [are] entitled to be treated on the same footing; the appropriate use of natural interactions can enormously extend the class of physical observables [italics added] accessible to us.... (p. xvii)

Once again we see the theme of appropriate levels of reality, this time directed at the question of what counts as an observable for physics.

We suspect that Walter would not be sympathetic to the above line of argument, countering that we ought to focus on what qualifies as a legitimate observable for psychology, instead of physics, for problems of knowing. This is apparent in his contrasting "physicalistic" with "psychodynamic" aspects of behavior, charging that the former are not "adequate to describe the brain." This is where his emphasis on vague states of human mind during thinking, rumination, and the like clashes most dramatically with our concern for the very unvague states of animal-environment systems during perceiving and acting. In his desire to understand brain (as the seat of mind), Walter holds thinking and, in particular, vague thinking as the focus of any theory of epistemic agents. But for us, reliable and reproducible behaviors must be the touchstone for any account of knowing. In infinitely varying settings, organisms are able to produce the same appropriate behavior consistently, adapting it to the particular circumstances. For example, countless times a day a bird will take off from a variety of surfaces of support at a wide range of heights and fly toward other surfaces of support at varying distances away, alighting on them gently. Sometimes it will steer around trees or pet cats and sometimes it will have a direct flight. Obstacles to and paths for locomotion and the appropriateness of accelerations and decelerations can be neither indistinctly specified in optical flow fields nor unreliably detected if the bird is to locomote through its cluttered terrain successfully. It is these kinds of behaviors, not indeterminate contemplations, that should provide the standard against which to judge the adequacy of theories of knowing.

The example of a bird in flight is an important one because it contains one feature--collisions with plate glass windows--of the sort that Walter, among others, uses to try to refute realism. The style of the argument can be characterized as follows: A bird who sees the window as an opening and flies into it has not perceived reality correctly and has not acted effectively. But in situations of so-called perceptual "mistakes," we embrace the distinction drawn by Lewis (1929)--ignorance of reality is not to be equated with erroneous knowledge of reality. A window does not structure the optic array at all points of observation so as to specify the substantiality of the transparent surface. The bird is ignorant of that aspect of reality because information about that aspect is not available to those points of observation along the bird's approach. Information about substantiality is available, however, to other points of observation, viz., on those paths where the optic array is structured by more reflective angles of the glass. When information about an obstacle to locomotion is not available, a bird will not change its path of locomotion. Perception in the first case is veridical; perception in the second case is "veridical but partial" (Lewis, 1929, p. 176).

A Final Note

The ecological approach addresses common behaviors under the general rubric of controlled collisions (Kugler et al., 1984) or controlled encounters

(Gibson, 1979). Such behaviors cut across species and allow us to highlight the very small number of design principles responsible for the wide range of activities that nervous systems support. While the processes that thinkers go through in conceiving and refining their ideas are intriguing, they should not provide the starting point for an explanation of perception in the service of activity. Putting them at the forefront of things to be explained is an apotheosis of the exotic and likely to be premature. As a parallel, consider the rainbow, which has fascinated philosophers and scientists for centuries. An adequate quantitative theory that accounts for all of the features and quirks of that phenomenon awaited the development of geometrical optics, and an understanding of the wave and particle-like properties of light, polarization, and the complex angular momentum method (Nussenzveig, 1977). We may have to be similarly thorough in uncovering those *fundamental* principles at the ecological scale on which the reliable and reproducible behaviors of epistemic agents are based and on which an acceptable account of thinking will rest.

References

Alley, T. R. (in press). Organism-environment mutuality, epistemics, and the concept of an ecological niche. *Synthese*.

Ashby, R. (1963). *An introduction to cybernetics*. New York: Wiley.

Ben-Zeev, A. (in press). The Kantian revolution in perception. *Journal for the Theory of Social Behavior*.

Blackmore, J. (1979). On the inverted use of the terms "realism" and "idealism" among scientists and historians of science. *British Journal of Philosophy of Science*, *30*, 125-134.

Bunge, M. (1973). On confusing "measure" with "measurement" in the methodology of behavioral science. In M. Bunge (Ed.), *The methodological unity of science*. Dordrecht: D. Reidel Publishing Co.

Carello, C., Turvey, M. T., Kugler, P. N., & Shaw, R. E. (1984). Inadequacies of the computer metaphor. In M. S. Gazzaniga (Ed.), *Handbook of cognitive neuroscience*. New York: Plenum.

Cutting, J. E., Proffitt, D. R., & Kozlowski, L. T. (1978). A biomechanical invariant for gait perception. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 357-372.

Dennett, D. C. (1969). *Content and consciousness*. London: Routledge & Kegan Paul.

Elton, C. (1927). *Animal ecology*. London: Sidgwick and Jackson.

Fodor, J. A. (1981). *Representations*. Cambridge, MA: MIT Press.

Gibson, E. J. (1983). Paper presened at meeting of the International Society for Ecological Psychology, Hartford, CT.

Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton-Mifflin.

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

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

Gibson, J. J. (1982). Notes on action. In E. Reed & R. Jones (Eds.), *Reasons for realism: Selected essays of James J. Gibson*. Hillsdale, NJ: Erlbum.

Gibson, J. J., & Crooks, L. E. (1938). A theoretical field analysis of automobile driving. *American Journal of Psychology*, *51*, 453-471.

Grinnel, J. (1917). The niche relationships of the California thrasher. *Auk*, *34*, 427-433.

Humphrey, C. (1933). *The nature of learning*. New York: Harcourt, Brace.

Iberall, A. S. (1974). *Bridges in science--from physics to social science*. Upper Darby, PA: General Technical Service, Inc.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*, 201-211.

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics*, *21*, 575-580.

Kugler, P. N. (1983). Unpublished doctoral dissertation, University of Connecticut.

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.

Kugler, P. N., & Turvey, M. T. (in press). *Information, natural law and the self-assembly of rhythmic movements: Theoretical and experimental investigations*. Hillsdale, NJ: Erlbaum.

Kugler, P. N., Turvey, M. T., Carello, C., & Shaw, R. E. (1984). The physics of controlled collisions: A reverie about locomotion. In W. H. Warren, Jr. & R. E. Shaw (Eds.), *Persistence and change: Proceedings of the first international conference on event perception*. Hillsdale, NJ: Erlbaum.

Lee, D. N. (1974) Visual information during locomotion. In R. B. MacLeod & H. L. Pick (Eds.), *Perception: Essays in honor of James J. Gibson*. Ithaca, NY: Cornell University Press.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception*, *5*, 437-459.

Lee, D. N. (1980). Visuo-motor coordination in space-time. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. New York: North-Holland.

Lee, D. N., & Reddish, P. E. (1981). Plummeting gannets: A paradigm of ecological optics. *Nature*, *293*, 293-294.

Lewis, C. I. (1929). *Mind and the world order*. New York: Dover Publications, Inc.

Lishman, R., & Lee, D. N. (1973). The autonomy of visual kinaesthesis. *Perception*, *2*, 287-294.

Loisos, C. (1969). An ethological study of chimpanzee play. In J. S. Bruner, A. Jolly, & K. Silva (Eds.), *Play - Its role in development and evolution*. New York: Basic Books, Inc.

Michaels, C. F., & Carello, C. (1981). *Direct perception*. New York: Prentice-Hall.

Newell, A. (1981). Physical symbol systems. In D. A. Norman (Ed.), *Perspectives on cognitive science*. Hillsdale, NJ: Erlbaum.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the Association for Computing Machinery*, *19*, 111-126.

Nussenzveig, H. M. (1977). The theory of the rainbow. *Scientific American*, *236*, 116-127.

Pattee, H. H. (1973). Physical problems of the origin of natural controls. In A. Locker (Ed.), *Biogenesis, evolution, homeostasis*. Heidelberg: Springer-Verlag.

Pattee, H. H. (1977). Dynamic and linguistic modes of complex systems. *International Journal of General Systems*, *3*, 259-266.

Patten, B. C. (1982). Environs: Relativistic elementary particles for ecology. *American Scientist*, *119*, 179-219.

Prigogine, I., & Stengers, I. (1984). *Order out of chaos*. New York: Bantam Books.

Reed, E., & Jones, R. (1982). *Reasons for realism: Selected essays of James J. Gibson*. Hillsdale, NJ: Erlbaum.

Rosen, R. (1978). *Fundamentals of measurement and representation of natural systems*. New York: North-Holland.

Runeson, S. (1983). On visual perception of dynamic events. *Acta Universitatis Upsaliensis: Studia Psychologica Upsaliensia*, Serial No. 9 (originally published, 1977).

Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 733-740.

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, *112*, 585-615.

Schiff, W. (1965). Perception of impending collision: A study of visually directed avoidant behavior. *Psychological Monographs*, *79*(604).

Searle, J. R. (1983). *Intentionality*. Cambridge: Cambridge University Press.

Shaw, R. E., & Turvey, M. T. (1982). Coalitions as models for ecosystems. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum.

Shaw, R. E., Turvey, M. T., & Mace, W. (1982). Ecological psychology: The consequences of a commitment to realism. In W. Weimer & D. Palermo (Eds.), *Cognition and the symbolic processes* (II). Hillsdale, NJ: Erlbaum.

Simon, H. A. (1981). Studying human intelligence by creating artificial intelligence. *American Scientist*, *69*, 300-309.

Solomon, J., Carello, C., & Turvey, M. T. (1984). Flow fields: The optical support for skilled activities. In W. F. Straub & J. M. Williams (Eds.), *Cognitive sport psychology*. Lansing, NY: Sport Science Associates.

Todd, J. T. (1981). Visual information about moving objects. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 795-810.

Turvey, M. T., & Carello, C. (1981). Cognition: The view from ecological realism. *Cognition*, *10*, 313-321.

Turvey, M. T., & Shaw, R. E. (1979). The primacy of perceiving: An ecological reformulation of perception for understanding memory. In L-G. Nilsson (Ed), *Perspectives on memory research: Essays in honor of Uppsala University's 500th anniversary*. Hillsdale, NJ: Erlbaum.

Turvey, M. T , Shaw, R. E., Reed, E. S., & Mace, W. M. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). *Cognition*, *9*, 237-304.

Walter, D. O. (1980). Choosing the right fictions of scientific law. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, *239*, R365-R371.

Walter, D. O. (1983). Creative vagueness in non-realist symbol systems. *Cognition and Brain Theory*, *6*, 227-236.

Warren, W. H., Jr. (1984). Perceiving affordances: The visual guidance of stair climbing. Journal of Experimental Psychology: Human Perception and Performance, 10, 683-703.

Warren, W. H., Jr., & Shaw, R. E. (1981). Psychophysics and ecometrics. The Behavioral and Brain Sciences, 4, 209-210.

Weiss, P. A. (1969). The living system: Determinism stratified. In A. Koestler & J. R. Smythies (Eds.), Beyond reductionism (pp. 3-42). Boston: Beacon.

THE INFORMATIONAL SUPPORT FOR UPRIGHT STANCE*

Claudia Carello,† M. T. Turvey,†† and Peter N. Kugler†††

Nashner and McCollum suggest that (1) perturbations of the body relative to the gravitational field and the surface of support parse into a small number of circumscribed kinetic states (regions of disequilibrium), and (2) a functional muscular organization, to restore upright posture, corresponds to each state. Though the authors talk about the sensing of these states, they give no indication of the relevant information. In a related way, we think, their references to neural signals that require interpretation, their appeals to memory (presumably of previous trajectories, previous initial conditions, previous sensory consequences, and previous postural achievements), and their supposition of anatomically defined senses uniquely tied to distinct frames of reference seem to run counter to the general Bernsteinian (1967) strategy that they are pursuing, that is, compressing in a principled fashion a movement problem of potentially very many degrees of freedom into a movement problem of very few degrees of freedom.

In contrast, we are inclined strongly toward Gibson's (1966, 1979) revision of the senses in terms of perceptual systems--active, interrelated systems (as opposed to senses) that detect information (rather than have sensations) about the perceiver-environment relation (rather than about their own states). Taking a Gibsonian stance, we ask whether there could be information _specific_ to a circumscribed disequilibrium state, regardless of etiology; whether there could be information _specific_ to approaching a region's boundary, regardless of the details of the trajectory; and whether such information can be independent of the mode of attention. We will start with Gibson's strict interpretation of information with respect to vision, demonstrate that equivalent information is obtainable by other perceptual systems, and conclude with speculation about properties that might generalize to the control of stance.

Information is optical structure lawfully generated by the persistent and changing layout of surfaces and by the displacements of the body (as a unit

---

*_The Behavioral and Brain Sciences_, 1985, _8_, 151-152. (Commentary on Nashner, L. M., & McCollum, G. The organization of human postural movements: A formal basis and experimental synthesis. _The Behavioral and Brain Sciences_, 1985, _8_, 135-150.

†State University of New York, Binghamton.

††Also University of Connecticut.

†††Crump Institute of Medical Engineering, UCLA.

Acknowledgment. The writing of this manuscript was supported in part by Office of Naval Research Contract N00014-83-C-0083 awarded to Haskins Laboratories.

relative to the surface layout and as parts relative to each other). Because the properties of the optic flow field are lawfully related to the properties of the kinetic field underlying them, they are said to specify those kinetic properties (see Runeson & Frykholm, 1983). Following Lee (1978), the optical flow field is exterospecific (specific to properties of surface layout), expropriospecific (specific to the orientational displacements of the point of observation relative to the surface layout), and propriospecific (specific to the relations among the parts of the body). And it can be specific in each of these ways simultaneously. How can this be? Each class of facts (extero, exproprio, proprio) imposes a distinct patterning--or structure, or form, or morphology (see Kugler & Turvey, in press)--on the optical flow field. These patternings are superposed on each other but differentiable from one another.

Consider one such patterning. An optical flow field can be treated as, roughly, a velocity vector field where the vectors represent angular velocities of the optical elements (see Gibson, 1979). When all vectors are undergoing a graduated magnification about a fixed point, then the point of observation is displacing rectilinearly toward the fixed point. It is suggested that any globally smooth velocity vector field specifies a displacement of the point of observation. (Note that the qualitative macroscopic properties of the field are what matter, not the individual vectors.) One can sketch a law at the ecological scale (see Turvey, Shaw, Reed, & Mace, 1981) roughly of the form:

```
displacement of point      LAWFULLY GENERATES      globally smooth velocity
   of observation         ------------------->          vector field
```

This law defines a particular kind of information in Gibson's specificational sense, that is,

```
globally smooth velocity     SPECIFIES        displacement of point of
     vector field          ------------> observation relative to surround
```

Note that the optical property in the foregoing law is a kinematic abstraction (dimensions: length and time) of an energy distribution (light) structured by properties of a kinetic field (dimensions: mass, length, and time), that is, the field determined by the animal and surface layout. Insofar as the same kinematic abstraction could be supported by other energy distributions modulated by the same kinetic facts, this analysis can be generalized to other modes of attention. For example, if a sound field with the same globally smooth morphology could be produced, according to Gibson's law-based/specificational interpretation of information, listeners should perceive themselves displacing relative to the surroundings (for confirming evidence, see Dodge, 1923; Lackner, 1977). Defining this morphology over deformations of the skin should yield the same impression of egomotion (again see Lackner, 1977).

This treatment of expropriospecification can be extended to extero- and propriospecification. It is suggested that distinct flow morphologies, now discontinuous rather than smooth, specify facts of surface layout and relations among joints (Gibson, 1966, 1979). Again, these morphologies can be instanced by different kinds of energy distributions. Note that is possible to describe vestibular stimulation--weights displacing in fluid-filled chambers relative to gravity's pull--and haptic-somatic stimulation--nonrigid mechanical deformations of the body's tissues--as kinematic or vector fields. And note further that, in principle, these velocity vector fields are characterizable alternatively as low-dimensional, macroscopic patternings. According to the ecological law formulation from above, if a given disequilibrium state gives rise to identical morphologies in the vector fields that are "attended to" vestibularly, haptically, and visually, then the same postural fact will be apprehended by each mode of attention.

Nashner and McCollum are puzzled by neural signals having equivalent postural consequences when the signals are different. In our view, their puzzlement is based on the wrong formulation: Information may be identical when neural signals, stimuli, etc., are different (see Gibson, 1966, p. 55). Nashner and McCollum feel that neural signals must be interpreted. Signal is a metaphor for sensations, and sensations strictly speaking can only be about states of nerves; hence the need for interpretation. Again, their formula is suspect. Information is about, in the sense of specific to, animal-environment facts. It needs to be detected, and its differentiation and pick up by a perceptual system improve with practice, but to interpret it would be superfluous.

We have suggested that the information about kinetic conditions (such as regions of postural equilibrium) is to be found in the morphology of kinematic fields. Moreover, the information is indifferent to the medium that has been structured kinematically. We conclude with a speculation about the morphological property specific to approaching a region's boundary--a generalization of the time-to-contact variable, T, and its derivative (Lee, 1980).

For the visual system, T is the inverse of the relative rate of dilation of, roughly, the optic array. It specifies when one will contact a surface on the path of locomotion. Its derivative specifies how hard the imminent collision will be (Lee, 1980). Our conjecture is that T may be a very general property of kinematic (flow) fields. Any kinetic field will have, as a rule, the equivalents of contactable "surfaces"; for example, attractors, basins, etc. Is there, as a rule, the equivalent of T in the kinematic abstraction of any kinetic field--for example, nonrigid mechanical distortions of body tissues? Suppose that the authors' regions of equilibrium are detected haptically. Then the proposed availability of T and its derivative would provide a principled haptic basis for regulating forces to prohibit crossing regions.

In sum, Gibson's treatment of information seems relevant to Nashner and McCollum in this sense: The low dimensionality of postural control they promise on the side of action could be reciprocated (as it must) on the side of perception.

References

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

Dodge, R. (1923). Thresholds of rotation. Journal of Experimental Psychology, 6, 107-137.

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

Gibson, J. J. (1979). The ecological approach to visual perception. Boston: Houghton-Mifflin.

Kugler, P. N., & Turvey, M. T. (in press). Information, natural law, and the self-assembly of rhythmic movement. Hillsdale, NJ: Erlbaum.

Lackner, J. R. (1977). Induction of nystagmus in stationary subjects with a rotating sound field. Aviation, Space and Environmental Medicine, 48, 129-131.

Lee, D. N. (1978). On the functions of vision. In H. L. Pick, Jr. & E. Saltzman (Eds.), Modes of perceiving and processing information. Hillsdale, NJ: Erlbaum.

Lee, D. N. (1980). Visuo-motor coordination in space-time. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. New York: North-Holland.

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. Journal of Experimental Psychology: General, 112, 585-615.

Turvey, M. T., Shaw, R. E., Reed, E. S., & Mace, W. M. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). Cognition, 9, 237-304.

DETERMINING THE EXTENT OF COARTICULATION: EFFECTS OF EXPERIMENTAL DESIGN*

Carole E. Gelfer,† Fredericka Bell-Berti,†† and Katherine S. Harris†

Abstract. Substantial differences in the reports of the extent of anticipatory coarticulation have made the task of deciding among unifying models of the process difficult. Two conceptuallly distinct groups of theories of coarticulation have emerged, one positing the migration of articulatory features to preceding segments and the other positing the temporal cohesiveness of the components of segmental articulations. In studies of anticipatory lip rounding, a possible source of the differences reported in its extent prior to a rounded vowel is that the alveolar consonants commonly employed in these studies are presumed to be unspecified with regard to lip configuration. Thus, the presence of EMG activity and/or protrusive lip movement during these consonants has been presumed to indicate vocalically conditioned lip activity. However, if this activity is directly related to the production of the consonant(s), then the interpretation of these results is problematic unless the experimental design allows for the differentiation of consonantal and vocalic effects. We offer here both data suggesting the need for such considerations and a paradigm that takes these considerations into account.

Introduction

The phenomena of anticipatory coarticulation have generally been presumed to reflect underlying aspects of speech motor control (e.g., Kozhevnikov & Chistovich, 1966; MacNeilage, 1970).[1] However, substantial differences in reports of the extent of anticipatory coarticulation make difficult the task of providing one model to account for these data. Two types of conceptually distinct theories of anticipatory coarticulation exist, both of which attempt to explain the apparently nondiscrete nature of speech output despite a presumed discrete input. According to one type of theory, upcoming phones are scanned for salient features, which then migrate to as many antecedent phones as are neutral for, or in no way antagonistic to, the migrating feature (e.g., Daniloff & Moll, 1968; Henke, 1966; Kozhevnikov & Chistovich, 1966; Sussman & Westbury, 1981). Thus, given some number of consonants unspecified for lip configuration immediately preceding a rounded vowel, these models predict that rounding will vary in its onset in direct proportion to the number and/or

---

*A version of the paper was presented at the 103rd Meeting of the Acoustical Society of America. Chicago, IL, May 1982.

†Also The Graduate Center, The City University of New York.

††Also St. John's University.

Acknowledgment. This work was supported by NINCDS grant NS-13617 and BRS grant RR-05596 to Haskins Laboratories.

duration of preceding segments. For example, Benguerel and Cowan (1974) reported that upper lip protrusion (in anticipation of a rounded vowel) begins as early as the first consonant in clusters of as many as six consonants. However, the second type of theory proposes that the observed co-occurrence of components of proximate segments results, not from feature migration, but from the overlapping of articulatory components of those segments (e.g., Bell-Berti & Harris, 1981, 1982; Fowler, 1980). Thus, in the absence of conflicting demands, the onsets of different components of the articulation of a given phone will bear a stable temporal relationship to each other. For example, Engstrand (1981) reported that lip protrusion activity for the rounded vowel /u/ occurs at a relatively fixed time before the onset of voicing for that vowel, regardless of the number of preceding consonants.

Despite their conceptual differences, however, a basic premise, having its roots in traditional linear generative phonology, is common to these models: namely, that a phone is neutral (i.e., unspecified) for a particular feature when that feature is not essential to its realization (Chomsky & Halle, 1968, pp. 402-403). Consequently, when activity associated with a given feature occurs during a segment that is "neutral" for that feature, that activity must be associated with another segment, and the time at which this activity begins is then assumed to reflect the extent of anticipatory coarticulation. In fact, however, it may be that feature descriptions are incomplete. For example, as Benguerel and Cowan (1974) have noted, American English /r/ is commonly produced with lip protrusion, although this protrusion often goes unmentioned in articulatory descriptions of /r/.

Upon closer consideration, it would appear that many of the differences in the existing literature might be reconciled, and thus allow the development of a single explanation for them, were these assumptions reconsidered. The work presented here is part of a study designed to account for the conflicting results of previous studies, and therefore to test the predictions of the different models of anticipatory coarticulation.

Methods

The alveolar consonants /t/ and /s/, whose articulation would be presumed to be neutral for lip constriction, were combined to form nine sequences designed to vary both in the number of consonants and in overall sequence durations.[2] The vowels in these utterances were /i/ and /u/, where V_1 was always /i/, while V_2 was either /i/ or /u/. Thus, there were two vowel conditions, the /iC_nu/ and /iC_ni/ conditions, each occurring with the nine different consonant string combinations, for a total of eighteen utterance types (Table 1). The sequences were made by combining "words," and were presented to the subjects in orthographic writing. The subjects were instructed to speak at a comfortable rate, in a conversational manner, without undue attention to marking word boundaries. Thus, the subjects could, and did, differ in the way in which they executed a given sequence (for example, leased tool (/list#tul/) was often realized as the sequence [list:ul]). Two native speakers of American English[3] produced between fifteen and twenty repetitions of each of the eighteen VC_nVs, spoken within the carrier phrase "It's a __________ again."

Surface electromyographic (EMG) recordings (Allen, Lubker, & Harrison, 1972) of orbicularis oris inferior (OOI), right and left, were made simultaneously with lip movement recording. Lip movements were tracked with an optoelectrical tracking system (Capstan Co. Model 400 Optical Tracking System) that sensed the position, in both the x and y planes, of an infrared light-emitting diode (LED) positioned on the lower lip. All data were simultaneously recorded on a 14-channel FM tape.

Table 1

Consonant Strings: Number and Duration

		Consonant String Duration (in milliseconds)	
Utterance	Number of Consonants	TB	CH
i#tu	1	75	68
i#su	1	220	160
i#stu	2	245	152
is#tu	2	230	163
is#su	2	300	238
is#stu	3	305	253
ist#tu	3	280	266
ist#su	3	385	331
ist#stu	4	360	355
i#ti	1	83	71
i#si	1	227	160
i#sti	2	240	136
is#ti	2	230	165
is#si	2	335	---
is#sti	3	331	245
ist#ti	3	284	272
ist#si	3	391	330
ist#sti	4	392	337

The EMG signals were rectified, and both the EMG and movement data were integrated and then digitized using a PDP 11/45 computer. The durations of the consonant strings were measured for each token of each utterance type, using a PCM waveform-editing program. The beginning of the consonant string was defined as the point at which either the frication appeared in the waveform (in consonant strings beginning with /s/), or the higher formants disappeared from the waveform (indicating the onset of closure in consonant strings beginning with /t/). The point in the acoustic signal corresponding to the release of the consonant occlusion immediately preceding V_2 was identified as the end of the consonant string and served as the acoustic reference, or line-up, point for subsequent ensemble averaging. Thus, when V_2 was preceded by /t/, the line-up point was the burst; when V_2 was preceded by /s/, the line-up point was the end of frication before the second vowel.

The beginning of OOI activity associated with the /VC$_n$V/ sequences was determined by identifying the time at which the EMG activity increased to a level equivalent to the baseline plus five percent of the difference between the baseline and the peak EMG levels. The beginning of the related movement was determined by identifying the onset of anteriorly-directed lip movement.

Results

Some representative EMG data are shown for each subject (Figure 1a). The EMG signals in each panel represent the ensemble average OOI EMG activity of an /iC$_n$u/ utterance, with consonant string length (i.e., both the number of segments and the durations of the sequences) differing across panels. The onset of EMG activity occurs earlier as consonant string duration increases, so that it would appear that there has been a migration of lip rounding back to the beginning of the consonant string. In fact, when the onset of OOI EMG activity for each of the nine /iC$_n$u/ utterances is plotted against the respective consonant string durations (Figure 1b), it seems that, for both subjects, these onsets bear an obvious relationship to consonant string duration. That is, they occur earlier as string duration increases, with correlation coefficients of r=.98 and .97 for TB and CH, respectively.

Although these results might be interpreted as evidence that lip rounding has spread to the beginning of the "neutral" consonant string, we believe that it is imperative to determine whether all of the EMG activity is actually vowel-related or, alternatively, if it reflects consonantal lip gestures. In other words, if the OOI activity during the consonant string is vowel-related, we would not expect to find such activity during the same consonant string when it is followed by an unrounded vowel. We therefore examined OOI activity for the minimally contrastive /iC$_n$i/ utterances, samples of which are shown in Figure 2a. It is clear that, even within this unrounded vowel environment, there is a significant amount of orbicularis oris activity during the consonant string articulation. In fact, if we treat these /iC$_n$i/ data as we did those for the /iC$_n$u/ utterances, identifying the onset of EMG activity for each utterance and plotting these times against consonant string durations (Figure 2b), the resulting scatter plots are strikingly similar to those for the /iC$_n$u/ utterance set (Figure 1b). That is, OOI activity begins earlier as consonant string duration increases. (Subject CH produced only eight of the nine /iC$_n$i/ utterances.) Obviously, then, this EMG activity cannot reflect the onset of vowel-related lip rounding (i.e., the migration of the vowel feature) since the relationship between consonant string duration and the onset of OOI activity is observed in both rounded and unrounded vowel environments. Indeed, correlation coefficients are as high or higher for these /iC$_n$i/ utterances (r=.98 and .99 for TB and CH, respectively) than they are for their rounded counterparts.

It is obvious, then, that the progressively earlier EMG activity must reflect consonant-related events. This is made more apparent when the EMG curves for the minimally contrastive /iC$_n$u/ and /iC$_n$i/ utterances are superimposed (Figure 3). The two signals diverge in the vicinity of the acoustic onset of V_2, with a second peak of activity evident when V_2 is /u/, while EMG activity is suppressed when V_2 is /i/. However, because the EMG signal never returns to a baseline level prior to /u/, the onset of the /u/-related

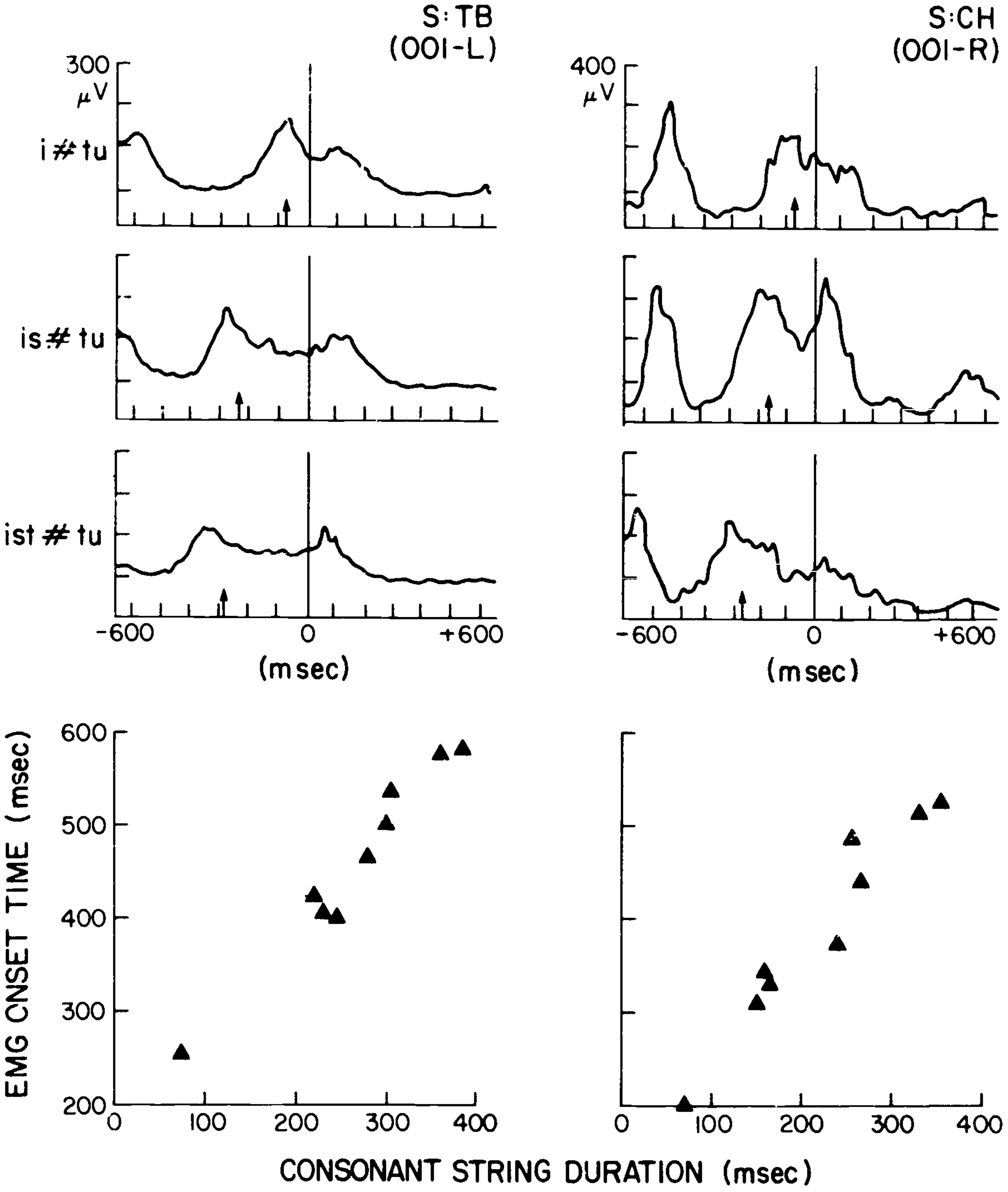


Figure 1. Upper panels (1a): Ensemble-average EMG data for subjects TB (left) and CH (right) recorded from orbicularis oris inferior (OOI) for three /iC$_n$u/ utterances. Lower panels (1b): EMG onset time (ms before line-up point) vs. consonant string duration for /iC$_n$u/ utterances. Time 0 represents the release of the consonant occlusion, determined from the acoustic waveform. The arrows indicate the average of the acoustic onsets of the consonant strings.

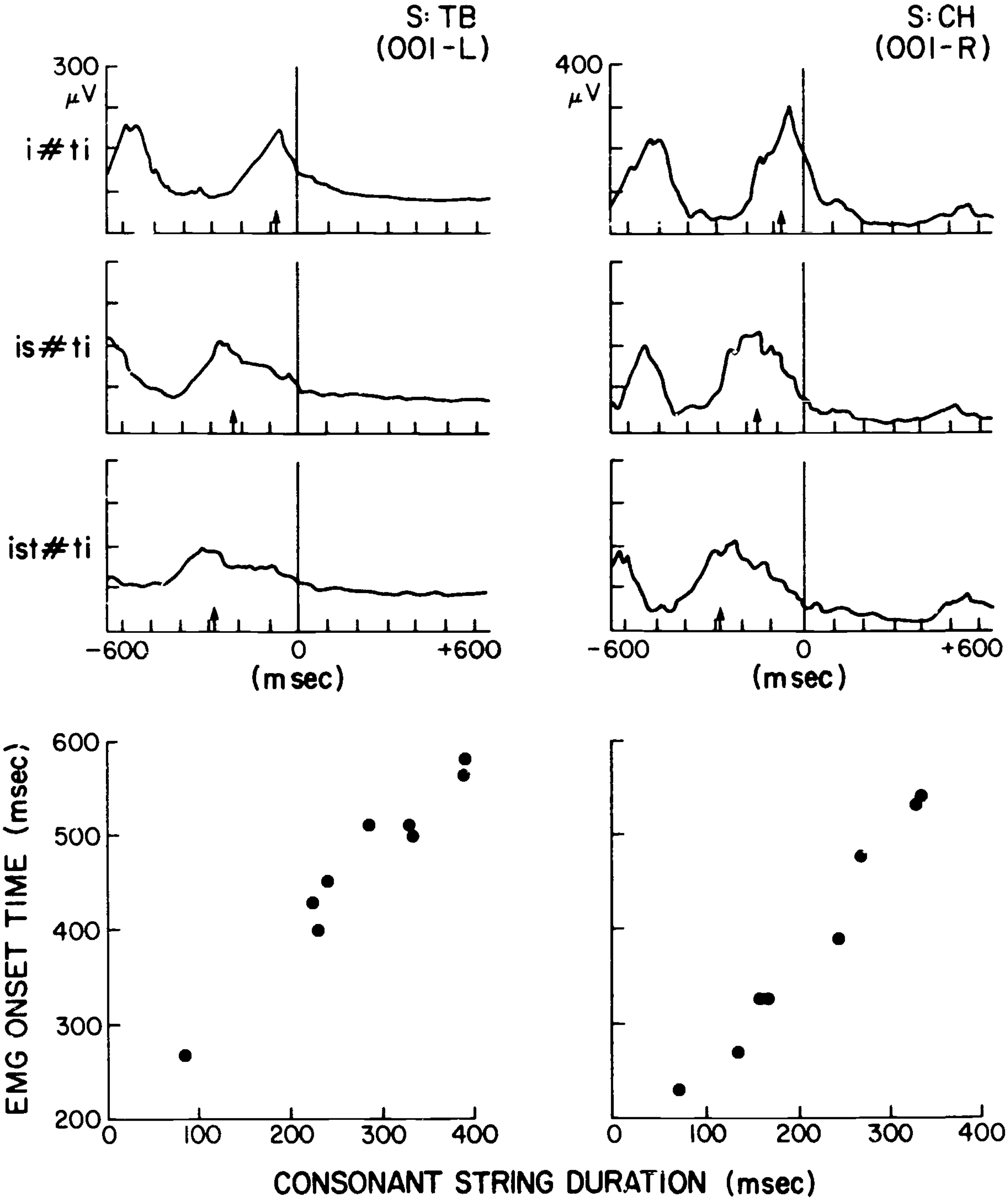


Figure 2. Upper panels (2a): Ensemble-average EMG data for subjects TB (left) and CH (right) recorded from orbicularis oris inferior (OOI) for three /iC$_n$i/ utterances. Lower panels (2b): EMG onset time (ms before line-up point) vs. consonant string duration for /iC$_n$i/ utterances. Time 0 represents the release of the consonant occlusion, determined from the acoustic waveform. The arrows indicate the average of the acoustic onsets of the consonant strings.

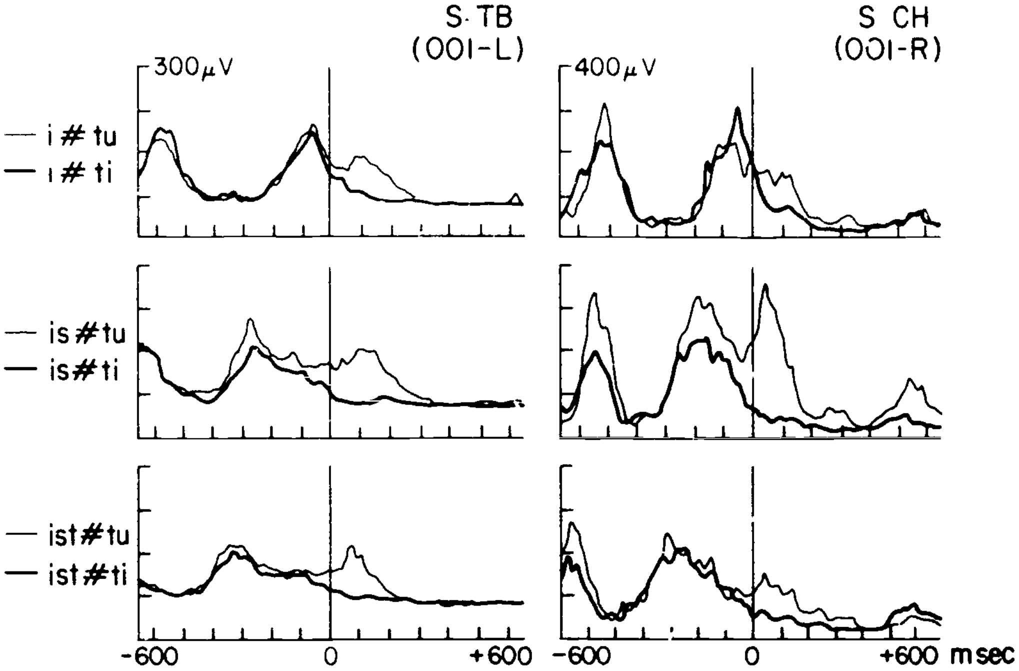


Figure 3. Ensemble-average EMG data for the two subjects, recorded from orbicularis oris inferior (OOI) for three minimally contrastive pairs of /iC_nV/ utterances.

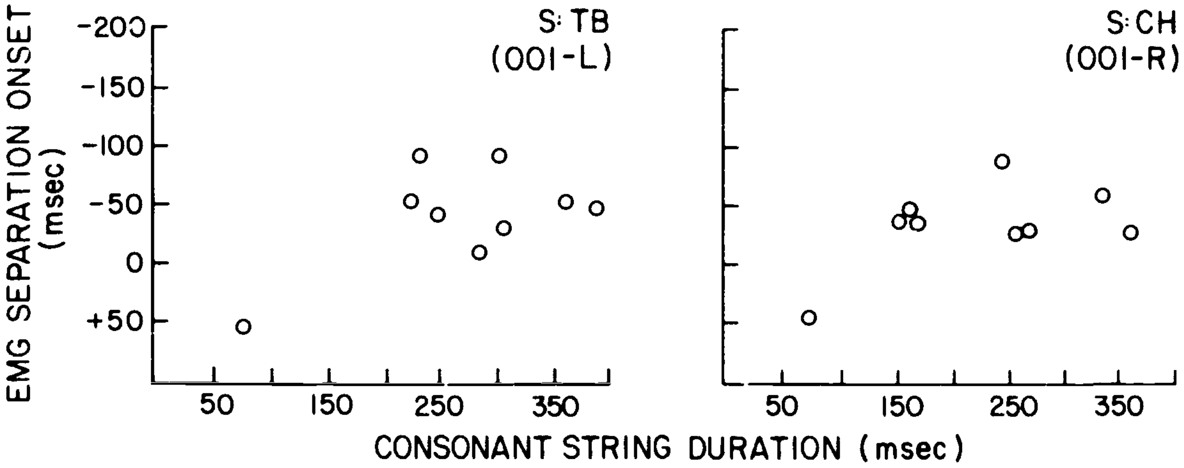


Figure 4. Statistically determined point of separation ("EMG separation onset") between minimally contrastive pairs of /iC_nV/ utterances vs. the average duration of the consonant sequences of the /iC_nu/ utterances of each pair.

EMG activity was determined statistically as the time at which the difference (in microvolts) following the divergence of the two signals reached significance ($p<.05$).

The statistically determined onsets of rounded vowel activity are plotted as a function of consonant string duration for the nine minimal pairs for subject TB, and for eight minimal pairs for subject CH (Figure 4). In contrast to the consonant-related EMG activity (see Figures 1 and 2), these onsets bear no obvious relation to the durations of the consonant strings.[4] Rather, with the exception of the /i#tu/ utterance, they occur within a fairly restricted range, bearing a stronger relationship to the onset of the rounded vowel than to the onset of the consonant string.

The EMG data thus show the following: First, for these two subjects, some lip activity appears to be inherent in the production of alveolar consonants. Second, the onset of EMG activity for /u/ appears to be related to the acoustic onset of that vowel, and not to the compatibility of the vowel and consonant articulations. Finally, even when there is lip activity for adjacent consonants and vowels, they appear to be organized as independent gestures, as the separate peaks of OOI activity for the /iC_nu/ utterances suggest.

Figure 5 shows movement data for both subjects, for the same /iC_nu/ utterances whose EMG data are presented above (Figure 1a). For TB, the data show a substantial forward lip movement in the vicinity of the acoustic onset of the consonant string, a position that is then sustained through V_2. However, while there is a less obvious separation between the consonant and vowel gestures in the movement than in the EMG records, there are troughs in the movement traces for all but the shortest utterance.[5] For subject CH, the anterior lip movement associated with the rounded vowel is more clearly separated from the anterior movement occurring earlier in the utterance.

When the movement traces for the /iC_nu/ and /iC_ni/ utterances are superimposed (Figure 6), the pattern is the same as that for the EMG records. That is, regardless of the identity of V_2, the curves are nearly identical through the consonant string, diverging in the vicinity of the onset of the second vowel. However, because of hardware limitations at the time of recording, the baselines for these data are not always aligned;[6] for this reason we were unable to determine statistically the times at which each minimally contrastive pair differed, as we had done for the EMG data. Furthermore, when the temporal relationships between the consonant-related EMC and the earliest anteriorly directed movements are examined, there are clearly differences for the two subjects. For subject TB, the earlier onset of OOI activity is associated with consonant-related forward lip movement. That is, there is an appropriate contraction time interval between the EMG and corresponding movement (Figure 7a). For subject CH, however, the earlier OOI activity is not associated with any significant anterior lip movement for the consonant string (Figure 7b). Rather, this movement is associated with the first vowel.

We are therefore faced with the question of what the consonant-related EMG activity means in terms of movement for subject CH. Figure 8 shows OOI activity for the three representative /iC_nu/ utterances, along with both the corresponding horizontal and vertical movement traces. It can be seen that, while the EMG and horizontal lip movements are poorly correlated in the vicinity of the consonant string, there is a good temporal correlation (i.e.,

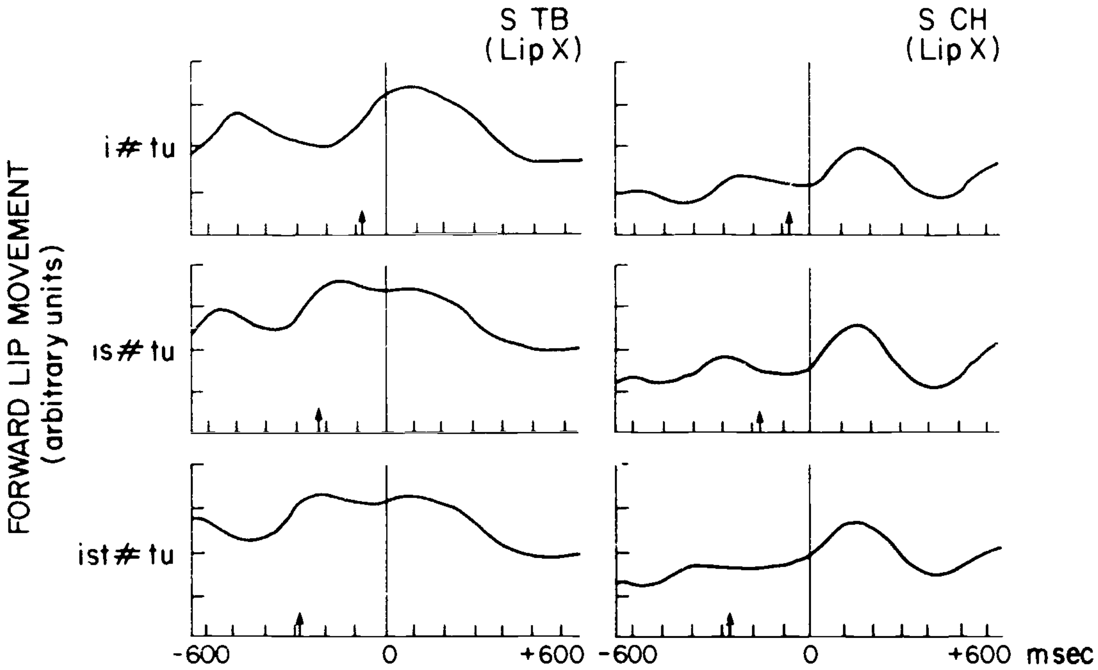


Figure 5. Antero-posterior lip position as a function of time for the two subjects for three /iC$_n$u/ utterances. The arrows indicate the average of the acoustic onsets of the consonant strings.

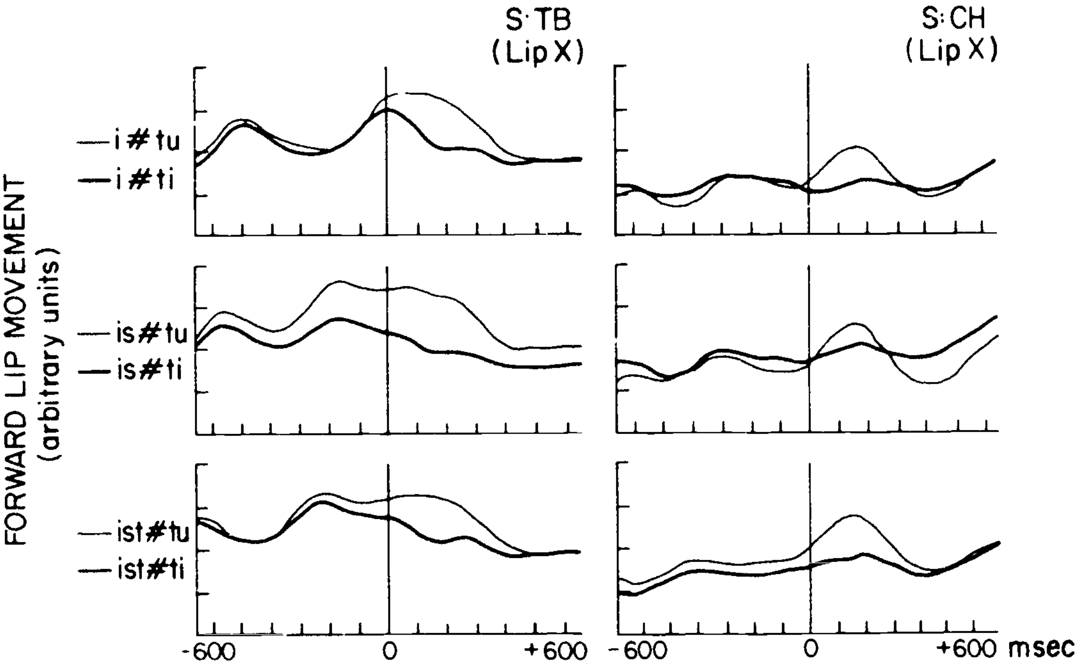


Figure 6. Antero-posterior lip position for both subjects as a function of time for three minimally contrastive pairs of /iC$_n$V/ utterances.

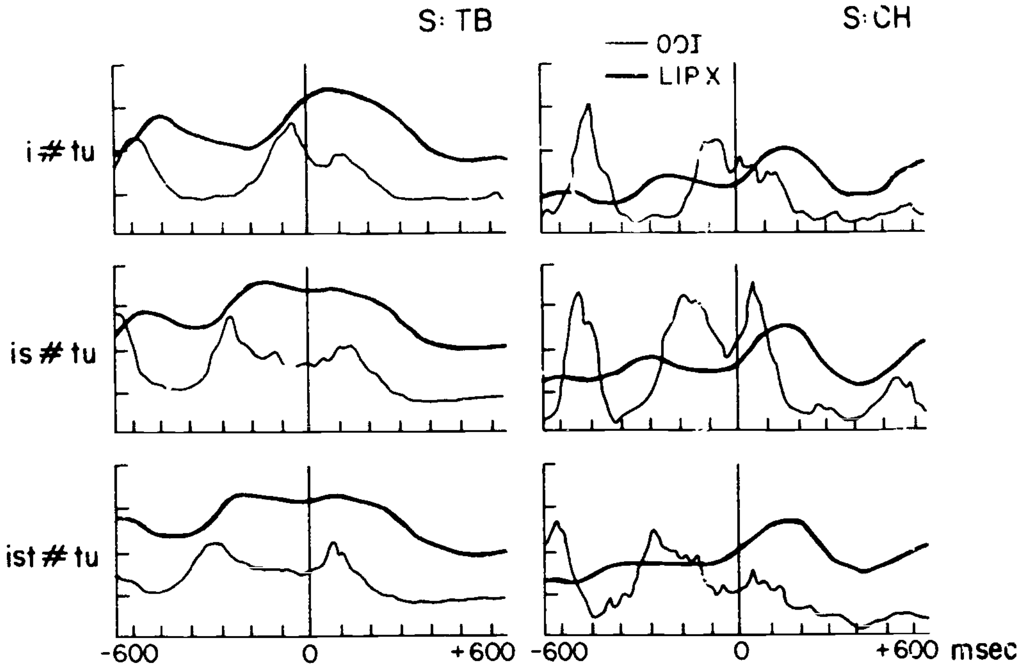


Figure 7. Ensemble-average EMG and lip position data as a function of time for both subjects for three /iC$_n$u/ utterances.

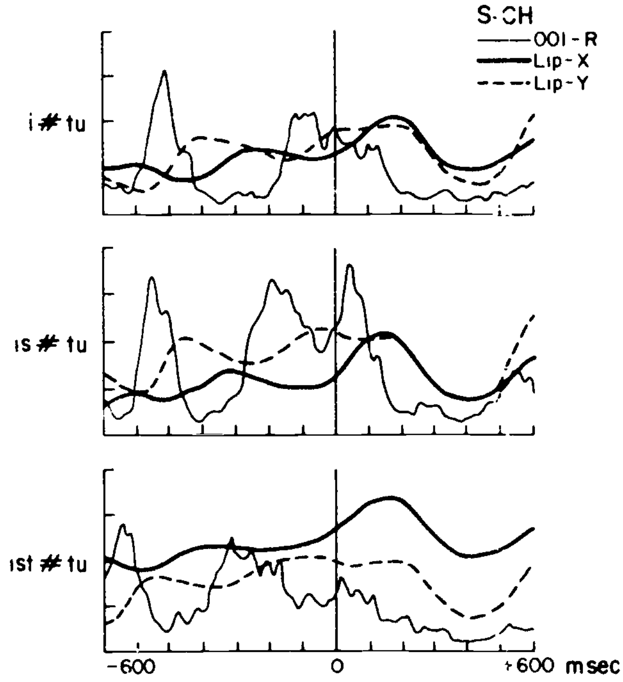


Figure 8. Ensemble-average EMG and lip position data for subject CH for three /iC$_n$u/ utterances. The thin line represents OOI-R data, the thick line anterior lip position data (lip X), and the dashed line vertical lip position data (lip Y).

contraction interval) between the consonant-related EMG and vertical lip movement. Thus, for this subject, the same muscle appears to be contributing to both vertical movement (in the production of the consonant string) and horizontal movement (in the production of the vowel), differences in orbicularis oris function that have been noted previously (cf. O'Dwyer, Quinn, Guitar, Andrews, & Neilson, 1981).

Discussion

The data offered here suggest that there are a number of reasons for the difficulty in reconciling the differences between sets of previously reported data on the extent of anticipatory coarticulation. One of these reasons resides in the unproven assumptions that, if a speech sound's articulation has not been described as including a particular gesture, then, first, that gesture has little, if any, consequence for the production of the sound and, second, that speech sound is "unspecified" for that gesture/feature. However, phoneticians have long known that the description of the articulation of speech sounds is incomplete (cf. Pike, 1943, p. 152); our data clearly indicate that, for some speakers at least, some alveolar consonants traditionally assumed to have no intrinsic lip gestures do in fact have such gestures as part of their natural production. Thus, the assumption that these consonants are neutral with regard to lip configu ion is untenable.

These data also provide evidence of the complexity of the electromyographic and kinematic data collected for studying coarticulation processes. First, it is impossible to separate active protrusion gestures from passive relaxation of lips that have been retracted, except by observing the activity of the muscles responsible for those protrusion gestures. Second, the EMG data may more closely reflect the underlying segmental structure of speech than do kinematic data. For example, while we see no trough in the movement traces of the /i#tu/ utterance for subject TB, there are clearly separate peaks of OOI activity for both the consonant and vowel segments, suggesting the segmental nature of the underlying articulatory organization.

In addition to providing insights into the causes of some of the apparent discepancies resulting from problems in experimental design, we would also suggest that another source of conflict in attempts to develop a single model of anticipatory phenomena stems from presupposing that the timing of the onset of rounding is an entirely anticipatory phenomenon. It is notable that in both this study and our earlier work (Bell-Berti & Harris, 1981), the onset of vowel-related lip rounding is closer to the acoustic onset of the rounded vowel for sequences of the form /i#tu/ than for any other sequence. This result might seem to provide some limited support for the feature migration hypothesis, if this sequence were compared with only one longer sequence (see, e.g., Sussman & Westbury, 1981). However, we believe that an equally plausible explanation is that the result reflects the suppression of lip rounding until the first vowel can be completed without distortion. That is, the onset of rounding may be constrained by the carryover effects of a preceding (unrounded) vowel. Thus, in a sequence like /i#tu/, where the vowel-to-vowel interval is fairly short, the rounding onset might be delayed relative to other sequences where the consonantal sequence occupies a longer time slot. In fact, Sussman and Westbury's (1981) observation of systematic differences in the onset of lip rounding as a function of the identity of the preceding unrounded vowel may be interpreted as evidence of the same carryover effect.

Summary

These data were part of a study designed to account for conflicting results of previously reported studies by suggesting that at least some of the apparent discrepancy arises from experimental design. Because our two subjects produced alveolar consonants with significant orbicularis oris activity in both rounded and unrounded vowel environments, we were able to establish that those gestures that were variable in their onsets on both the EMG and movement levels were clearly tied to something that was acoustically variable as well--namely, the onsets of consonant strings of differing durations. We also observed separate consonant and vowel-related activity, as in the EMG records of the /iC$_{n}$u/ utterances, where there were almost always distinct peaks for each. Furthermore, our EMG data may be interpreted as reflecting a stable onset of lip rounding independent of consonant string duration, except for the case of the shortest consonant string. And, while the tendency has been to view all of these phenomena as reflecting only anticipatory coarticulation, we believe it more likely that they represent the combined effect of carryover and anticipatory processes.

References

Allen, G. D., Lubker, J. F., & Harrison, G. (1972). New paint-on electrodes for surface electromyography. *Journal of the Acoustical Society of America*, *52*, 124(A).

Bell-Berti, F., & Harris, K. S. (1974). More on the motor organization of speech gestures. *Haskins Laboratories Status Report on Speech Research*, *SR-37/38*, 73-77.

Bell-Berti, F., & Harris, K. S. (1981). A temporal model of speech production. *Phonetica*, *38*, 9-20.

Bell-Berti, F., & Harris, K. S. (1982). Temporal patterns of coarticulation: Lip rounding. *Journal of the Acoustical Society of America*, *71*, 449-454.

Benguerel, A.-P., & Cowan, H. A. (1974). Coarticulation of upper lip protrusion in French. *Phonetica*, *30*, 41-55.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Cambridge, MA: MIT Press.

Daniloff, R. G., & Moll, K. L. (1968). Coarticulation of lip-rounding. *Journal of Speech and Hearing Research*, *11*, 707-721.

Engstrand, O. (1981). Acoustic constraints of invariant input representation? An experimental study of selected articulatory movements and targets. *RUUL* (Reports from Uppsala University Department of Linguistics), *7*, 67-94.

Engstrand, O. (1983). *Articulatory coordination in selected VCV utterances: A means-end view*. Unpublished doctoral dissertation, Department of Linguistics, University of Uppsala, Uppsala, Sweden.

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

Gay, T. J. (1977). Cinefluorographic and electromyographic studies of articulatory organization. In M. Sawashima & F. S. Cooper (Eds.), *Dynamic aspects of speech production*. Tokyo: University of Tokyo Press.

Henke, W. L. (1966). *Dynamic articulatory model of speech production using computer simulation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Kozhevnikov, V. A., & Chistovich, L. A. (1966). *Rech, Artikulyatsiya, i vospriyatiye*, [*Speech: Articulation and perception*] (*30*, p. 543). Washington, DC: Joint Publications Research Service. (Originally published 1965.)

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. F. (1970). Motor control of serial ordering of speech. Psychological Review, 77, 182-196.

O'Dwyer, N. J., Quinn, P. T., Guitar, B. E., Andrews, G., & Neilson, P. D. (1981). Procedures for verification of electrode placement in EMG studies of orofacial and mandibular muscles. Journal of Speech and Hearing Research, 24, 273-288.

Pike, K. (1943). Phonetics. Ann Arbor: University of Michigan Press.

Sussman, H. M., & Westbury, J. R. (1981). The effects of antagonistic gestures on temporal and amplitude parameters of anticipatory labial coarticulation. Journal of Speech and Hearing Research, 24, 16-24.

Footnotes

[1] We have limited ourselves here primarily to a consideration of anticipatory phenomena. This limitation was imposed because most theoretical discussions have focused on anticipatory coarticulation.

[2] The literature in this area contains two different indices to consonant string length: the number of consonant segments (e.g., Daniloff & Moll, 1968; Lubker & Gay, 1982) and the duration of the consonant sequence (e.g., Bell-Berti & Harris, 1974, 1982; Engstrand, 1981). Although these two measures are related, the relationship is not isomorphic (see, for example, Table 1).

[3] Subject TB is a speaker of educated Greater Metropolitan New York City English. Subject CH is a speaker of educated Central Florida English.

[4] This result is compatible with results of other studies using subjects known to produce the alveolar consonants /s/ and /t/ without lip rounding (cf. Bell-Berti & Harris, 1982; Engstrand, 1981), although these studies clearly still subscribe to the possibility that alveolar consonants have inherently neutral lip specifications.

[5] The observation of "troughs" in EMG and movement records is not new (cf. Bell-Berti & Harris, 1974; Engstrand, 1983; Gay, 1977). The fact that a trough is absent in movement records when the intervocalic consonant is short may not reflect differences in gestural organization, but, rather, biomechanical constraints that could influence the response characteristics of the lips. That is, with movement being rather slow relative to EMG activity, it is hardly surprising that the lips do not have time to protrude, retract, and protrude again for the rounded vowel during the 75 ms /t/ closure.

[6] We would note, however, that there was no consistent pattern of DC offsets between the /iC_ni/ and /iC_nu/ utterances, suggesting that these differences were independent of vowel rounding.

THE ROLES OF PHONEME FREQUENCY, SIMILARITY, AND AVAILABILITY IN THE EXPERIMENTAL ELICITATION OF SPEECH ERRORS*

Andrea G. Levitt† and Alice F. Healy††

Abstract. In two experiments subjects read aloud pairs of nonsense syllables rapidly presented on a display screen or repeated the same syllables presented auditorily. The error patterns in both experiments showed significant asymmetry, thus lending support to explanations of the error generation process that consider certain phonemes to be "stronger" than others. Further error analyses revealed substantial effects of phoneme frequency in the language and effects of phoneme similarity, which depended on the feature system used to index similarity. Phoneme availability (the requirement that an intruding phoneme be part of the currently presented stimulus) was also important but not essential. We argue that the experimental elicitation of errors provides critical tests of hypotheses generated by the analysis of naturally occurring speech errors.

Recent interest in speech errors has focused largely on the evidence such errors provide about levels of linguistic analysis and psychological models of the speech production process. For example, Fromkin (1971), basing her analysis on a corpus of naturally occurring speech errors, found evidence in support of the independence of various levels of linguistic analysis, including both phonemes and phonetic features. On the other hand, Garrett (1980), also basing his analysis on spontaneous-error collections, examined speech error distributions for the constraints they provide about a model of sentence production.

---

*In press, Journal of Memory and Language.

†Also Wellesley College.

††University of Colorado.

Acknowledgment. This research was supported in part by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories, by NSF Grant BNS 80-00263 to Yale University, and by NSF Grant BNS 80-25020 to the Institute of Cognitive Science, University of Colorado, Boulder. The experimental work was conducted while the second author was supported by a Senior Faculty Fellowship from Yale University. This work was completed while the first author was a NATO postdoctoral fellow at the Centre d'etude des processus cognitifs et du langage, Ecole des hautes etudes en sciences sociales, Paris France. The authors are grateful to Mary LaRue for her help in testing subjects, transcribing tapes, and tabulating the data for Experiment 1, to Michele Noe for help in testing subjects and transcribing tapes for Experiment 2, to Harriet Magen for help in transcribing tapes and tabulating data for Experiment 2, and to Tony Hatoun for help in analyzing the results of both experiments. We also thank Bruno Repp and Stefanie Shattuck-Hufnagel for helpful comments on an earlier version of this manuscript.

The development of experimental techniques for the elicitation of speech errors (see, for example, Motley & Baars, 1976) provides a new source of data, which, when used in conjunction with the evidence from naturally occurring errors, greatly facilitates the modeling of speech error generation. As Fowler (1983) points out, the experimental elicitation of speech errors permits tape recording of subjects' responses so that errors are less likely to be misheard or overlooked. Furthermore, experimental elicitation provides more thorough tests of hypotheses generated by the analysis of spontaneous error collections, especially when portions of the error pattern in the naturally occurring corpus are based on relatively few examples. On the other hand, there is always the danger of introducing influences in the laboratory that do not apply in more natural settings.

Shattuck-Hufnagel and Klatt (1979) analyzed collections of naturally occurring segment substitution errors and contrasted two types of error generation explanations. In the case of the first type of explanation, it is assumed that some segments are "strong" whereas others are "weak." Strong segments might be those that occur more frequently in the language, are acquired earlier, are unmarked in phonological theory, or are easier to articulate. The precise definition of segment strength is less important than the role strong segments play. Each segment substitution error has an intended, or target, segment source for the intruding error. The explanation predicts that strong segments appear more often as intrusions, whereas weak segments appear more often as targets in segmental substitution errors. A confusion matrix of such speech errors should thus be asymmetrical. This asymmetry would reflect the pattern of strength versus weakness of the segments involved.

In the case of the second type of explanation, on the other hand, the tendency of one segment (y) to substitute for another segment (x) would be related to their degree of similarity, but substitutions of x to y and y to x would be equally frequent. A confusion matrix of speech errors, if such errors arose as predicted by this type of explanation, should thus be symmetrical.

Shattuck-Hufnagel and Klatt (1979) analyzed the confusion matrix generated by 1620 substitution errors. The matrix proved to be asymmetrical. However, further analysis revealed that the asymmetry was due almost exclusively to four consonant segments /s, š, č, t/, such that errors of the type /s/ to /š/, /s/ to /č/, and /t/ to /č/ were all more frequent, respectively, than /š/ to /s/, /č/ to /s/, and /č/ to /t/. Once this source of asymmetry was removed, the confusion matrix of segmental errors was no longer significantly asymmetrical. However, the pattern of errors for /s, š, č, t/, which contributed most to the asymmetry of the matrix, could not be accounted for by stronger segments intruding more often, since, according to Shattuck-Hufnagel and Klatt, /š/ and /č/, for example, are less frequent and acquired later than /s/ (i.e., they are weaker), yet they intruded more often.

Shattuck-Hufnagel and Klatt proposed to account for the asymmetrical pattern of their confusion matrix in terms of a palatalization mechanism. They checked their corpus for factors that might "palatalize" the pronunciation of a non-palatal consonant (e.g., /s/ becoming /š/), but no difference was found between the source consonant environments in which palatalizing and non-palatalizing errors occurred. When the vowel environments of the target utterances were examined, Shattuck-Hufnagel and Klatt found that a palatalizing error occurred proportionately more often before a high vowel (e.g., /i/), but that this difference was not statistically significant. However, their calculations were based on a relatively small number of observations. The effect of the following vowel might indeed be reliable given a larger number of observations.

The authors concluded that the evidence from their data suggests that errors arise during the speech production process when one of two simultaneously available segments is mis-selected for a slot in an utterance, with the two segments generally being equally likely to be mis-selected.

Notice, however, that an explanation assuming that phonemes are not equal in strength, in particular one for which a strong segment is defined as a more frequent segment in the language,[1] does not receive a fair test in a corpus of naturally collected errors, because the prior probabilities of occurrence for all the segments are not equal. Imagine an explanation of the error generation process according to which segment strength is defined by segment frequency and similar segments are likely to substitute for one another. Such an explanation would predict that the rate that a frequent segment would be mispronounced given that it was intended would be lower than the rate that an infrequent segment would be mispronounced given that it was intended. So, for example, for /s/ and /š/, similar segments that might easily be confused, with /s/ as the stronger because it is more frequent, the rate of /s/ being mispronounced given that it was intended should be lower than the rate of /š/ being mispronounced. But the collection of naturally generated speech errors reflects the frequency of occurrence of phonemes in English, not just the error rates given that the phonemes are intended. Thus, since /s/ is much more frequent in the language than /š/, it will occur much more often as an intended phoneme so that it will occur more frequently as a target than /š/, even if its rate of occurrence as a target given that it was intended is lower. Furthermore, /š/, which is likely to substitute for /s/ because it is very similar, will appear more often as an intrusion than as a target, because of the high prior probability or frequency of /s/ as an intended phoneme. Note that the asymmetry arises because of the segmental similarity of /s/ and /š/ and a great discrepancy in their relative frequencies of occurrence in English. An experimental elicitation of errors using these segments in source utterances provides a good way of avoiding the problem of unequal frequencies of occurrence, because in the experimental situation, the intended utterances can be assigned equal prior probabilities. If frequency contributes to segment strength and if strength is a factor in the error generation process, then /s/ should appear more often as an intrusion and /š/ more often as a target, in the controlled experimental situation.

Intuitively, /s/ and /š/ seem quite similar, but similarity between two segments has not been clearly defined in the speech error context, although several investigators (Fromkin, 1971; MacKay, 1970; Nooteboom, 1969) have discussed the role of features in the error generation process. One way of defining segment similarity might be on the basis of the number of shared features. Clearly, the choice of a particular feature system can be crucial. Given a particular feature system, segments might need to share all or almost all features and only differ in some single individual feature (e.g., anterior or high) or type of feature (e.g., features for place of articulation) for errors to occur frequently. The role of segment similarity can be assessed in two ways: 1) Does the similarity of two segments in an utterance affect the tendency of subjects to make errors on those segments and 2) Given that an error has occurred, how similar is the intruding phoneme to its intended target?

Another issue is whether it is necessary for the target and intruding segments to be simultaneously available for a substitution error to occur that involves them. In a very broad view, the availability of a segment as an error source should be a function of its frequency in the language. A narrower view might define segment availability such that the source of an error need occur within a relatively constrained portion of the intended utterance. One could assess this narrower view of availability experimentally by seeing whether substitutions of y for x are more likely to occur when y is part of the stimulus.

Finally, it may be that Shattuck-Hufnagel and Klatt's observed asymmetry involving /s, š, č, t/ does reflect a palatalizing mechanism but there were insufficient observations in the environment of high vowels or palatal consonants. Again, the experimental situation permits a direct test of this hypothesis.

The basic technique for the experimental elicitation of speech errors involves what Baars (1980) calls the "competing plans framework." Essentially, the subject is given two alternative plans for the production of an utterance and is required to make a rapid response. For example, the subject might see the series of word pairs "give book, go back, get boot, bad goof" flashed rapidly on a screen. Notice that the fourth word pair, the test pair, "bad goof" involves a reversal of the initial consonant pattern found in the first three pairs, the bias pairs. After the test pair, at the sound of a buzzer, the subject would be expected to say the now-occluded final pair as quickly as possible. Under these conditions, a number of subjects will produce a speech error and may even spoonerize the test pair, reversing the initial consonants, and say "gad boof" instead.

We adapted this basic technique for the purposes of our study. Since previous work (Baars, Motley, & MacKay, 1975) has shown that there is output monitoring for the lexical status of spoonerized words (e.g., that "gad boof," which contains two non-lexical items, will occur less often as an error for "bad goof" than "darn bore," which contains two lexical items, will occur as an error for "barn door" in a similar sequence), we chose pairs of nonsense CV syllables as stimuli.[2] In pilot work, we found that subjects tended to make a greater number of errors when they were asked to pronounce both the bias and test items than when they pronounced only the test items. Hence, we required subjects to pronounce all of the items flashed before them on a screen.[3] Furthermore, pilot work indicated that when the bias pairs had a consistent vowel pattern (e.g., compare the bias series "right lean, ripe leap, ride leak" with the one given above), more errors tended to occur than when the vowel pattern was inconsistent (see also Dell, 1984). Thus, we restricted our bias pairs to those with consistent vowel patterns. We created our CV stimuli from the four consonants in Shattuck-Hufnagel and Klatt's data base that had been responsible for the initial asymmetry /s, š, č, t/, plus the additional consonant phoneme /θ/. The addition of /θ/ allowed us to test whether similarity, defined as a single feature difference, depends on a specific feature, since the consonants in the pairs /s, č/ and /θ, t/ differ on the single feature <u>continuant</u>, according to Chomsky and Halle (1968), whereas the consonants in the pair /s, θ/ differ on the single feature <u>strident</u>. The consonant /θ/ also provides another relatively infrequent, but non-palatal phoneme to test against the infrequent palatal set /š, č/. We chose the vowels /a, i, u for the test set, so as to be able to assess whether vowel height, high /i, u/ versus low /a/, or vowel height and frontness, front high /i/ versus /a, u/, might be the possible source of palatalizing errors.

Experiment 1

In Experiment 1, pairs of CV nonsense stimuli were presented visually, and subjects were asked to read all presented items as rapidly as possible.

Method

Materials. Using the set of consonant phonemes /s, š, č, t, θ/, written as *s*, *sh*, *ch*, *t*, and *th*, respectively, and the set of vowels /a, i, u/, we constructed pairs of CV nonsense syllables. Since we eliminated pairs with matched consonants (e.g., *ta ti*) as well as those with matched vowels (e.g., *sa ta*), there were twenty possible consonant permutations and six possible vowel permutations for a total of 120 test stimuli. A set of 120 filler pairs of CV nonsense syllables was analogously constructed using another set of consonants /r, l, b, v, m/ and the same set of vowels /a, i, u/.

Design. Each of the 120 test stimuli was preceded by three identical bias pairs of nonsense syllables that were constructed analogously to the test CV pair set and in which the order of the vowels was preserved but that of the consonants was switched. For example, for the test stimulus *su ti*, the presentation order was *tu si*, *tu si*, *tu si*, *su ti*. In order to prevent subjects from anticipating a switch after three identical CV nonsense pairs, 30 of the test CV nonsense syllables were also presented as distractors in groups of four (e.g., *tu si*, *tu si*, *tu si*, *tu si*), 30 in groups of three, 30 in pairs, and 30 singly. The 120 filler CV nonsense pairs also served to divert subjects' attention from the test stimuli consonants and pattern of presentation. Thirty of the filler CV nonsense syllables were presented in groups of four, 30 in groups of three, 30 in groups of two, and 30 singly. For half the trials with the filler syllables, the last item preserved the consonant order (e.g., *ra li*, *ra li*, *ra li*, *ra li*) and for half the trials the last item reversed the consonant order (e.g., *ra li*, *ra li*, *ra li*, *la ri*). The presentation of the test stimuli, distractors, and filler sequences was in pseudorandom order with the constraint that there were four test sequences, four filler sequences, and four distractor sequences in every block of twelve sequences. There was a total of 1080 pairs of CV nonsense syllables presented to subjects.

Subjects. Thirteen men and women participated in the experiment. Four were volunteers from the Haskins Laboratories staff (who were relatively knowledgeable phonetically), and nine were Yale University undergraduates receiving course credit for their participation. (Five additional subjects [one volunteer and four students] were tested, but their data were not analyzed because they failed to read a substantial number of the syllable pairs, and it was often not possible to determine what syllable pair they were responding to when they did utter something.)

Apparatus and procedure. The pairs of CV nonsense syllables were projected under program control onto the self-refreshing screen of a Decgraphic 11 GT-40 computer terminal hooked up to a PDP 11/45 computer at the rate of two syllable pairs a second. Subjects were asked to pronounce each syllable pair aloud as accurately as possible. During this task, subjects listened to white noise presented over Grason-Stadler TDH 39-300Z headphones in order to encourage them to speak up as loudly as possible and to minimize their ability to monitor their own utterances. Subjects' responses to the stimuli were recorded via a Sony F-27S microphone onto a Sony cassette tape recorder model TC-110B for later analysis.

Subjects were told that the nonsense syllables they would see would be composed of three vowel sounds, spelled as *i*, *a*, and *u*. They were instructed to pronounce the letter *i* as /i/ as in the word *eat*, the letter *a* as /a/ as in the word *father*, and the letter *u* as /u/ as in the word *boot*. They were also told to pronounce the letter pair *th* as in the word *think*, *sh* as in *shoe*, and *ch* as in *church*. Subjects were then shown CV nonsense syllable pairs typewritten on a sheet of paper and asked to read them aloud. Their pronunciation was checked, and if they did not pronounce the letters as instructed, they were asked to do so. There were 29 CV nonsense pairs from the filler set presented first to subjects as practice with the computer apparatus.

Results

Subjects' responses to all 1080 CV stimulus pairs were transcribed by one listener and then checked by another. Across the 13 subjects, there were 185 disagreements (1.3%), which were resolved by relistening to the disputed pairs

Table 1

Feature Differences Separating Consonants in a Pair and Error Frequencies for Test Stimuli in Experiment 1 as a Function of Consonant Pair and Vowel Pair

Feature Differences			Vowel Pair						
C&H	B&G	Consonant Pair	ai	ia	au	ua	ui	iu	Total
1	1	sh - ch	7	7	3	9	8	5	39
		ch - sh	7	5	4	8	6	7	37
1	3	t - th	3	3	2	1	3	0	12
		th - t	2	4	2	2	3	1	14
1	1	s - th	3	2	1	2	2	1	11
		th - s	3	3	5	4	3	7	25
2	1	s - sh	1	2	6	2	3	1	15
		sh - s	3	4	5	8	8	7	35
2	2	t - s	4	2	3	1	1	3	14
		s - t	2	3	1	4	0	0	10
3	1	sh - th	6	5	3	5	2	7	28
		th - sh	8	4	2	4	6	10	34
3	2	t - ch	1	2	4	3	2	2	14
		ch - t	2	3	5	2	4	2	18
3	2	s - ch	2	0	7	3	3	3	18
		ch - s	3	6	3	4	1	2	19
4	3	t - sh	3	2	4	2	1	4	16
		sh - t	3	5	0	7	4	1	20
4	2	ch - th	2	5	4	4	3	4	22
		th - ch	5	2	5	6	5	5	28
		TOTAL	70	69	69	81	68	72	429

until a consensus was reached. A response was scored as an error if the pair deviated in any way from the stimulus; thus, null responses were scored as errors. The results for the 120 test stimuli are summarized in Table 1 in terms of error frequencies as a function of consonant pair and vowel pair.

As is clear from Table 1, the pairs did not have consistent effects on error rates. An analysis of variance was conducted on the error data summed across vowel pairs in order to determine the significant effects due to consonant pairs. Two factors were included in the analysis, one for the ten different combinations of consonants, and the second to assess the effect of consonant frequency on error rates, such that the first permutation of the consonant pair had the more frequent of the two consonants preceding the less frequent consonant (with frequency determined by Dewey, 1923), and the second permutation had the less frequent consonant preceding the more frequent one, as revealed in the ordering of Table 1. Both main effects were significant. The consonant pairs were significantly different from one another, *F*(9,108)=6.89, *p*<.0001, and consonant pairs for which the less frequent consonant preceded the more frequent consonant had a significantly greater number of errors, *F*(1,12)=5.76, *p*=.0335. The interaction of consonant pairs and frequency was not significant, *F*(9,108)=1.50, *p*=.1560.

Feature analysis 1. A further analysis was performed on the same data in order to test the hypothesis that the number of feature differences between each consonant in a target pair was crucial in determining the error rate. Since there are a variety of competing feature analyses and since the choice of a single feature system could bias our results, we chose to contrast two phonetic feature systems: the well-known system devised by Chomsky and Halle (1968), henceforth *C & H*, and another one derived from a corpus of speech errors in English and German by van den Broecke and Goldstein (1980), henceforth *B & G*. First, the consonant pairs were divided into four feature difference classes according to *C & H* (see Table 1), and errors were averaged across consonant pairs in each class. The main effect of feature difference class was not significant, *F*(3,36)=1.09, *p*=.3672. Furthermore, the error rate did not monotonically increase or decrease with the number of feature differences, and the error rate for the consonant pair *sh-ch* differed greatly from that for *th-t*, though both consonant pairs differ on the same single feature.

Next, the consonants were divided into three feature difference classes according to *B & G* (see Table 1). With this feature set, the main effect of feature difference class was significant, *F*(2,24)=14.22, *p*=.0002. The mean number of errors per subject for consonant pairs differing on one feature was 2.2, on two features, 1.4, and on three features, 1.2.

Substitution errors. A separate analysis was made of substitution errors, in which the correct consonant in a syllable of a test stimulus was replaced by another consonant in the stimulus set. The resulting confusion matrix is presented in Table 2.

In order to determine whether the relative frequency with which each consonant segment intrudes is the same as the frequency with which it appears as a target, we computed a χ statistic comparing the two distributions and found that they were in fact significantly different from one another, χ (4)=69.1, p<.01. One striking discrepancy between the previous study by Shattuck-Hufnagel and Klatt (1980) and ours concerns the asymmetrical pattern of

Table [illegible]

Substitution Errors in Experiment 1 as a Function of Target Consonant and Intrusion Consonant

	Target					
Intrusion	T	S	SH	CH	TH	TOTAL
T	-	10	6	5	27	48
S	6	--	54	6	10	76
SH	4	28	--	49	5	86
CH	6	6	26	--	18	56
TH	5	5	10	9	--	29
TOTAL	21	49	96	69	60	295

substitutions involving *sh* and *s*. In the earlier study, there were more replacements of *s* by *sh* than vice versa, whereas the opposite was found in the present study. This discrepancy may be attributable in part to visual factors. Perhaps, consonant segments that contain the same letters (*sh*/*s* and *th*/*t*) are particularly likely to be confused, especially in the direction of letter deletion. An analysis that eliminates such confusions, by combining the *sh* and *s* segments and the *th* and *t* segments, yields a marginally significant difference between the target and intrusion distributions, $\chi\ (2) = 4.8$, $p < .10$.

Frequency analysis. To determine whether the incidence of errors for each target consonant phoneme is related to the log frequency of that segment in English, we computed a Pearson Product-Moment correlation coefficient relating the frequency with which each of the five consonants occurred as a target to its log frequency in English (Dewey, 1923). As expected according to the strength explanation, there was a negative correlation, although it did not reach standard levels of statistical significance, $r(3) = -.696$, $p > .10$. A significant negative correlation was found when the frequency analysis of Shattuck-Hufnagel and Klatt (1979) was used instead of that of Dewey (1923), $r(3) = -.887$, $p < .05$. This new frequency analysis, henceforth the content count, was derived from the speech sample of Carterette and Jones (1974) and includes only content words, not function words or common bound morphemes.*

A similar analysis was conducted to compare intrusion frequency and log frequency in the language. The correlations in this case were not significant for the Dewey (1923) count, $r(3) = .284$, $p > .10$, nor for the content count, $r(3) = -.054$, $p > .10$.

In view of the high correlations for target frequency and despite the low correlations for intrusion frequency, frequency in the language in addition to visual confusions may be a source of the asymmetry in intrusions noted earlier. In order to test this hypothesis, for the ten consonant pairs (e.g., *ch*-*t*), we compared how often the more frequent phoneme intruded for the less frequent phoneme (*t* for *ch*) rather than vice versa (*ch* for *t*). For one test we used the Dewey count, which yielded a significant difference, $t(9) = 2.41$, $p < .05$, and for a second test we used the more recent content count, which was not significant $t(9) < 1$. By both counts, the more frequent phoneme in the pair intruded more often on the average than did the less frequent phoneme, in accord with a strength explanation of speech errors.

Feature analysis 2. A second feature analysis was performed on the substitution data to see whether more substitutions of y for x occur when x and y differ by a single phonetic feature than when they differ by more. For the C & H features, the mean number of substitution errors involving a change of one feature was 20, of two features 24, of three features 6, and of four features 9. Clearly though one- and two-feature changes are more frequent than three- and four-feature changes, there is not a monotonic decrease in the number of substitution errors as the number of feature changes increases. Indeed, sh-ch and th-t, which differ on the same single feature according to C & H, show mean substitution rates of 38 and 16, respectively. Furthermore, there are complementary asymmetries in the substitution rates for these two pairs (see Table 2) such that the feature change to [+ continuant] involves fewer errors for the pair t-th but more errors for the pair ch-sh.

For the B & G features, the mean number of substitution errors involving a change of one feature was 23, of two features, 8, and of three features, 10. Although there is not a perfect monotonic decrease in the number of substitution errors as the number of feature changes increases, it is clear that the single feature substitution errors are most frequent.

Availability analysis. A further analysis was performed on the substitution errors to assess the role of segment availability. We determined the number of times a substitution error of y for x occurred in the environment of y (i.e., how often did the intrusion phoneme /t/ occur for the target phoneme /s/ when the test consonant pair was t-s or s-t). By comparing that number to the overall number of y for x substitutions, we determined the percentage of times that a substitution occurred when the error was part of the intended utterance (see Table 3). For substitution errors of y for x, y was part of the intended utterance 47.5% of the time. Since x was paired with phonemes other than y three times as often as it was paired with y, the appropriate chance percentage is 25%. Hence, segment availability in the stimulus does seem to influence error rate. However, it clearly is not necessary for the intruding phoneme to be part of the intended utterance, since the majority of the substitutions of y for x occur when y is not part of the intended utterance, defined narrowly here as the test CV nonsense syllable pair.

Furthermore, phoneme frequency seems to influence the importance of availability. When the direction of the substitution error involves a change from a relatively more frequent (strong or +) to a relatively less frequent (weak or -) phoneme (see Table 3), then it is more important that the infrequent segment be available, than when the direction of the substitution involves a change from a relatively weak to a relatively strong phoneme. Thus, by the Dewey count of phoneme frequency, when a change involves strong (+) to weak (-), the weak segment is available 58.1% of the time, whereas when the change involves weak (-) to strong (+), the strong segment is available only 41.6% of the time, $t(9)=3.19$, $p<.05$. The same pattern obtains with the content count (53.8% from strong (+) to weak (-), 42.3% from weak (-) to strong (+)), although the latter set of differences is not significant, $t(9)<1$.

On the other hand, the availability of the intruding phoneme did not vary regularly with the number of feature differences separating each consonant pair. By the C & H feature set, the intruding phoneme was available 42.6% of the time when there was a single feature difference between the consonants in a pair, 41.8% of the time when there were two feature differences, 55.3% of the time when there were three feature differences, and 70.3% of the time when there were four feature differences. Although this pattern suggests the

---

Table 3

Relative Frequency of Target Phoneme (x) and Intruding Phoneme (y) and Percentage of Errors of the Type x Changes to y When y Was Available in the Stimulus in Experiment 1

Target Phoneme	Intruding Phoneme	Relative Freq. Dewey		Content		Number of x to y Errors y available	Total	%
x	y	x	y	x	y			
sh	ch	+	-	-	+	13	26	50.0
ch	sh	-	+	+	-	24	49	49.0
t	th	+	-	+	-	2	5	40.0
th	t	-	+	-	+	5	27	18.5
s	th	+	-	+	-	3	5	60.0
th	s	-	+	-	+	5	10	50.0
s	sh	+	-	+	-	15	28	53.6
sh	s	-	+	-	+	16	54	29.6
t	s	+	-	+	-	4	6	66.7
s	t	-	+	-	+	6	10	60.0
sh	th	+	-	-	+	8	10	80.0
th	sh	-	+	+	-	3	5	60.0
t	ch	+	-	+	-	2	6	33.3
ch	t	-	+	-	+	2	5	40.0
s	ch	+	-	+	-	3	6	50.0
ch	s	-	+	-	+	3	6	50.0
t	sh	+	-	+	-	4	4	100.0
sh	t	-	+	-	+	4	6	66.7
ch	th	+	-	-	+	7	9	77.8
th	ch	-	+	+	-	11	18	61.1
				TOTAL		140	295	47.5

---

possibility that it is more important that the intruding phoneme be available when consonant pairs differ by three or more features, it is not confirmed in the pattern of availability for the B & G features. In that case, the intruding phoneme was available 46.5% of the time when the consonants in a pair differed on a single feature, 57.6% of the time when they differed on two features, but only 35.7% of the time when the consonants differed on three features.

Discussion

The results of Experiment 1 show that the likelihood of an error occurring for a given segment in a test pair depends in part on the relative frequency in English of the individual segments in the pair. Thus, the matrix generated by the substitution errors showed significant asymmetry. There was a high negative correlation between the frequency of an error occurring for a target segment and its log frequency of occurrence in English as well as evidence that a more frequent segment is more likely to intrude for a less frequent segment than vice versa.

Segment similarity clearly influences the generation of speech errors, although the pattern of errors and substitutions is more interpretable when segment similarity is based on the B & G rather than the C & H feature set.

Finally, availability of the source segment along with the target segment within the intended utterance, although important, does not seem to be a necessary factor, but its role increases when the intended segment is higher in frequency than the one that replaces it.

In order to assure that the results of Experiment 1, in which the stimuli were visually presented, were not an artifact of the visual modality, we redesigned our materials for auditory presentation in Experiment 2.

Experiment 2

Tongue twisters (e.g., "she sells sea shells") often result from conflicting vowel and consonant patterns. For example, there is an ABBA (/š/-/s/-/s/-/š/) consonant pattern and an ABAB (/i/-/ɛ/-/i/-/ɛ/) vowel pattern in the well-known tongue twister cited above. Our CV nonsense test syllables were presented auditorily to subjects in this tongue twister format, four syllables at a time, such that the consonant pattern of presentation was ABBA and the vowel pattern ABAB, and subjects were asked to repeat the sequence of four syllables as quickly and as accurately as possible.

Method

Stimuli. The test consonant phonemes /s, š, č, t, θ/ and vowels /a, i, u/ of Experiment 1 were used in Experiment 2. Each of the possible CV nonsense pairs (eliminating all identical consonant and identical vowel possibilities) was joined with a CV nonsense pair in which the order of the consonants changed but the vowels remained the same (e.g., sa tu ta su). There was a total of 120 such four-syllable stimuli. Each of the original 15 syllables (5 consonants x 3 vowels) was recorded by one of the investigators (AGL), digitized at 20 kH and stored on tape. All of the four-syllable nonsense CV stimuli were thus produced from the same original 15 syllables. There were 300 ms between syllables in a four-syllable string and a 5 s ISI between stimuli. There were no distractor or filler sequences.

Design. The stimuli were presented in pseudorandom order in six blocks of 20 each with the following constraints: No consonant occurred on two successive trials, each of the 20 consonant pairs occurred once in each block, and each vowel pair occurred once with each consonant pair in the test and either 3 (4 pairs) or 4 (2 pairs) times per 20-trial block.

Subjects. Eighteen men and women from the University of Colorado participated in the experiment and received course credit in an introductory psychology class.

Apparatus and procedure. The stimuli were transmitted to the subject binaurally through a pair of Telephonics earphones (Model TDH-39). The stimulus tape was played with a TEAC A-3300S tape recorder at a comfortable listening level. The subjects spoke into a Superscope Model EC-1 condenser microphone that was attached to an optisonics Sound-O-Matic II cassette tape recorder.

The subjects were told that they would hear a series of four-syllable sequences. They were instructed that the four syllables in each sequence would all be composed of a consonant sound followed by a vowel sound and that the consonants would always be presented in an ABBA pattern, and the vowels would always be in an ABAB pattern. They were given as an example the four-syllable sequence _ta_-_si_-_sa_-_ti_, which has a _t_-_s_-_s_-_t_ (or ABBA) consonant pattern and an _a_-_i_-_a_-_i_ (or ABAB) vowel pattern. The subjects were further told that there were only five different initial consonants (_s_ as in _sigh_; _t_ as in _tie_; _th_ as in _thigh_; _sh_ as in _shy_; and _ch_ as in _child_) and only three different vowels (/a/ as in _cot_; /i/ as in _eat_; and /u/ as in _boot_).

The subjects' task was to repeat aloud into the microphone each four-syllable sequence they heard as quickly as possible without making errors. They were told to try to say all four syllables and guess if necessary. They were instructed not to worry if they made a mistake or had trouble repeating a sequence but to listen carefully for the sequence following the one they missed and to try and keep up with the tape. The subjects were then given three practice trials spoken by the experimenter (_sa_-_ti_-_ta_-_si_; _chu_-_tha_-_thu_-_cha_; _shi_-_su_-_si_-_shu_).

Results

Subjects' responses to all 120 test stimulus quadruples were transcribed by one listener and then checked by another. Across the 18 subjects, there were 340 discrepancies (3.9%), which were resolved by a third listener. However, since a great number of these disagreements involved confusions of /θ/ and /f/ and since /f/ was not a possible stimulus, all responses of /f/ were replaced by /θ/ (there were 718 /f/ responses [8.3%] that were replaced in this way). Each syllable was scored separately and was determined to be an error if it deviated in any way from the stimulus. The results for the 120 test stimuli are summarized in Table 4 in terms of error frequencies as a function of consonant pair (ABBA) and vowel pair (ABAB).

As in Experiment 1, the vowel pairs did not have consistent effects on error rates. An analysis of variance was conducted on the error data summed across vowel pairs to assess the effects due to consonant pairs. The consonant pairs differed significantly from one another, $F(9,153)=14.17$, $p<.0001$, and the quadruples for which the less frequent sound was heard first had significantly more errors, $F(1,17)=15.92$, $p=.0009$.

Feature analysis 1. As for Experiment 1, the consonant pairs were first divided into four feature-difference classes by the _C & H_ feature system (see Table 4), and errors were averaged across consonant pairs in each class. The main effect of feature-difference class was marginally significant, $F(3,51)=2.59$, $p=.0632$, but the error rate again did not monotonically increase or decrease with the number of feature differences.

Next, the consonant pairs were divided into three feature difference classes by the _B & G_ feature system (see Table 4). The main effect of feature difference class was significant, $F(2,34)=16.24$, $p<.0001$. The mean number of errors per subject for consonant pairs differing on one feature was 11.2, on two features, 9.2, and on three features, 8.1.

Table 4

Feature Differences Separating Consonants in a Pair and Error Frequencies in Experiment 2 as a Function of Consonant Pair (ABBA) and Vowel Pair (ABAB)

Feature Differences			Vowel Pair						
C&H	B&G	Consonant Pair	ai	ia	au	ua	ui	iu	Total
1	1	sh - ch	42	40	32	39	36	40	229
		ch - sh	31	45	39	37	51	43	246
1	3	t - th	21	16	17	24	17	20	115
		th - t	17	27	26	35	17	18	140
1	1	s - th	22	24	21	15	27	23	132
		th - s	14	32	20	27	12	32	137
2	1	s - sh	41	41	36	45	31	30	224
		sh - s	37	45	42	39	42	43	248
2	2	t - s	18	32	13	23	27	35	148
		s - t	24	23	24	16	28	18	133
3	1	sh - th	36	27	31	31	38	27	190
		th - sh	43	29	19	37	47	33	208
3	2	t - ch	34	27	21	18	20	32	152
		ch - t	33	28	34	34	29	40	198
3	2	s - ch	29	25	32	31	21	33	171
		ch - s	20	37	21	30	37	39	184
4	3	t - sh	29	17	40	30	22	24	162
		sh - t	23	32	29	25	27	27	163
4	2	ch - th	20	23	24	33	23	29	152
		th - ch	20	31	23	31	39	42	186
		TOTAL	554	601	544	600	591	628	3518

Table 5

Substitution Errors in Experiment 2 as a Function of Target Consonant and Intrusion Consonant

	Target					
Intrusion	T	S	SH	CH	TH	TOTAL
T	---	132	135	259	150	676
S	95	---	289	143	188	715
SH	63	171	---	112	69	415
CH	166	162	312	---	126	766
TH	131	151	144	134	---	560
TOTAL	455	616	880	648	533	3132

Substitution errors. As in Experiment 1, a separate analysis was made of the substitution errors in which the correct consonant sound was replaced by another consonant in the stimulus set (see Table 5). To evaluate the extent to which the relative frequency that each consonant segment intruded corresponds to the frequency that it appeared as a target, we computed a χ statistic comparing the two distributions and found that they were in fact significantly different from each other, $\chi(4)=391.8$, $p<.01$, as in Experiment 1. Also in agreement with Experiment 1, but unlike the study by Shattuck-Hufnagel and Klatt (1980), we found more replacements of sh by s than vice versa.

Frequency analysis. As for Experiment 1, we computed two sets of correlation coefficients to determine the relation between the log frequency in the language of a given consonant segment and its frequency of occurrence as a target or intrusion. For targets, the correlations were negative, as expected, but nonsignificant for both the Dewey, $r(3)=-.352$, $p>.10$, and the content count, $r(3)=-.658$, $p>.10$. For intrusions, the correlations were positive but not significant, for Dewey, $r(3)=.331$, $p>.10$, and for the content count, $r(3)=.505$, $p>.10$. To evaluate whether frequency in the language may account for the asymmetry in intrusion errors, we compared how often the more frequent phoneme in a pair intruded for the less frequent phoneme rather than vice versa. The more frequent phoneme intruded more often on the average for both counts. This difference was significant by the content count, $t(9)=3.20$, $p<.05$, but not by the Dewey count, $t(9)<1$.

Feature analysis 2. For the C & H features, the mean number of substitution errors involving a change of one feature was 174, of two features 172, of three features 157, and of four features 114. Although substitution errors monotonically decreased as feature differences increased, again, sh-ch and th-t, which differ on the same single feature according to C & H, show mean substitution rates of 212 and 140, respectively.

For the B & G features, the mean number of substitution errors involving a change of one feature was 180, of two features, 152, and of three features, 120. Again, there is a monotonic decrease as the number of feature differences increases.

Availability analysis. For substitutions of y for x, y was part of the intended utterance 41.6% of the time (see Table 6), a percentage which is substantially higher than that expected on the basis of chance alone (25%).

Phoneme frequency again appears to have an effect on the importance of availability. When the direction of substitution goes from a strong (+) to a weak (-) segment, the weak segment is available 47.9% of the time by the content count, and when the direction of substitution goes from a weak segment (-) to a strong segment (+), the strong segment is available 37.3% of the time by the content count, $t(9)=2.93$, $p<.05$. The same pattern obtains with the Dewey count (42.1% from strong (+) to weak (-), 41.0% from weak (-) to strong (+)), although the latter set of differences was not significant, $t(9)<1$.

We found only a slight trend indicating that the intruding phoneme is less available when consonant pairs differ by a single feature than when they differ by more features for either feature set. For C & H, the intruding phoneme was available 37.9% of the time when there was a single feature difference between the consonants in a pair, 45.7% of the time when there were two feature differences, 42.2% of the time when there were three feature differences, and 42.4% of the time when there were four feature differences.

Table 6

Relative Frequency of Target Phoneme (x) and Intruding Phoneme (y) and Percentage of Errors of the Type x Changes to y When y Was Available in the Stimulus in Experiment 2

Target	Intruding	Relative Freq.				Number of x to y Errors		
Phoneme	Phoneme	Dewey		Content		y available	Total	%
x	y	x	y	x	y			
sh	ch	+	–	–	+	107	312	34.3
ch	sh	–	+	+	–	55	112	49.1
t	th	+	–	+	–	47	131	35.9
th	t	–	+	–	+	67	150	44.7
s	th	+	–	+	–	47	151	31.1
th	s	–	+	–	+	73	188	38.8
s	sh	+	–	+	–	96	171	56.1
sh	s	–	+	–	+	112	289	38.8
t	s	+	–	+	–	48	95	50.5
s	t	–	+	–	+	58	132	43.9
sh	th	+	–	–	+	52	144	36.1
th	sh	–	+	+	–	44	69	63.8
t	ch	+	–	+	–	88	166	53.0
ch	t	–	+	–	+	92	259	35.5
s	ch	+	–	+	–	72	162	44.4
ch	s	–	+	–	+	50	143	35.0
t	sh	+	–	+	–	40	63	63.5
sh	t	–	+	–	+	47	135	34.8
ch	th	+	–	–	+	47	134	35.1
th	ch	–	+	+	–	60	126	47.6
				TOTAL		1302	3132	41.6

For *B* & *G*, the intruding phoneme was available 40.8% of the time when there was a single feature difference between the consonants in a pair, 42.3% of the time when there were two feature differences and 42.0% of the time when there were three feature differences.

General Discussion

Comparison of Experiments 1 and 2. In Experiment 1 we considered the possibility that visual confusions contributed to the error pattern in that experiment. Experiment 2 provides an important control, since the stimuli in Experiment 2 were presented auditorily. Once we had corrected in Experiment 2 for the common auditory confusion of /f/ and /θ/, we found that the results of the two experiments were very similar. In fact, the Pearson Product-Moment correlation coefficient comparing the target phoneme frequencies in Experiments 1 and 2 showed a significant correlation, $r(3)=.915$, $p<.05$. When the intrusion phoneme frequencies of the two experiments were compared, we found a nonsignificant negative correlation, $r(3)=-.250$, $p>.10$. Although the exact patterns of intrusions for the two experiments did not correspond, the t tests reported earlier did show an effect of phoneme frequency on intrusions for both experiments. The error frequencies for the twenty consonant pairs themselves were also highly correlated in the two experiments. Recall that the test stimuli in Experiment 1 were CV nonsense pairs (e.g., *ta si*) and in Experiment 2 they were CV nonsense quadruples (e.g., *ta si sa ti*). Thus error frequencies for the stimuli in Experiment 1 were compared separately for the first two and the second two syllables in Experiment 2. When the number of errors for each CV pair in Experiment 1 was compared with the number of errors for the first two syllables of the CV nonsense quadruple in Experiment 2, the resulting correlation was statistically significant, $r(18)=.772$, $p<.01$. When the error frequencies of Experiment 1 were compared with those of the second two syllables of the CV nonsense quadruples of Experiment 2, the correlation was again statistically significant, $r(18)=.539$, $p<.05$. Finally, the error frequencies for the first two and the second two syllables of the CV nonsense quadruples in Experiment 2 were compared and again the correlation was significant, $r(18)=.772$, $p<.01$. Though visual confusions may have had a small effect on the error pattern of Experiment 1 and auditory confusions (most clearly those involving /f/ and /θ/) did occur in Experiment 2, the patterns of errors in the two experiments are clearly very similar. These patterns point to the importance of phoneme frequency in the error generation process.

The role of phoneme frequency. We can see two ways in which phoneme frequency had an effect on our results. In the first place, when we examined our data for errors of any type, we found in both experiments that consonant pair stimuli in which the first consonant was less frequent than the second (e.g., *ch-t*) tended to produce more errors than consonant pair stimuli in which the first consonant was more frequent (e.g., *t-ch*).

In the second place, when we examined substitution errors restricted to the test consonant set, we found that phoneme frequency in English showed a negative correlation with target frequencies. We also found, when we looked at the ten consonant combinations, that the more frequent phoneme of the pair was more likely to intrude as an error for the other member than vice versa. These findings lend support to an explanation of the error generation process in which phoneme strength is determined by phoneme frequency. Thus we find a negative correlation between target phoneme frequency and frequency of occurrence in English because more frequent or stronger phonemes are less likely to function as targets or mispronounced segments. On the other hand, more frequent or stronger phonemes are somewhat more likely to function as intrusions.

These effects of frequency emerge in the experimental elicitation of errors because we were able to control the prior probabilities of occurrence of the individual phonemes. With equal prior probabilities, we find an asymmetrical pattern of substitution errors. However, the asymmetrical pattern that emerges from our data is different from the one found initially by Shattuck-Hufnagel and Klatt: We find no evidence for a palatalizing mechanism, since we find more non-palatalizing (e.g., *sh* to *s*) than palatalizing (e.g., *s* to *sh*) substitution errors in both experiments.

There is always the danger in an experimental situation that some factor that does not operate in the spontaneous error generation process was introduced. We used nonsense syllables as stimuli rather than English words, in order to eliminate effects of lexical frequency and lexical bias in the error generation process, but nonsense syllables may behave differently than English words. For example, in an experiment designed to elicit speech errors, in which she had subjects read or recall tongue twisters, Shattuck-Hufnagel

(1982) found a differential pattern of errors in the recall condition when she compared CVC nonsense syllables CVC English words, CVC nonsense syllables embedded in short phrases, and CVC English words embedded in short phrases. Only the CVC English words embedded in short phrases showed a higher percentage of word-initial errors (as is found in naturally occurring speech errors), whereas all the other eliciting sets showed a higher percentage of word-final errors. However, this result was obtained largely through a reduction in the number of word-final errors in the CVC English words embedded in phrases as compared to the other conditions. Furthermore, since we used only CV syllables in our study, and we found very few vowel errors, our errors were almost entirely in word-initial position. Thus, we believe that the differences between our findings and those of Shattuck-Hufnagel and Klatt (1980) are due largely to the differences in prior probabilities of the phoneme targets and are not due to factors introduced by our experimental method or use of nonsense syllables.

In our view, phoneme strength is a function of phoneme frequency rather than ease of articulation, age of acquisition, or status in phonological theory of the phoneme in question. With respect to articulation, in comparing /s/ and /š/, Borden and Harris (1980) point out that "a wide range of openings beyond those for /s/ result in /š/ type sounds" (p. 121). So we see that articulation of the phoneme /s/ is more precise and therefore presumably more difficult (see also Anderson, 1942). In contrast, there are claims in the literature (e.g., Lester & Skousen, 1974) that /s/ is acquired earlier than /š/. Closer examination of the data reveals that children often produce an /s/-like fricative phoneme or stop where the adult model has an /s/ (Ferguson, 1978; Moskowitz, 1973) before they produce a phoneme for words in which the adult model has an /š/, probably because of the higher frequency of /s/. However, that correct articulation of /s/ is often acquired rather late is clear from reports of speech therapists (Anderson, 1942; Berry & Eisenson, 1947) and others (Ingram, Christensen, Veach, & Webster, 1980; Sander, 1972; Velleman, 1983) who attest to its difficulty. Finally, although in phonological markedness theory, as outlined by Chomsky and Halle (1968), /s/ is less marked than /š/, in a more general test of phonological markedness in the elicitation of speech errors, Motley and Baars (1975) did not find markedness to be a significant factor.

Frequency in the language is then for us the best index of a phoneme's strength. We believe that frequent phonemes are "stronger" than infrequent ones because they are the more common of highly overlearned motor patterns. In this view, we see single segment errors involving similar segments as examples of Norman's (1981) capture errors: "when a sequence being performed is similar to another more frequent or better learned sequence, the latter may capture control" (p. 6). The initial gestures relevant to the pronunciation of /s/ and /š/ are no doubt very similar, if not identical. It is easy to see how the gestures to produce an /š/ could be "captured" by the more frequent /s/ gestures.

Segment similarity. Do speech error rates or patterns of substitutions depend on minimal feature differences between consonant pairs? The answer to such a question depends on the feature system one chooses. Ideally, one would like to find that a system motivated on independent grounds, such as the one devised by Chomsky and Halle (1968), also captures in a principled way the structural relationships in speech errors. Indeed, van den Broecke and Goldstein (1980) compared a number of feature systems, along with the one they devised on the basis of English and German speech errors, and found that "feature systems designed without incorporating evidence from speech errors are all capable of showing meaningful structure in phonological speech errors as they occur" (p. 63). Nonetheless, segment similarity emerges as a significant effect in our data only when we use the B & G features to determine segment similarity. That the segment similarity effects in our data are best demonstrated by the B & G features, derived from the analysis of naturally occurring speech errors in English and German, suggests that the errors we find in our experimental situation are analogous to those occurring in collections of naturally occurring utterances.

Availability. When naturally occurring speech errors are analyzed, the assumption is often made that errors are most likely to occur when similar segments are simultaneously available. Yet the results of our experiments suggest that availability, here defined in narrow terms as a substitution of x for y when x is part of the stimulus, is important but not necessary, since the percentage of the x for y substitutions in both experiments that occur when x is part of the stimulus is substantially greater than the chance value but no greater than 50%. Indeed, the substitution errors in the corpus examined by Shattuck-Hufnagel and Klatt (1979) include 30% with no known source word. It is possible that the actual proportion of naturally occurring speech errors that have no source in the surrounding context might be higher than that estimated by Shattuck-Hufnagel and Klatt, and it might be wrong to assume in such cases that the intruding error was part of the intended utterance (see Harley, 1984, for a discussion of higher level non-plan-internal errors). Finally, we find that segment availability becomes increasingly important as the frequency of the intruded phoneme decreases and perhaps, to a lesser extent, as the featural similarity between the intruded and target phonemes decreases.

However, it is difficult to compare the relative magnitudes of the effects of phoneme frequency and availability (see Sechrest & Yeaton, 1982). Moreover, the influence of phoneme frequency on the importance of availability suggests that both effects may stem from the same activation mechanism. The frequency effect may be reflecting differences in the base activation levels of phonemes, whereas the availability effect may reflect transient increases in phoneme activation that result from being part of the intended utterance.[5]

Conclusions. The results of our two experiments provide support for an explanation of the speech error generation process in which a segment's strength is a function of its frequency of occurrence in English: Weak (or infrequent) segments tend to serve as targets whereas strong (or frequent) segments tend to serve as intrusions. The role of phoneme frequency is a consistently important one. Phoneme availability also plays a role, though perhaps more restricted than expected. Furthermore, availability may be reflecting the same activation mechanism responsible for the frequency effect. Finally, the notion that the segments that interact in speech errors are likely to be similar is best supported by our data when segment similarity is defined in terms of a feature set derived from naturally occurring speech errors.

References

Anderson, V. A. (1942). Training the speaking voice. New York: Oxford University Press.

Baars, B. J. (1980). On eliciting predictable speech errors in the laboratory. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the eye, ear, pen and hand* (pp. 307-318). New York: Academic Press.

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14*, 382-391.

Berry, M. L., & Eisenson, J. (1947). *The defective in speech*. New York: F.S. Crofts.

Borden, G. J., & Harris, K. S. (1980). *A speech science primer*. Baltimore: Williams & Wilkins.

Carterette, E. C., & Jones, M. H. (1974). *Informal speech*. Berkeley: University of California Press.

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

Cohen, A. (1973). Errors of speech and their implication for understanding the strategy of language users. In V. A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 88-92). The Hague: Mouton.

Dell, G. S. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 222-233.

Dewey, G. (1923). *Relative frequency of English speech sounds*. Cambridge, MA: Harvard University Press.

Ferguson, C. A. (1978). Fricatives in child language acquisition. In V. Honsa & M. J. Hardman-de-Bautista (Eds.), *Papers on linguistics and child language* (pp. 73-115). The Hague: Mouton.

Fowler, C. A. (1983). Review of V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand*. In *Linguistics, 19*, 819-840.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27-52.

Garrett, M. F. (1980). The limits of accommodation: Arguments for independent processing levels in sentence production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand* (pp. 263-271). New York: Academic Press.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science, 8*, 191-219.

Ingram, D., Christensen, L., Veach, S., & Webster, B. (1980). The acquisition of word initial fricatives and affricates in English by children two and six years. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child psychology* (Vol. 1, pp. 169-192). London: Academic Press.

Lester, L., & Skousen, R. (1974). The phonology of drunkenness. *Papers from the parasession on natural phonology* (pp. 233-239). Chicago: Chicago Linguistics Society.

MacKay, D. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8*, 323-350.

Moskowitz, A. J. (1973). The acquisition of phonology and syntax. In K. J. J. Hintikka, J. M. E. Moravcsik, & P. Suppes (Eds.), *Approaches to natural language* (pp. 48-84). Boston: D. Reidel.

Motley, M. T., & Baars, B. J. (1975). Encoding sensitivities to phonological markedness and transitional probability: Evidence from spoonerisms. *Human Communication Research, 1*, 353-361.

Motley, M. T., & Baars, B. J. (1976). Semantic bias effects on the outcomes of verbal slips. *Cognition, 4*, 177-187.

Nooteboom, S. G. (1969). The tongue slips into patterns. In A. G. Sciarone, A. J. van Essen, and A. A. van Raad (Eds.), Nomen Society, Leyden studies in linguistics and phonetics (pp. 114-132). The Hague: Mouton.

Norman, D. A. (1981). Categorization of action slips. Psychological Review, 88, 1-15.

Sanaer, E. K. (1972). When are speech sounds learned? Journal of Speech and Hearning Disorders, 37, 55-63.

Sechrest, L., & Yeaton, W. H. (1982). Magnitudes of experimental effects in social science research. Evaluation Research, 6, 579-600.

Shattuck-Hufnagel, S. (1982). Position of errors in tongue twisters and spontaneous speech: Evidence for two processing mechanisms? Working Papers, Speech Communications Group, Research Laboratory of Electronics, M.I.T., 1, 1-8.

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. Journal of Verbal Learning and Verbal Behavior, 18, 41-55.

Shattuck-Hufnagel, S., & Klatt, D. H. (1980). How single phoneme error data rule out two models of error generation. In V. A. Fromkin (Ed.), Errors in linguistic performance: Slips of the tongue, ear, pen and hand (pp. 35-46). New York: Academic Press.

van den Broecke, M. P. R., & Goldstein, L. (1980). Consonant features in speech errors. In V. A. Fromkin (Ed.), Errors in linguistic performance: Slips of the tongue, ear, pen and hand (pp. 47-65). New York: Academic Press.

Velleman, S. L. (1983). Children's production and perception of English voiceless fricatives. Unpublished doctoral dissertation, University of Texas at Austin.

Footnotes

[1] Motley and Baars (1975) found experimental evidence that consonant frequency in initial position affects the tendency of initial consonants in pairs of CVC nonsense words to interchange. Hence, frequency in the language seems like an appropriate initial index of phoneme strength.

[2] Although none of the CV nonsense pairs represented common lexical items as visually presented, six of them did represent common lexical items as pronounced: si = 'see'; shi = 'she'; ti = 'tea'; su = 'sue'; shu = 'shoe'; tu = 'two.'

[3] It is possible that this rapid reading procedure is influenced by articulatory interference of the type involved in tongue twisters as well as by the factors producing higher-level slips of the tongue. However, Cohen (1973) found that the pattern of speech errors induced via a rapid reading procedure was of a very similar nature to that of a naturally collected corpus.

[4] The rank order of the consonant phonemes by the Dewey (1923) count is t>s>sh>ch>th, whereas by the content count it is t>s>th>ch>sh.

[5] We are indebted to Marcel Just for making this point.

ON LEARNING TO SPEAK*

Michael Studdert-Kennedy†

Abstract. Every language, spoken or signed, deploys a large lexicon, made possible by permutation and combination of a small set of linguistic elements. In speech, rapid interleaving of the gestures that form these elements (consonants and vowels) leads to a complex acoustic signal in which the boundaries between elements are lost. However, for the child learning to speak, the initial task is not to recover these elements, but simply to imitate the sound pattern that it hears. Studies of "lipreading" in adults and infants suggest that imitation is mediated by an amodal representation, closely related to the dynamics of articulation, and that a left-hemisphere perceptuomotor mechanism specialized to make use of this representation develops during the first six months of life. By drawing on this specialized mechanism, the infant learns the recurrent patterns of acoustic structure and articulatory gestures from which linguistic segments must be presumed to emerge.

As a system of animal communication, language has the distinctive property of being open, that is, fitted to carrying messages on an unlimited range of topics. Human cognitive capacity is, of course, greater than that of other animals, but this may be a consequence as much as a cause of linguistic range. Other primate communication systems have a limited referential scope--sources of food or danger, personal and group identity, sexual inclination emotional state, and so on--and a limited set of no more than 10-40 signals (Wilson, 1975, p. 183). In fact, 10-40 holistically distinct signals may be close to the upper range of primate perceptual and motor capacity. The distinctive property of language is that it has finessed that upper limit, by developing a double structure, or dual pattern (Hockett, 1958).

The two levels of patterning are phonology and syntax. The first permits us to develop a large lexicon, the second permits us to deploy the lexicon in predicating relations among objects and events. My present concern is entirely with the first level. A six-year-old middle-class American child already recognizes some 13,000 words (Templin, 1957), while an adult's recognition vo-

---

*Human Neurobiology, 1983, 2, 191-195.

†Also Queens College and Graduate Center, City University of New York.

Acknowledgment. Preparation of this paper was supported in part by NICHD Grant HD-01994 to Haskins Laboratories. The argument of the paper extends and elaborates one first put forward as part of a paper read at a conference at the University of Connecticut, Storrs, in June 1980, and published in: W. H. Warren & R. E. Shaw (Eds.), (1985) Persistence and change: Proceedings of the First International Conference on Event Perception. Hillsdale, NJ: Lawrence Erlbaum Associates.

cabulary may be well over 100,000. Every language, however primitive the culture of its speakers by Western standards, deploys a large lexicon. This is possible because the phonology, or sound pattern, of a language draws on a small set (roughly between 20 and 100 elements) of meaningless units--consonants and vowels--to construct a very large set of meaningful units, words (or morphemes). These meaningless units may themselves be described in terms of a smaller set of recurrent, contrasting phonetic properties or features. Evidently, there emerged in our hominid ancestors a combinatorial principle (later, perhaps, extended into syntax) by which a finite set of articulatory gestures could be repeatedly permuted to produce a very large number of distinctively different patterns.

"Articulatory gesture" refers, at a gross level, to opening and closing the mouth. Repeated constriction of the vocal tract, somewhere between lips and glottis to form consonants, and repeated opening of the tract by lowering the jaw to form vowels, give rise to the basic consonant vowel syllable from which the sound patterns of all spoken languages are formed. The varying phonetic qualities of consonants and vowels are determined by the precise shape of the vocal tract through which sound--the buzz of vocal fold vibration or the hiss of air blown through a narrow constriction--is filtered. The shape of the resonating cavities of the vocal tract is determined by fine positioning of the articulators: raising, lowering, fronting or backing the tip, blade or body of the tongue, raising or lowering the velum, rounding or spreading the lips, and so on.

Thus, permutation and combination of some two dozen gestures provide "...a kind of impedance match between an open-ended set of meaningful symbols and a decidedly limited set of signaling devices" (Studdert-Kennedy & Lane, 1980, p. 35). Yet permutation and combination alone would not suffice for a flexible and open-ended system of communication, if the gestures were not executed rapidly enough to evade the limits of short-term memory and to match the natural rate of thought and action.

What this "natural rate" may be we do not know. But for English, at least, a typical rate of speech is of the order of 150 words/min. This reduces to roughly 10 to 15 phonemes (consonants and vowels)/s. As Cooper has remarked, such rates can be achieved "...only if separate parts of the articulatory machinery--muscles of the lips, tongue, velum, etc.--can be separately controlled, and if...a change of state for any one of these articulatory entities, taken together with the current state of others, is a change to...another phoneme.... It is this kind of parallel processing that makes it possible to get high-speed performance with low-speed machinery" (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967, p. 446). Thus, repeated use of a small set of interleaved gestures may not only expand the potential lexicon, but also ensure rapid execution of its elements.

Let me conclude this brief introduction by noting that the dual motoric structure of spoken language has no known parallel in any other system of animal behavior, except manual-facial sign languages. Over the past 15-20 years we have learned that American Sign Language (ASL), the first language of over 100,000 deaf persons, and the fourth most common language in the United States (Mayberry, 1978), is a fully independent language with its own characteristic formational ("phonological") structure and syntax (Klima & Bellugi, 1979). Whether signed language is a mere analog of spoken language or a true homolog, drawing on the same neural structures, we do not yet know--although studies of sign language deficits following left hemisphere lesion reveal remarkable parallels with aphasic deficits of spoken language users (e.g., Kimura, Battison, & Lubert, 1976; Poizner, Bellugi, & Iragui, in press).

In any event, my point here is simply that each ASL sign is formed by combining four intrinsically meaningless components: a hand configuration, a palm orientation, a place in the body space where it is formed, and a movement. These four classes of component, like the two segmental classes of spoken language (consonants and vowels), may also be described in terms of a smaller set of recurrent, contrasting features (e.g., Klima & Bellugi, 1979, Chapter 7). There are some fifty values, or "primes," distributed across the four dimensions and their combination in a sign follows "phonological rules," analogous to those that constrain the structure of a syllable in spoken languages. In short, both spoken and signed languages exploit combinatorial principles of lexical formation. Moreover, it would seem that short-term memory and cognitive capacity have constrained signed and spoken languages to similar rates of communication. For, although each ASL sign takes roughly three times as long to form as an English word, the proposition rates in the two languages are almost identical (Klima & Bellugi, 1979). This is possible because, while the phonological and syntactic structures of a spoken language are largely implemented by sequential organization over time, a signed language can exploit simultaneous manual and facial gestures distributed in space. Thus, both types of languages are grounded in a capacity for rapid, precise, and precisely coordinated movements of a small set of articulators.

In what follows, I shall have little further to say about signed languages. Here, I simply note two points. First, we do not talk with our toes, and we may doubt whether any imaginable system of human articulators, other than those of the hand and mouth, would be capable of the motor speed and precision necessary to implement language, as we know it. Second, whatever the evolutionary sequence may have been, the well-established (albeit imperfect) correlation between hemispheric specializations for language and manual praxis is, I assume, not mere coincidence. In all likelihood, the two modes of language draw on closely related neural structures.

I have dwelt so far on motor requirements. But there are perceptual demands also. If spoken language is indeed constructed from rapid sequences of consonants and vowels, the listener must somehow extract these recurrent elements from the signal. Yet, from the earliest spectrographic studies (Joos, 1948) it has been known that the acoustic flow of speech cannot be readily divided into an alphabetic sequence of invariant segments corresponding to the invariant segments of linguistic description. The reason for this is simply that we do not speak segment by segment, or even syllable by syllable. At any instant, the several articulators are executing a complex, interleaved pattern of movements, of which the spatio-temporal coordinates reflect the influence of several neighboring segments (The reader may test this by slowly uttering, for example, the words <u>call</u> and <u>keel</u>. The reader will find that the position of the tongue on the palate during closure for the first consonant, /k/, is slightly farther back for the first word than for the second.) The consequence of this imbricated pattern of movement is, of course, an imbricated pattern of sound, such that any particular acoustic segment typically specifies more than one linguistic segment, while any particular linguistic segment is specified by more than one acoustic segment (Fant, 1962; Liberman et al., 1967). This lack of isomorphism between acoustic and linguistic structure is the central unsolved problem of speech perception. Its continued recalcitrance is reflected in the fact that we are little closer to automatic phonetic transcription of speech now than we were thirty years ago (Levinson & Liberman, 1981).

Many different approaches to the problem have been proposed, but I will not review them here (see Studdert-Kennedy, 1980, for fuller discussion). Instead, I will attempt to recast the problem by setting aside, for the moment, the discrepancy between acoustic signal and linguistic description, and simply asking what we know about how a child learns to speak. I shall assume that, whatever the process, it is sufficiently general to permit the deaf child to learn to sign with as much ease as a hearing child learns to speak. I note, further, that when a child learns to sign or speak, it learns a specific dialect. That is to say, it gradually discovers, in the detailed acoustic or optic patterns of its caretakers' signals, specifications for a no less detailed pattern of motor organization.

Stated in this way, the problem becomes a special case of the general problem of imitation. Relatively few species imitate. The higher primates imitate general bodily actions, but vocal imitation is peculiar to a few species of songbirds, certain marine mammals, and humans. The capacity to imitate is evidently a rare, specialized capacity for discovering links between perceived movements and their corresponding motor controls.

We may gain insight into the bases of speech imitation from recent studies of "lip-reading" in adults and infants. That adults can learn to lip-read is, of course, a commonplace of aural rehabilitation, but the theoretical implications of this capacity have only recently begun to emerge. McGurk and MacDonald (1976) demonstrated that listeners' perceptions of a spoken syllable often change, if they simultaneously watch a video display of a speaker pronouncing a different syllable. For example, if listeners are presented with the acoustic syllable [ba] repeated four times, while watching a synchronized optic display of a speaker articulating [ba, va, ða, da], they will typically report the latter, optically specified sequence. That the effect is not simply a matter of visual dominance in a sensory hierarchy (Marks, 1978) is evidenced by the fact that certain combinations (e.g., acoustic [ba] with optic [ga] may be perceived as clusters ([bga] or [gba]), or even as syllables corresponding to neither display ([da]). Thus listeners' percepts seem to arise from a process by which two distinct sources of information, acoustic and optic, are actively combined at an abstract level where each has already lost its distinctive sensory quality. (For fuller discussion, see Summerfield, 1979).

Further evidence for a amodal representation of speech comes from a cross-modal study of the so-called suffix effect by Campbell and Dodd (1980). A standard finding of short-term memory studies is that listeners, recalling a list of auditorily presented words, recall those at the end better than those in the middle (recency effect). The effect is reduced if the list is presented graphically. Moreover, Crowder and Morton (1969) demonstrated that the effect could be abolished, or significantly reduced, if a spoken word was appended to the list, not for recall but simply as a signal to begin recall (suffix effect). Presumably, the suffix "interferes" in some way with the representation of recent items. That this representation is at some relatively "low," yet structured, level is argued by the facts that the effect (1) is unaffected by degree of semantic similarity between suffix and list, (2) is reduced if suffix and list are presented to opposite ears, (3) does not occur if the suffix is a tone or burst of noise.

Campbell and Dodd (1980) used this paradigm to test listeners' recall of digits, either lip-read (without sound) or presented graphically, with and without the spoken suffix, "ten" (heard, but not seen). They found significant recency and suffix effects for the lip-read, but not for the graphic, lists. In a complementary study, Spoehr and Corin (1978) demonstrated that a lip-read suffix reduced recall of auditorily presented lists. Evidently, speech heard, but not seen, and speech seen, but not heard, share a common representation. Moreover, the fact that Campbell and Dodd did not find a suffix effect for graphically presented lists suggests that this shared representation is not at some abstract, phonological level where spoken and written language converge. Rather, these studies, like that of McGurk and MacDonald (1976), hint at a representation in some form common to both the light reflected and the sound radiated from mouth and lips.

Consider, now, that infants are also sensitive to structural correspondences between the acoustic and optic specifications of an event. Spelke (1976) showed that four-month-old infants preferred to watch the film (of a woman playing "peekaboo," or of a hand rhythmically striking a wood block and a tambourine with a baton) that matched the sound track they were hearing. Dodd (1979) showed that four-month-old infants watched the face of a woman reading nursery rhymes more attentively when her voice was synchronized with her facial movements than when it was delayed by 400 ms. If these preferences were merely for synchrony, we might expect infants to be satisfied with any acoustic-optic pattern in which moments of abrupt change are arbitrarily synchronized. Thus, in speech they might be no less attentive to an articulating face whose closed mouth was synchronized with syllable amplitude peaks and open mouth with amplitude troughs than to the (natural) reverse. However, Kuhl and Meltzoff (1982) showed that four- to five-month-old infants looked longer at the face of a woman articulating the vowel they were hearing (either [i] or [a]) than at the same face articulating the other vowel <u>in synchrony</u>. Moreover, the preference disappeared when the signals were pure tones, matched in amplitude and duration to the vowels, so that the infant preference was evidently for a match between a mouth shape and a particular spectral structure. Similarly, MacKain et al. (1983) showed that five- to six-month-old infants preferred to look at the face of a woman repeating the disyllable they were hearing (e.g., [zuzi]) than at the synchronized face of the same woman repeating another disyllable (e.g., [vava]).

In both these studies, the infants' preferences were for natural structural correspondences between acoustic and optic information. Both studies hint at infant sensitivity to intermodal correspondences that could play a role in learning to speak. However, I am not suggesting that optic information is necessary, since the blind infant also learns to speak.[1] My intent rather is to gain leverage on the puzzle of imitation. What we need therefore is to establish that the underlying metric of auditory-visual correspondence is related to that of the auditory-motor correspondence required for an individual to imitate the utterances of another.

To this end we may note, first, the visual-motor link evidenced in the capacity to imitate facial expression and, second, the association across many primate species between facial expression and pattern of vocalization (Hooff, 1976; Marler, 1975; Ohala, 1983). Recently, Field et al. (1982) reported that 36-hour-old infants could imitate the "happy, sad and surprised" expressions of a model. However, these are relatively stereotyped emotional responses that might be evoked without recourse to the visual-motor link required for imitation of novel movements. More striking is the work of Meltzoff and Moore (1977) who showed that 12- to 21-day-old infants could imitate both arbitrary mouth movements, such as tongue protrusion and mouth opening, and (of particular interest for the acquisition of ASL) arbitrary hand movements, such as opening and closing the hand by serially moving the fingers. Here mouth opening was elicited without vocalization; but had vocalization occurred, its structure would, of course, have reflected the shape of the mouth. Kuhl and Meltzoff (1982) do, in fact, report as an incidental finding of their study that 10 of their 32 four- to 5-month-old infants "...produced sounds that resembled the adult female's vowels. They seemed to be imitating the female talker, 'taking turns' by alternating their vocalizations with hers" (p. 1140). If we accept the evidence that the infants of this study were recognizing acoustic-optic correspondences, and add to it the results of the adult lipreading studies, calling for a metric in which acoustic and optic information are combined, then we may conclude that the perceptual structure controlling the infants' imitations was specified in this common metric.

Evidently, the desired metric must be "...closely related to that of articulatory dynamics" (Summerfield, 1979, p. 329). Following Runeson and Frykholm (1981) (see also Summerfield, 1980), we may suppose that in the visual perception of an event we perceive not simply the surface kinematics (displacement, velocity, acceleration), but also the underlying biophysical properties that define the structure being moved and the forces that move it (mass, force, momentum, elasticity, and so on). Similarly, in perceiving speech, we perceive not only its "kinematics," that is, the changes and rates of change in spectral structure, but also the underlying dynamic forces that produce these changes. In other words, to perceive speech is to perceive movements of the articulators, specified by a pattern of radiated sound, just as we perceive movements of the hand, specified by a pattern of reflected light.

The close link, for the infant, between perceiving speech and producing it, is further suggested by a curious aspect of the study by MacKain et al. (1983), cited earlier. This is the fact that infants' preferences for a match between the facial movements they were watching and the speech sounds they were hearing was statistically significant only when they were looking to their right sides. Fourteen of the eighteen infants in the study preferred more matches on their right sides than on their left. Moreover, in a follow-up investigation of familial handedness, MacKain and her colleagues have learned that six of the infants have left-handed first or second order relatives. Of these six, four are the infants who displayed more left-side than right-side matches.

These results can be interpreted in the light of studies by Kinsbourne and his colleagues. Kinsbourne (1972) found that right-handed adults tended to shift their gaze to the right while solving verbal problems, to the left while visualizing spatial relations; left-handers tended to shift gaze in the same direction for both types of task, with each direction roughly equally represented across the subject group. Lempert and Kinsbourne (1982) showed that the effect was reversible for right-handed subjects on a verbal task: subjects who rehearsed sentences with head and eyes turned right recalled the sentences better than subjects who rehearsed while turned left. Thus, attention to one side of the body may facilitate processes for which the contralateral hemisphere is specialized.

Extending this interpretation to the infants of MacKain et al. (1983), we may infer that infants with a preference for matches on the right side, rather than the left, were revealing a left hemisphere capacity for recognizing acoustic-optic correspondence in speech. If, further, the metric specifying these correspondences is the same as that specifying the auditory-motor correspondences necessary for imitation (as was argued above), we may conclude that five- to six-month-old infants already display a speech perceptuo-motor link in the left hemisphere.

How early this link may develop we do not yet know. However, Best et al. (1982), testing, two-, three-, and four-month-old infants dichotically, in a cardiac habituation paradigm, found a right-ear advantage for speech and a left-ear advantage for music in the three- and four-month olds, but only a left-ear advantage for music in the two-month olds. We may suspect, then, that the perceptual component of the speech link begins to develop between the second and third months of life. By five to six months, close to the onset of babbling, the motor component is beginning to emerge. By the end of the first year, as babbling fades, the infant would be equipped with the perceptuo-motor mechanisms necessary for imitating the sounds of the language it is going to learn.

In conclusion, let me recall the paradoxical discrepancy between the speech signal and its linguistic description with which I began. The approach to imitation I have sketched deliberately sidesteps this problem. Yet it may ultimately contribute to its solution by focusing on the infant for whom the discrepancy does not yet exist, for the simple reason that the infant has not yet learned the phonetic categories of its language. Tracing the process by which the recurrent patterns of infant articulation coalesce into categorical linguistic units, evidenced by spoonerisms and other adult speech errors (Shattuck-Hufnagel, 1979) is a task for the future. However, the task may be easier, if we see it as a problem in the development of a unique mode of motor control, characteristic of human language.

References

Best, C. T., Hoffman, H., & Glanville, B. B. (1982). Development of infant ear asymmetries for speech and music. Perception & Psychophysics, 31, 75-85.

Campbell, R., & Dodd, B. (1980). Hearing by eye. Quarterly Journal of Experimental Psychology, 32, 85-99.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). Perception & Psychophysics, 5, 365-373.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. Cognitive Psychology, 11, 478-484.

Fant, G. (1962). Descriptive analysis of the acoustic aspects of speech. Logos, 5, 3-17.

Field, T. M., Woodson, R., Greenberg, R., & Cohen, D. (1982). Discrimination and imitation of facial expressions by neonates. Science, 218, 179-181.

Hockett, C. (1958). A course in modern linguistics. New York: Macmillan.

Hoof, J. van (1976). The comparison of facial expression in man and higher primates. In M. von Cranach (Ed.), Methods of inference from human to animal behavior (pp. 165-196). Chicago: Aldine.

Joos, M. (1948). Acoustic phonetics. Language Monograph 23, Vol. 24(Suppl.)

Kelso, J. A. S., Cook, E., Olson, M. E., & Epstein, W. (1975). Allocation of attention and the locus of adaptation to displaced vision. Journal of Experimental Psychology: Human Perception and Performance, 1, 237-245.

Kimura, D., Battison, R., & Lubert, B. (1976). Impairment of nonlinguistic hand movements in a deaf aphasic. Brain and Language, 4, 566-571.

Kinsbourne, M. (1972). Eye and head turning indicates cerebral lateralization. Science, 176, 539-541.

Klima, E. S., & Bellugi, U. (1979). The signs of language. Cambridge, MA: Harvard University Press.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. Science, 218, 1138-1144.

Lempert, H., & Kinsbourne, M. (1982). Effect of laterality of orientation on verbal memory. Neuropsychologia, 20, 211-214.

Levinson, S. E., & Liberman, M. Y. (1981). Speech recognition by computer. Scientific American, 244, 64-87.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.

MacKain, K. S., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left hemisphere function. Science, 219, 1347-1349.

Marks, L. E. (1978). Multimodal perception. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception (Vol. VIII, pp. 321-339). New York: Academic Press.

Marler, P. (1975). On the origin of speech from animal sounds. In J. F. Kavanagh & J. E. Cutting (Eds.), The role of speech in language (pp. 11-37). Cambridge, MA: MIT Press.

Mayberry, R. I. (1978). Manual communication. In H. Davis & S. R. Silverman (Eds.), Hearing and deafness (4th ed.). New York: Holt, Rinehart & Winston.

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

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. Science, 198, 175-178.

Ohala, J. (1983). Cross-language uses of pitch. Phonetica, 40, 1-18.

Poizner, H., Bellugi, U., & Iragui, V. (in press). Apraxia and aphasia for a visual-gestural language. American Journal of Physiology.

Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. Journal of Experimental Psychology: Human Perception and Performance, 7, 733-740.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial ordering mechanism in speech production. In W. E. Cooper & E. C. T. Walker (Eds.), Sentence processing: Psycholinguistic studies presented to Merrill Garrett. Hillsdale, NJ: Lawrence Erlbaum Associates.

Spelke, E. (1976). Infants' intermodal perception of events. Cognitive Psychology, 8, 553-560.

Spoehr, K. T., & Corin, W. J. (1978). The stimulus suffix effect as a memory coding phenomenon. Memory & Cognition, 6, 583-589.

Studdert-Kennedy, M. (1980). Speech perception. Language and Speech, 23, 45-66.

Studdert-Kennedy, M., & Lane, H. (1980). Clues from the differences between signed and spoken language. In U. Bellugi & M. Studdert-Kennedy (Eds.), Signed and spoken language: Biolcgical constraints on linguistic form (pp. 29-39). Deerfield Park, FL: Verlag Chemie.

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

Templin, M. (1957). Certain language skills of children. Minneapolis: University of Minnesota Press.

Wilson, E. O. (1975). Sociobiology. Cambridge: The Belknap Press.

Footnote

[1] I have often heard it said that blind children develop language more slowly than their sighted peers, but I know of no systematic study on the topic.

THE MOTOR THEORY OF SPEECH PERCEPTION REVISED*

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

Abstract. A motor theory of speech perception, initially proposed to account for results of early experiments with synthetic speech, is now extensively revised to accommodate recent findings, and to relate the assumptions of the theory to those that might be made about other perceptual modes. According to the revised theory, phonetic information is perceived in a biologically distinct system, a "module" specialized to detect the intended gestures of the speaker that are the basis for phonetic categories. Built into the structure of this module is the unique but lawful relationship between the gestures and the acoustic patterns in which they are variously overlapped. In consequence, the module causes perception of phonetic structure without translation from preliminary auditory impressions. Thus, it is comparable to such other modules as the one that enables an animal to localize sound. Peculiar to the phonetic module are the relation between perception and production it incorporates and the fact that it must compete with other modules for the same stimulus variations.

Together with some of our colleagues, we have long been identified with a view of speech perception that is often referred to as a "motor theory." Not *the* motor theory, to be sure, because there are other theories of perception that, like ours, assign an important role to movement or its sources. But the theory we are going to describe is only about speech perception, in contrast to some that deal with other perceptual processes (e.g., Berkeley, 1709; Festinger, Burnham, Ono, & Bamber, 1967) or, indeed, with all of them (e.g., Washburn, 1926; Watson, 1919). Moreover, our theory is motivated by considerations that do not necessarily apply outside the domain of speech. Yet even there we are not alone, for several theories of speech perception, being more or less "motor," resemble ours to varying degrees (e.g., Chistovich, 1960; Dudley, 1940; Joos, 1948; Ladefoged & McKinney, 1963; Stetson,

---

*Cognition, in press.

†Also Yale University and University of Connecticut

††Also University of Connecticut

Acknowledgment. The writing of this paper was supported by a grant to Haskins Laboratories (NIH-NICHD HD-01994). We owe a special debt to Harriet Magen for invaluable help with the relevant literature, and to Alice Dadourian for coping with an ever-changing manuscript. For their patient responses to our frequent requests for information and criticism, we thank Franklin Cooper, Jerry Fodor, Carol Fowler, Scott Kelso, Charles Liberman, Robert Remez, Bruno Repp, Arthur Samuel, Michael Studdert-Kennedy, Michael Turvey, and Douglas Whalen. We also acknowledge the insightful comments of an anonymous reviewer.

1951). However, it is not relevant to our purposes to compare these, so, for convenience, we will refer to our motor theory as the motor theory.

We were led to the motor theory by an early finding that the acoustic patterns of synthetic speech had to be modified if an invariant phonetic percept was to be produced across different contexts (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952; Liberman, Delattre, & Cooper, 1952). Thus, it appeared that the objects of speech perception were not to be found at the acoustic surface. They might, however, be sought in the underlying motor processes, if it could be assumed that the acoustic variability required for an invariant percept resulted from the temporal overlap, in different contexts, of correspondingly invariant units of production. In its most general form, this aspect of the early theory survives, but there have been important revisions, including especially the one that makes perception of the motor invariant depend on a specialized phonetic mode (Liberman, 1982; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Studdert-Kennedy, 1978; Mattingly & Liberman, 1969). Our aim in this paper is to present further revisions, and so bring the theory up to date.

The Theory

The first claim of the motor theory, as revised, is that the objects of speech perception are the intended phonetic gestures of the speaker, represented in the brain as invariant motor commands that call for movements of the articulators through certain linguistically significant configurations. These gestural commands are the physical reality underlying the traditional phonetic notions--for example, "tongue backing," "lip rounding," and "jaw raising"--that provide the basis for phonetic categories. They are the elementary events of speech production and perception. Phonetic segments are simply groups of one or more of these elementary events; thus [b] consists of a labial stop gesture and [m] of that same gesture combined with a velum-lowering gesture. Phonologically, of course, the gestures themselves must be viewed as groups of features, such as "labial," "stop," "nasal," but these features are attributes of the gestural events, not events as such. To perceive an utterance, then, is to perceive a specific pattern of intended gestures.

We have to say "intended gestures," because, for a number of reasons (coarticulation being merely the most obvious), the gestures are not directly manifested in the acoustic signal or in the observable articulatory movements. It is thus no simple matter (as we shall see in a later section) to define specific gestures rigorously or to relate them to their observable consequences. Yet, clearly, invariant gestures of some description there must be, for they are required, not merely for our particular theory of speech perception, but for any adequate theory of speech production.

The second claim of the theory is a corollary of the first: if speech perception and speech production share the same set of invariants, they must be intimately linked. This link, we argue, is not a learned association, a result of the fact that what people hear when they listen to speech is what they do when they speak. Rather, the link is innately specified, requiring only epigenetic development to bring it into play. On this claim, perception of the gestures occurs in a specialized mode, different in important ways from the auditory mode, responsible also for the production of phonetic structures, and part of the larger specialization for language. The adaptive function of the perceptual side of this mode, the side with which the motor theory is directly concerned, is to make the conversion from acoustic signal to gesture automatically, and so to let listeners perceive phonetic structures without mediation by (or translation from) the auditory appearances that the sounds might, on purely psychoacoustic grounds, be expected to have.

A critic might note that the gestures do produce acoustic signals, after all, and that surely it is these signals, not the gestures, which stimulate the listener's ear. What can it mean, then, to say it is the gestures, not the signals, that are perceived? Our critic might also be concerned that the theory seems at first blush to assign so special a place to speech as to make it hard to think about in normal biological terms. We should, therefore, try to forestall misunderstanding by showing that, wrong though it may be, the theory is neither logically meaningless nor biologically unthinkable.

An Issue That Any Theory of Speech Perception Must Meet. The motor theory would be meaningless if there were, as is sometimes supposed, a one-to-one relation between acoustic patterns and gestures, for in that circumstance it would matter little whether the listener was said to perceive the one or the other. Metaphysical considerations aside, the proximal acoustic patterns might as well be the perceived distal objects. But the relation between gesture and signal is not straightforward. The reason is that the timing of the articulatory movements--the peripheral realizations of the gestures--is not simply related to the ordering of the gestures that is implied by the strings of symbols in phonetic transcriptions: the movements for gestures implied by a single symbol are typically not simultaneous, and the movements implied by successive symbols often overlap extensively. This coarticulation means that the changing shape of the vocal tract, and hence the resulting signal, is influenced by several gestures at the same time. Thus, the relation between gesture and signal, though certainly systematic, is systematic in a way that is peculiar to speech. In later sections of the paper we will consider how this circumstance bears on the perception of speech and its theoretical interpretation. For now, however, we wish only to justify consideration of the motor theory by identifying it as one of several choices that the complex relation between gesture and signal faces us with. For this purpose, we will describe just one aspect of the relation, that we may then use it as an example.

When coarticulation causes the signal to be influenced simultaneously by several gestures, a particular gesture will necessarily be represented by different sounds in different phonetic contexts. In a consonant-vowel syllable, for example, the acoustic pattern that contains information about the place of constriction of the consonantal gesture will vary depending on the following vowel. Such context-conditioned variation is most apparent, perhaps, in the transitions of the formants as the constriction is released. Thus, place information for a given consonant is carried by a rising transition in one vowel context and a falling transition in another (Liberman, Delattre, Cooper, & Gerstman, 1954). In isolation, these transitions sound like two different glissandi or chirps, which is just what everything we know about auditory perception leads us to expect (Mattingly, Liberman, Syrdal, & Halwes, 1971); they do not sound alike, and, just as important, neither sounds like speech. How is it, then, that, in context, they nevertheless yield the same consonant?

Auditory theories and the accounts they provide. The guiding assumption of one class of theories is that ordinary auditory processes are sufficient to explain the perception of speech; there is no need to invoke a further specialization for language, certainly not one that gives the listener access to gestures. The several members of this class differ in principle, though they are often combined in practice.

One member of the class counts two stages in the perceptual process: a first stage in which, according to principles that apply to the way we hear all sounds, the auditory appearances of the acoustic patterns are registered, followed by a second stage in which, by an act of sorting or matching to prototypes, phonetic labels are affixed (Crowder & Morton, 1969; Fujisaki & Kawashima, 1970; Oden & Massaro, 1978; Pisoni, 1973). Just why such different acoustic patterns as the rising and falling transitions of our example deserve the same label is not explicitly rationalized, it being accounted, presumably, a characteristic of the language that the processes of sorting or matching are able to manage. Nor does the theory deal with the fact that, in appropriate contexts, these transitions support phonetic percepts but do not also produce such auditory phenomena as chirps. To the contrary, indeed, it is sometimes made explicit that the auditory stage is actually available for use in discrimination. Such availability is not always apparent because the casual (or forgetful) listener is assumed to rely on the categorical labels, which persist in memory, rather than on the context-sensitive auditory impressions, which do not; but training or the use of more sensitive psychophysical methods is said to give better access to the auditory stage and thus to the stimulus variations--including, presumably, the differences in formant transition--that the labels ignore (Carney, Widin, & Viemeister, 1977; Pisoni & Tash, 1974; Samuel, 1977).

Another member of the class of auditory theories avoids the problem of context-conditioned variation by denying its importance. According to this theory, speech perception relies on there being at least a brief period during each speech sound when its short-time spectrum is reliably distinct from those of other speech sounds. For an initial stop in a stressed syllable, for example, this period includes the burst and the first 10 ms after the onset of voicing (Stevens & Blumstein, 1978). That a listener is nevertheless able to identify speech sounds from which these invariant attributes have been removed is explained by the claim that, in natural speech, they are sometimes missing or distorted, so that the child must learn to make use of secondary, context-conditioned attributes, such as formant transitions, which ordinarily co-occur with the primary, invariant attributes (Cole & Scott, 1974). Thus, presumably, the different-sounding chirps develop in perception to become the same-sounding (non-chirpy) phonetic element with which they have been associated.

The remaining member of this class of theories is the most thoroughly auditory of all. By its terms, the very processes of phonetic classification depend directly on properties of the auditory system, properties so independent of language as to be found, perhaps, in all mammals (Kuhl, 1981; Miller, 1977; Stevens, 1975). As described most commonly in the literature, this version of the auditory theory takes the perceived boundary between one phonetic category and another to correspond to a naturally-occurring discontinuity in perception of the relevant acoustic continuum. There is thus no first stage in which the (often) different auditory appearances are available, nor is there a process of learned equivalence. An example is the claim that the distinction between voiced and voiceless stops--normally cued by a complex of acoustic differences caused by differences in the phonetic variable known as voice-onset-time--depends on an auditory discontinuity in sensitivity to temporal relations among components of the signal (Kuhl & Miller, 1975; Pisoni, 1977). Another is the suggestion that the boundary between fricative and affricate on a rise-time continuum is the same as the rise-time boundary in the analogous nonspeech case--that is, the boundary that separates the nonspeech percepts "pluck" and "bow" (Cutting & Rosner, 1974; but see Rosen & Howell, 1981). To account for the fact that such discontinuities move as a function of phonetic context or rate of articulation, one can add the assumption that the several components of the acoustic signal give rise to interactions of a purely auditory sort (Hillenbrand, 1984; but see Summerfield, 1982). As for the rising and falling formant transitions of our earlier example, some such assumption of auditory interaction (between the transitions and the remainder of the acoustic pattern) would presumably be offered to account for the fact that they sound like two different glissandi in isolation, but as the same (non-glissando-like) consonant in the context of the acoustic syllable. The clear implication of this theory is that, for all phonetic contexts and for every one of the many acoustic cues that are known to be of consequence for each phonetic segment, the motivation for articulatory and coarticulatory maneuvers is to produce just those acoustic patterns that fit the language-independent characteristics of the auditory system. Thus, this last auditory theory is auditory in two ways: speech perception is governed by auditory principles, and so, too, is speech production.

The account provided by the motor theory. The motor theory offers a view radically different from the auditory theories, most obviously in the claim that speech perception is not to be explained by principles that apply to perception of sounds in general, but must rather be seen as a specialization for phonetic gestures. Incorporating a biologically based link between perception and production, this specialization prevents listeners from hearing the signal as an ordinary sound, but enables them to use the systematic, yet special, relation between signal and gesture to perceive the gesture. The relation is systematic because it results from lawful dependencies among gestures, articulator movements, vocal-tract shapes, and signal. It is special because it occurs only in speech.

Applying the motor theory to our example, we suggest what has seemed obvious since the importance of the transitions was discovered: the listener uses the systematically varying transitions as information about the coarticulation of an invariant consonant gesture with various vowels, and so perceives this gesture. Perception requires no arbitrary association of signal with phonetic category, and no correspondingly arbitrary progression from an auditory stage (e.g., different sounding glissandi) to a superseding phonetic label. As Studdert-Kennedy (1976) has put it, the phonetic category "names itself."

By way of comparison with the last of the auditory theories we described, we note that, just as this theory is in two ways auditory, the motor theory i in two ways motor. First, because it takes the proper object of phonetic perception to be a motor event. And, second, because it assumes that adaptations of the motor system for controlling the organs of the vocal tract took precedence in the evolution of speech. These adaptations made it possible, not only to produce phonetic gestures, but also to coarticulate them so that they could be produced rapidly. A perceiving system, specialized to take account of the complex acoustic consequences, developed concomitantly. Accordingly, the theory is not indifferently perceptual or motor, implying simply that the basis of articulation and the object of perception are the same. Rather, the emphasis is quite one-sided; therefore the theory fully deserves the epithet "motor."

How the Motor Theory Makes Speech Perception Like Other Specialized Perceiving Systems. The specialized perceiving system that the motor theory assumes is not unique; it is, rather, one of a rather large class of special systems or "modules." Accordingly, one can think about it in familiar biological terms. Later, we will consider more specifically how the phonetic module fits the concept of modularity developed recently by Fodor (1983); our concern now is only to compare the phonetic module with others.

The modules we refer to have in common that they are special neural structures, designed to take advantage of a systematic but unique relation between a proximal display at the sense organ and some property of a distal object. A result in all cases is that there is not, first, a cognitive representation of the proximal pattern that is modality-gereral, followed by translation to a particular distal property; rather, perception of the distal property is immediate, which is to say that the module has done all the hard work. Consider auditory localization as an example. One of several cues is differences in time of arrival of particular frequency components of the signal at the two ears (see Hafter, 1984, for a review). No one would claim that the use of this cue is part of the general auditory ability to perceive, as such, the size of the time interval that separates the onsets of two different signals. Certainly, this kind of general auditory ability does exist, but it is no part of auditory localization, either psychologically or physiologically. Animals perceive the location of sounding objects only by means of neural structures specialized to take advantage of the systematic but special relation between proximal stimulus and distal location (see, for example, Knudsen, 1984). The relation is systematic for obvious reasons; it is special because it depends on the circumstance that the animal has two ears, and that the ears are set a certain distance apart. In the case of the human, the only species for which the appropriate test can be made, there is no translation from perceived disparity in time because there is no perceived disparity.

Compare this with the voicing distinction (e.g., [ba] vs. [pa]) referred to earlier, which is cued in part by a difference in time of onset of the several formants, and which has therefore been said by some to rest on a general auditory ability to perceive temporal disparity as such (Kuhl & Miller, 1975; Pisoni, 1977). We believe, to the contrary, that the temporal disparity is only the proximal occasion for the unmediated perception of voicing, a distal gesture represented at the level of articulation by the relative timing of vocal-tract opening and start of laryngeal vibration (Lisker & Abramson, 1964). So we should expect perceptual judgments of differences in signal onset-time to have no more relevance to the voicing distinction than to auditory localization. In neither case do general auditory principles and procedures enlighten us. Nor does it help to invoke general principles of auditory interaction. The still more general principle that perception gives access to distal objects tells us only that auditory localization and speech perception work as they are supposed to; it does not tell us how. Surely the "how" is to be found, not by studying perception, even auditory perception, in general, but only by studying auditory localization and speech perception in particular. Both are special systems; they are, therefore, to be understood only in their own terms.

Examples of such biologically specialized perceiving modules can be multiplied. Visual perception of depth by use of information about binocular disparity is a well-studied example that has the same general characteristics we have attributed to auditory localization and speech (Julesz, 1960, 1971; Poggio, 1984). And there is presumably much to be learned by comparison with such biologically coherent systems as those that underlie echolocation in bats (Suga, 1984) or song in birds (Marler, 1970; Thorpe, 1958). But we will not elaborate, for the point to be made here is only that, from a biological point of view, the assumptions of the motor theory are not bizarre.

How the Motor Theory Makes Speech Perception Different from other Specialized Perceiving Systems. Perceptual modules, by definition, differ from one another in the classes of distal events that form their domains and in the relation between these events and the proximal displays. But the phonetic module differs from others in at least two further respects.

Auditory and phonetic domains. The first difference is in the locale of the distal events. In auditory localization, the distal event is "out there," and the relation between it and the proximal display at the two ears is completely determined by the principles of physical acoustics. Much the same can be said of those specialized modules that deal with the primitives of auditory quality, however they are to be characterized, and that come into play when people perceive, for example, whistles, horns, breaking glass, and barking dogs. Not so for the perception of phonetic structure. There, the distal object is a phonetic gesture or, more explicitly, an "upstream" neural command for the gesture from which the peripheral articulatory movements unfold. It follows that the relation between distal object and proximal stimulus will have the special feature that it is determined not just by acoustic principles but also by neuromuscular processes internal to the speaker. Of course, analogues of these processes are also available as part of the biological endowment of the listener. Hence, some kind of link between perception and production would seem to characterize the phonetic module, but not those modules that provide auditory localization or visual perception of depth. In a later section, we will have more to say about this link. Now we will only comment that it may conceivably resemble, in its most general characteristics, those links that have been identified in the communication modules of certain nonhuman creatures (Gerhardt & Rheinlaender, 1982; Hoy, Hahn, & Paul, 1977; Hoy & Paul, 1973; Katz & Gurney, 1981; Margoliash, 1983; McCasland & Konishi, 1983; Nottebohm, Stokes, & Leonard, 1976; Williams, 1984).

The motor theory aside, it is plain that speech somehow informs listeners about the phonetic intentions of the talker. The particular claim of the motor theory is that these intentions are represented in a specific form in the talker's brain, and that there is a perceiving module specialized to lead the listener effortlessly to that representation. Indeed, what is true of speech in this respect is true for all of language, except, of course, that the more distal object for language is some representation of linguistic structure, not merely of gesture, and that access to this object requires a module that is not merely phonetic, but phonological and syntactic as well.

Competition between phonetic and auditory modes. A second important difference between the phonetic module and the others has to do with the question: how does the module cooperate or compete with others that use stimuli of the same broadly defined physical form? For auditory localization, the key to the answer is the fact that the module is turned on by a specific and readily specifiable characteristic of the proximal stimulus: a particular range of differences in time of arrival at the two ears. Obviously, such differences have no other utility for the perceiver but to provide information about the distal property, location: there are no imaginable ecological circumstances in which a person could use this characteristic of the proximal stimuli to specify some other distal property. Thus, the proximal display and the distal property it specifies only complement the other aspects of what a listener hears; they never compete.

In phonetic perception, things are quite different because important acoustic cues are often similar to, even identical with, the stimuli that inform listeners about a variety of nonspeech events. We have already remarked that, in isolation, formant transitions sound like glissandi or chirps. Now surely we don't want to perceive these as glissandi or chirps when we are listening to speech, but we do want to perceive them so when we are listening to music or to birdsong. If this is true for all of the speech cues, as in some sense it presumably is, then it is hard to see how the module can be turned on by acoustic stigmata of any kind--that is, by some set of necessary cues defined in purely acoustic terms. We will consider this matter in some greater detail later. For now, however, the point is only that cues known to be of great importance for phonetic events may be cues for totally unrelated nonphonetic events, too. A consequence is that, in contrast to the generally complementary relation of the several modules that serve the same broadly defined modality (e.g., depth and color in vision), the phonetic and auditory modules are in direct competition. (For a discussion of how this competition might be resolved, see Mattingly & Liberman, 1985.)

Experimental Evidence for the Theory

Having briefly described one motive for the motor theory--the context-conditioned variation in the acoustic cues for constant phonetic categories--we will now add others. We will limit ourselves to the so-called segmental aspects of phonetic structure, though the theory ought, in principle, to apply in the suprasegmental domain as well (cf. Fowler, 1982).

The two parts of the theory--that gestures are the objects of perception and that perception of these gestures depends on a specialized module--might be taken to be independent, as they were in their historical development, but the relevant data are not. We therefore cannot rationally apportion the data between the parts, but must rather take them as they come.

A result of articulation: The multiplicity, variety, and equivalence of cues for each phonetic percept. When speech synthesis began to be used as a tool to investigate speech perception, it was soon discovered that, in any specific context, a particular local property of the acoustic signal was sufficient for the perception of one phonetic category rather than another and, more generally, that the percept could be shifted along some phonetic dimension by varying the synthetic stimulus along a locally-definable acoustic dimension. For example, if the onset frequency of the transition of the second formant during a stop release is sufficiently low, relative to the frequency of the following steady state, the stop is perceived as labial; otherwise, as apical or dorsal (Liberman et al., 1954). A value along such an acoustic dimension that was optimal for a particular phonetic category, or, more loosely, the dimension itself, was termed an "acoustic cue."

Of course, the fact that particular acoustic cues can be isolated must, of itself, tell us something about speech perception, for it might have been otherwise. Thus, it is possible to imagine a speech-perception mechanism, equipped, perhaps, with auditory templates, that would break down if presented with anything other than a wholly natural and phonetically optimal stimulus. Listeners would either give conflicting and unreliable phonetic judgments or else not hear speech at all. Clearly, the actual mechanism is not of this kind, and the concept of cue accords with this fact.

Nevertheless, the emphasis on the cues has, perhaps, been unfortunate, for the term "cue" might seem to imply a claim about the elemental units of speech perception. But "cue" was simply a convenient bit of laboratory jargon referring to acoustic variables whose definition depended very much on the design features of the particular synthesizers that were used to study them. The cues, as such, have no role in a theory of speech perception; they only describe some of the facts on which a theory might be based (cf. Bailey & Summerfield, 1980). There are, indeed, several generalizations about the cues--some only hinted at by the data now available, others quite well founded--that are relevant to such a theory.

One such generalization is that every "potential" cue--that is, each of the many acoustic events peculiar to a linguistically significant gesture--is an <u>actual</u> cue. (For example, every one of eighteen potential cues to the voicing distinction in medial position has been shown to have some perceptual value; Lisker, 1978.) All possible cues have not been tested, and probably never will be, but no potential cue has yet been found that could not be shown to be an actual one.

A closely related generalization is that, while each cue is, by definition, more or less sufficient, none is truly necessary. The absence of any single cue, no matter how seemingly characteristic of the phonetic category, can be compensated for by others, not without some cost to naturalness or even intelligibility, perhaps, but still to such an extent that the intended category is, in fact, perceived. Thus, stops can be perceived without silent periods, fricatives without frication, vowels without formants, and tones without pitch (Abramson, 1972; Inoue, 1984; Remez & Rubin, 1984; Repp, 1984; Yeni-Komshian & Soli, 1981).

Yet another generalization is that even when several cues are present, variations in one can, within limits, be compensated for by offsetting variations in another (Dorman, Raphael, & Liberman, 1979; Dorman, Studdert-Kennedy, & Raphael, 1977; Hoffman, 1958; Howell & Rosen, 1983; Lisker, 1957; Summerfield & Haggard, 1977). In the case of the contrast between fricative-vowel and fricative-stop-vowel (as in [ʃa] vs. [ʃta]), investigators have found that two important cues, silence and appropriate formant transitions, engage in just such a trading relation. That this bespeaks a true equivalence in perception was shown by experiments in which the effect of variation in one cue could, depending on its "direction," be made to "add to" or "cancel out" the effect of the other (Fitch, Halwes, Erickson, & Liberman, 1980). Significantly, this effect can also be obtained with sine-wave analogues of speech, but only for subjects who perceive these signals as speech, not for those who perceive them as nonspeech tones (Best, Morrongiello, & Robson, 1981).

Putting together all the generalizations about the multiplicity and variety of acoustic cues, we should conclude that there is simply no way to define a phonetic category in purely acoustic terms. A complete list of the cues--surely a cumbersome matter at best--is not feasible, for it would necessarily include all the acoustic effects of phonetically distinctive articulations. But even if it were possible to compile such a list, the result would not repay the effort, because none of the cues on the list could be deemed truly essential. As for those cues that might, for any reason, be finally included, none could be assigned a characteristic setting, since the effect of changing it could be offset by appropriate changes in one or more of the others. This surely tells us something about the design of the phonetic module. For if phonetic categories were acoustic patterns, and if, accordingly, phonetic perception were properly auditory, one should be able to describe quite straightforwardly the acoustic basis for the phonetic category and associated percept. According to the motor theory, by contrast, one would expect the acoustic signal to serve only as a source of information about the gestures; hence the gestures would properly define the category. As for the perceptual equivalence among diverse cues that is shown by the trading relations, explaining that on auditory grounds requires ad hoc assumptions. But if, as the motor theory would have it, the gesture is the distal object of perception, we should not wonder that the several sources of information about it are perceptually equivalent, for they are products of the same linguistically significant gesture.

A result of coarticulation: I. Segmentation in sound and percept. Traditional phonetic transcription represents utterances as single linear sequences of symbols, each of which stands for a phonetic category. It is an issue among phonologists whether such transcriptions are really theoretically adequate, and various alternative proposals have been made in an effort to provide a better account. This matter need not concern us here, however, since all proposals have in common that phonetic units of some description are ordered from left to right. Some sort of segmentation is thus always implied, and what theory must take into account is that the perceived phonetic object is thus segmented.

Segmentation of the phonetic percept would be no problem for theory if the proximal sound were segmented correspondingly. But it is not, nor can it be, if speech is to be produced and perceived efficiently. To maintain a straightforward relation in segmentation between phonetic unit and signal would require that the sets of phonetic gestures corresponding to phonetic units be produced one at a time, each in its turn. The obvious consequence would be that each unit would become a syllable, in which case talkers could speak only as fast as they could spell. A function of coarticulation is to evade this limitation. There is an important consequence, however, which is that there is now no straightforward correspondence in segmentation between the phonetic and acoustic representations of the information (Fant, 1962; Joos, 1948). Thus, the acoustic information for any particular phonetic unit is typically overlapped, often quite thoroughly, with information for other units. Moreover, the span over which that information extends, the amount of overlap, and the number of units signalled within the overlapped portion all vary according to the phonetic context, the rate of articulation, and the language (Magen, 1984; Manuel & Krakow, 1984; Öhman, 1966; Recasens, 1984; Repp, Liberman, Eccardt, & Pesetsky, 1978; Tuller, Harris, & Kelso, 1982).

There are, perhaps, occasional stretches of the acoustic signal over which there is information about only one phonetic unit--for example, in the middle of the frication in a slowly articulated fricative-vowel syllable and in vowels that are sustained for artificially long times. Such stretches do, of course, offer a relation between acoustic patterns and phonetic units that would be transparent if phonetic perception were merely auditory. But even in these cases, the listener automatically takes account of, not just the transparent part of the signal, but the regions of overlap as well (Mann & Repp, 1980, 1981; Whalen, 1981). Indeed, the general rule may be that the phonetic percept is normally made available to consciousness only after all the relevant acoustic information is in, even when earlier cues might have been sufficient (Martin & Bunell, 1981, 1982; Repp et al.. 1978).

What wants explanation, then, is that the percept is segmented in a way that the signal is not, or, to put it another way, that the percept does not mirror the overlap of information in the sound (cf. Fowler, 1984). The motor theory does not provide a complete explanation, certainly not in its present state, but it does head the theoretical enterprise in the right direction. At the very least, it turns the theorist away from the search for those unlikely processes that an auditory theory would suggest: How listeners learn phonetic labels for what they hear and thus re-interpret perceived overlap as sequences of discrete units; or how discrete units emerge in perception from interactions of a purely auditory sort. The first process seems implausible on its face, the second because it presupposes that the function of the many kinds and degrees of coarticulation is to produce just those combinations of sounds that will interact in accordance with language-independent characteristics of the auditory system. In contrast, the motor theory begins with the assumption that coarticulation, and the resulting overlap of phonetic information in the acoustic pattern, is a consequence of the efficient processes by which discrete phonetic gestures are realized in the behavior of more or less independent articulators. The theory suggests, then, that an equally efficient perceptual process might use the resulting acoustic pattern to recover the discrete gestures.

A result of coarticulation: II. Different sounds, different contexts, same percept. That the phonetic percept is invariant even when the relevant acoustic cue is not was the characteristic relation between percept and sound that we took as an example in the first section. There, we observed that variation in the acoustic pattern results from overlapping of putatively invariant gestures, an observation that, as we remarked, points to the gesture, rather than the acoustic pattern itself, as the object of perception. We now add that the articulatory variation due to context is pervasive: in the acoustic representation of every phonetic category yet studied there are context-conditioned portions that contribute to perception and that must, therefore, be taken into account by theory. Thus, for stops, nasals, fricatives. liquids, semivowels, and vowels, the always context-sensitive transitions are cues (Harris, 1958; Jenkins, Strange, & Edman, 1983; Liberman et al., 1954; O'Connor, Gerstman, Liberman, Delattre, & Cooper, 1957; Strange, Jenkins, & Johnson, 1983). For stops and fricatives, the noises that are produced at the point of constriction are also known to be cues, and, under some circumstances at least, these, too, vary with context (Dorman et al., 1977; Liberman et al., 1952; Whalen, 1981).

An auditory theory that accounts for invariant perception in the face of so much variation in the signal would require a long list of apparently arbitrary assumptions. For a motor theory, on the other hand, systematic stimulus variation is not an obstacle to be circumvented or overcome in some arbitrary way; it is, rather, a source of information about articulation that provides important guidance to the perceptual process in determining a representation of the distal gesture.

A result of coarticulation: III. Same sound, different contexts, different percepts. When phonetic categories share one feature but differ in another, the relation between acoustic pattern and percept speaks, again, to the motor theory and its alternatives. Consider, once more, the fricative [s] and the stop [t] in the syllables [sa] and [sta]. In synthesis, the second- and third-formant transitions can be the same for these two categories, since they have the same place of articulation; and the first-formant transition, normally a cue to manner, can be made ambiguous between them. For such stimuli, the perception of [sta] rather than [sa] depends on whether there is an interval of silence between the noise for the [s] and the onsets of the transitions.

Data relevant to an interpretation of the role of silence in thus producing different percepts from the same transition come from two kinds of experiments. First are those that demonstrate the effectiveness of the transitions as cues for the place feature of the fricative in fricative-vowel syllables (Harris, 1958). The transitions are not, therefore, masked by the noise of the [s] friction, and thus the function of silence in a stop is not, as it might be in an auditory theory, to protect the transitions from such masking. The second kind of experiment deals with the possibility of a purely auditory interaction--in this case, between silence and the formant transitions. Among the findings that make such auditory interaction seem unlikely is that silence affects perception of the formant transitions differently in and out of speech context and, further, that the effectiveness of silence depends on such factors as continuity of talker and prosody (Dorman et al., 1979; Rakerd, Dechovitz, & Verbrugge, 1982). But perhaps the most direct test for auditory interaction is provided by experiments in which such interaction is ruled out by holding the acoustic context constant. This can be done by exploiting "duplex perception," a phenomenon to be discussed in greater detail in the next section. Here it is appropriate to say only that duplex perception provides a way of presenting acoustic patterns so that, in a fixed context, listeners hear the same second- or third-formant transitions in two phenomenally different ways simultaneously: as nonspeech chirps and as cues for phonetic categories. The finding is that the presence or absence of silence determines whether formant transitions appropriate for [t] or for [p], for example, are integrated into percepts as different as stops and fricatives; but silence has no effect on the perception of the nonspeech chirps that these same transitions produce (Liberman, Isenberg, & Rakerd, 1981). Since the latter result eliminates the possibility of auditory interaction, we are left with the account that the motor theory would suggest: that silence acts in the specialized phonetic mode to inform the listener that the vocal tract was completely closed to produce a stop consonant, rather than merely constricted to produce a fricative. It follows, then, that silence will, by its presence or absence, determine whether identical transitions are cues in percepts that belong to the one manner or the other.

An acoustic signal diverges to phonetic and auditory modes. We noted earlier that a formant transition is perceptually very different depending on whether it is perceived in the auditory mode, where it sounds like a chirp, or in the phonetic mode, where it cues a "nonchirpy" consonant. Of course, the comparison is not entirely fair, since acoustic context is not controlled: the transition is presented in isolation in the one case, but as an element of a larger acoustic pattern in the other. We should, therefore, call attention to the fact that the same perceptual difference is obtained even when, by resort to a special procedure, acoustic context is held constant (Liberman, 1979; Rand, 1974). This procedure, which produces the duplex percept referred to earlier, goes as follows. All of an acoustic syllable except only the formant transition that decides between, for example, [da] and [ga] is presented to one ear. By itself, this pattern, called the "base," sounds like a stop-vowel syllable, ambiguous between [da] and [ga]. To the other ear is presented one or the other of the transitions appropriate for [d] or [g]. In isolation, these sound like different chirps. Yet, when base and transition are presented dichotically, and in the appropriate temporal relationship, they give rise to a duplex percept: [da] or [ga], depending on the transition, and, simultaneously, the appropriate chirp. (The fused syllable appears to be in the ear to which the base had been presented, the chirp in the other.)

Two related characteristics of duplex perception must be emphasized. One is that it is obtained only when the stimulus presented to one ear is, like the "chirpy" transition, of short duration and extremely unspeechlike in quality. If that condition is not met, as, for example, when the first two formants are presented to one ear and the entire third formant to the other, perception is not duplex. It is, on the contrary, simplex; one hears a coherent syllable in which the separate components cannot be apprehended. (A very different result is obtained when two components of a musical chord are presented to one ear, a third component to the other. In that case, listeners can respond to the third component by itself and also to that component combined with the first two [Pastore, Schmuckler, Rosenblum, & Szczesiul, 1983].)

The other, closely related characteristic of duplex perception is that it is precisely duplex, not triplex. That is, listeners perceive the nonspeech chirp and the fused syllable, but they do not also perceive the base--i.e., the syllable, minus one of the formant transitions--that was presented to one ear (Repp, Milburn, & Ashkenas, 1983). (In the experiment with musical chords by Pastore et al., referred to just above, there was no test for duplex, as distinguished from triplex, perception.)

The point is that duplex perception does not simply reflect the ability of the auditory system to fuse dichotically presented stimuli and also, as in the experiment with the chords, to keep them apart. Rather, the duplex percepts of speech comprise the only two ways in which the transition, for example, can be heard: as a cue for a phonetic gesture and as a nonspeech sound. These percepts are strikingly different, and, as we have already seen, they change in different, sometimes contrasting ways in response to variations in the acoustic signals--variations that must have been available to all structures in the brain that can process auditory information. A reasonable conclusion is that there must be two modules that can somehow use the same input to produce simultaneous representations of two distal objects. (For speculation about the mechanism that normally prevents perception of this ecologically impossible situation, and about the reason why that highly adaptive mechanism might be defeated by the procedures used to produce duplex perception, see Mattingly & Liberman, 1985.)

Acoustic and optical signals converge on the phonetic mode. In duplex perception, a single acoustic stimulus is processed simultaneously by the phonetic and auditory modules to produce perception of two distal objects: a phonetic gesture and a sound. In the phenomenon to which we turn now, something like the opposite occurs: two different stimuli--one acoustic, the other optical--are combined by the phonetic module to produce coherent perception of a single distal event. This phenomenon, discovered by McGurk and McDonald (1976), can be illustrated by this variant on their original demonstration. Subjects are presented acoustically with the syllables [ba], [ba], [ba] and optically with a face that, in approximate synchrony, silently articulates [bɛ], [vɛ], [ðɛ]. The resulting and compelling percept is [ba], [va], [ða], with no awareness that it is in any sense bimodal--that is, part auditory and part visual. According to the motor theory, this is so because the perceived event is neither; it is, rather, a gesture. The proximal acoustic signal and the proximal optical signal have in common, then, that they convey information about the same distal object. (Perhaps a similar convergence is implied by the finding that units in the optic tectum of the barn owl are bimodally sensitive to acoustic and optical cues for the same distal property, location in space; Knudsen, 1982).

Even prelinguistic infants seem to have some appreciation of the relation between the acoustic and optical consequences of phonetic articulation. This is to be inferred from an experiment in which it was found that infants at four to five months of age preferred to look at a face that articulated the vowel they were hearing rather than at the same face articulating a different vowel (Kuhl & Meltzoff, 1982). Significantly, this result was not obtained when the sounds were pure tones matched in amplitude and duration to the vowels. In a related study it was found that infants of a similar age looked longer at a face repeating the disyllable they were hearing than at the same face repeating another disyllable, though both disyllables were carefully synchronized with the visible articulation (MacKain, Studdert-Kennedy, Spieker, & Stern, 1983). Like the results obtained with adults in the McGurk-MacDonald kind of experiment, these findings with infants imply a perception-production link and, accordingly, a common mode of perception for all proper information about the gesture.

The general characteristics that cause acoustic signals to be perceived as speech. The point was made in an earlier section that acoustic definitions of phonetic contrasts are, in the end, unsatisfactory. Now we would suggest that acoustic definitions also fail for the purpose of distinguishing in general between acoustic patterns that convey phonetic structures and those that do not. Thus, speech cannot be distinguished from nonspeech by appeal to surface properties of the sound. Surely, natural speech does have certain characteristics of a general and superficial sort--for example, formants with characteristic bandwidths and relative intensities, stretches of waveform periodicities that typically mark the voiced portion of syllables, peaks of intensity corresponding approximately to syllabic rhythm, etc.--and these can be used by machines to detect speech. But research with synthesizers has shown that speech is perceived even when such general characteristics are absent. This was certainly true in the case of many of the acoustic patterns that were used in work with the Pattern Playback synthesizer, and more recently it has been shown to be true in the most extreme case of patterns consisting only of sine waves that follow natural formant trajectories (Remez, Rubin, Pisoni, & Carrell, 1981). Significantly, the converse effect is also obtained. When reasonably normal formants are made to deviate into acoustically continuous but abnormal trajectories, the percept breaks into two categorically distinct parts: speech and a background of chirps, glissandi, and assorted noises (Liberman & Studdert-Kennedy, 1978). Of course, the trajectories of the formants are determined by the movements of the articulators. Evidently, those trajectories that conform to possible articulations engage the phonetic module; all others fail.

We conclude that acoustic patterns are identified as speech by reference to deep properties of a linguistic sort: if a sound can be "interpreted" by the specialized phonetic module as the result of linguistically significant gestures, then it is speech; otherwise, not. (In much the same way, grammatical sentences can be distinguished from ungrammatical ones, not by lists of surface properties, but only by determining whether or not a grammatical derivation can be given.) Of course, the kind of mechanism such an "interpretation" requires is the kind of mechanism the motor theory presumes.

Phonetic and auditory responses to the cues. Obviously, a module that acts on acoustic signals cannot respond beyond the physiological limits of those parts of the auditory system that transmit the signal to the module. Within those limits, however, different modules can be sensitive to the signals in different ways. Thus, the auditory-localization module enables listeners to perceive differences in the position of sounding objects given temporal disparity cues smaller by several orders of magnitude than those required to make the listener aware of temporal disparity as such (Brown & Deffenbacher, 1979, Chap. 7; Hirsh, 1959). If there is, as the motor theory implies, a distinct phonetic module, then in like manner its sensitivities should not, except by accident, be the same as those that characterize the module that deals with the sounds of nonspeech events.

In this connection, we noted in the first section of the paper that one form of auditory theory of speech perception points to auditory discontinuities in differential sensitivity (or in absolute identification), taking these to be the natural bases for the perceptual discontinuities that characterize the boundaries of phonetic categories. But several kinds of experiments strongly imply that this is not so.

One kind of experiment has provided evidence that the perceptual discontinuities at the boundaries of phonetic categories are not fixed; rather, they move in accordance with the acoustic consequences of articulatory adjustments associated with phonetic context, dialect, and rate of speech. (For a review, see Repp & Liberman, in press.) To account for such articulation-correlated changes in perceptual sensitivities by appeal to auditory processes requires, yet again, an ultimately countless set of ad hoc assumptions about auditory interactions, as well as the implausible assumption that the articulators are always able to behave so as to produce just those sounds that conform to the manifold and complex requirements that the auditory interactions impose. It seems hardly more plausible that, as has been suggested, the discontinuities in phonetic perception are really auditory discontinuities that were caused to move about in phylogenetic or ontogenetic development as a result of experience with speech (Aslin & Pisoni, 1980). The difficulty with this assumption is that it presupposes the very canonical form of the cues that does not exist (see above) and, also, that it implies a contradiction in assuming, as it must, that the auditory sensitivities underwent changes in the development of speech, yet somehow also remained unchanged and nonetheless manifest in the adult's perception of nonspeech sounds.

Perhaps this is the place to remark about categorical perception that the issue is not, as is often supposed, whether nonspeech continua are categorically perceived, for surely some do show tendencies in that direction. The issue is whether, given the same (or similar) acoustic continua, the auditory and phonetic boundaries are in the same place. If there are, indeed, auditory boundaries, and if, further, these boundaries are replaced in phonetic perception by boundaries at different locations (as the experiments referred to above do indicate), then the separateness of phonetic and auditory perception is even more strongly argued for than if the phonetic boundaries had appeared on continua where auditory boundaries did not also exist.

Also relevant to comparison of sensitivity in phonetic and auditory modes are experiments on perception of acoustic variations when, in the one case, they are cues for phonetic distinctions, and when, in some other, they are perceived as nonspeech. One of the earliest of the experiments to provide data about the nonspeech side of this comparison dealt with perception of frequency-modulated tones--or "ramps" as they were called--that bear a close resemblance to the formant transitions. The finding was that listeners are considerably better at perceiving the pitch at the end of the ramp than at the beginning (Brady, House, & Stevens, 1961). Yet, in the case of stop consonants that are cued by formant transitions, perception is better syllable-initially than syllable-finally, though in the former case it requires information about the beginning of the ramp, while in the latter it needs to know about the end. Thus, if one were predicting sensitivity to speech from sensitivity to the analogous nonspeech sounds, one would make exactly the wrong predictions. More recent studies have made more direct comparisons and found differences in discrimination functions when, in speech context, formant transitions cued place distinctions among stops and liquids, and when, in isolation, the same transitions were perceived as nonspeech sounds (Mattingly et al., 1971; Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 197_).

More impressive, perhaps, is evidence that has come from experiments in which listeners are induced to perceive a constant stimulus in different ways. Here belong experiments in which sine-wave analogues of speech, referred to earlier, are presented under conditions that cause some listeners to perceive them as speech and others not. The perceived discontinuities lie at different places (on the acoustic continuum) for the two groups (Best et al., 1981; Best & Studdert-Kennedy, 1983; Studdert-Kennedy & Williams, 1984; Williams, Verbrugge, & Studdert-Kennedy, 1983). Here, too, belongs an experiment in which the formant-transitions appropriate to a place contrast between stop consonants are presented with the remainder of a syllable in such a way as to produce the duplex percept referred to earlier: the transitions cue a stop consonant and, simultaneously, nonspeech chirps. The result is that listeners yield quite different discrimination functions for exactly the same formant transitions in exactly the same acoustic context, depending on whether they are responding to the speech or nonspeech sides of the duplex percept; only on the speech side of the percept is there a peak in the discrimination function to mark a perceptual discontinuity at the phonetic boundary (Mann & Liberman, 1983).

Finally, we note that, apart from differences in differential sensitivity to the transitions, there is also a difference in absolute-threshold sensitivity when, in the one case, these transitions support a phonetic percept, and when, in the other, they are perceived as nonspeech chirps. Exploiting, again, the phenomenon of duplex perception, investigators found that the transitions were effective (on the speech side of the percept) in cueing the contrast between stops at a level of intensity 18 db lower than that required for comparable discrimination of the chirps (Bentin & Mann, 1983). At that level, indeed, listeners could not even hear the chirps, let alone discriminate them; yet they could still use the transitions to identify the several stops.

The Several Aspects of the Theory

For the purpose of evaluating the motor theory, it is important to separate it into its more or less independent parts. First, and fundamentally, there is the claim that phonetic perception is perception of gesture. As we have seen, this claim is based on evidence that the invariant source of the phonetic percept is somewhere in the processes by which the sounds of speech are produced. In the first part of this section we will consider where in those processes the invariant might be found.

The motor theory also implies a tight link between perception and production. In the second part of this section we will ask how that link came to be.

Where is the Invariant Phonetic Gesture? A phonetic gesture, as we have construed it, is a class of movements by one or more articulators that results in a particular, linguistically significant deformation, over time, of the vocal-tract configuration. The linguistic function of the gesture is clear enough: phonetic contrasts, which are of course the basis of phonological categories, depend on the choice of one particular gesture rather than another. What is not so clear is how the gesture relates to the actual physical movements of articulators and to the resulting vocal-tract configurations, observed, for example, in x-ray films.

In the early days of the motor theory we made a simplifying assumption about this relation: that a gesture was effected by a single key articulator. On this assumption, the actual movement trajectory of the articulator might vary, but only because of aerodynamic factors and the physical linkage of this articulator with others, so the neural commands in the final common paths (observable with electromyographic techniques) would nevertheless be invariant across different contexts. This assumption was appropriate as an initial working hypothesis, if only because it was directly testable. In the event, there proved be a considerable amount of variability that the hypothesis could not account for.

In formulating this initial hypothesis, we had overlooked several serious complications. One is that a particular gesture typically involves not just one articulator, but two or more; thus "lip rounding," for example, is a collaboration of lower lip, upper lip, and jaw. Another is that a single articulator may participate in the execution of two different gestures at the same time; thus, the lips may be simultaneously rounding and closing in the production of a labial stop followed by a rounded vowel, e.g., [bu]. Prosody makes additional complicating demands, as when a greater displacement of some or all of the active articulators is required in producing a stressed syllable rather than an unstressed one; and linguistically irrelevant factors, notably speaking rate, affect the trajectory and phasing of the component movements.

These complications might suggest that there is little hope of providing a rigorous physical definition of a particular gesture, and that the gestures are hardly more satisfactory as perceptual primitives than are the acoustic cues. It might, indeed, be argued that there is an infinite number of possible articulatory movements, and that the basis for categorizing one group of such movements as "lip rounding" and another as "lip closure" is entirely a priori.

But the case for the gesture is by no means as weak as this. Though we have a great deal to learn before we can account for the variation in instances of the same gesture, it is nonetheless clear that, despite such variation, the gestures have a virtue that the acoustic cues lack: instances of a particular gesture always have certain topological properties not shared by any other gesture. That is, for any particular gesture, the same sort of distinctive deformation is imposed on the current vocal-tract configuration, whatever this "underlying" configuration happens to be. Thus, in lip rounding, the lips are always slowly protruded and approximated to some appreciable extent, so that the anterior end of the vocal tract is extended and narrowed, though the relative contributions of the tongue and lips, the actual degrees of protrusion and approximation, and the speed of articulatory movement vary according to context. Perhaps this example seems obvious because lip rounding involves a local deformation of the vocal-tract configuration, but the generalization also applies to more global gestures. Consider, for example, the gesture required to produce an "open" vowel. In this gesture, tongue, lips, jaw, and hyoid all participate to contextually varying degrees, and the actual distance between the two lips, as well as that between the tongue blade and body and the upper surfaces of the vocal tract, are variable; but the goal is always to give the tract a more open, horn-shaped configuration than it would otherwise have had.

We have pointed out repeatedly that, as a consequence of gestural overlapping, the invariant properties of a particular gesture are not manifest in the spectrum of the speech signal. We would now caution that a further consequence of this overlapping is that, because of their essentially topological character, the gestural invariants are usually not obvious from inspection of a single static vocal-tract configuration, either. They emerge only from consideration of the configuration as it changes over time, and from comparison with other configurations in which the same gesture occurs in different contexts, or different gestures in the same context.

We would argue, then, that the gestures do have characteristic invariant properties, as the motor theory requires, though these must be seen, not as peripheral movements, but as the more remote structures that control the movements. These structures correspond to the speaker's intentions. What is far from being understood is the nature of the system that computes the topologically appropriate version of a gesture in a particular context. But this problem is not peculiar to the motor theory; it is familiar to many who study the control and coordination of movement, for they, like us, must consider whether, given context-conditioned variability at the surface, motor acts are nevertheless governed by invariants of some sort (Browman & Goldstein, 1985; Fowler, Rubin, Remez, & Turvey, 1980; Tuller & Kelso, 1984; Turvey, 1977).

The Origin of the Perception-Production Link. In the earliest accounts of the motor theory, we put considerable attention on the fact that listeners not only perceive the speech signal but also produce it. This, together with doctrinal behaviorist considerations, led us to assume that the connection between perception and production was formed as a wholly learned association, and that perceiving the gesture was a matter of picking up the sensory consequences of covert mimicry. On this view of the genesis of the perception-production link, the distinguishing characteristic of speech is only that it provides the opportunity for the link to be established. Otherwise, ordinary principles of associative learning are adequate to the task; no specialization for language is required.

But then such phenomena as have been described in this paper were discovered, and it became apparent that they differed from anything that association learning could reasonably be expected to produce. Nor were these the only relevant considerations. Thus, we learned that people who have been pathologically incapable from birth of controlling their articulators are nonetheless able to perceive speech (MacNeilage, Rootes, & Chase, 1967). From the research pioneered by Eimas, Siqueland, Jusczyk, and Vigorito (1971), we also learned that prelinguistic infants apparently categorize phonetic distinctions much as adults do. More recently, we have seen that even when the distinction is not functional in the native language of the subjects, and when, accordingly, adults have trouble perceiving it, infants nevertheless do quite well up to about one year of age, at which time they begin to perform as poorly as adults (Werker & Tees, 1984). Perhaps, then, the sensitivity of infants to the acoustic consequences of linguistic gestures includes all those gestures that could be phonetically significant in any language, acquisition of one's native language being a process of losing sensitivity to gestures it does not use. Taking such further considerations as these into account, we have become even more strongly persuaded that the phonetic mode, and the perception-production link it incorporates, are innately specified.

Seen, then, as a view about the biology of language, rather than a comment on the coincidence of speaking and listening, the motor theory bears at several points on our thinking about the development of speech perception in the child. Consider, first, a linguistic ability that, though seldom noted (but see Mattingly, 1976), must be taken as an important prerequisite to acquiring the phonology of a language. This is the ability to sort acoustic patterns into two classes: those that contain (candidate) phonetic structures and those that do not. (For evidence, however indirect, that infants do so sort, see Alegria & Noirot, 1982; Best, Hoffman, & Glanville, 1982; Entus, 1977; Molfese, Freeman, & Palermo, 1975; Segalowitz & Chapman, 1980; Witelson, 1977; but see Vargha-Khadem & Corballis, 1979). To appreciate the bearing of the motor theory on this matter, recall our claim, made in an earlier section, that phonetic objects cannot be perceived as a class by reference to acoustic stigmata, but only by a recognition that the sounds might have been produced by a vocal tract as it made linguistically significant gestures. If so, the perception-production link is a necessary condition for recognizing speech as speech. It would thus be a blow to the motor theory if it could be shown that infants must develop empirical criteria for this purpose. Fortunately for the theory, such criteria appear to be unnecessary.

Consider, too, how the child comes to know, not only that phonetic structures are present, but, more specifically, just what those phonetic structures are. In this connection, recall that information about the string of phonetic segments is overlapped in the sound, and that there are, accordingly, no acoustic boundaries. Until and unless the child (tacitly) appreciates the gestural source of the sounds, it can hardly be expected to perceive, or ever learn to perceive, a phonetic structure. Recall, too, that the acoustic cues for a phonetic category vary with phonetic factors such as context and with extra-phonetic factors such as rate and vocal-tract size. This is to say, once again, that there is no canonical cue. What, then, is the child to learn? Association of some particular cue (or set of cues) with a phonetic category will work only for a particular circumstance. When circumstances change, the child's identification of the category will be wrong, sometimes grossly, and it is hard to see how it could readily make the appropriate correction. Perception of the phonetic categories can properly be generalized only if the acoustic patterns are taken for what they really are: information about the underlying gestures. No matter that the child sometimes mistakes the phonological significance of the gesture, so long as that which is perceived captures the systematic nature of its relation to the sound; the phonology will come in due course. To appreciate this relation is, once again, to make use of the link between perception and production.

How "Direct" is Speech Perception?

Since we have been arguing that speech perception is accomplished without cognitive translation from a first-stage auditory register, our position might appear similar to the one Gibson (1966) has taken in regard to "direct perception." The similarity to Gibson's views may seem all the greater because, like him, we believe that the object of perception is motoric. But there are important differences, the bases for which are to be seen in the following passage (Gibson, 1966, p. 94):

> An articulated utterance is a source of a vibratory field in the air. The source is biologically 'physical' and the vibration is acoustically 'physical'. The vibration is a potential stimulus, becoming effective when a listener is within range of the vibratory field. The listener then perceives the articulation because the invariants of vibration correspond to those of articulation. In this theory of speech perception, the units and parts of speech are present both in the mouth of the speaker and in the air between the speaker and listener. Phonemes are in the air. They can be considered physically real if the higher-order invariants of sound waves are admitted to the realm of physics.

The first difference between Gibson's view and ours relates to the nature of the perceived events. For Gibson, these are actual movements of the articulators, while for us, they are the more remote gestures that the speaker intended. The distinction would be trivial if an articulator were affected by only one gesture at a time, but, as we have several times remarked, an articulatory movement is usually the result of two or more overlapping gestures. The gestures are thus control structures for the observable movements.

The second difference is that, unlike Gibson, we do not think articulatory movements (let alone phonetic structures) are given directly (that is, without computation) by "higher-order invariants" that would be plain if only we had a biologically appropriate science of physical acoustics. We would certainly welcome any demonstration that such invariants did exist, since, even though articulatory movement is not equivalent to phonetic structure, such a demonstration would permit a simpler account of how the phonetic module works. But no higher-order invariants have thus far been proposed, and we doubt that any will be forthcoming. We would be more optimistic on this score if it could be shown, at least, that articulatory movements can be recovered from the signal by computations that are purely analytic, if nevertheless complex. One might then hope to reformulate the relationship between movements and signal in a way that would make it possible to appeal to higher-order invariants and thus obviate the need for computation. But, given the many-to-one relation between vocal-tract configurations and acoustic signal, a purely analytic solution to the problem of recovering movements from the signal seems to be impossible unless one makes unrealistic assumptions about excitation, damping, and other physical variables (Sondhi, 1979). We therefore remain skeptical about higher-order invariants.

The alternative to an analytic account of speech perception is, of course, a synthetic one, in which case the module compares some parametric description of the input signal with candidate signal descriptions. As with any form of "analysis-by-synthesis" (cf. Stevens & Halle, 1967), such an account is plausible only if the number of candidates the module has to test can be kept within reasonable bounds. This requirement is met, however, if, as we suppose, the candidate signal descriptions are computed by an analog of the production process--an internal, innately specified vocal-tract synthesizer, as it were (Liberman, Mattingly, & Turvey, 1972; Mattingly & Liberman, 1969)--that incorporates complete information about the anatomical and physiological characteristics of the vocal tract and also about the articulatory and acoustic consequences of linguistically significant gestures. Further constraints become available as experience with the phonology of a particular language reduces the inventory o possible gestures and provides information about the phonotactic and temporal restrictions on their occurrence. The module has then merely to determine which (if any) of the small number of gestures that might have been initiated at a particular instant could, in combination with gestures already in progress, account for the signal.

Thus, we would claim that the processes of speech percepti are, like other linguistic processes, inherently computational and quite indirect. If perception seems nonetheless immediate, it is not because the process is in fact straightforward, but because the module is so well-adapted to its complex task.

The Motor Theory and Modularity

In attributing speech perception to a "module," we have in mind the notion of modularity proposed by Fodor (1983). A module, for Fodor, is a piece of neural architecture that performs the special computations required to provide central cognitive processes with representations of objects or events belonging to a natural class that is ecologically significant for the organism. This class, the "domain" of the module, is apt also to be "eccentric," for the domain would be otherwise merely a province of some more general domain, for which another module must be postulated anyway. Besides domain-specificity and specialized neural architecture, a module has other characteristic properties. Because the perceptual process it controls is not cognitive, there is little or no possibility of awareness of whatever computations are carried on within the module ("limited central access"). Because the module is specialized, it has a "shallow" output, consisting only of rigidly definable, domain-relevant representations: accordingly, it processes only the domain-relevant information in the input stimulus. Its computations are thus much faster than those of the less specialized processes of central cognition. Because of the ecological importance of its domain for the organism, the operation of the module is not a matter of choice, but "mandatory"; for the same reason, its computations are "informationally encapsulated," that is, protected from cognitive bias.

Most psychologists would agree that auditory localization, to return to an example we have mentioned several times, is controlled by specialized processes of some noncognitive kind. They might also agree that its properties are those that Fodor assigns to modules. At all events, they would set auditory localization apart from such obviously cognitive activities as playing chess, proving theorems, and recognizing a particular chair as a token of the type called "chair." As for perception of language, the consensus is that it qualifies as a cognitive process par excellence, modular only in that it is supported by the mechanisms of the auditory modality. But in this, we and Fodor would argue, the consensus is doubly mistaken: the perception of language is neither cognitive nor auditory. The events that constitute the domain of linguistic perception, however they may be defined, must certainly be an ecologically significant natural class, and it has been recognized since Broca that linguistic perception is associated with specialized neural architecture. Evidently, linguistic perception is fast and mandatory; arguably, it is informationally encapsulated--that is, its phonetic, morphological and syntactic analyses are not biassed by knowledge of the world--and its output is shallow--that is, it produces a linguistic description of the utterance, and only this. These and other considerations suggest that, like auditory localization, perception of language rests on a specialization of the kind that Fodor calls a module.

The data that have led us in the past to claim that "speech is special" and to postulate a "speech mode" of perception can now be seen to be consistent with Fodor's claims about modularity, and especially about the modularity of language. (What we have been calling a phonetic module is then more properly called a linguistic module.) Thus, as we have noted, speech perception uses all the information in the stimulus that is relevant to phonetic structures: every potential cue proves to be an actual cue. This holds true even across modalities: relevant optical information combines with relevant acoustic information to produce a coherent phonetic percept in which, as in the example described earlier, the bimodal nature of the stimulation is not detectable. In contrast, irrelevant information in the stimulus is not used: the acoustic properties that might cause the transitions to be heard as chirps are ignored--or perhaps we should say that the auditory consequences of those properties are suppressed--when the transitions are in context and the linguistic module is engaged. The exclusion of the irrelevant extends, of course, to stimulus information about voice quality, which helps to identify the speaker (perhaps by virtue of some other module) but has no phonetic importance, and even to that extraphonetic information which might have been supposed to help the listener distinguish sounds that contain phonetic structures from those that do not. As we have seen, even when synthetic speech lacks the acoustic properties that would make it sound natural, it will be treated as speech if it contains sufficiently coherent phonetic information. Moreover, it makes no difference that the listener knows, or can determine on auditory grounds, that the stimulus was not humanly produced; because linguistic perception is informationally encapsulated and mandatory, the listener will hear synthetic speech as speech.

As might be expected, the linguistic module is also very good at excluding from consideration the acoustic effects of unrelated objects and events in the environment; the resistance of speech perception to noise and distortion is well known. These other objects and events are still perceived, because they are dealt with by other modules, but they do not, within surprisingly wide limits, interfere with speech perception (cf. Darwin, 1984). On the other hand, the module is not necessarily prepared for non-ecological conditions, as the phenomenon of duplex perception illustrates. Under the conditions of duplex perception the module makes a mistake it would never normally make: it treats the same acoustic information both as speech and as nonspeech. And, being an informationally encapsulated and mandatorily operating mechanism, it keeps on making the same mistake, whatever the knowledge or preference of the listener.

Our claim that the invariants of speech perception are phonetic gestures is much easier to reconcile with a modular account of linguistic perception than with a cognitive account. On the latter view, the gestures would have to be inferred from an auditory representation of the signal by some cognitive process, and this does not seem to be a task that would be particularly congenial to cognition. Parsing a sentence may seem to bear some distant resemblance to the proving of theorems, but disentangling the mutually confounding auditory effects of overlapping articulations surely does not. It is thus quite reasonable for proponents of a cognitive account to reject the possibility that the invariants are motoric and to insist that they are to be found at or near the auditory surface, heuristic matching of auditory tokens to auditory prototypes being perfectly plausible as a cognitive process.

Such difficulties do not arise for our claim on the modular account. If the invariants of speech are phonetic gestures, it merely makes the domain of linguistic perception more suitably eccentric; if the invariants were auditory, the case for a separate linguistic module would be the less compelling. Moreover, computing these invariants from the acoustic signal is a task for which there is no obvious parallel among cognitive processes. What is required for this task is not a heuristic process that draws on some general cognitive ability or on knowledge of the world, but a special-purpose computational device that relates gestural properties to the acoustic patterns.

It remains, then, to say how the set of possible gestures is specified for the perceiver. Does it depend on tacit knowledge of a kind similar, perhaps, to that which is postulated by Chomsky to explain the universal constraints on syntactic and phonological form? We think not, because knowledge of the acoustic-phonetic properties of the vocal tract, unlike other forms of tacit knowledge, seems to be totally inaccessible: no matter how hard they try, even post-perceptually, listeners cannot recover aspects of the process--for example, the acoustically different transitions--by which they might have arrived at the distal object. But, surely, this is just what one would expect if the specification of possible vocal-tract gestures is not tacit knowledge at all, but rather a direct consequence of the eccentric properties of the module itself. As already indicated, we have in earlier papers suggested that speech perception is accomplished by virtue of a model of the vocal tract that embodies the relation between gestural properties and acoustic information. Now we would add that this model must be part of the very structure of the language module. In that case, there would be, by Fodor's account, an analogy with all other linguistic universals.

Perception and Production: One Module or Two?

For want of a better word, we have spoken of the relation between speech perception and speech production as a "link," perhaps implying thereby that these two processes, though tightly bonded, are nevertheless distinct. Much the same implication is carried, more generally, by Fodor's account of modularity, if only because his attention is almost wholly on perception. We take pains, therefore, to disown the implication of distinctness that our own remarks may have conveyed, and to put explicitly in its place the claim that, for language, perception and production are only different sides of the same coin.

To make our intention clear, we should consider how language differs from those other modular arrangements in which, as with language, perception and action both figure in some functional unity: simple reflexes, for example; or the system that automatically adjusts the posture of a diving gannet in accordance with optical information that specifies the time of contact with the surface of the water (Lee & Reddish, 1981). The point about such systems is that the stimuli do not resemble the responses, however intimate the connection between them. Hence, the detection of the stimulus and the initiation of the response must be managed by separate components of the module. Indeed, it would make no great difference if these cases were viewed as an input module hardwired to an output module.

Language is different: the neural representation of the utterance that determines the speaker's production is the distal object that the listener perceives; accordingly, speaking and listening are both regulated by the same structural constraints and the same grammar. If we were to assume two modules, one for speaking and one for listening, we should then have to explain how the same structures evolved for both, and how the representation of the grammar acquired by the listening module became available to the speaking module.

So, if it is reasonable to assume that there is such a thing as a language module, then it is even more reasonable to assume that there is only one. And if, within that module, there are subcomponents that correspond to the several levels of linguistic performance, then each of these subcomponents must deal both with perception and production. Thus, if sentence planning is the function of a particular subcomponent, then sentence parsing is a function of the same subcomponent, and similarly, mutatis mutandis, for speech production and speech perception. And, finally, if all this is true, then the corresponding input and output functions must themselves be as computationally similar as the inherent asymmetry between production and perception permits, just as they are in man-made communication devices.

These speculations do not, of course, reveal the nature of the computations that the language module carries out, but they do suggest a powerful constraint on our hypotheses about them, a constraint for which there is no parallel in the case of other module systems. Thus, they caution that, among all plausible accounts of language input, we should take seriously only those that are equally plausible as accounts of language output; if a hypothesis about parsing cannot be readily restated as a hypothesis about sentence-planning, for example, we should suppose that something is wrong with it.

Whatever the weaknesses of the motor theory, it clearly does conform to this constraint, since, by its terms, speech production and speech perception are both inherently motoric. On the one side of the module, the motor gestures are not the means to sounds designed to be congenial to the ear; rather, they are, in themselves, the essential phonetic units. On the other side, the sounds are not the true objects of perception, made available for linguistic purposes in some common auditory register; rather, they only supply the information for immediate perception of the gestures.

References

Abramson, A. S. (1972). Tonal experiments with whispered Thai. In A. Valdman (Ed.), Papers in linguistics and phonetics to the Memory of Pierre Delattre (pp. 31-44). The Hague: Mouton.

Alegria, J., & Noirot, E. (1982). Oriented mouthing activity in neonates: Early development of differences related to feeding experiences. In J. Mehler, S. Franck, E. C. T. Walker & M. Garrett (Eds.), Perspectives on mental representation. Hillsdale, NJ: Erlbaum

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.), Child phonology. New York: Academic Press.

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.

Bentin, S., & Mann, V. A. (1983). Selective effects of masking on speech and nonspeech in the duplex perception paradigm. Haskins Laboratories Status Report on Speech Research, SR-76, 65-85.

Berkeley, G. (1709). An essay towards a new theory of vision. Dublin: Printed by Aaron Rhames for Jeremy Pepyal.

Best, C. T., Hoffman, H., & Glanville, B. B. (1982). Development of infant ear asymmetries for speech and music. Perception & Psychophysics, 31, 75-85.

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

Best, C. T., & Studdert-Kennedy, M. (1983). Discovering phonetic coherence in acoustic patterns. In A. Cohen & M. P. R. van den Broecke (Eds.), Abstracts of the Tenth International Congress of Phonetic Sciences. Dordrecht, The Netherlands: Foris Publications.

Brady, P. T., House, A. S., & Stevens, K. N. (1961). Perception of sounds characterized by a rapidly changing resonant frequency. Journal of the Acoustical Society of America, 33, 1357-1362.

Browman, C. P., & Goldstein, L. M. (1985). Dynamic modeling of phonetic structure. In V. Fromkin (Ed.), Phonetic linguistics. New York: Academic Press.

Brown, E. L., & Deffenbacher, K. (1979). Perception and the senses. New York: Oxford University Press.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. Journal of the Acoustical Society of America, 62, 961-970.

Chistovich, L. A. (1960). Classification of rapidly repeated speech sounds. Akustichneskii Zhurnal, 6, 392-398. Tr. in Soviet Physics-Acoustics, 6, 393-398 (1961).

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. Psychological Review, 81, 348-374.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America*, *24*, 597-606.

Crowder, R. G., & Morton, J. (1969). Pre-categorical acoustic storage (PAS). *Perception & Psychophysics*, *5*, 365-373.

Cutting, J. E., & Rosner, B. S. (1974). Categories and boundaries in speech and music. *Perception & Psychophysics*, *16*, 564-570.

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.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. (1979). Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, *65*, 1518-1532.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, *22*, 109-122.

Dudley, H. (1940). The carrier nature of speech. *Bell Systems Technical Journal*, *19*, 495-515.

Eimas, P., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in early infancy. *Science*, *171*, 304-306.

Entus, A. K. (1977). Hemispheric asymmetry in processing dichotically presented speech and nonspeech stimuli by infants. In S. J. Segalowitz & F. A. Greber (Eds.), *Language development and neurological theory*. New York: Academic Press.

Fant, C. G. M. (1962). Descriptive analysis of the acoustic aspects of speech. *Logos*, *5*, 3-17.

Festinger, L., Burnham, C. A., Ono, H., & Bamber, D. (1967). Efference and the conscious experience of perception. *Journal of Experimental Psychology Monograph*, *74* (4, Pt. 2).

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (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.

Fowler, C. A. (1982). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General*, *112*, 386-412.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, *36*, 359-368.

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.

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.

Gerhardt, [illegible]. C., & Rheinlaender, J. (1982). Localization of an elevated soun source by the green tree frog. *Science*, *217*, 663-664.

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

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.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, *1*, 1-7.

Hillenbrand, J. (1984). Perception of sine-wave analogs of voice onset time stimuli. *Journal of the Acoustical Society of America, 75*, 231-240.

Hirsh, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America, 31*, 759-767.

Hoffman, H. S. (1958). Study of some cues in the perception of the voiced stop consonants. *Journal of the Acoustical Society of America, 30*, 1035-1041.

Howell, P., & Rosen, S. (1983). Closure and frication measurements and perceptual integration of temporal cues for the voiceless affricate/fricative contrast. *Speech Hearing and Language Work in Progress*. University College London, Department of Phonetics and Linguistics.

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.

Inoue, A. (1984). *A perceptual study of Japanese voiceless vowels and its implications for the phonological analysis of voiceless consonants.* Unpublished manuscript.

Jenkins, J. J., Strange, W., & Edman, T. R. (1983). Identification of vowels in 'voiceless' syllables. *Perception & Psychophysics, 34*, 441-450.

Joos, M. (1948). Acoustic phonetics. *Language Monograph 23*, Supplement to *Language, 24*.

Julesz, B. (1960). Binocular depth perception of computer-generated patterns. *Bell System Technical Journal 39*, 1125-1162.

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago: University of Chicago Press.

Katz, L. C., & Gurney, M. E. (1981). Auditory responses in the zebra finch's motor system for song. *Brain Research, 221*, 192-197.

Knudsen, E. I. (1982). Auditory and visual maps of space in the optic tectum of the owl. *Journal of Neuroscience, 2*, 1117-1194.

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, *Dynamic aspects of neocortical function*. New York: Wiley.

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., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138-1144.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190*, 69.

Ladefoged, P., & McKinney, N. (1963). Loudness, sound pressure, and subglottal pressure in speech. *Journal of the Acoustical Society of America, 35*, 454-460.

Lee, D. N., & Reddish, P. E. (1981). Plummeting gannets: A paradigm of ecological optics. *Nature, 293*, 293-294.

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

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., Delattre, P. C., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. American Journal of Psychology, 65, 497-516.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychological Monographs, 68, 1-13.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. Perception & Psychophysics, 30, 133-143.

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.

Liberman, A. M., & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), Handbook of sensory physiology, Vol. VIII: Perception. New York: Springer-Verlag.

Lisker, L. (1957). Closure duration, first-formant transitions, and the voiced-voiceless contrast of intervocalic stops. Haskins Laboratories Quarterly Progress Report, 23, Appendix 1.

Lisker, L. (1978). Rapid vs. rabid: A catalogue of acoustic features that may cue the distinction. Haskins Laboratories Status Report on Speech Research, SR-54, 127-132.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. Word, 20, 384-422.

MacKain, K. S., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left hemisphere function. Science, 219, 1347-1349.

MacNeilage, P. F., Rootes, T. P., & Chase, R. A. (1967). Speech production and perception in a patient with severe impairment of somesthetic perception and motor control. Journal of Speech and Hearing Research, 10, 449-468.

Magen, H. (1984). Vowel-to-vowel coarticulation in English and Japanese. Journal of the Acoustical Society of America 75, S41.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. Cognition, 14, 211-235.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on the perception of [ʃ]-[s] distinction: 1. Temporal factors. Perception & Psychophysics, 28, 213-228.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. Journal of the Acoustical Society of America, 69, 548-558.

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.

Margolish, D. (1983). Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. Journal of Neuroscience, 3, 1039-1057.

Marler, P. (1970). Birdsong and speech development: Could there be parallels? American Scientist, 58, 669-673.

Martin, J. G., & Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects in /stri, stru/ sequences. Journal of the Acoustical Society of America, 69, S92.

Martin, J. G., & Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 473-488.

Mattingly, I. G. (1976). Phonetic prerequisites for first-language acquisition. In W. Von Raffler-Engel & Y. Lebrun (Eds.), *Baby talk and infant speech*. Lisse, Netherlands: Swets & Zeitlinger.

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-Verlag.

Mattingly, I. G., & Liberman, A. M. (1985). Verticality unparalleled. *The Behavioral and Brain Sciences*, *8*, 24-26.

Mattingly, I. G., Liberman, A. M., Syrdal, A. M., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology*, *2*, 131-157.

McCasland, J. S., & Konishi, M. (1983). Interaction between auditory and motor activities in an avian song control nucleus. *Proceedings of the National Academy of Sciences*, *78*, 7815-7819.

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

Miller, J. D. (1977). Perception of speech sounds in animals: Evidence for speech processing by mammalian auditory mechanisms. In T. H. Bullock (Ed.), *Recognition of complex acoustic signals* (Life Sciences Research Report 5) (p. 49). Berlin: Dahlem Konferenzen.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, 18, 331-340.

Molfese, D. L., Freeman, R. B., & Palermo, D. S. (1975). The ontogeny of brain lateralization for speech and nonspeech stimuli. *Brain and Language*, *2*, 356-368.

Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in the canary, Serinus canarius. *Journal of Comparative Neurology*, *165*, 457-486.

O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1957). Acoustic cues for the perception of initial /w,r,l/ in English. *Word*, *13*, 25-43.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, *85*, 172-191.

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

Pastore, R. E., Schmuckler, M. A., Rosenblum, L., & Szczesiul, R. (1983). Duplex perception with musical stimuli. *Perception & Psychophysics*, *33*, 469-474.

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

Pisoni, D. B. (1977). Identification and discrimination of the relative onset of two component tones: Implications for the perception of voicing in stops. *Journal of the Acoustical Society of America*, *61*, 1352-1361.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*, 285-290.

Poggio, G. F. (1984). Processing of stereoscopic information in primate visual cortex. In G. M. Edelman, W. E. Gall & W. M. Cowan (Ed.), *Dynamic aspects of neocortical function*. New York: Wiley.

Rakerd, B., Dechovitz, D. R., & Verbrugge, R. R. (1982). An effect of sentence finality on the phonetic significance of silence. *Language and Speech*, *25*, 7-282.

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

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

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

Remez, R. E., & Rubin, P. E. (1984). On the perception of intonation from sinusoidal signals: Tone height and contour. *Journal of the Acoustical Society of America* *75*, S39.

Repp, B. H. (1984). The role of release bursts in the perception of [s]-stop clusters. *Journal of the Acoustical Society of America*, *75*, 1219-1230.

Repp, B. H., & Liberman, A. M. (in press). Phonetic categories are flexible. In S. Harnad (Ed.), *Categorical perception*. Cambridge, Cambridge University Press.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 621-637.

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

Rosen, S. M., & Howell, P. (1981). Plucks and bows are not categorically perceived. *Perception & Psychophysics*, *30*, 156-168.

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

Segalowitz, S. J., & Chapman, J. S. (1980). Cerebral asymmetry for speech in neonates: A behavioral measure. *Brain and Language*, *9*, 281-288.

Sondhi, M. M. (1979). Estimation of vocal-tract areas: the need for acoustical measurements. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *ASSP-27*, 268-273.

Stetson, R. H. (1951). *Motor phonetics: A study of speech movements in action*. Amsterdam: North Holland.

Stevens, K. N. (1975). The potential role of property detectors in the perception of consonants. In G. Fant & M. A. Tatham (Eds.), *Auditory analysis and perception of speech*. New York: Academic Press.

Stevens, K. N., & Halle, M. (1977). Remarks on analysis by synthesis and distinctive features. In W. Wathen-Dunn (Ed.). *Models for the perception of speech and visual form*. Cambridge, MA: MIT Press.

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.

Strange, W., Jenkins, J. J., & Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America*, *74*, 695-705.

Studdert-Kennedy, M. (1976). Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press.

Studdert-Kennedy, M., & Williams, D. R. (1984). Range effects for speech and nonspeech judgments of sine wave stimuli. *Journal of the Acoustical Society of America*, *75*, S64.

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

Summerfield, Q. (1982). Differences between spectral dependencies in auditory and phonetic temporal processing: Relevance to the perception of voicing in initial stops. Journal of the Acoustical Society of America, 72, 51-61.

Summerfield, Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. Journal of the Acoustical Society of America, 62, 436-448.

Thorpe, W. H. (1958). The learning of song patterns by birds, with especial reference to the song of the chaffinch, Fringilla coelebs. Ibis, 100, 535-570.

Tuller, B., Harris, K., & Kelso, J. S. (1982). Stress and rate: Differential transformations of articulation. Journal of the Acoustical Society of America, 71, 1534-1543.

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.

Turvey, M. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting, and knowing: Toward an ecological physiology. Hillsdale, NJ: Erlbaum.

Vargha-Khadem, F., & Corballis, M. (1979). Cerebral asymmetry in infants. Brain and Language, 8, 1-9.

Washburn, M. F. (1926). Gestalt psychology and motor psychology. American Journal of Psychology, 37, 516-520.

Watson, J. B. (1919). Psychology from the standpoint of a behaviorist. Philadelphia: J. B. Lippincott Co.

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

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. Journal of the Acoustical Society of America, 69, 275-282.

Williams, D. R., Verbrugge, R. R., & Studdert-Kennedy, M. (1983). Judging sine wave stimuli as speech and nonspeech. Journal of the Acoustical Society of America, 74, S66.

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

Witelson, S. (1977). Early hemisphere specialization and interhemispheric plasticity: An empirical and theoretical review. In S. J. Segalowitz & F. A. Gruber (Eds.), Language development and neurological theory. New York: Academic Press.

Yeni-Komshian, G. H., & Soli, S. D. (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. Journal of the Acoustical Society of America, 70, 966-975.

LINGUISTIC AND ACOUSTIC CORRELATES OF THE PERCEPTUAL STRUCTURE FOUND IN AN INDIVIDUAL DIFFERENCES SCALING STUDY OF VOWELS*

Brad Rakerd† and Robert R. Verbrugge††

Abstract. Subjects judged the similarities among a set of American English vowels (/i,ɪ,ɛ,æ,ʌ,ɑ,ɔ,o,ʊ,u/) presented in isolation or in a /dVd/ consonantal frame. Individual differences scaling was employed to analyze these similarities data for each of the conditions separately and for the two conditions combined. In all cases, perceptual dimensions corresponding to the advancement, height, and tenseness vowel features were recovered. Given the determinacy of individual differences scaling, this finding is taken to provide strong evidence for the perceptual significance of those features. The perceptual dimensions are considered in relation to various acoustic parameters of the stimuli employed in this study. They are also considered in relation to perceptual dimensions that have been observed in other vowel scaling studies.

Introduction

Multidimensional scaling provides a means of modeling the psychological structure that is reflected in perceptual judgments. Scaling is particularly useful because judgments regarding a large number of stimuli can very often be modeled with a structure of relatively few dimensions, and because those dimensions can then be interpreted in terms of properties familiar to an investigator (Carroll & Wish, 1974; Kruskal & Wish, 1978). In the domain of vowel perception, investigators have frequently found that the dimensions revealed by scaling can be related to various phonological features, a fact which is taken to imply that those features play a significant perceptual role (e.g., Fox, 1983; Singh & Woods, 1970; Shepard, 1972).

The strength of that implication is, however, contingent on the type of scaling method that is used in a study. One class of scaling techniques yields solutions for which no single interpretation is possible. This owes to the fact that the models of psychological structure, which are spatial in

---

*Journal of the Acoustical Society of America, 1985, 77, 296-301

†Michigan State University, Department of Psychology.

††Bell Laboratories.

Acknowledgment. This report is based on portions of a doctoral dissertation submitted by the first author to the University of Connecticut. The research was conducted at Haskins Laboratories and funded by grants awarded to that institution (NIH grants HD-01994, RR-05596). Support was also provided by the University of Connecticut Research Foundation. We thank Alvin Liberman, Michael Turvey, and Benjamin Sachs for their comments on the manuscript.

character, lack a fixed orientation for their axes. One must therefore rotate these structures in search of an orientation that permits interpretation of the dimensions. There are an infinite number of possible rotations and the search must be constrained by an investigator's *a priori* notions regarding interpretation. Any conclusions drawn are correspondingly vulnerable to the challenge that some alternative interpretation would have been equally supported by the data had some other rotation been carried out.

A second class of scaling techniques cannot be challenged on these grounds because they specify a fixed orientation of dimension axes for their models of psychological structure. This class, the *individual differences* scaling techniques, achieve their added determinacy by modeling multiple sets of data simultaneously, each set reflecting the performance of a different subject.[1] An important underlying assumption of individual differences scaling is that when judging a common set of stimuli subjects can differ from one another in terms of the relative weights they attach to a set of shared perceptual dimensions, but not in terms of the identity of the dimensions themselves (Carroll & Chang, 1970). Except in unusual cases, there is one and only one orientation in which the shared dimensions can be weighted so as to account optimally for the variance in those subjects' data. That is the orientation recovered by individual difference scaling. It has been conjectured that with a well-defined perceptual task the dimensions revealed by individual differences scaling will correspond to fundamental sensory or judgmental processes (Carroll & Chang, 1970). There are a number of instances in which that conjecture has been supported (Wish & Carroll, 1974).

In this paper, we report on an individual differences scaling study of vowel perception. It was conducted to address questions about the potential influences that consonantal context can exert on vowel perception, and elsewhere (Rakerd, 1984) we have considered the results in that regard. We did so by comparing the weights that subjects attached to a set of shared perceptual dimensions, depending on whether they heard vowels in or out of a consonantal frame. Our concern here is not with the weights, however, but with the shared dimensions themselves. Those dimensions can be usefully compared with linguistic features that have been found to be related to perceptual structure in other scaling studies (e.g., Fox, 1982, 1983; Terbeek, 1977), particularly those conducted with less determinate scaling techniques (Hanson, 1967; Pols, van der Kamp, & Plomp, 1969; Shepard, 1972; Singh & Woods, 1970). That is the first purpose of this paper. We examine subjects who judged vowels in consonantal context and subjects who judged isolated vowels, analyzing their data both separately and in combination.

The second purpose of this paper is to report on correlations between the perceptual structure revealed by individual differences scaling and various acoustic parameters of our vowel stimuli. Though based on a limited number of stimulus tokens, those correlations are suggestive in that they speak to hypotheses that previous investigators have put forth regarding relationships between vowel features and the acoustic signal.

I. Methods

A. *Subjects*

Twenty-three subjects participated in this experiment. All of them were native speakers of English with normal hearing according to self-report. Twelve of the subjects were randomly assigned to make perceptual judgments regarding vowels in consonantal context. The remaining 11 subjects judged vowels in isolation.

B. Stimuli

The stimuli for the experiment were ten different American English vowels (/i,ɪ,ɛ,æ,ʌ,ɑ,ɔ,o,ʊ,u/) spoken by a male talker with a general American dialect. For the consonantal context condition, he produced those vowels in the trisyllabic frame /hədVdə/, with stress placed on the second syllable (/dVd/). For the isolated condition, he produced them with no surrounding phonetic context (/#V#/). Two tokens of each vowel were produced in each condition. Recordings of these tokens were digitized at a sampling rate of 10 kHz and stored in separate computer files.

C. Procedure

Subjects were tested individually. Their task was to judge the similarity relations that they perceived among the ten different vowels. They were instructed to base those judgments on properties of the vowel sounds that seemed to them to distinguish words in English (Carlson & Granström, 1979; Klatt, 1979). The similarity judgments were made with a triadic comparisons method that has been employed in previous vowel perception studies (Pols et al., 1969; Terbeek, 1977; Terbeek & Harshman, 1971). According to this procedure, three of the ten vowels were rated on each experimental trial. Subjects listened to these vowels in any order that they chose and as often as they chose.[2] They then reported which two of the three vowels sounded most alike to them and which two least alike. Over trials, all possible triads were judged. The judgments were then summed across trials, with a score of +1 assigned to all most-alike pairs and -1 to all least-alike pairs. This yielded a single (symmetrical) matrix of similarity judgments for each subject.

D. Data analysis

The matrices for all 23 subjects who participated in the experiment were submitted to nonmetric individual differences scaling, using the ALSCAL procedure developed by Takane, Young, and Leuuw (1977). It was determined that a three-dimensional scaling solution was most appropriate for the data. That decision was based on several factors. First, modeling in three dimensions accounted for a substantially greater percentage of variance (an average of 70% for each subject) than modeling in two dimensions (60%), and only marginally less than modeling in four dimensions (72%). Second, the three dimensions were readily interpretable from a linguistic standpoint. And finally, those dimensions were quite stable, in that they were also found in separate analyses of the two experimental conditions (see Sec. II) and, with certain modeling constraints, in the scaling solution for a memory study (Rakerd, 1984) that complemented this perceptual study.

For additional details concerning the data analysis, as well as other aspects of the experimental method, see Rakerd (1984).

II. The Scaling Solutions

We first consider the perceptual dimensions that emerged from an analysis of data matrices for all 23 subjects. Although these dimensions have been described elsewhere by Rakerd (1984), they are examined here in greater detail, with particular attention paid to comparisons with phonological features of linguistic description for vowels, and with dimensions that have been reported in previous scaling studies of vowels. In the second part of Sec. II, we describe the perceptual dimensions that resulted from *separate* analyses of the consonantal-context and isolated conditions of the study.

A. The two conditions combined

1. Dimensions 1 and 2

Dimension 2 (D2) of the scaling solution for all subjects is plotted against dimension 1 (D1) in the top half of Fig. 1. The distribution of vowels in this plane is clearly related to the traditional "vowel quadrilateral" (Ladefoged, 1975; Lindau, 1978), with D1 corresponding to the *advancement* feature of vowels,[3] and D2 to the *height* feature. There is considerable precedent for observing correlates of these two phonological features in vowel scaling studies (Fox, 1982, 1983; Hanson, 1967; Pols et al., 1969; Shepard, 1972; Singh & Woods, 1970). Those findings, together with the results of the present study, strongly support the view that the advancement and height features play a significant role in the perception of vowels in English. The findings are also consistent with the larger view that advancement and height enjoy a special status in all languages (Lindau, 1978).

2. Dimension 3

The third dimension of the combined group space (D3) is plotted against D1 in the bottom half of Fig. 1. The vowels are ordered along it such that /ɪ,æ,ɛ,ʌ,ʊ/ have negative values and /i,ɑ,ɔ,o,u/ have positive values. The former are lax vowels, the latter tense. Hence, D3 can be interpreted as corresponding to the *tenseness* feature. Unlike advancement and height, a tenseness dimension has very rarely been recovered in vowel scaling studies. To our knowledge, only Anglin (1971; cited in Singh, 1976), who scaled similarity judgments for vowels in /hVd/ context, has recovered a dimension similar to D3. In that analysis, the scaling method did not yield a single, interpretable orientation for the model of psychological structure. The present, more determinate scaling result might therefore be taken to provide the strongest available evidence for perceptual significance of the tenseness feature.

B. Separate analyses of the conditions

When perceptual judgments for the isolated and consonantal-context conditions were scaled separately, in three dimensions, the amount of variance that could be accounted for in the data (VAF) improved marginally over its corresponding value in the combined analysis. (VAF for analysis of the isolated condition was 74%, that for the consonantal-context condition was 72%. This compares with 70% in the combined analysis.) This marginal improvement resulted from some local shifts in the positioning of vowels in the separate scaling solutions. As will be seen, the global structure nevertheless remained quite similar to that of the combined analysis.

1. The isolated condition

The perceptual dimensions for the isolated condition are shown in Fig. 2. Only D2 is notably different from the corresponding dimensions of the combined analysis (see Fig. 1). Along this dimension, the vowels /ɛ/ and /o/ have as-

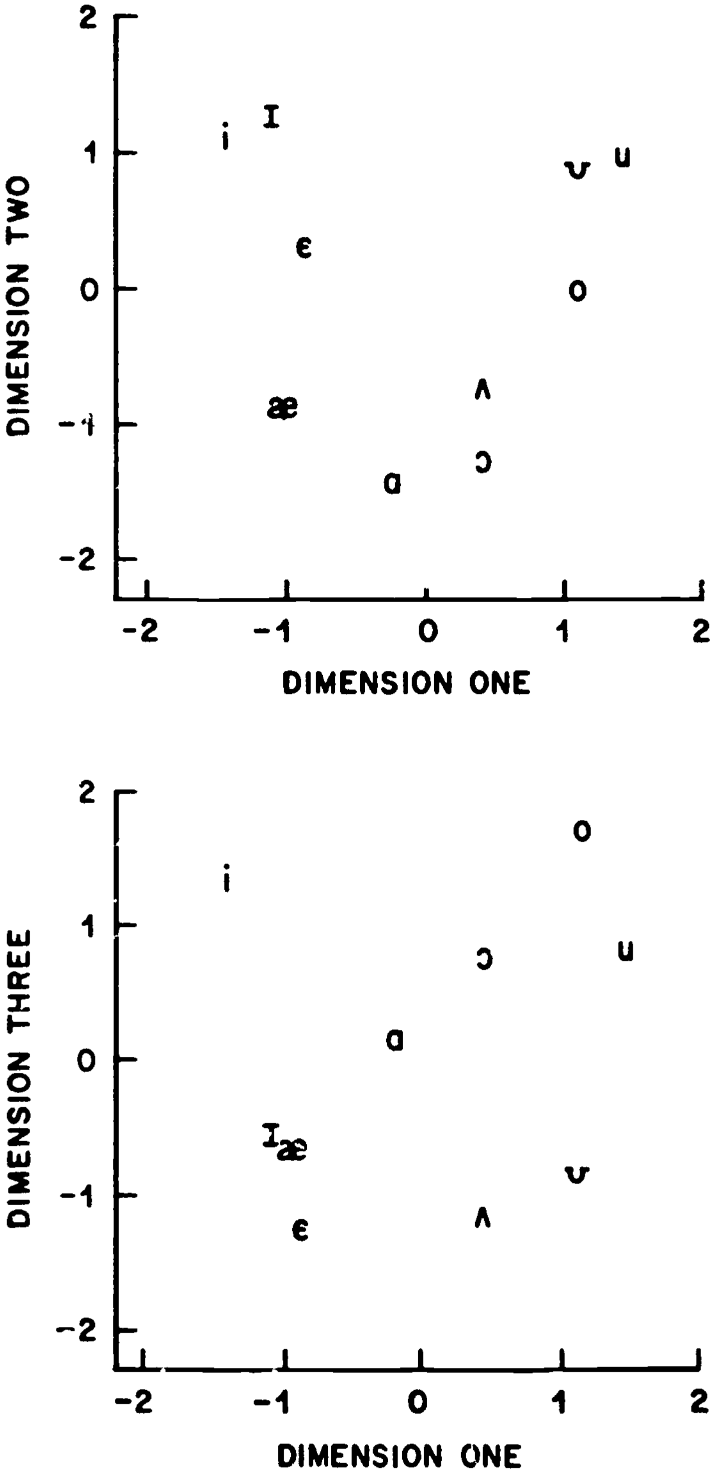
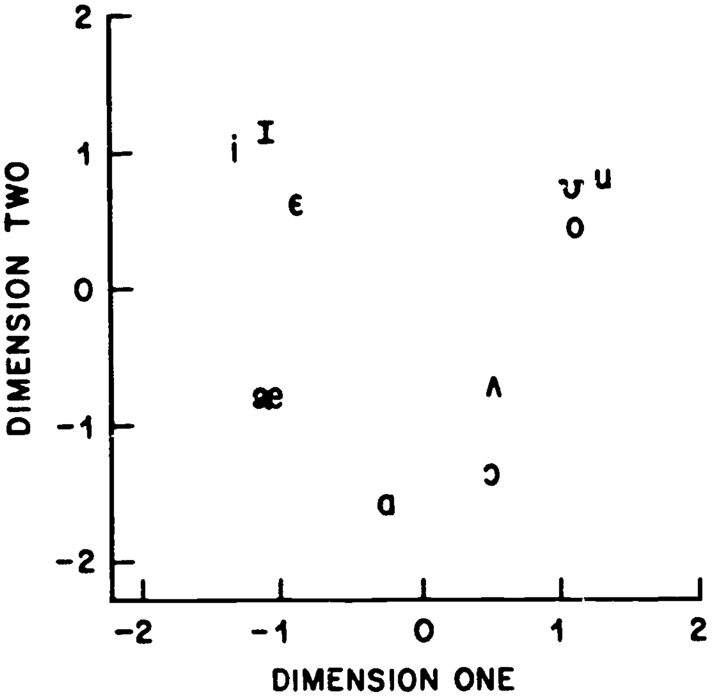


Figure 1. Perceptual structure for subjects from the two experimental conditions combined. This figure is reproduced from Rakerd (1984) by permission of The Psychonomic Society.



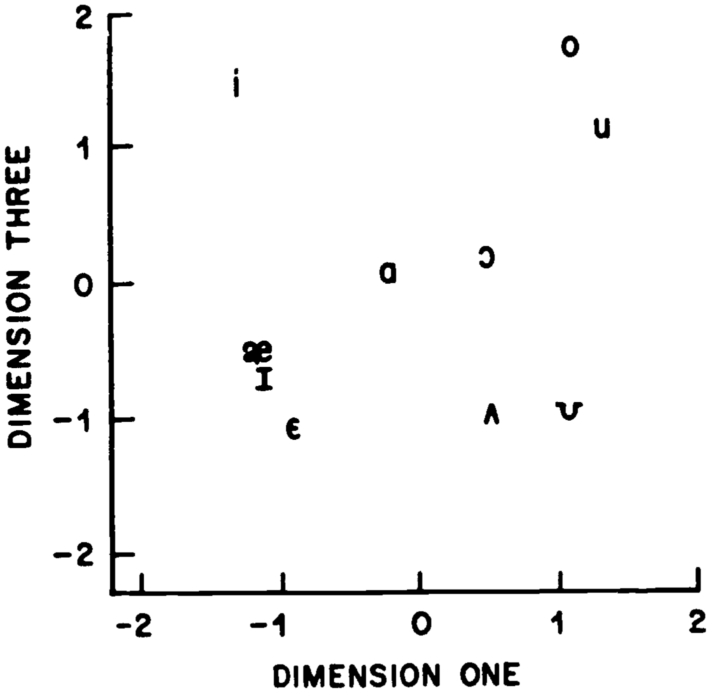


Figure 2. Perceptual structure for subjects from the isolated condition.

sumed values that are somewhat more positive than they had been previously. The movement of /ɛ/ principally reflects the fact that /ɛ/ and /ɪ/ were judged to be highly similar, indeed, the most similar of all vowel pairs in the isolated condition. Likewise, the movement of /o/ is largely dictated by a single vowel pairing; /ɔ/ and /u/ were judged to be extremely similar in isolation, perhaps reflecting the fact that they were the only two diphthongized vowels in the isolated set.

Despite repositioning of these two vowels, the dimensions of the isolated solution maintain a strong correspondence with the advancement, height, and tenseness features, respectively.

This analysis can be usefully compared with one by Singh and Woods (1970) in which it was found that tenseness had no perceptual significance for listeners who rated the relative similarity of isolated vowels. Those investigators attributed their finding to listeners' knowledge that isolated lax vowels are phonologically impermissible in English. The outcome of the present study indicates that there may have been other factors at work as well. For several of our isolated-vowels subjects, the tenseness dimension (D3) did, indeed, have little or no perceptual salience, but for others it was the most heavily weighted dimension (Rakerd, 1984). Perhaps talkers produced their isolated vowels differently in the Singh and Woods study, or perhaps, by averaging their data over subjects prior to scaling, Singh and Woods lost any statistical evidence of the significance of tenseness. Whatever the case, it is apparent that under certain conditions listeners can attend to the tenseness dimension of isolated vowels, despite the phonological restriction.

2. The consonantal-context condition

Perceptual dimensions for the separate analysis of the consonantal-context condition are shown in Fig. 3. D1 and D2 are quite similar to their counterparts in the combined analysis (Fig. 1), again reflecting sensitivity to advancement and vowel height, respectively. Along the third dimension, there is some divergence from the combined solution, with the vowel /ɪ/ moving in a more positive direction. This movement resulted from the fact that the /i-ɪ/ vowel pair was judged highly similar in consonantal context. Nevertheless, D3 retains a correspondence with tenseness.

3. Stability of the scaling solutions

The agreement among these separate scaling solutions and the combined solution is evidence of the stability of this modeling outcome. Perceptual dimensions closely related to advancement, height, and tenseness were recovered in all cases, which makes it extremely unlikely that their emergence in any individual case was a coincidental consequence of the scaling analysis itself.

III. Acoustic Correlates of the Perceptual Dimensions

We computed correlations to assess the strength of relationships between the perceptual dimensions revealed by our combined scaling analysis and various acoustic parameters of the vowel stimuli. The acoustic measurements were made from wideband spectrograms. In the case of isolated vowels, center frequencies of the first three formants (F1, F2, and F3) were measured at a point approximately halfway through each token. Duration of voicing was also measured for the isolated vowels.

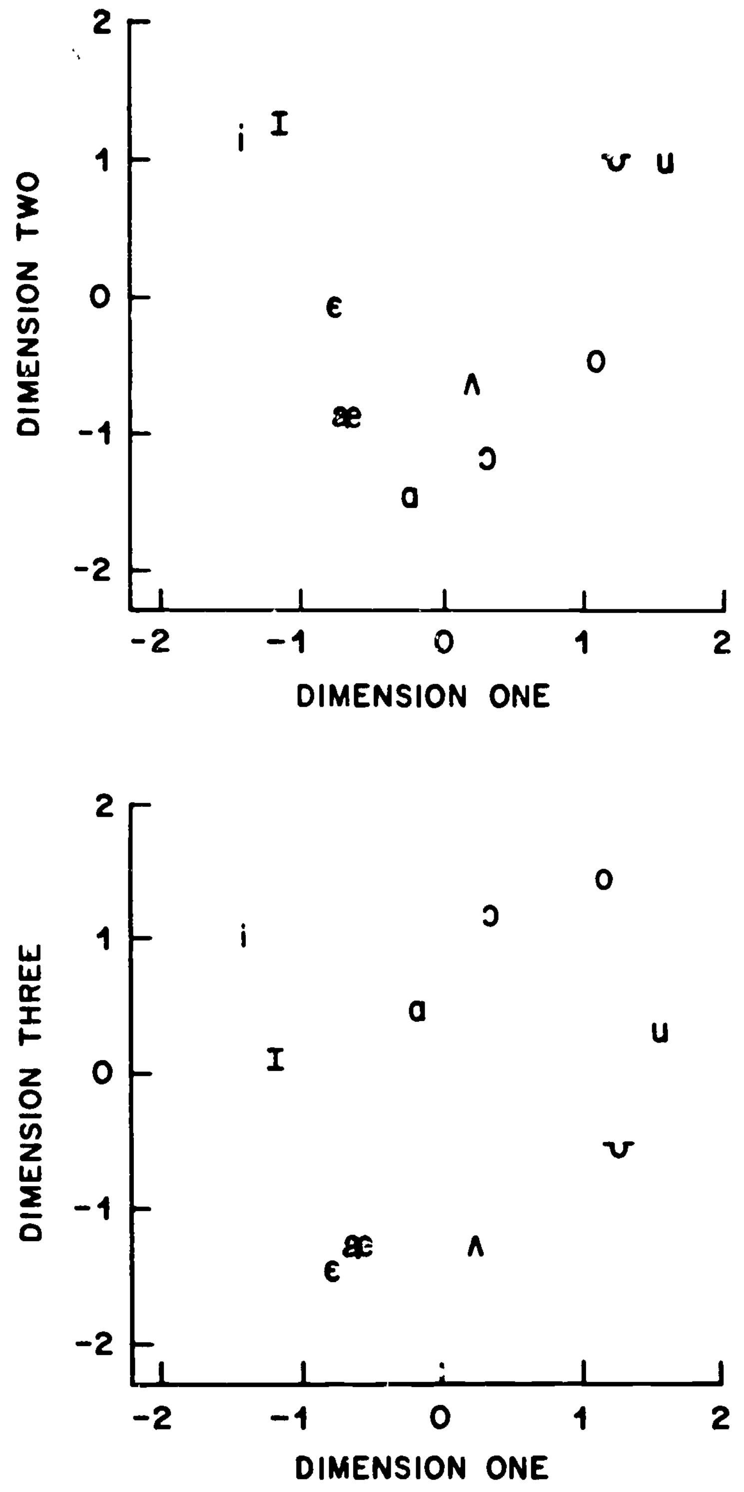


Figure 3. Perceptual structure for subjects from the consonantal-context condition.

The acoustic structure of the /dVd/ syllables comprised an onglide (or period of syllable-initial formant transition) and an offglide (syllable-final transition), with little or no region in which the formants could be described as maintaining a steady-state frequency. Therefore, we adopted the convention of measuring F1, F2, and F3 at the end of the onglide (a point also representing the beginning of the offglide). Duration was measured from the first evidence of voicing following initial-/d/ release to the last prior to final-/d/ closure. Last, we computed the proportion of total syllable duration that was taken up by the offglide.

Recall that there were two tokens of each vowel in each context. The mean parameter values for those two tokens are listed in Table 1. Isolated-vowel parameters appear in the top half of the table, /dVd/ parameters in the bottom half. An examination of Table 1 shows that the stimuli were acoustically "normal" in the sense that their parameters were roughly comparable to those that other investigators have reported for much larger data bases (Klatt, 1975; Peterson & Barney, 1952; Peterson & Lehiste, 1960; Umeda, 1975). The data also provide evidence of vowel reduction (Joos, 1948; Lindblom, 1963) in consonantal context. Formant frequency differences among the vowels were smaller in the /dVd/ condition than in isolation.

Rank-order correlations (Spearman's <u>rho</u>) were computed between the acoustic data reported in Table 1 and coordinates for the perceptual dimensions of the combined analysis. The results are reported in Table 2. First consider correlations for the isolated vowels, which appear in the top half of the table. The following correlations (and no others) proved significant: D1 (which we have interpreted as advancement) with F2 and F3, D2 (height) with F1, and D3 (tenseness) with duration. The findings regarding D1 and D2 are anticipated by a number of previous scaling studies (Fox, 1982, 1983; Pols et al., 1969; Shepard, 1972). The finding for D3 is consistent with the report that vowel tenseness is related to duration (Peterson & Lehiste, 1960).

The bottom half of Table 2 shows correlations for vowel in /dVd/ context. Note that relative to the isolated vowels there is a substantial reduction in the strength of the correlation between D1 and F2 (0.72, down from 0.95) and between D1 and F3 (0.66, down from 0.84). These statistical changes reflect the fact that the high-back vowels /o,ʊ,u/ were radically reduced in /dVd/ context, as might be expected given the alveolar place of articulation of the consonants. Though not unusual, this circumstance merits comment in that it calls into question strong statements to the effect that the relationship between the advancement feature and the formant structure of vowels is a simple one (see, e.g., Lindau, 1978; Singh, 1976). Our finding is one of the sort that shows that this relationship is affected by the phonetic context in which a vowel occurs.

It can also be seen in Table 2 that, in the consonantal-context condition, duration was not significantly correlated with D3 (tenseness), as it had been with isolated vowels. It appears that judgments regarding D3 could not have been made on the basis of vowel duration in this condition. Apparently, subjects' perceptions of tenseness were cued by some other acoustic property in the /dVd/ context. A likely candidate is offglide proportion, which was significantly correlated with D3. Indeed, it is possible to account perfectly for at least the macrostructure of D3 ordering on the basis of offglide proportion alone. Table 1 shows that the tense vowels, which all had positive D3 coordinates, also had offglide proportions of 50% or less, and that the lax

Table 1

Acoustic Parameters of the Stimuli

Condition	Vowel	Formant frequencies F1	F2	F3	Duration	Offglide proportion
Isolated vowels	i	225	2210	2835	235	...
	ɪ	465	1920	2600	165	...
	ɛ	555	1620	2135	180	...
	æ	665	1200	2225	180	...
	ʌ	640	1640	2170	155	...
	ɑ	780	1180	2090	235	...
	ɔ	680	1000	2175	195	...
	o	515	950	2110	225	...
	ʊ	565	1125	2020	135	...
	u	395	875	2085	220	...
Consonantal context	i	330	2060	2595	120	0.50
	ɪ	455	1795	2435	90	0.61
	ɛ	545	1640	2515	125	0.61
	æ	620	1595	2280	165	0.67
	ʌ	575	1375	2000	125	0.60
	ɑ	730	1300	2200	175	0.42
	ɔ	655	1005	2125	175	0.45
	o	530	1160	2045	160	0.37
	ʊ	460	1460	2380	95	0.60
	u	420	1355	2110	130	0.36

Table 2

Rank-order Correlations Between Acoustic Parameters of the Stimuli and Perceptual Dimensions of the Combined Analysis

Condition	Acoustic parameter	Perceptual dimensions D1	D2	D3
Isolated vowels	F1	0.12	-0.92^{b}	-0.37
	F2	-0.95^{b}	0.30	-0.31
	F3	-0.84^{b}	0.19	0.09
	Duration	-0.13	-0.17	0.76^{b}
Consonantal context	F1	0.04	-0.95^{b}	-0.31
	F2	-0.72^{a}	0.67^{a}	-0.32
	F3	-0.66^{a}	0.47	-0.18
	Duration	0.23	-0.87^{b}	0.27
	Offglide prop.	-0.61^{a}	0.15	-0.73^{a}

$^{a}p < 0.05$
$^{b}p < 0.01$

vowels, which all had negative D3 coordinates, also had offglide proportions of 60% or more. This finding is reminiscent of an observation made by Lehiste and Peterson (1961), although our measurement procedures were somewhat different from theirs. In both instances, tense vowels were found to be marked by a relatively brief period of offglide into a following consonant, and lax vowels by an offglide that was more substantial in duration.

A number of investigators have reported that vowels in consonantal context are identified with greater accuracy than isolated vowels (Gottfried & Strange, 1980; Rakerd et al., 1984; Strange, Edman, & Jenkins, 1976; Strange, Verbrugge, Shankweiler, & Edman, 1979). It has been suggested that one reason for this perceptual advantage may be that the dynamic acoustic structure of syllables is a unique source of vowel information (Strange et al., 1976; Strange, Jenkins, & Johnson, 1983). Our observation of an association between offglide proportion and the tenseness feature is certainly consistent with this view.

IV. Summary and Conclusions

A stable, interpretable individual differences scaling solution was found for subjects' similarity judgments regarding a set of American English vowels. This solution had three dimensions which corresponded, respectively, to the linguistic features of advancement, height, and tenseness. Those correspondences provide particularly strong evidence for the perceptual significance of the features due to the determinacy of individual differences scaling.

While the results regarding the advancement and height features confirm expectations based on a number of previous scaling studies, recovery of a tenseness dimension is more surprising. One reason for its recovery in the present instance may have to do with the individual differences scaling method itself. Across subjects, there was wide variability in the perceptual salience of tenseness, particularly among those who rated isolated vowels (Rakerd, 1984). With individual differences scaling, this variability was manifest in the different weighting that each subject attached to D3. However, had the data been averaged over subjects prior [illegible] analysis, as required by many scaling methods, it is likely that the [illegible]ity would have made it impossible to recover a tenseness dimension. [illegible] also be relevant that we instructed subjects to attend to those aspects of the vowel sounds that seemed to them to distinguish words in English. Previous investigators (Carlson & Granström, 1979; Klatt, 1979) have reported that an instruction of this type can strengthen the linguistic character of subjects' perceptual judgments.

There were two noteworthy findings regarding correlations between the scaling results and acoustic parameters of the vowel stimuli. The first was that vowel duration was not significantly correlated with the tenseness dimension in /dVd/ context. Hence, the emergence of this dimension, particularly in the separate analysis of the consonantal-context condition, cannot be attributed to subjects having attended to durational differences among the vowels.

The second observation was that in /dVd/ context tenseness was significantly correlated with offglide proportion. Tense vowels had an internal syllable structure in which the offglide constituted 50% or less of the vocalic region. For lax vowels, the offglide made up 60% or more of the vocalic region. This finding is similar to one reported by Lehiste and Peterson

Singh, S. (1976). Distinctive features: Theory and validation. Baltimore: University Park Press.

Singh, S., & Woods, D. R. (1970). Perceptual structure of 12 American English vowels. Journal of the Acoustical Society of America, 49, 1861-1866.

Strange, W., Edman, T. R., & Jenkins, J. J. (1979). Acoustic and phonological factors in vowel identification. Journal of Experimental Psychology: Human Perception and Performance, 5, 643-656.

Strange, W., Jenkins, J. J., & Johnson, T. (1983). Dynamic specification of coarticulated vowels. Journal of the Acoustical Society of America, 74, 695-705.

Strange, W., Verbrugge, R., Shankweiler, D. P., & Edman, T. R. (1976). Consonantal environment specifies vowel identity. Journal of the Acoustical Society of America, 60, 213-224.

Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. Psychometrika, 42, 7-67.

Terbeek, D. (1977). A cross-language multidimensional scaling study of vowel perception. UCLA Working Papers in Phonetics, 37, 1-271.

Terbeek, D., & Harshman, D. (1971). Cross-language differences in the perception of natural vowel sounds. UCLA Working Papers in Phonetics, 19, 26-38.

Umeda, N. (1975). Vowel duration in American English. Journal of the Acoustical Society of America, 58, 434-445.

Wish, M., & Carroll, J. D. (1974). Applications of individual differences scaling to studies of human perception and judgment. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception (Vol. 2, pp. 449-491). New York: Academic Press.

Footnotes

[1] Although this is most commonly the case, and was the case in the present study, each of the several data matrices submitted to an individual differences scaling analysis need not represent the performance of a single subject. As alternatives, there could, for example, be one matrix for each of the several conditions of an experiment, or one for each of the several experiments in a study. From a computational standpoint, it is only required that there be multiple matrices.

[2] The subjects had complete control over the ordering and pacing of stimulus presentation. They directed presentation of the triad of stimuli for each trial by pressing three different buttons on a computer terminal.

[3] The term advancement is used to be consistent with the earlier work of Singh and Woods (1970), and with Rakerd (1984). An alternative, and perhaps more common term for this feature would be backness (Ladefoged, 1975).

Singh, S. (1976). Distinctive features: Theory and validation. Baltimore: University Park Press.

Singh, S., & Woods, D. R. (1970). Perceptual structure of 12 American English vowels. Journal of the Acoustical Society of America, 49, 1861-1866.

Strange, W., Edman, T. R., & Jenkins, J. J. (1979). Acoustic and phonological factors in vowel identification. Journal of Experimental Psychology: Human Perception and Performance, 5, 643-656.

Strange, W., Jenkins, J. J., & Johnson, T. (1983). Dynamic specification of coarticulated vowels. Journal of the Acoustical Society of America, 74, 695-705.

Strange, W., Verbrugge, R., Shankweiler, D. P., & Edman, T. R. (1976). Consonantal environment specifies vowel identity. Journal of the Acoustical Society of America, 60, 213-224.

Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. Psychometrika, 42, 7-67.

Terbeek, D. (1977). A cross-language multidimensional scaling study of vowel perception. UCLA Working Papers in Phonetics, 37, 1-271.

Terbeek, D., & Harshman, D. (1971). Cross-language differences in the perception of natural vowel sounds. UCLA Working Papers in Phonetics, 19, 26-38.

Umeda, N. (1975). Vowel duration in American English. Journal of the Acoustical Society of America, 58, 434-445.

Wish, M., & Carroll, J. D. (1974). Applications of individual differences scaling to studies of human perception and judgment. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception (Vol. 2, pp. 449-491). New York: Academic Press.

Footnotes

[1] Although this is most commonly the case, and was the case in the present study, each of the several data matrices submitted to an individual differences scaling analysis need not represent the performance of a single subject. As alternatives, there could, for example, be one matrix for each of the several conditions of an experiment, or one for each of the several experiments in a study. From a computational standpoint, it is only required that there be multiple matrices.

[2] The subjects had complete control over the ordering and pacing of stimulus presentation. They directed presentation of the triad of stimuli for each trial by pressing three different buttons on a computer terminal.

[3] The term advancement is used to be consistent with the earlier work of Singh and Woods (1970), and with Rakerd (1984). An alternative, and perhaps more common term for this feature would be backness (Ladefoged, 1975).

PERCEPTUAL COHERENCE OF SPEECH: STABILITY OF SILENCE-CUED STOP CONSONANTS*

Bruno H. Repp

Abstract. A series of experiments was conducted to examine the perceptual stability of stop consonants cued by silence alone, as when [s]+silence+[læt] is perceived as "splat." Following a replication of this perceptual integration phenomenon (Exp. 1), attempts were made to block it by instructing subjects to disregard the initial [s] and to focus instead on the onset of the following signal, which was varied from [plæt] to [læt]. However, these instructions had little effect at short silence durations (Exp. 2), and they reduced stop percepts for only two subjects at longer silence durations (Exp. 3). That is, subjects were generally unable to dissociate the [s] noise from the following signal voluntarily and thus to perceive the silent interval as silence rather than as a carrier of phonetic information. A low-uncertainty paradigm facilitated the task somewhat (Exp. 4). However, when the [s] frication was replaced with broadband noise (Exp. 5), listeners had no trouble at all in the selective-attention task, except at very short silence durations (< 40 ms). This last finding suggests that, except for the shortest durations, the effect of silence on phonetic perception does not arise at the level of psychoacoustic stimulus interactions. Rather, the results support the hypothesis that perceptual integration of speech components, including silence, is a largely obligatory perceptual function driven by the listener's tacit knowledge of phonetic regularities.

When listening to speech we perceive a coherent stream of sound, not a sequence of clicks, whistles, buzzes, and hisses. In view of the many abrupt changes of excitation and spectral structure that take place in normal speech, this apparent auditory coherence might seem like a remarkable perceptual accomplishment. However, it may well reflect the fact that the ordinary listener's attention is not focused on the detailed physical properties of the speech signal but on the underlying, linguistically relevant information. That is, auditory coherence of speech may be inferred from the perceived actual continuity of certain underlying articulatory events. If so, then there may be a more analytic level of perception that is sensitive to physical discontinuities in the speech signal.

Speech does possess certain acoustic features that promote auditory coherence of otherwise disparate signal portions. For example, formant transitions have been considered to provide a kind of "perceptual glue" that

---

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

Acknowledgment. This research was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. I am grateful to Richard E. Pastore for extensive critical comments that helped improve the manuscript.

holds successive sounds together and helps preserve their temporal order (Cole & Scott, 1973; Dorman, Cutting, & Raphael, 1975). This can hardly be the whole story, however. If perceptual coherence and integration were determined entirely by properties of the acoustic signal and their auditory transforms, it would be impossible for a listener to decompose the speech signal into its components deliberately. Nevertheless, this is possible, at least to a certain extent, by focusing one's attention on the level of auditory qualities (see e.g., Pilch, 1979). For example, it is not difficult even for a naive listener to attend selectively to the series of high-pitched hisses that represent repeated occurrences of [s] in the speech stream. Under special conditions, the perceptual isolation of such auditory components may be facilitated: Cole and Scott (1973) rapidly repeated the syllable [sa] over and over, and listeners soon reported hearing two separate streams of sounds, one consisting of hisses (the fricative noises) and the other of syllables sounding like [ta] (the vowel with its initial formant transitions). In this unnatural situation, the segregation may take place at a relatively early perceptual stage; similar "streaming" can be induced in repetitive multicomponent nonspeech signals (Bregman, 1978).

Under more natural circumstances, the perceptual integration of certain disparate acoustic components of speech may still not be completely obligatory, though it reflects the normal mode of speech perception. If perceptual integration of these speech components could be disengaged by manipulating listeners' interpretation of the stimulus, this would suggest that the normally perceived coherence of the speech signal is contingent on a nonobligatory, central function characteristic of phonetic perception. If the integrative function proved difficult to disengage, and if low-level psychoacoustic interactions can be ruled out as the cause of the integration, then the conclusion would be that perceptual integration of speech components is not only a characteristic but also an obligatory function of phonetic perception.[1]

Evidence in favor of the hypothesis that certain types of perceptual integration are speech-specific has been obtained in several recent studies concerned with "trading relations" among acoustic cues. Thus, Best, Morrongiello, and Robson ([illegible]) have shown that, in noise-plus-sinewave analogs of utterances of the type "say" versus "stay," the silent closure interval following the noise and the onset frequency of the tone mimicking the first formant (F1) both contribute to a stop consonant percept as long as the stimuli are perceived as speech; however, when the stimuli are perceived as nonspeech, the two acoustic cues are no longer integrated and are perceived as unrelated auditory properties. In another study, Repp (1981) trained subjects to discriminate the pitch of fricative noises preceding different vowels containing one of two sets of formant transitions. There was no effect of the vocalic context on the subjects' pitch judgments, even though the phonetic identification of the fricative consonant was influenced by both vowel quality and formant transitions. Furthermore, Dorman, Raphael, and Liberman (1979) and Rakerd, Dechovitz, and Verbrugge (1982) experimented with utterances whose precise phonetic interpretation depended on the duration of a silent closure interval occurring at a syllable boundary. When either fundamental frequency (Dorman et al., 1979) or the intonation contour (Rakerd et al., 1982) was changed abruptly across syllables, the silence lost its perceptual effect. Although spectral discontinuity could have played a role here, circumstantial evidence suggests that subjects' perception of one versus two speakers or utterances was responsible for the effect. Thus, all the studies cited provide evidence for a central level of perceptual integration that can be disengaged in at least three ways: by leaving the speech mode altogether, by selectively attending to specific auditory properties of the speech signal, or by perceiving a change of source or of linguistic structure.

In the present research, the focus is on the perceptual integration occurring in [spl] clusters. Acoustic cues to the perception of a labial stop consonant in this context include, first and foremost, an interval of silence following the [s] noise (Bastian, Eimas, & Liberman, 1961; Fitch, Halwes, Erickson & Liberman, 1980), but also spectral changes in the fricative noise and the amplitude contour at noise offset (Summerfield, Bailey, Seton, & Dorman, 1981), the duration of the [s] noise (Repp, 1984c), the presence and amplitude of a release burst following the silent closure (Repp, 1984b, 1984d), formant onset frequencies and transitions in the following voiced portion (Fitch et al., 1980; see also Bailey & Summerfield, 1980), and the duration and possibly the amplitude envelope of the voiced portion (Repp, 1984c). Of special interest here is the finding (Dorman et al., 1979) that a percept of "split" can be elicited by simply concatenating an [s] noise and a [lIt] syllable, with an appropriate interval of silence (about 100-300 ms) in between; in other words, in this context silence alone can be a sufficient cue for the perception of a "p," as long as there are no contradictory cues from the surrounding signal portions. Since neither of the energy-carrying signal portions in isolation contains sufficient cues to a "p," and the silence by itself naturally does not either, the stop consonant percept in this case is a pure product of perceptual integration over time and thus constitutes an ideal test case for our purposes.

The question addressed in the present study is: How robust is this perceptual integration effect--that is, can a listener deliberately avoid the stop consonant percept and hear the stimulus components the way they sound in isolation, for example, as "s" followed by "lit"? This question is not unreasonable because a stop cued by silence alone does not sound perfectly natural and might be expected to be perceptually unstable, almost an illusion. The answer to the question also bears on two contrasting hypotheses that have been put forward to account for perceptual integration and cue trading relations in phonetic perception (see Pastore, 1981; Repp, 1982): If these phenomena are a function of purely psychoacoustic stimulus properties that emerge in peripheral auditory processing, then it should be extremely difficult to disengage them through acts of selective attention or linguistic restructuring. If they are a function of speech-specific mechanisms, however, it might be possible to change them by manipulating listeners' interpretation of the stimulus, without necessarily leaving the speech mode. A positive result would simultaneously refute the psychoacoustic hypothesis and support the existence of a special integrative level of perception, whereas a negative result, to be interpretable, would require an additional demonstration that psychoacoustic interactions are not the cause of the subjects' difficulty.

Accordingly, this paper reports several attempts to "get rid of the stop" in subjects' perception of [s]+silence+[lIt] = "split" type utterances by directing their attention to the stimulus portion following the silence. A replication of the basic phenomenon of silence-cued stop consonant perception (Exp. 1) is followed by experiments that investigate the effect of selective attention instructions for stimuli with different absolute silence durations (Exps. 2 and 3), and with some subsequent changes in test format to reduce stimulus uncertainty (Exp. 4). Since, as will be seen, the stop consonant percepts proved unexpectedly resistant to these manipulations, the last experiment (Exp. 5) aimed at ruling out psychoacoustic interactions as the cause of the silence-cued stop percept. On the assumption that this last study succeeded in its aim, the conclusion will be that perceptual integration of speech components, in this instance at least, is a relatively compulsory function of phonetic perception.

Experiment 1

Experiment 1 was an attempt to replicate an earlier striking demonstration of the perceptual integration phenomenon of interest, owing to Dorman et al. (1979, Exp. 3). These authors concatenated natural [s] and [lIt] utterances that had been recorded in isolation and that were considered to contain no traces of any [p]. When the silent interval between the stimulus components was shorter than 60 ms, listeners uniformly reported "slit." At silent intervals between 80 and 450 ms, however, listeners reported predominantly "split," with a maximum of over 90 percent around 300 ms of silence. This optimal closure interval was much longer than a typical [p] closure in this context (about 90 ms; see Morse, Eilers, & Gavin, 1982); moreover, it took as much as 650 ms of silence before subjects uniformly reported hearing "s-lit" (i.e., "s" followed by "lit"), rather than "split." Since the "p" percepts in such stimuli are sometimes not very convincing, a replication of the Dorman et al. study seemed advisable, to verify that their subjects' "p" percepts were not just phantoms.

The long optimal closure duration (300 ms) in the Dorman et al. experiment may have been due to perceptual compensation for the absence of other cues to stop manner. However, there is also the possibility that the use of a wide range of closure durations (0-650 ms), combined with a higher relative frequency of short intervals, promoted a bias toward reporting "split" at atypically long closure durations. Therefore, two different stimulus ranges were employed here to assess the effect of this variable on the "sl"-"spl" and "spl"-"s-l" boundaries. The stimuli in this part of the experiment (1a) began with a fricative noise that contained some positive stop manner cues and that was also used in Experiments 2-4. To approximate the conditions of the Dorman et al. (1979) study even more closely, the test employing a wide range of closure durations was later repeated (1b) using a fricative noise without positive stop manner cues.

Method

Subjects. Nineteen paid volunteers served as subjects, 10 in Experiment 1a and 9 in 1b. They were Yale undergraduates and native speakers of American English.

Stimuli. A female speaker recorded several repetitions of the utterance [splæt] ("splat"). One good token was low-pass filtered (-3 dB at 9.6 kHz, -55 dB at 10 kHz) and digitized at a 20 kHz sampling rate. Because this speaker's fricative noises contained significant energy at frequencies above 10 kHz, which caused some digitization artifacts, digitization and subsequent recording of audio tapes were done at half speed. The [s] noise was 125 ms long. The silent closure interval and the initial 11.5 ms of the following stimulus portion, corresponding to the labial release burst (and perhaps including a weak first glottal pulse), were removed. The remaining portion in isolation elicited over 90 percent "lat" responses (see Exps. 2 and 3, pretest). Thus it did not seem to contain any sufficient cues to a preceding labial stop. The fricative noise from [splæt], however, may have contained such cues. Therefore, Experiment 1b used a fricative noise derived from an utterance of [slæt] produced by the same speaker, 190 ms in duration.[2]

Two identification tests were assembled for Experiment 1a. In one, the [s] noise was followed by the [læt] portion at each of 14 different closure durations: 0, 20, 40, 60, 80, 100, 150, 200, 250, 300, 400, 500, 600, and 700 ms. This test was also duplicated in Experiment 1b with the different [s] noise. In the other test used in Experiment 1a, only the 9 closure durations up to 250 ms were included. Each test contained 10 successive randomizations of the stimuli, with interstimulus intervals (ISIs) of 2.5 s and interblock intervals of 6 s. The stimulus sequences were recorded at half speed on audio tape using high-quality equipment, with closure durations and ISIs at twice their nominal values; thus they had the intended values at playback speed.

Procedure. The subjects listened individually or in small groups over TDH-39 earphones in a quiet room. They identified each stimulus in writing as beginning with "sl," "spl," or "s-l" (i.e., "s" followed by silence and "lat").

Results and Discussion

The average percentage of stop (i.e., "spl") responses is plotted in Figure 1 as a function of closure duration (on a logarithmic scale). Filled and open circles represent the data from the two conditions of Experiment 1a. It is evident that stimuli with short closure intervals were perceived as beginning with "sl." The "sl"-"spl" boundary fell at about 70 ms of closure duration. "Spl" responses were obtained for closure intervals ranging from 60-300 ms, with the peak occurring at 100-150 ms of silence. At longer closure durations, an increasing number of "s-l" responses was obtained.[3] Truncation of the stimulus range did not affect the "sl"-"spl" boundary but shortened the "spl"-"s-l" boundary by about 80 ms. At closure intervals of 200 and 250 ms combined, there were significantly fewer "spl" responses in the narrow-range than in the wide-range condition (one-way repeated-measures ANOVA: $F(1,9) = 26.25$, $p = .0006$). The "spl"-"s-l" distinction is not very categorical and was expected to be affected by stimulus range. The fixed "sl"-"spl" boundary, on the other hand, suggests that the silence-cued "p" percepts at closure durations below 150 ms were relatively stable and insensitive to range effects.

The results from Experiment 1b are represented by the triangles in Figure 1. They confirm that the fricative noise in Experiment 1a contained some positive stop manner cues. The "sl"-"spl" boundary was at a longer silent interval here (close to 100 ms), the maximum of "spl" responses was less pronounced and occurred at longer silences (150-250 ms), and the subjects experienced more uncertainty at the longest intervals, giving more "spl" responses here than in Experiment 1a. All these differences are at least in part due to the longer duration of the fricative noise used in Experiment 1b (cf. Repp, 1984c), but spectral differences at noise offset may also have played a role.

The general pattern of these results is consistent with the findings of Dorman et al. (1979). That is, even without any strong stop manner cues in the surrounding signal portions, "p" percepts are obtained in a certain range of closure durations. The 70 ms boundary separating "sl" from "spl" responses

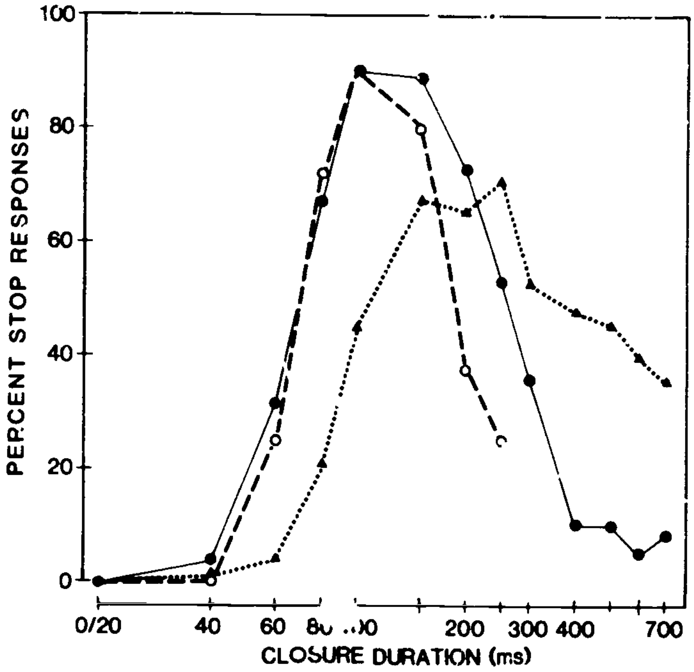


Figure 1. Percent stop (i.e., "spl") responses as a function of closure duration in Experiments 1a (filled and open circles) and 1b (triangles). The open circles represent the results from the condition with a reduced range of closure durations.

in Experiment 1a is very close to that obtained by Dorman et al. The results of Experiment 1b resemble the Dorman et al. findings in terms of the optimal closure duration for hearing "p"; they suggest that listeners need exceptionally long closure intervals for stop perception when closure duration is the sole stop manner cue, perhaps to compensate for the absence of other cues. The optimal closure duration in Experiment 1a, however, is shorter than in than in the Dorman et al. study, and so is the longest closure at which "p" percepts were still obtained. These results are somewhat closer to reflecting the typical closure durations observed in natural speech.

Experiment 2

Even though Experiment 1 demonstrated the perceptual reality of silence-cued stop consonants, it did not tell us how obligatory these percepts are. The fact that the percentage of "spl" responses did not reach 100 percent at any closure duration suggests a certain amount of ambiguity. Subjects may also have felt compelled to apply the "spl" response category supplied by the experimenter. How easy would it be to convince listeners that what they are hearing is really "s" followed by "lat," and not "splat"? The technique adopted to investigate this issue in the following experiments was to construct a continuum from [plæt] to [læt], to prefix it with an [s] noise plus a varying silent interval, and to instruct listeners either to identify the whole stimulus ("integrative" condition) or to ignore the [s] and identify only the part following the silence ("analytic" or selective-attention condition). Since the test included clear [splæt] (i.e., [s]+silence+[plæt]) stimuli, there was no pressure to give any stop responses to [s]+silence+[læt] stimuli. On the contrary, contrast among stimuli in the test should reduce any such tendencies. The analytic instructions were reinforced by the use of the response "b" (actually, "bl") for the syllable-initial labial stop, if one was perceived, as contrasted with "p" (actually, "spl") in the integrative condition.[4] Note that the analytic instructions required a perceptual reinterpretation within the linguistic domain, without leaving the speech mode (although thinking of the [s] as some extraneous noise might help). If the instructions were effective, fewer stop responses should be obtained in the analytic than in the integrative condition at closure durations beyond 100 ms, particularly for those stimuli whose final portion was perceived as beginning with "l" in isolation.

The "stop generation effect" discussed so far--the introduction of a stop percept by appropriate amounts of silence in the absence of any other sufficient cues--may be contrasted with a "stop suppression effect" due to an absence of a sufficient interval of silence in the presence of other sufficient cues. Thus, earlier observations (e.g., Fitch et al., 1980; Mann & Repp, 1980) lead to the expectation that stimuli perceived as beginning with "bl" in isolation will lead to "sl" responses when preceded by an [s] noise with little or no silence in between. If this stop suppression effect reflected the same higher-level, integrative mechanisms as the stop generation effect, and if analytic listening instructions were effective, then more stop responses should be obtained in the analytic than in the integrative condition at short closure durations, particularly for those stimuli whose final portion was perceived as beginning with "bl" in isolation.

Thus, the strongest prediction for Experiment 2 is that silent closure duration will have a marked effect on stop perception in the integrative listening condition but no effect at all in the analytic condition: Stimuli should be labeled as if there were no preceding [s]. However, apart from the fact that it is more realistic to expect only a more or less pronounced tendency in the predicted direction, the stop generation and suppression effects may well be differentially sensitive to attentional strategies. The stop suppression effect, which results from signal components occurring in close succession, is much more likely to involve auditory interactions (such as forward masking) than the stop generation effect, which results from components that are more widely separated in time. If this notion is correct, then the prediction should be that selective attention instructions, if effective, will lead to a reduction of stop percepts at longer silences but not to an increase of stop percepts at short silences.

Method

Subjects. The same 10 subjects as in Experiment 1a participated.

Stimuli. A continuum from [plæt] to [læt] was constructed from the source utterance used in Experiment 1a, [splæt]. The original 11.5 ms labial release burst was truncated by 0, 2, 4, 7.5, or 11.5 ms, yielding five stimuli intended to range perceptually from "blat" to "lat" in the absence of a preceding [s].[5] The cutpoints were placed at zero-crossings in the digitized waveform. A brief pretest was assembled in which these five stimuli (without any preceding [s]) occurred 10 times in random sequence, with ISIs of 2.5 s.

Two additional identification tests were assembled. In one, designed for integrative listening, each stimulus from the [plæt]-[læt] continuum was preceded by [s] at silent intervals of 0, 40, 80, 120, and 160 ms, for a total of 25 stimuli that were recorded 10 times in random sequence with ISIs of 2.5 s. The other test, designed for analytic listening, contained 10 random sequences of the same 25 stimuli plus 10 x 2 replications of the 5 stimuli without a preceding [s] interspersed among them, resulting in 10 35-item blocks. The "no-[s]" stimuli were intended to remind the subjects of the stimulus portion to attend to, and perhaps to facilitate selective attention.

Procedure. All subjects listened first to the tapes of Experiment 1a. Subsequently, in the same session, the integrative listening test was presented. As in Experiment 1a, the task was to label the stimuli as beginning with "sl" or "spl." The pretest followed, with instructions to label the stimuli as beginning with "bl" or "l." Finally, the analytic listening test was presented, in which the labels "bl" and "l" were again to be used. Subjects were told to ignore the [s], if present, to the best of their ability. They were informed about the structure of the stimuli and about the perceptual effect to be avoided.

Results and Discussion

The [plæt]-[læt] continuum was perceived as intended. In the pretest, the average percentages of "bl" responses to the 5 stimuli were 100, 100, 90, 9, and 3, respectively. (Note the listeners' remarkable sensitivity to the 3.5 ms release burst cutback occurring between stimuli 3 and 4; for comparable results, see Repp, 1984b: Exp. 1.) The same no-[s] stimuli interspersed in the analytic listening test received 99, 99, 92, 24, and 20 percent "bl" responses, respectively. Thus, stimuli 4 and 5 were sometimes perceived as beginning with "bl" in this environment, but they still were clearly distinguished from stimuli 1, 2, and 3, which sufficed for the purposes of this experiment.

In both the integrative and analytic listening conditions, stimuli with no closure silence at all never elicited labial stop responses. Clearly, analytic listening instructions were totally ineffective here--not an unexpected result. Therefore, those data were excluded from further analysis, reducing the number of closure durations to 4. Figure 2 shows the percentages of labial stop responses in the two listening conditions as a function of closure duration and of stimulus number on the continuum. The responses to no-[s] stimuli in the analytic test are plotted on the far right.

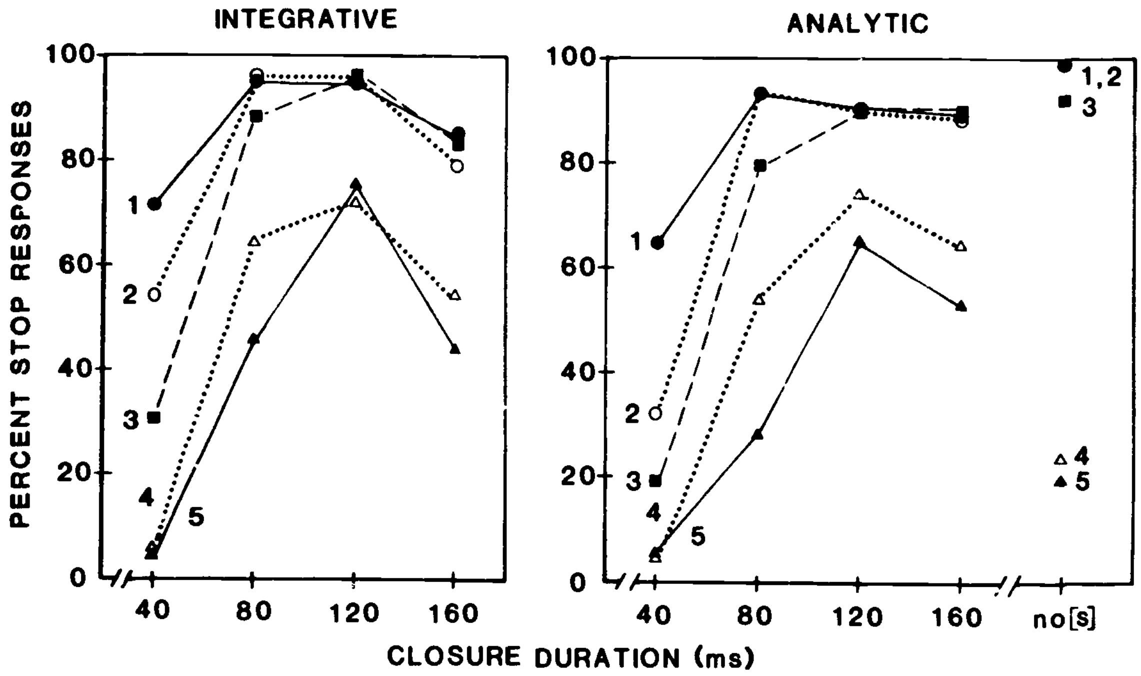


Figure 2. Percent stop responses in the integrative and analytic conditions of Experiment 2, separately for the five stimuli from the [plæt]-[læt] continuum. Data for the 0 ms closure duration are omitted.

It is evident that the response patterns in the integrative and analytic conditions were highly similar. A repeated-measures ANOVA showed the expected significant main effects of closure duration and stimulus continuum, and also an interaction between these factors (all p's < .0001), but no significant main effect of conditions. The conditions by closure duration interaction was significant, $F(3,27) = 5.45$, $p < .005$, due to a slight reduction in labial stop percepts at the shorter closure durations in the analytic condition relative to the integrative condition, and a relative increase at the longest closure duration, where perceptual segregation of the [s] noise from the rest of the stimulus might have been expected to be relatively easier. This pattern of results is the opposite of the predicted one. Thus there is no evidence that the analytic listening instructions had the desired effect. Instead of selectively attending to the stimulus portion following the silence, the subjects apparently responded by parsing off the "s" and changing the "p" to "b" in their phonological (or orthographic) representation of the whole stimulus.

The peak rate of labial stop responses to stimuli 4 and 5 preceded by [s] (about 70 percent at 120 ms of silence in both conditions) clearly exceeded that for stimuli 4 and 5 in isolation, but was lower than that in Experiment 1a (about 90 percent). This may suggest unstable "p" percepts, but the results of the analytic condition do not bear this out. That is, the instability was only in the choice of response from one trial to the next, not in the percept on which it was based.

It is interesting to note that stimuli 1, 2, and 3, which tended to give very similar results at longer closures and in isolation (probably due to a ceiling effect), elicited different response rates at the 40 ms closure duration. In fact, an orderly trading relation can be seen between stimulus number (i.e., degree of release burst truncation) and silent closure duration, as previously demonstrated by Repp (1984b, Exp. 1) for alveolar stops in the "say"-"stay" contrast. The "sl"-"spl" boundary (50 percent intercept) ranged from approximately 30 ms (stimulus 1, extrapolated) to over 90 ms of silence (stimulus 5)--a remarkable range, considering that the release burst being truncated was only 11.5 ms long. A lot of silence was needed to compensate for the loss of a small piece of plosive noise.

Experiment 3

Experiment 2 suggests that, at least without special training, subjects are unable to dissociate an [s] noise perceptually from the following speech signal. In part, this may have been due to the relatively short silent intervals used. Experiment 3 examined the same issue at longer closure durations, where selective attention to the stimulus portion following the [s] might be facilitated by the increased temporal separation and the consequent reduction of any potential auditory stimulus interactions across the silence. Experiment 3 used only an analytic listening condition, taking the integrative identification data of Experiment 1a for comparison. Since the closure intervals used were all in the range beyond the stop suppression effect, the expectation was that stop responses would be reduced relative to Experiment 1a and would approximate the percentages for no-[s] stimuli.

Method

Subjects. Ten paid volunteers participated, four of whom had taken part in Experiments 1a and 2.

Stimuli. The test sequence contained the five stimuli from the [plæt]-[læt] continuum preceded by the [s] noise at silent intervals of 100, 150, 200, 250, 300, 400, and 500 ms. The resulting 35 stimuli were augmented by 4 repetitions of the 5 stimuli without preceding [s], and all 55 stimuli were recorded in 5 randomized orders with ISIs of 2.5 s. The pretest of Experiment 2 (no-[s] stimuli only) was also used.

Procedure. Six of the subjects first listened to the pretest, labeling each stimulus as beginning with "bl" or "l." (The four remaining subjects had received the pretest in an earlier session in connection with Experiment 2.) Following the pretest, all subjects went through Experiment 4 (described below) before embarking on Experiment 3. The instructions were to ignore the initial [s], if present, and to label each stimulus as beginning with either "bl" or "l." The subjects were informed about the purpose of the experiment and about the nature of the stimuli.

Results and Discussion

The average percentages of labial stop responses to the five stimuli in the pretest were 100, 100, 89, 16, and 10, respectively. For the same stimuli in the analytic identification test, subjects' average percentages were 99, 99, 78, 13, and 8. Unlike Experiment 2, there was no increase in "bl" responses to stimuli 4 and 5 in the environment of stimuli with initial [s], perhaps because there were no contextual stimuli that sounded like "slat."

Figure 3 plots "bl" responses to stimuli preceded by [s] as a function of silent closure duration. The response percentages for the interspersed no-[s] stimuli are plotted on the far right. Several patterns are evident in the results: (1) Stimuli 1, 2, and 3 elicited fewer stop responses when preceded by [s] than when presented in isolation. (2) At closure durations shorter than 300 ms, stimuli 4 and 5 elicited more stop responses when preceded by [s] than when presented in isolation. (3) The percentage of stop responses increased as closure duration decreased, reaching a peak at 150 ms for stimuli 3, 4, and 5. Responses to stimuli 1 and 2, on the other hand, were not sensitive to changes in closure duration. In the analysis of variance, this was reflected in a significant closure duration by stimulus number interaction, $F(24,216) = 2.09$, $p < .005$.

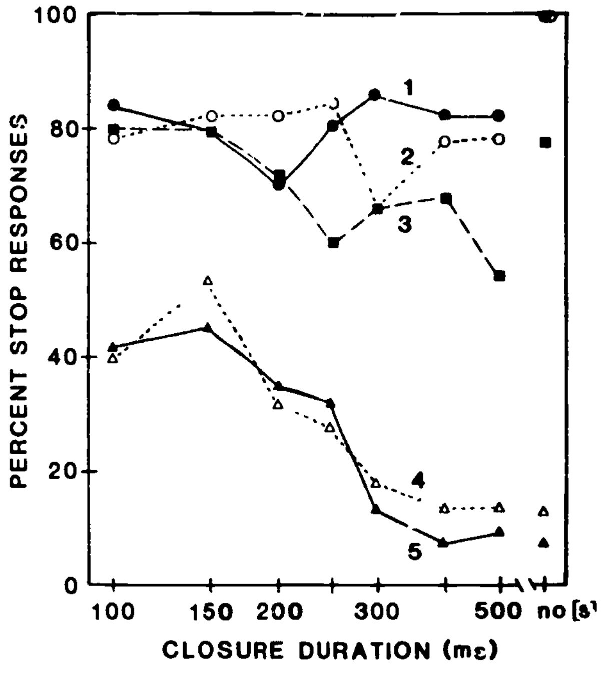


Figure 3. Percent stop responses in the analytic task that constituted Experiment 3.

The main result of this study is the increase in stop responses when [læt]-like stimuli were preceded by [s] at closure durations of less than 300 ms. This increase resembles the results of Experiment 1a, obtained with stimulus 5 in a standard (integrative) labeling task. Thus, as in Experiment 2, subjects were not able to get rid of stop percepts by ignoring the [s] precursor and focusing their attention on the onset of the stimulus portion following the closure silence. Some measure of success in the selective-attention task is indicated, perhaps, by the fact that stop responses to stimulus 5 preceded by [s] reached a maximum of only 50 percent, whereas the same stimulus elicited as much as 90 percent stop responses in Experiment 1a. However, in the integrative condition of Experiment 2, there was also a relatively low percentage of stop responses to stimulus 5 at comparable closure durations (about 60 percent). Moreover, since subjects had been told that a preceding [s] tended to generate labial stop percepts that were to be avoided, a bias against responding "bl" may have operated. This is strongly suggested by the lowered rate of "bl" responses (around 80 percent) to stimuli 1 and 2 preceded by [s], which certainly would have been labeled "spl" 100 percent of the time in an integrative task. Thus, the effect of the selective-attention instructions on perceptual organization may actually have been rather small (see discussion of Figure 5 below).

This conclusion must be qualified immediately, however, because closer inspection of the data revealed considerable individual differences (in contrast to Experiment 2). In particular, there were 2 (out of 10) subjects who appeared to be totally successful in ignoring the [s] precursor, whose labeling responses were not influenced by closure duration, and who exhibited no response bias.[6] Four or five other subjects showed patterns of which Figure 3 is representative, and the remaining subjects exhibited idiosyncratic patterns and showed large response biases against "bl." These individual differences are reminiscent of those observed by Repp (1981) in a study that required listeners to dissociate a fricative noise perceptually from a following vocalic portion. The success of two subjects in the present study suggests that analytic listening to speech components is not an impossible task, at least not when the closure durations are fairly long. These observations are consistent with the hypothesis that silence-induced stop percepts are products of a higher-level integrative process, and not of psychoacoustic interactions among stimulus components. Nevertheless, the fact remains that the perceptual strategy for performing the selective attention task was not available to most listeners, even though they had received a moderate amount of training by performing the low-uncertainty task of Experiment 4 before Experiment 3.

Experiment 4

Experiments 2 and 3 have provided only very limited evidence that subjects can perceptually dissociate the two stimulus components, even at relatively long temporal separations. In part, subjects' difficulties in carrying out the selective-attention instructions may reflect ingrained habits of integrative phonetic processing when listening to speech. At very short temporal separations, however, psychoacoustic interactions among the stimulus components may come into play, and these interactions may be truly impossible to disengage by acts of selective attention or other perceptual strategies. To investigate this issue further, Experiment 4 employed a low-uncertainty paradigm to test subjects' ability to distinguish between clear instances of [plæt] and [læt] when preceded by [s] at various fixed intervals of silence. It was expected that a reduction in stimulus uncertainty would facilitate the selective attention task.

Method

Subjects. The same 10 subjects as in Experiment 3 participated.

Stimuli. Only stimuli 1 and 5 from the [plæt]-[læt] continuum were used, as well as the [s] noise derived from the natural [splæt]. Seven stimulus sequences were recorded, each containing 20 repetitions of stimuli 1 and 5 in random order, with ISIs of 2 s. In the first sequence, there was no preceding [s] noise. In the subsequent sequences, each stimulus was preceded by [s] at a fixed silent interval. Over these six sequences, the closure interval decreased from 500 to 200, 100, 50, 20, and finally 0 ms.

Procedure. The subjects were told that, in each block of 40 stimuli, half were "blat" and half were "lat." They were asked to label each stimulus as beginning with "bl" or "l," guessing if necessary, and to ignore the [s] precursors. Note that Experiment 4 preceded Experiment 3.

Results and Discussion

Figure 4 shows the effect of [s] precursors at various closure durations, with the no-[s] stimuli on the far right. Labeling of the two stimuli without the [s] precursor was virtually perfect. Reading the graph from right to left, it can be seen that discrimination of stimuli 1 and 5 (in terms of the difference in "bl" responses) was unaffected at the 500-ms interval, then decreased but stayed fairly high up to the 50 ms separation; then it declined rapidly and reached chance at 0 ms (51.5 percent correct responses in terms of identification of stimulus 1 as "bl" and of stimulus 5 as "l"). Although the subjects had been encouraged to guess even if all stimuli sounded like "lat," few followed these instructions. The low percentage of stop responses at the shortest closure durations reflects the fact that [s] + [plæt] sounds like "slat" when there is no closure silence.[7]

Individual differences were evident in this task also. Three subjects, including the two who stood out in Experiment 3, performed almost perfectly down to 20 ms of silence, where they suddenly gave only "l" responses and thus performed at chance level. The other subjects were more error-prone at silent intervals of 50-200 ms, and one subject seemed to reverse the response categories.

To determine how subjects' performance in the low-uncertainty task of Experiment 4 compared with the performance obtained in Experiments 2 and 3, d' values for the stimulus 1 vs. stimulus 5 discrimination (treating the binary category labels as if they were "yes" and "no" responses in a signal detection task) were computed from the overall response percentages--a rough measure that, however, is adequate for an informal graphic comparison.[8] These d' values are plotted in Figure 5. The figure suggests that discrimination was more accurate in Exp. 4 than in Exp. 3, presumably due to the paradigm that reduced stimulus uncertainty and thus facilitated selective attention. It also seems, however, that at silent intervals in the range of 40-100 ms, there was no difference in accuracy between Exps. 2 and 4. (It is also clear that there was no difference between the integrative and analytic conditions in Exp. 2.) Since performance in Exps. 2 and 3 matched at intervals of 100-160 ms, there is no reason to assume that the subjects in Exp. 2 were especially accurate.

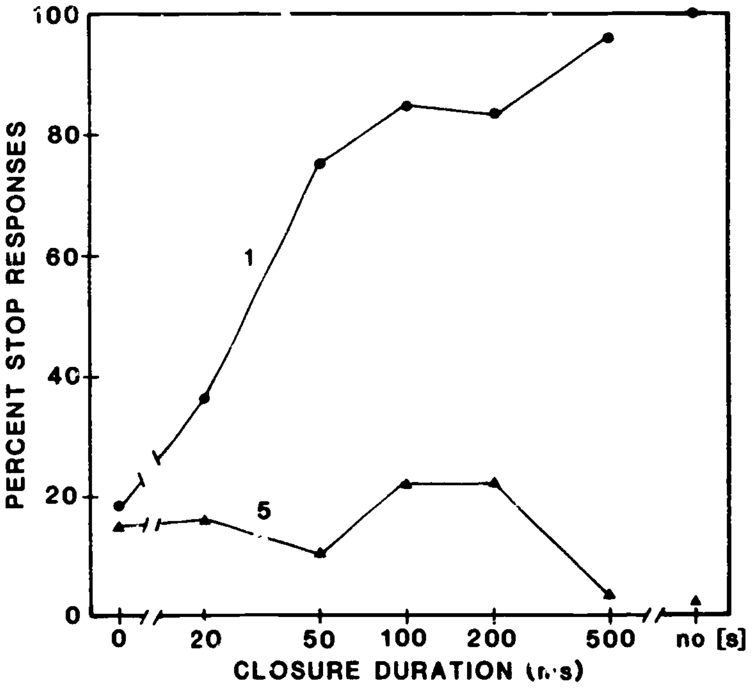


Figure 4. Percent stop responses in the low-uncertainty task that constituted Experiment 4 (stimuli 1 and 5 only).

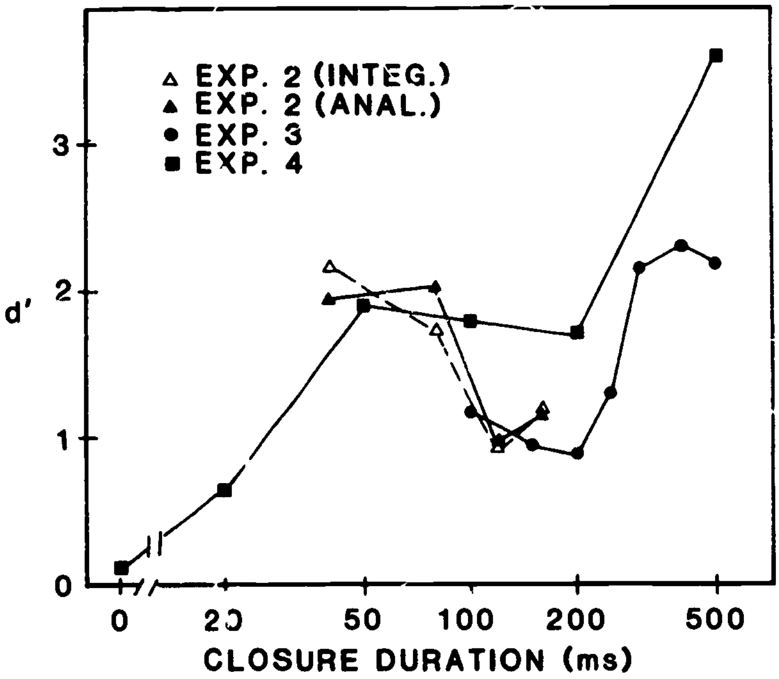


Figure 5. Discriminability of stimuli 1 and 5, expressed as d', in Experiments 2-4.

Rather, it seems that the procedure of Exp. 4, though it was beneficial at longer closure durations, conferred no advantage in the vicinity of the "sl"-"spl" category boundary (between 40-100 ms of silence; see Fig. 1).

This observation, together with subjects' extremely poor performance at very short closure durations, is compatible with the hypothesis that the stop suppression effect, and with it the "sl"-"spl" category distinction, rests on a psychoacoustic interaction that cannot be disengaged through selective attention. The silence-cued "p" percepts (the stop generation effect) at intervals beyond 100 ms, on the other hand, are sensitive, to some extent, to listeners' strategies and thus may represent a higher-level integrative process peculiar to phonetic perception. The comparisons in Figure 5 suggest, furthermore, that discriminative sensitivity is heightened in the category boundary region, whereas discrimination at silent intervals characteristic of strong "p" percepts (i.e., within-category discrimination) is less accurate and requires the overcoming of integrative phonetic processing strategies. This pattern of results is similar to that obtained in many studies of categorical perception (see Repp, 1984a).

Experiment 5

The hypothesis that the "sl"-"spl" boundary--more specifically, the suppression of a stop percept at short closure durations--has a psychoacoustic origin, although consistent with the data so far, is contradicted by a recent study of Pastore, Szczesiul, and Rosenblum (1984). These researchers employed binaural phase shifts to differentially lateralize the [s] and [plIt] components of their "slit"-"split" stimuli. This manipulation left the category boundary (located at 68 ms of closure silence in their study) completely unaffected. The authors argued that differential lateralization should reduce psychoacoustic interactions between the stimulus components and that, therefore, the absence of an effect suggests that the "sl"-"spl" boundary does not rest on a psychoacoustic criterion. However, apart from the possibility that the phase shift technique was too weak a manipulation to remove psychoacoustic interactions, these results do not rule out such interactions at closure intervals shorter than the boundary value.

An additional experiment probing the possible psychoacoustic basis of silence-cued stop consonant perception is also necessitated by the fact that Experiments 2-4 were relatively unsuccessful in disengaging subjects' integrative processing strategies. The evidence for a higher-level, speech-specific basis for the stop generation effect is suggestive at best, and a demonstration that psychoacoustic interactions are not involved would strengthen the argument considerably.

For the present stimuli, in which the difference between [plæt] and [læt] rests entirely on a brief release burst, the most obvious psychoacoustic hypothesis is that, at short temporal separations, the burst suffers from forward masking by the preceding fricative noise, and therefore becomes difficult to detect. If so, then this masking effect should occur also when a burst of white noise is substituted for the [s] frication, provided that the energy of the white noise is not substantially below that of the frication. From the viewpoint of phonetic perception, however, the white noise is less speech-like and therefore should be more easily filtered out in a selective-attention task. If the "sl"-"spl" boundary does not rest on a psychoacoustic interaction, subjects should be more successful in identifying "blat" and "lat" when white noise replaces the [s] precursor.

Method

Subjects. The same 9 subjects as in Experiment 1b participated.

Stimuli. The five stimuli from the [plæt]-[læt] continuum were again used. Instead of a natural [s] noise, however, a burst of white noise was used as a precursor. The white noise was recorded from a General Radio 1390-A random noise generator, low-pass filtered and digitized at half speed at a 20 kHz sampling rate. It differed from the [s] noise used previously (Exp. 1a and Exps. 2-4) in three respects: (1) Its duration was 200 ms, versus 125 ms for the [s] noise. (2) It was gated on and off abruptly, whereas the [s] noise had gradual on- and offsets. (3) It had a flat spectrum, whereas the spectrum of the [s] noise had a pronounced peak at about 8.6 kHz, which projected by about 20 dB above a relative energy plateau ranging from 4 to 10 kHz. The spectral energy of the white noise matched that of the plateau; its energy was higher than that of the [s] noise below 4 kHz and above 10 kHz, and lower between about 8-9 kHz. Its energy at offset was considerably higher than that of the fricative noise across the whole spectrum. All these differences led to the expectation that the white noise would have a more pronounced forward masking effect than the [s] noise, if such a psychoacoustic effect is involved at all. On the other hand, relatively long duration, abrupt offset, and flat spectrum are all uncharacteristic of natural fricative noises preceding a stop closure.[9]

The stimulus tape matched that of the analytic condition in Experiment 2. That is, silent intervals ranged from 0 to 160 ms, and "no-noise" stimuli were interspersed.

Procedure. All subjects listened first to the tape of Experiment 1b (an integrative labeling task) and then to the pretest, as used in Experiments 2 and 3 (stimuli without preceding noise). Instructions for the main test were the same as in the analytic condition of Experiment 2: Ignore the noise and label the stimuli as beginning with "bl" or "l."

Results and Discussion

Figure 6 shows the results, which are strikingly different from those of Experiment 2 (cf. Fig. 2, right-hand panel). Over the range from 40-160 ms of silence, the white noise precursor had no effect at all on subjects' ability to identify the stimuli from the [plæt]-[læt] continuum, except for introducing a slight bias against stop responses.[10] In particular, the white noise did not induce any stop percepts when it preceded stimuli 4 and 5. Only when there was no silent interval between the noise and the speech did the noise exert a perceptual effect, rendering stimuli 2-5 indiscriminable, while stimulus 1 continued to receive a higher rate of stop responses. Note also that, in this condition, subjects were equally willing to respond "bl" or "l," whereas in the corresponding condition of Experiment 2 (not shown in Fig. 2) responses were exclusively "l." This suggests that the subjects in Experiment 5 considered the white noise as an extraneous signal that might obscure stop consonant cues present in the speech signal, whereas the subjects in Experiment 2 perceived the [s] noise as part of the utterance, even when asked not to do so, and thus were unwilling to consider the possibility of an inaudible stop consonant.

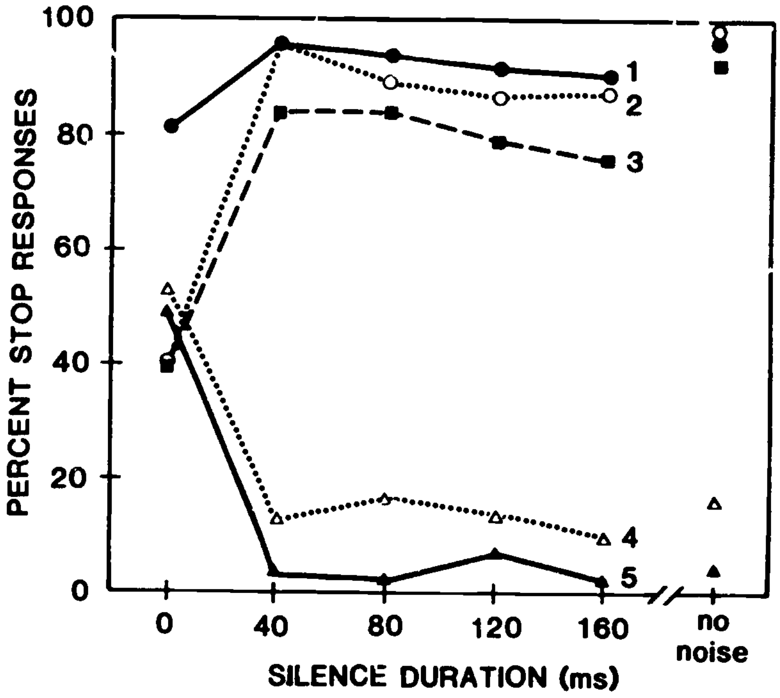


Figure 6. Percent stop responses in Experiment 5.

It seems extremely unlikely that spectral or other properties of the white noise were responsible for its reduced masking power, since it was a more powerful signal than the [s] noise by most acoustic criteria. Although the [s] noise was more intense between 8 and 9 kHz, the spectral peaks of the labial release burst were in a region (below 4.5 kHz) where the white noise exceeded the [s] noise in energy. Therefore, the results suggest that psychoacoustic interference (i.e., forward masking) was involved only at the very shortest closure intervals (less than 40 ms). Consequently, the reduction in stop responses when an [s] noise precedes [plæt] stimuli by 40-80 ms (see Fig. 2) probably does not represent psychoacoustic interference, but rather a specifically phonetic effect reflecting the listener's tacit knowledge about the minimal permissible duration of stop consonant closures in this context. Apparently, listeners are compelled to apply this knowledge as long as they perceive a coherent stream of speech. This conclusion is consistent with that reached by Pastore et al. (1984), and it suggests that the two effects of closure silence (stop suppression at short durations, stop generation at longer durations) can be accounted for within a single theoretical framework, that of perception in the "speech mode" (Liberman, 1982; Repp, 1982).

Summary and Conclusions

The present series of studies addressed the question of the origin of the auditory coherence of speech by focusing on one particularly striking phenomenon--that of silence-cued labial stop consonants in fricative-liquid context. This phenomenon illustrates both the coherence of acoustically heterogeneous speech components in general and the perceptual integration of disparate cues to the perception of a particular phonetic contrast. Between the fricative noise and the resonances resulting from production of the liquid consonant, there is an abrupt change in the nature and location of the sound source (from voiceless and dental to voiced and laryngeal) and in spectral composition (from higher to lower frequencies). Nevertheless, with or without an intervening brief silent interval, listeners usually perceive both sounds as part of a coherent speech stream. This coherence in turn gives rise to a stop consonant percept when a silent interval of appropriate duration (roughly, 80-200 ms) is present. Thus the silence itself becomes part of the speech stream; rather than interrupting the continuity and contributing to the perceptual segregation of acoustically disparate signal components, the silence functions as a carrier of phonetic information. Only when the silence duration clearly exceeds the acceptable limits of a stop consonant closure does it lead to perceptual segregation of the signal components.

It was hypothesized that the integrative function that gives rise to these phenomena is a characteristic of perception in the speech mode--that is, of perceiving the information that is most useful for linguistic communication. One way of testing this hypothesis would be to lead listeners to perceive the same stimuli as either speech or nonspeech. Some evidence favoring the hypothesis has already been obtained using variants of that method (Best et al., 1981; Repp, 1981). A somewhat different approach was taken here. It was argued that, if perceptual integration of the form studied here is a speech-specific function, it might be possible to influence its operation by directly manipulating the listeners' interpretation of the speech stimulus, staying entirely within the speech mode. The success of this approach was not guaranteed, of course, since manipulation of listeners' strategies through instructions may simply be ineffective. In the absence of a convincing psychoacoustic explanation for the perceptual integration of speech components, however, negative findings may tell us that certain perceptual strategies are not easily modified or abandoned--not that they are not speech-specific.

In a series of experiments (Exps. 2-5) following a basic demonstration of silence-cued stop consonants (Exp. 1), it was attempted to alter subjects' interpretation of the stimulus by instructing them to mentally separate the fricative noise from the following signal portion. The relative ineffectiveness of the selective-attention instructions with stimuli of seemingly minimal acoustic coherence is interpreted as evidence for the relative stability of the perceptual integration function. Experiment 3 indicated, however, that some subjects can be successful in this task, and Experiment 4 showed that a low-uncertainty paradigm also facilitates selective attention. These results parallel those obtained in studies of categorical perception (see Repp, 1984a, for a review), where subjects frequently need to disengage or ignore another basic function of the speech mode, that of phonetic classification, in order to discriminate speech stimuli. In these studies, it seems that success in within-category discrimination often requires perceptual strategies that operate outside the speech mode. The present task, too, could in principle have been accomplished by listening specifically for the release burst, though there was no evidence that the subjects used this "auditory" strategy. Rather, the few successful subjects appeared to be able to do what the instructions asked for: to ignore the fricative noise and listen to the remainder of the stimulus as speech--a skill that trained phoneticians presumably would have in their repertoire.

One way of ignoring a fricative noise is to think of it as a nonspeech hiss arising from a source outside the speaker's vocal tract. That this strategy could be effective is clear from Experiment 5 which, by substituting a nonspeech noise for the frication, actually created the situation that subjects otherwise might try to imagine. The ease with which the subjects carried out the selective-attention instructions in this situation argues against a psychoacoustic account of perceptual integration and of the effect of the silent interval on stop consonant perception. This latter effect has two aspects, which were termed "stop suppression" (short intervals) and "stop generation" (longer intervals). On the basis of the results of Experiment 5 it was concluded that both of these effects are likely reflections of speech-specific perceptual criteria, with only the suppression effect at extremely short closure silences having a psychoacoustic origin.[11]

In conclusion, then, the results of the present experiments are consistent with a theoretical view of speech perception that postulates a number of specific--though not necessarily unique--functions. These perceptual functions, which include the perceptual integration of speech components, are assumed to be driven by an internal representation of the regularities of spoken language. How this representation should be characterized and how it is acquired are fundamental questions for future research.

References

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.

Bastian, J., Eimas, P., & Liberman, A. M. (1961). Identification and discrimination of a phonemic contrast induced by silent interval. _Journal of the Acoustical Society of America_, _33_, 842. (Abstract)

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. _Perception & Psychophysics_, _29_, 191-211.

Bregman, A. S. (1978). The formation of auditory streams. In J. Requin (Ed.), _Attention and Performance VII_ (pp. 63-76). New York: Wiley.

Cole, R. A., & Scott, B. (1973). Perception of temporal order in speech: The role of vowel transitions. _Canadian Journal of Psychology_, _27_, 441-449.

Delgutte, B., & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. _Journal of the Acoustical Society of America_, _75_, 897-907.

Dorman, M. F., Cutting, J. E., & Raphael, L. (1975). Perception of temporal order in vowel sequences with and without formant transitions. _Journal of Experimental Psychology: Human Perception and Performance_, _1_, 121-129.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. (1979). Some experiments on the sound of silence in phonetic perception. _Journal of the Acoustical Society of America_, _65_, 1518-1532.

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

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

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. Perception & Psychophysics, 28, 213-228.

Morse, P. A., Eilers, R. E., & Gavin, W. J. (1982). The perception of the sound of silence in early infancy. Child Development, 53, 189-195.

Nooteboom, S. G., & Doodeman, G. J. N. (1980). Production and perception of vowel length in spoken sentences. Journal of the Acoustical Society of America, 67, 276-287.

Pastore, R. E. (1981). Possible psychoacoustic factors in speech perception. In P. D. Eimas & J. L. Miller (Eds.), Perspectives on the study of speech. Hillsdale, NJ: Erlbaum.

Pastore, R. E., Szczesiul, R., & Rosenblum, L. (1984). Does silence simply separate speech components? Journal of the Acoustical Society of America, 75, 1904-1907.

Pilch, H. (1979). Auditory phonetics. Word, 29, 148-160.

Rakerd, B., Dechovitz, D. R., & Verbrugge, R. R. (1982). An effect of sentence finality on the phonetic significance of silence. Language and Speech, 25, 267-282.

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

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

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

Repp, B. H. (1984b). The role of release bursts in the perception of [s]-stop clusters. Journal of the Acoustical Society of America, 75, 1219-1230.

Repp, B. H. (1984c). Effects of temporal stimulus properties on perception of the [sl]-[spl] distinction. Phonetica, 41, 117-124.

Repp, B. H. (1984d). Closure duration and release burst amplitude cues to stop consonant manner and place of articulation. Language and Speech, 27, 245-254.

Summerfield, Q., Bailey, P. J., Seton, J., & Dorman, M. F. (1981). Fricative envelope parameters and silent intervals in distinguishing 'slit' and 'split.' Phonetica, 38, 181-192.

Footnotes

[1] The question posed here is similar in many ways to that underlying categorical perception research (see Repp, 1984a), but the methodology is different. Categorical perception experiments examine subjects' ability to discriminate stimulus differences within phonetic categories; here, the focus is on listeners' ability to ignore one part of a stimulus (a skill that may play a role in some discrimination tasks). Both tasks are difficult because listeners tend to adhere to their habitual mode of phonetic perception, which is categorical and integrative. No claim is made here that this type of perceptual mode is specific to speech; it is called "phonetic" only because the stimuli happen to be speech. That being so, however, many specific instances of perceptual integration may indeed be speech-specific, simply because they have no parallels in other domains of experience.

[2] To be sure, the [s] noise must not be too long, and its offset and the [l] onset not too gradual; otherwise, no stop percepts will be obtained. The presence of stop manner cues in the [s] noise was irrelevant in Experiments 2-4, because subjects' attention was directed toward the stimulus portion following the silence. As far as that portion is concerned, it was sufficient that it not elicit any stop percepts in isolation. No claim is being made that either signal portion contained no cues whatsoever to stop consonant perception (see also Footnote 6).

[3] Some subjects, especially in Experiment 1a, spontaneously gave "s^{p}-l" responses, indicating that they detected stop manner cues in the frication, while at the same time perceiving a gap between the [s] and the rest of the stimulus. These responses were treated as equivalent to "s-l"; thus they are not included in the "spl" percentages plotted in Figure 1.

[4] The phonetic symbol [p] represents a voiceless unaspirated labial stop consonant, which in English orthography is rendered as "p" in some contexts (e.g., following a voiceless fricative in the same syllable) but as "b" in others. Throughout this paper, phonetic symbols in brackets denote stimuli or the speaker's intentions, whereas orthographic symbols in quotes refer to responses or the listeners' percepts.

[5] For the author and most subjects, excision of the natural labial release burst in [plæt] resulted in elimination of the stop percept. Some listeners, however, still claimed to hear a "b," which may reflect a special sensitivity to weak coarticulatory cues in the [l] portion. These coarticulatory cues may reside in spectral or amplitude properties of the signal immediately following the release burst or, perhaps more likely, in the shorter duration of the [l] as compared to one articulated in absolute utterance-initial position. One additional subject in Experiment 2 and two additional subjects in Experiment 3 were excluded because they perceived all stimuli from the [plæt]-[læt] continuum as "blat."

[6] One of these two subjects had participated in Experiments 1a and 2. In the labeling task of Experiment 1a, which used stimulus 5 of the [plæt]-[læt] continuum, she gave 90 percent stop responses at closure durations of 100 and 150 ms. In Experiment 2, for stimuli 4 and 5 with 120 and 160 ms of silence, she gave 63 percent stop responses in the integrative condition, 70 percent in the analytic condition, and 0 percent when there was no preceding [s] noise. In Experiment 3, however, she gave not a single stop response to the same stimuli with silent intervals of 100 and 150 ms. Clearly, she had discovered an effective selective attention strategy in Experiment 3, perhaps as a result of going through the task of Experiment 4 (where she likewise did not give any stop responses in the comparable stimulus conditions).

[7] It might be noted that while the inexperienced subjects performed at chance level in the 0 ms condition, the author as a pilot subject obtained a score of 85 percent correct. Thus, it is not impossible to discriminate the [plæt] and [læt] components in this condition, but a different perceptual strategy seems to be required (viz., listening for a certain difference in auditory quality caused by the presence versus absence of a release burst). Note that this strategy is nonphonetic in character, unlike the phonetic

dissociation strategy requested by the analytic listening instructions. Indeed, those few subjects who seemed to be successful analytic listeners in Experiments 3 and 4 still failed to discriminate the stimuli at the very shortest closure durations. It is likely that nonphonetic strategies would be fostered by extensive training with feedback, which is one reason why this method was not used to induce analytic phonetic strategies.

[8] The d' values were also computed for individual subjects and then averaged. The results were not substantially different from the global d' values shown in Figure 5. Although certain distortions in the global values may have occurred due to different degrees of criterion variability in different experiments, the individual subjects' values are even more distorted because of the many occurrences of response percentages of 0 and 100, which necessitate setting an arbitrary upper limit for d'. For this reason, the d' values computed from the average response percentages were preferred for this informal comparison among experiments.

[9] Of course, the white noise did not sound like a fricative noise (at best, it sounded remotely [f]-like). For this reason, an integrative listening condition, in which subjects try to interpret the noise as a fricative, was not considered. The point here is that, if psychoacoustic interactions are involved, they should not depend on the speechlikeness of the noise.

[10] This tendency, as well as its apparent increase with closure duration, was due to two subjects' data only.

[11] Another possible auditory interaction that was not considered seriously here, but that may warrant some further investigation, is auditory short-term adaptation (see Delgutte & Kiang, 1984). The [s] precursor should adapt high-frequency neurons more than low-frequency neurons, so that the auditory response to the following signal portion would be more vigorous in the low-frequency regions, which might favor labial stop percepts. There are several problems with that hypothesis, however: (a) The long temporal range of the stop generation effect (Exp. 1) exceeds the range of auditory adaptation. (b) The stop suppression effect remains unexplained. (c) The ability of some subjects to disengage the stop generation effect argues against peripheral auditory factors. (d) The [s] noise spectrum is not differentiated enough in the low-frequency region to substantially alter the shape of the "auditory spectrum" at the onset of the following signal. (e) The stop generation effect is reduced by an increase in fricative noise duration (Exp. 1b; Repp, 1984c).

DEVELOPMENT OF THE SPEECH PERCEPTUOMOTOR SYSTEM*

Michael Studdert-Kennedy†

Introduction

The intent of the present paper is to reflect on the development of the speech perceptuomotor system in light of the infant's evident capacity for intermodal (or, better, amodal) perception, discussed by Meltzoff and by Kuhl in this volume. The central issue is imitation. How does a child (or, for that matter, an adult) transform a pattern of light or sound into a pattern of muscular controls that serves to reproduce a structure functionally equivalent to the model? The hypothesis to be outlined is that imitation is a specialized mode of action, in which the structure of an amodal percept directly specifies the structure of the action to be performed (cf. Meltzoff & Moore, 1983).

The General Function of Perception

Let us begin by considering briefly the function of perception from an ethological perspective (Gibson, 1966, 1979; von Uexküll, 1934). The general function of perception is to control action. Perception and action are two terms in a functional system that permits an animal to survive. To survive, an animal must constantly negotiate a physical world, moving around, over or under objects in its path, seeking food or mates, escaping from predators. The actions that an animal takes, its coordinated patterns of goal-seeking movements, are more or less precisely matched to the world it perceives; and the world it perceives is constantly modulated by the actions it takes. Thus, action and perception are mutually entailed components of a single system: each fits the other as key fits lock.

How is the fit achieved? How are the varying patterns of light, sound, temperature, pressure that determine perception transduced into the neuromuscular patterns that determine action? Can we find a single set of descriptive terms that will match all the various sensory modalities with the single modality of action? We may approach an answer to these questions by asking another: What information do light, sound and other modes of energy convey? Following Gibson (1966, 1979) we answer quite generally: Information that specifies the structures of objects and events to which action must adapt.

We may note two properties of perceived object-event structures. First, they are amodal. We perceive a desk, say, through a pattern of light struc-

---

*In B. Lindblom and R. Zetterstrom (Eds.), Early precursors of speech. Basingstoke: MacMillan, 1985.

†Also Queens College and Graduate Center, City University of New York.

Acknowledgment. Preparation of this paper was supported in part by NICHD Grant HD-01994 to Haskins Laboratories.

tured by its light-reflecting properties, or by touch through the pattern of mechanical resistance it offers to our fingers. A bat, being equipped with sonar, might perceive the desk by virtue of the desk's sound-reflecting properties. Similarly, we normally perceive a spoken word through a pattern of sound, structured by the coordinated articulations of a speaker. To the extent that these articulations reflect radiant energy within the visible spectrum, we may also perceive the word by virtue of its optical structure. The deaf-blind, using the Tadoma method, may even perceive the word by touch (Norton, Schultz, Reed, Braida, Durlach, Rabinowitz, & Chomsky, 1977). What we perceive, then, are objects and events, independent, in principle, of the sensory modalities through which we perceive them.

The second point to note about object-event structures is that their perceived qualities vary with the perceiving organism. The "same" object has different utilities for different animals, or for the same animal at different times. Objects and events differ in what von Uexküll (1934) termed their "functional tones," what Gibson (1966) termed their "affordances." The puddle that a person steps over affords a dog an opportunity to drink; the desk that offers support for a writing pad on one occasion may serve as a seat on another; a word spoken in Mandarin is merely a vocalization to someone who knows no Chinese. Thus, different animals perceive different worlds (von Uexküll's Umwelten), each structured by the animal's potential actions, just as its actions are structured by its perceived world.

The Function of Speech Perception

The Speech Percept as Amodal

The first function of speech perception is social and communicative, a pragmatic function analogous to the general function of perception discussed above. As the carrier of language, speech offers meaning, that is to say (very broadly), information conveying the structure of a social world within which an individual may act. The individual, by acting in response, whether linguistically or non-linguistically, then modulates the perceived structure of her social world.

A second function of speech perception, ontogenetically prior to the first and of more immediate interest here, is in language acquisition. While the adult may listen simply for meaning, the learning child must listen both for meaning and for information specifying a talker's articulatory gestures. This second perceptual function therefore controls action in the more limited sense of providing a model for imitation.

Before we consider imitation, let us explicate and justify the claim that speech carries information specifying a talker's articulatory gestures. Notice, first, that this is not the customary account. For example, Abercrombie (1967) characterizes one form of the information conveyed by speech as linguistic and segmental, intending by this a sequence of phonetic elements, the consonants and vowels of a phonetic transcription. This is certainly correct, at one level of description, as our ability to read and write alphabetically demonstrates. However, a transcription is so far removed from the signal that most people in the world who can speak and understand speech cannot read or write.

What, then, is the difference between the information in a spoken utterance and the information in its written counterpart? Following Carello, Turvey, Kugler, and Shaw (1984) (see also Turvey & Kugler, 1984) we may say that the difference is between information that specifies and information that indicates. The information in a spoken utterance is not arbitrary: its acoustic structure is a lawful consequence of the articulatory gestures that shaped it. In other words, its acoustic structure is specific to those gestures, so that a human listener (who knows the language spoken) has no difficulty in following the specifications and organizing her own articulations to reproduce the utterance. By contrast, the form of a written transcription is an arbitrary convention, a string of symbols that indicate to the reader what she is to do, but do not specify how she is to do it. The important point here is that indicational information cannot control action in the absence of information specific to the act to be performed. For example, a road sign indicates that we are to stop, but we can only follow the instruction if we have information specifying our velocity and our distance from the required stopping point (Turvey & Kugler, 1984). Similarly, we can only reproduce an utterance from its transcription, if we have information specifying the correspondences between the symbol string and the motor control structures that must be engaged for speaking. It is these correspondences that the illiterate has not discovered. Just how these two forms of linguistic information are related is, of course, a central issue of speech research. My concern here is merely to make the distinction. For we shall be led astray in our study of speech perception (and so of speech acquisition), if we strive to equate the linguist's description of speech as a string of symbols with the dynamic structure of the speech signal itself.

Consider, here, an early interpretation of the lip-reading studies of McGurk and MacDonald (197 These authors discovered that listeners' perceptions of a syllable presented over a loudspeaker could be changed, if they simultaneously watched a videotape of a speaker producing another syllable. For example, presented with audio [ba] and video [da], subjects typically report the latter, optically specified syllable; presented with audio [na] and video [ba], subjects typically report [ma], a combination of the two. Such observations are consistent with the notion that subjects engage two independent phonetic systems, drawing manner and voicing features from the acoustic structure, place of articulation features from the optic structure (MacDonald & McGurk, 1978). This interpretation assumes that we perceive speech by extracting phonetic features and combining them to form phonetic segments--in other words, it assumes that the speech signal carries information about a string of linguistic symbols. As already remarked, this is true at one level of description. However, this interpretation bypasses the actual event specified by the dynamic acoustic-optic structure and does not address the puzzle of its transformation into a static linguistic symbol.

Moreover, the featural interpretation breaks down in the face of other findings. For example, presented with audio [ga] and video [ba], subjects typically report a cluster [b'ga] or [g'ba]; presented with the reverse arrangement, audio [ba] and video [ga], subjects often report a sort of acoustic-optic blend, [da]. In these instances, the percept corresponds either to both inputs or to neither, so that the notion of two independent and additive phonetic systems breaks down.

While much remains to be done before we have a satisfactory account of such findings, the effect seems to arise from a process by which two continuous sources of information, acoustic and optic, are actively combined at a precategorical level where each has already lost its distinctive sensory quality (Summerfield, 1979). In other words, the McGurk effect (and, indeed, normal lipreading as practiced in aural rehabilitation) is only possible because acoustic and optic structures specify an amodal event: a coordinated pattern of articulatory action.

Imitation

A general capacity to imitate is rare among animals. The specialized capacity to imitate vocalizations is confined to a few species of birds and of marine mammals, and to man. Here we should distinguish between mimicry and repetition, or reproduction. The Indian mynah bird, for example, mimics human speech quite precisely, within the limits of its vocal apparatus (Klatt & Stefanski, 1974). However, a human speaker repeats the utterances of another (when not deliberately attempting mimicry) by producing a functionally equivalent, though acoustically distinct, pattern of sound. Given within-species individual differences in size and structure, we may reasonably suppose that the production of distinct, yet functionally equivalent, acts is the normal mode of animal imitation, whether in human speech or in, say, the nest-building of a young chimpanzee. In any event, both mimicry and reproduction call on a specialized capacity for finding in the perceptual array an organized pattern of information specific to an organized pattern of action. To find a pattern the imitator must find both the <u>pieces</u> of an act and their spatio-temporal <u>relations</u> (Fentress, 1984).

Consider, for example, the following transcription of ten attempts by a 15-month old girl to say <u>pen</u>, within a single half-hour's recording session: [ma^{ə},ˇʌ,dɛ^{dn}.,hɪn,^{m}bõ, p^{h}ɪn, t^{h}nt^{h}nt^{h}n,ba^{h},d^{h}au^{N},buã] (Ferguson & Farwell, 1975). Note once again that the transcriptions are merely convenient (and approximate) indicators of what the child did. For what the child evidently did, in each case, was to extract from the sound pattern of <u>pen</u> information specific to certain articulatory gestures, such as lip closure, lingua-alveolar closure, velum lowering, glottal narrowing and spreading. Thus the child analyzed the word (with varying success) into its component gestures, or pieces, but could not discover, at least motorically, their spatio-temporal relations. Perhaps we have here an instance of the necessary sequence in learning to speak, or indeed in learning to reproduce any act performed by another: first perceptual analysis, then motor synthesis. We can hardly doubt that a capacity to perceive the pieces of an act and their relations, and to reproduce them in our own behavior, rests on some form of structural (anatomical, physiological) correspondence between imitator and model. This observation leads us to a brief digression.

Can Non-human Animals Perceive Speech?

The answer to this question must depend on what we mean by "perceive speech." Here we have been misled, it would seem, by the behaviorist view of perception as a mere matter of psychophysical capacity. We have tended to describe speech in purely acoustic terms as a collection of "cues," without regard to the articulatory events that the cues specify, and then to suppose that any animal able to discriminate these cues can perceive speech. Yet the psychophysical capacities of an unlimited set of animals--from the human infant to the chinchilla--may suffice to discriminate among formant transitions, formant onset frequencies, brief silences, patches of noise, and so on. However, these capacities may not suffice to discover the functional relations among the perceptual pieces.

In fact, the perceptual status of communicative signals varies even for closely related species. For example, while two species of macaque (pig-tail and bonnet) and an African vervet may learn an arbitrary discriminative response to contrasting calls of the Japanese macaque, the latter learns the response significantly more rapidly (Zoloth, Petersen, Beecher, Green, Marler, Moody, & Stebbins, 1979). Moreover, the processes underlying the Japanese macaque's response to its own calls are evidently localized in the left cerebral hemisphere, while those of the other two species of macaque are not (Peterson, Beecher, Zoloth, Moody, & Stebbins, 1978; cf. Heffner & Heffner, 1984). Whether this hemispheric specialization has a perceptuomotor origin (as in the human: see below), we do not yet know. The point here is that, if we show a particular discriminative task to be within the psychophysical competences of two different species, we have not thereby shown their percepts to be equivalent.

In short, if the structure of perception can properly be said to be tuned to the structure of the perceiver's capacity for action, a non-human animal's perception of speech must differ radically from a human's. What actions of a macaque, say, are controlled by its perception of speech? What events do the acoustic patterns of speech specify for a macaque? Presumably, the patterns do not specify articulatory gestures, and the actions brought under control in the laboratory (such as lever holding or escape from shock) are the arbitrary choices of an experimenter, adventitious and ethologically empty. In other words, the information in speech may indicate to a non-human animal what it should do in a particular situation, but (pace the mynah bird) the information cannot specify for the animal, as it does for a human, the speaker's pattern of articulatory gestures.

Perceptuomotor Relations in the Infant

Since the infant, by definition, does not speak, our understanding of perceptuomotor development over the first year of life must be largely inferential. Here I will consider three classes of evidence, concerning: (1) the adult perceptuomotor system, particularly its cerebral locus; (2) infant perceptual capacity; (3) infant behavior, reflecting hemispheric specialization for speech perception.

The Adult Perceptuomotor System

Aphasia studies for over a century have suggested that the right cerebral hemisphere of most right-handed individuals is essentially mute (see, for example, Milner, 1974). Differential anesthesia of left and right hemispheres by intracarotid sodium amytal injection (preparatory to possible brain surgery) has confirmed this fact experimentally (Borchgrevink, 1982; Milner, Branch, & Rasmussen, 1964). Thus, speech motor control is vested in the left hemisphere of most individuals (roughly 90% of the population). (The origins of a population diversity, such that speech motor control is vested in the left hemisphere for some 90%, in the right hemisphere for some 10% of the population, are not yet understood.)

Since any imitative behavior calls for close neurophysiological connections between perceptual and motor processes, we might predict that left hemisphere control of articulation would be coupled with left hemisphere specialization for speech perception. Numerous monotic and dichotic studies of normal subjects have confirmed this prediction, and have demonstrated a double dissociation of left and right hemispheres for the perception of speech and non-speech (e.g., Kimura, 1961a, 1961b; Studdert-Kennedy & Shankweiler, 1970). Furthermore, studies of split-brain patients (whose cerebral hemispheres have been surgically separated for relief of epilepsy) have shown that, while the right hemisphere may recognize the meaning of a word from its overall auditory shape, only the left hemisphere can carry out the phonetic analysis necessary to establish a new word in an individual's lexicon (Zaidel, 1974, 1978). (Phonetic analysis refers, of course, to analysis of a word into its articulatory components and to recognition of the relations among them, as discussed above.) Thus, we have solid evidence that the adult speech perceptuomotor system is a left hemisphere function.

Infant Perceptual Capacity

As is well known, infants in the first six months of life can discriminate virtually any adult speech contrast on which they are tested (for reviews, see Aslin, Pisoni, & Jusczyk, 1983; Eimas, 1982). Much of the infant research has been carried out with synthetic speech continua on which adults typically display "categorical perception," that is, good discrimination between sounds that fall into different adult phonetic categories, but poor discrimination between sounds that fall into the same phonetic category. Infants have generally displayed a similar pattern, and this outcome has been interpreted as evidence that infants are prepared at birth, or very soon after, to perceive speech in terms of adult phonetic categories (Eimas, 1982).

This interpretation has been weakened by two sets of findings. First, we now know that categorical perception is not peculiar to speech, nor even to audition (e.g., Pastore et al., 1977). Second, Kuhl and her colleagues (Kuhl, 1978; Kuhl & Miller, 1978; Kuhl & Padden, 1983) have demonstrated categorical discrimination along synthetic speech continua for macaques and chinchillas. The issue is complicated by the fact that speakers of different languages may display different boundaries between the phonetic categories of a continuum (see Repp, 1984) and we may suspect (following the argument of the previous section) that quite different processes underlie the seemingly equivalent human and animal behavior. However, let us assume that categorical perception is essentially a psychophysical phenomenon, susceptible perhaps to effects of learning and attention, but based on the psychoacoustic tuning of the mammalian auditory system.

Nonetheless, we have ample other evidence that speech already has a unique status for the infant within a few hours or days of birth. For example, neonates can discriminate speech from non-speech (Alegria & Noirot, 1978, 1982), prefer speech to non-speech (Hutt, Hutt, Lenard, Bernuth, & Muntjewerff, 1968), and prefer their mother's voice to a stranger's (DeCasper & Fifer, 1980), provided she speaks with normal intonation rather than in word-by-word citation (Mehler, Barrière, & Jasik-Gerschenfeld, 1978). However, the strongest evidence for the unique status of speech comes from studies of infant hemispheric specialization.

Cerebral Asymmetry for Speech in Infants

A number of studies have demonstrated dissociation of the left and right sides of the brain for perceiving speech and non-speech sounds at, or very shortly after, birth. These include both physiological and behavioral studies. For example, Molfese, Freeman and Palermo (1975) measured auditory evoked responses, over left and right temporal lobes, of 10 infants aged from one week to 10 months. Their stimuli were four naturally spoken monosyllables, a C-Major piano chord, and a 250-4000 Hz burst of noise. Median amplitude of response was higher over the left hemisphere for all four syllables in nine out of ten infants, higher over the right hemisphere for the chord and the noise in all ten infants. Molfese (1977) has reported similar asymmetries for syllables and pure tones in neonates.

Dissociation between responses to speech and non-speech has also been demonstrated by Best, Hoffman, and Glanville (1982). These authors tested forty-eight 2- 3- and 4-month-old infants for ear differences in a memory-based dichotic task. They used a cardiac orienting response to measure recovery from habituation to synthetic stop-vowel syllables and to Minimoog simulations of concert A (440 Hz), played on different instruments. In the speech task, a single dichotic habituation pair (either /ba-da/ or /pa-ta/) was presented nine times at randomly varying intervals. On the tenth presentation, one ear again received its habituation syllable, while the other received a test syllable (either /ga/ or /ka/), differing in place of articulation from both habituation syllables. An analogous procedure was followed in the musical note task. The results showed significantly greater recovery of cardiac response for right ear test syllables in the 3- and 4-month-olds, and for left ear musical notes in all age groups. The authors propose that right-hemisphere memory for musical sounds develops before left-hemisphere memory for speech sounds, and that the latter begins to develop between the second and third months of life.

A further, particularly telling result, in light of the presumed amodal nature of the speech percept, comes from a study by MacKain, Studdert-Kennedy, Spieker, and Stern (1983). These authors showed that 5- to 6-month-old infants preferred to look at the face of a woman repeating the disyllable they were hearing (e.g., [zuzi]) than at the synchronized face of the same woman repeating another disyllable (e.g., [vava]). Thus, as in the study of Kuhl and Meltzoff (1982; Kuhl, this volume), infant preferences were for natural structural correspondences between acoustic and optic information, specifying the same articulatory event.

However,the most remarkable aspect of the study by MacKain et al. (1983) was that infant preferences for a match between the facial movements they were watching and the speech sounds they were hearing were only significant when the infants were looking to their right sides. We can interpret this result in the light of work by Kinsbourne and his colleagues (e.g., Kinsbourne, 1972; Lempert & Kinsbourne, 1982). Their work suggests that attention to one side of the body may facilitate processes for which the contralateral hemisphere is specialized. If this is so, we may infer that infants with a preference for matches on their right side were revealing a left hemisphere sensitivity to articulation specified by acoustic and optic information.

The work by MacKain and her colleagues has not yet been replicated. But if it proves reliable, we have some evidence that 5- to 6-month-old infants, close to the onset of babbling, already display a left hemisphere sensitivity to the amodal structure of speech events. For the moment, this seems to be close as we have come to detecting an incipient capacity for imitation on which spoken language is based.

Summary and Conclusions

Perception and action are mutually entailed components of a single system. Their interlocking operation is possible because the information picked up by a perceptual system is amodal and directly specifies, within the constraints of the actor's goal, the action to be performed.

Imitation is a specialized mode of action, requiring the imitator to find in the act of a model both the pieces of the act and their spatio-temporal relations. Imitation also calls for close neurophysiological connections between perception and motor control. For speech these perceptuomotor connections are localized in the left cerebral hemisphere.

Studies of infant speech perception have shown that infants are sensitive to structural correspondences between acoustic and optic specifications of speech, and that their left cerebral hemispheres are differentially activated by speech sounds soon after birth. We also have preliminary evidence for left hemisphere sensitivity to the amodal structure of speech by the fifth or sixth month of life.

The approach to speech perceptuomotor development outlined above also promises an ontogenetic solution to the vexed problem of the incommensurability of the speech acoustic signal and its linguistic description. The approach distinguishes between the dynamic information conveyed by an act and the static information in a symbol string. Thus, linguistic units are not postulated as part of the infant's native endowment. Rather they are seen as elements that emerge from a self-organizing system of perceptuomotor control (cf. Lindblom, MacNeilage, & Studdert-Kennedy, 1983).

References

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine.

Alegria, J., & Noirot, E. (1978). Neonate orientation behavior toward human voice. *International Journal of Behavioral Development*, *1*, 291-312.

Alegria, J., & Noirot, E. (1982). Oriented mouthing activity in neonates: Early development of differences related to feeding experiences. In J. Mehler, S. Franck, E. C. Walker, & M. Garrett (Eds.), *Perspectives on mental representation: Experimental and theoretical studies of cognitive processes and capacities*. Hillsdale, NJ: Erlbaum.

Aslin, R. N., Pisoni, D. B., & Jusczyk, P. W. (1983). Auditory development and speech perception in infancy. In M. M. Haith & J. J. Campos (Eds.), Infancy and the biology of development. Vol. II of *Carmichael's manual of child psychology*, 4th edition. New York: Wiley and Sons.

Best, C. T., Hoffman, H., & Glanville, B. B. (1982). Development of infant ear asymmetries for speech and music. *Perception & Psychophysics*, *31*, 75-85.

Borchgrevink, H. M. (1982). Mechanisms of speech and musical sound perception. In R. Carlson & B. Granström (Eds.), *The representation of speech* *in the peripheral auditory system* (pp. 251-258). New York: Elsevier Biomedical Press.

Carello, C., Turvey, M. T., Kugler, P. N., & Shaw, R. E. (1984). Inadequacies of the computer metaphor. In M. S. Gazzaniga (Ed.), *Handbook of cognitive neuroscience*. New York: Plenum.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mother's voices. *Science*, *208*, 1174-1176.

Eimas, P. D. (1982). Speech perception: A view of the initial state and perceptual mechanics. In J. Mehler, E. C. T. Walker, & M. Garrett (Eds.), *Perspectives on mental representation* (pp. 339-360). Hillsdale, NJ: Erlbaum.

Fentress, J. C. (1984). The development of coordination. *Journal of Motor Behavior*, *16*, 99-134.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition: English initial consonants in the first fifty words. *Language*, *51*, 419-439.

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

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

Heffner, H. E., & Heffner, R. S. (1984). Temporal lobe lesions and perception of species-specific vocalizations by macaques. *Science*, *226*, 75-76.

Hutt, S. J., Hutt, C., Lenard, H. G., Bernuth, H., & Muntjewerff, W. J. (1968). Auditory responsivity in the human neonate. *Nature*, *218*, 888-890.

Kimura, D. (1961a). Some effects of temporal lobe damage on auditory perception. *Canadian Journal of Psychology*, *15*, 156-165.

Kimura, D. (1961b). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, *15*, 166-171.

Kinsbourne, M. (1972). Eye and head turning indicates cerebral lateralization. *Science*, *176*, 539-541.

Klatt, D. H., & Stefanski, R. A. (1974). How does a mynah bird imitate human speech? *Journal of the Acoustical Society of America*, *55*, 82-89.

Kuhl, P. K. (1978). Predispositions for the perception of speech-sound categories: A species-specific phenomenon? In F. D. Minifie & L. L. Lloyd (Eds.), *Communicative and cognitive abilities. Early behavioral assessment.* (NICHD Series) (pp. 229-155). Baltimore, MD: University Park Press.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138-1144.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, *63*, 905-917.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America*, *73*, 1003-1010.

Lempert, H., & Kinsbourne, M. (1982). Effect of laterality of orientation on verbal memory. *Neuropsychologia*, *20*, 211-214.

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

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, *24*, 253-257.

MacKain, K. S., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left hemisphere function. *Science*, *219*, 1347-1349.

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

Mehler, J., Barrière, M., & Jasik-Gerschenfeld, D. (1976). La reconnaissance de la voix maternelle par le nourrisson. *La Recherche*, *70*, 787-788.

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, *574*, 702-709.

Milner, B. (1974). Hemispheric specialization: Scope and limitations. In F. O. Schmitt & F. G. Worden (Eds.), *The neurosciences: Third study program*. Cambridge, MA: MIT Press.

Milner, B., Branch, D., & Rasmussen, T. (1964). Observations on cerebral dominance. In V. S. DeRenck & M. O'Connor (Eds.), *Disorders of language* (Ciba Foundation Symposium, pp. 200-214). London: J. & A. Churchill.

Molfese, D. L. (1977). Infant cerebral asymmetry. In S. J. Segalowitz & F. A. Gruber (Eds.), *Language development and neurological theory*. New York: Academic Press.

Molfese, D. L., Freeman, R. B., & Palermo, D. S. (1975). The ontogeny of brain lateralization for speech and nonspeech stimuli. *Brain and Language*, *2*, 356-368.

Norton, S. J., Schultz, M. C., Reed, C. M., Braida, L. D., Durlach, N. I., Rabinowitz, W. M., & Chomsky, C. (1977). Analytic study of the Tadoma method: Background and preliminary results. *Journal of Speech and Hearing Research*, *20*, 574-595.

Pastore, R. E., Ahroon, W. A., Baffuto, K. J., Friedman, C., Puleo, J. S., & Fink, E. A. (1977). Common factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 686-696.

Petersen, M. R., Beecher, M. D., Zoloth, S. R., Moody, D. B., & Stebbins, W. C. (1978). Neural lateralization of species-specific vocalizations by Japanese macaques (*Macaca fuscata*). *Science*, *202*. 324.

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

Studdert-Kennedy, M., & Shankweiler, D. P. (1970). Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, *48*, 579-594.

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

Turvey, M. T., & Kugler, P. N. (1984). A comment on equating information with symbol strings. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, *246*, R925-R927.

von Uexküll, J. (1934). A stroll through the worlds of animals and men. In C. H. Schiller (Ed.), *Instinctive behavior* (reprinted 1957). New York: International Universities Press.

Zaidel, E. (1976). Language, dichotic listening and the disconnected hemispheres. In D. O. Walter, L. Rogers, & J. M. Finzi-Fried (Eds.), *Conference on Human Brain Function, Brain Information Service, BRI*. Los Angeles: UCLA Publications Office.

Zaidel, E. (1978). Lexical organization in the right hemisphere. In P. A. Buser & A. Rougeul-Buser (Eds.), *Cerebral correlates of conscious experience* (pp. 177-197). Amsterdam: Elsevier/North Holland Biomedical Press.

Zoloth, S. R., Petersen, M. R., Beecher, M. D., Green, S., Marler, P., Moody, D. B., & Stebbins, W. (1979). Species-specific perceptual processing of vocal sounds by monkeys. *Science*, *204*, 870-873.

DEPENDENCE OF READING ON ORTHOGRAPHY: INVESTIGATIONS IN SERBO-CROATIAN*

Claudia Carello† and M. T. Turvey††

1 Introduction

The relation between script and speech differs among the various orthographic categories. In general, alphabets maintain a closer link than do logographies. Comparisons between instances of each category, say between English and Chinese, are instigated in order to uncover whether or not different orthographic styles might be reflected in differing processing strategies used by readers. A number of investigators have pointed out, however, that "alphabet" does not constitute a monolithic category and English is, in no sense, to be taken as typical of all alphabets. Nonetheless, a majority of the reading data have been collected for English and the conclusions they suggest have been accepted, more or less by default, for alphabets in general. But a growing body of data for Serbo-Croatian, the (alphabetically transcribed) language of Yugoslavia, reveals important differences with English. We will summarize these data and elaborate their implications for linguistic issues, particularly the role of phonology in reading, that may be important for Chinese.

2 Linguistic Issues in Cross-language Comparisons

Orthographies can be distinguished along a number of dimensions, two of which will concern us here. First, they differ with respect to the particular units that are overtly represented, be they morphemes or syllables or the more (linguistically) abstract phonemes. Second, orthographies can be considered deep or shallow depending on their relative remoteness from the sounds to be read. As will be illustrated in the following characterizations of Serbo-Croatian, English, and Chinese, these dimensions are orthogonal--orthographies of "equal depth" can differ in the unit represented.

Serbo-Croatian uses an alphabet that represents phonemes in a straightforward symbol-to-sound mapping: Each letter has only one pronunciation. A novel word or pseudoword can be named (in the sense of pronounced) simply by generating the sounds from the letters. A letter such as a will be pronounced /a/ regardless of the letters that precede or follow it (ignoring, of course, subtle changes as a consequence of coarticulation). In order to

---

*To appear in the Proceedings of the Conference on Psychological Studies in the Chinese Language, Hong Kong, June 1984.

†State University of New York at Binghamton.

††Also University of Connecticut.

Acknowledgment. The research was supported in part by National Institute of Health Grants HD-08495 to the University of Belgrade, Yugoslavia, and HD-01994 to Haskins Laboratories.

preserve this mapping, the etymological relationships among words are sacrificed. Wherever the spoken language has imparted phonological variation in, say, declensions of a given noun, the variations are enforced in the spelling (e.g., nominative singular RUK+A, dative singular RUCI; nominative singular SNAHA, dative singular SNASI). It is, therefore, considered to be a shallow orthography (Liberman, Liberman, Mattingly, & Shankweiler, 1980).

In contrast, English uses an alphabet that also represents phonemes but enforces morphological continuity. Where the spoken language changes the pronunciation of a root morpheme, its spelling does not necessarily change. The sounds are determined by phonological rules with the result that etymological hints are retained (e.g., the relationship between "bomb" and "bombard" is preserved in their spellings despite alteration in the sound of the second "b"). A novel word or pseudoword can be named by generating the sounds from the letters and phonological rules. An alphabet that does not represent phonological variations that are determined by phonological rule[1] can be said to be deep.

Finally, Chinese uses a logography to represent morphemes. Although a large proportion of characters are phonograms--comprising both a semantic and a phonetic component--the hints to sound are not completely reliable (Wang, 1973). Using the phonetic component to sound out a character yields only 39% accuracy (Tzeng & Hung, 1980). By and large, therefore, the character names must be memorized in order to be read. Because of the opacity of the phonology, Chinese can be considered a deep orthography.

The fact that orthographies differ with respect to both the units they represent and the phonological transparency of those representations suggests that orthographies might also vary in the linguistic demands that they place on the reader, particularly the beginner. In other words, the effective use of orthographies might depend on how much readers know about the structure of their languages, with certain orthographies requiring an explicit understanding of the more abstract (and, presumably, harder to come by) aspects. Limiting our discussion to structural units, speaker-hearers can become aware of the words, morphemes, syllables, and phonemes that comprise their spoken language. If they are to become readers of that language, alphabets *require* an appreciation of the phonemic structure that logographies do not. Whatever the orthography, the level of linguistic awareness (Mattingly, 1972) must be compatible with the units represented, while using the orthography might be said to tune one to the level of awareness demanded. By this reasoning, fluent readers of Chinese are less likely to be aware of the phonemic structure of their language than are fluent readers of English because fluency in the morpheme-based orthography does not demand such awareness.

A similar circular causality is found in what has been termed phonological maturity (Liberman et al., 1980), the appreciation that readers have, to varying degrees, of the (morpho-)phonological rules which rationalize spellings that are related complexly to sound. That is to say, phonological maturity helps in reading words where phonological variation is determined by rule rather than orthographic representation (e.g., real is read /rēl/, reality is read /rē.al`.ət.ē/); reading experience, in turn, promotes phonological development. The demands of linguistic awareness and phonological maturity can be said to parallel, more or less, the dimensions we identified as distinguishing orthographies--the represented unit and its phonological transparency, respectively.[2]

3 Serbo-Croatian: A Bi-alphabetic, Inflected Language

Phonological transparency is only one characteristic that distinguishes Serbo-Croatian from English. The major language of Yugoslavia is also highly inflected. Nouns, pronouns, and adjectives are declined in seven plural and seven singular cases (nominative, locative, dative, instrumental, genitive, accusative, and vocative). Verbs are conjugated by person and number in six forms. But, because of the dictum to "Write as you speak and read as it is written" (the guiding principle behind the mid-19th century alphabet reforms directed by the Serbian language scholar Vuk Karadzić), root morphemes often are varied orthographically when an inflectional element is added.

Of primary relevance to transforming the linguistic issues of the last section into experimental questions, however, is the fact that Serbo-Croatian is written in two alphabets. Both the Cyrillic script (learned first in eastern parts of the country) and the Roman script (learned first in the West) map onto the same set of 30 phonemes but in an interesting way. While most letters are unique to one or the other alphabet, seven are common (i.e., are read the same way in the two scripts) and four are ambiguous (i.e., receive a different phonetic interpretation in each script). Since Yugoslavs are typically facile with both alphabets, the letters can be combined in a variety of ways for experimental purposes, which will become apparent in Section 5.0.

4 Assessing Lexical Access

We are interested in whether or not variations in the speech-script relationship promote differing processing strategies in reading. Since reading involves recognizing words, one process that has received considerable scrutiny is the pattern recognition step--how is a written letter string matched to its lexical representation? This question of lexical access has been addressed with (primarily) two paradigms: (1) In lexical decision tasks, subjects must decide as rapidly as possible whether or not a given letter string is a word; (2) In naming tasks, subjects must simply read the letter string aloud as rapidly as possible. In both tasks, the time transpiring between onset of the stimulus and initiation of the response is measured. Visual and phonological characteristics of the letter strings are varied to ascertain what effect, if any, they have on the response latencies.

Effects on lexical decision time are taken to have implications for the nature of lexical access, models of which include linguistic processes (phonological recoding of letter strings), nonlinguistic processes (simple figural analyses), and combinations of both (dual processing). Effects on naming may be consistent with one or another lexical routes or may suggest, further, that the lexicon need not be accessed at all in order to pronounce a letter string. These implications rest on two logical underpinnings. First, if a letter string is phonologically ambiguous (i.e., can be pronounced in more than one way), then any phonological analysis (if it exists) ought to be hindered in comparison to such an analysis on phonologically unique letter strings. This would be true in both lexical decision and naming. If phonological ambiguity produces no effect, the case for phonological analysis is undermined. Second, while the three general models of word processing all suggest that words should be named faster than pseudowords, a phonologically analytic strategy ought to yield a fairly small difference that is relatively constant for ambiguous and unambiguous letter strings. An interaction between lexicality and phonological ambiguity, however, would seem to support one of the other models. These will be elaborated in Section 6.0.

Obviously, a great deal hinges on the manipulation of phonological ambiguity. In English, two methods have been used. In one, pseudowords are constructed to be homophonic with words. While lexical rejection of pseudohomophones takes longer than rejection of pseudowords (Coltheart, Davelaar, Jonasson, & Besner, 1977), at least for good readers (Barron, 1978), interpretation of this fact is tricky because the appropriateness of pseudohomophones has been questioned on a number of grounds (Feldman, Lukatela, & Turvey, 1985; Martin, 1982). These include (i) the possibility that phonetic representations may be sensitive to orthographic differences between letter strings that sound alike when spoken aloud; (ii) the formal distinction, in English, between phonetic and morphophonological representations; and (iii) the suspicion that pseudohomophones are structurally odd.

The second way in which phonological ambiguity has been manipulated in English is through a comparison of words with regular and irregular (or exceptional) pronunciations. Whether or not differences are found, however, depends on how regularity is defined (Parkin, 1982). For example, words in which each graphemic unit receives the major phonemic correspondence (as detailed in Venezky's [1970] rules) are considered regular while those that receive a minor correspondence may be treated as irregular (Coltheart, Besner, Jonasson, & Davelaar, 1979). A finer distinction reveals that words can be classified as regular and consistent (i.e., they and all words that are visually similar to them receive the major phonemic correspondences) or regular and inconsistent (i.e., they receive major correspondences but other exemplars receive minor correspondences and, thus, are irregular [Glushko, 1979]). Some irregular words might be considered especially exceptional, however, if only because lexicographers provide pronunciation guides for them (but not for all minor correspondence words [Parkin, 1982]). Moreover, a particular grapheme-phoneme correspondence will be considered minor and, therefore, exceptional because there are fewer instances of it when, in fact, those instances might occur with greater frequency than the so-called major grapheme-phoneme correspondences (Parkin, 1982). Lastly, phonologically irregular words may differ with respect to whether or not they are orthographically irregular as well (Parkin & Underwood, 1983). Depending on which of these characterizations of regularity is used, one will or will not find differences between regular and irregular words, either supporting or belying claims for phonological analysis.

As important as the phonological manipulation is to evaluating lexical properties, it is not clear that studies in English have been successful in providing unequivocal tests. The task is much more straightforward in Serbo-Croatian, however, where the unique properties of the orthography can be exploited. In the following review, we will focus on the bi-alphabetism of fluent readers.

5 Reading in Serbo-Croatian Is Phonologically Analytic

Because Serbo-Croatian is phonologically shallow, there are no minor phonemic correspondences, no irregular words nor inconsistent regular words, and no orthographically irregular words. Phonological ambiguity is manipulated by choosing words (or nonwords) that combine common letters with unique letters (unambiguous letter strings) or common letters with ambiguous letters (ambiguous letter strings). The lexical status of letter strings so chosen will depend on their phonemic interpretation--that is, in which alphabet they are read. For example, an ambiguous string could be a word in Cyrillic but a pseudoword in Roman (or vice versa). Or it could be one word in Cyrillic but a different word in Roman (or pseudowords in both). An unambiguous string could be a word in one alphabet and impossible in the other (or a pseudoword in one and impossible in the other). Finally, if composed exclusively of common letters, a string would be the same word in both alphabets (or the same pseudoword).

In lexical decision tasks, comparisons of response times to the variety of letter string types reveals a phonological ambiguity effect--an ambiguous letter string takes longer to decide about than an unambiguous letter string. This is true when it is (i) a word in one reading and a pseudoword in the other; (ii) a word, though different, in both readings; and (iii) a pseudoword, though different, in both readings (Lukatela, Popadić, Ognjenović, & Turvey, 1980; Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978). The effect is more pronounced with words than pseudowords (Feldman & Turvey, 1983; Lukatela et al., 1978). The greater the number of ambiguous letters in the string, the longer lexical decision takes (Feldman, Kostić, Lukatela, & Turvey, 1983; Feldman & Turvey, 1983). While attempts to bias subjects toward a Roman reading by instructions or task (i.e., uniquely Cyrillic letters never appear) did not eliminate the effect, the presence of a single unique character did (Feldman et al., 1983; Lukatela et al., 1978). Finally, the effect is more pronounced in good readers than in poor readers (Feldman et al., 1985), suggesting that those who more effectively exploit the phonologically analytic strategy are harmed more by ambiguity.

It is important to note that the phonological ambiguity effect is not an artifact of the frequency of ambiguous letter strings. These occur regularly in the Serbo-Croatian language. But the point is underscored nicely by two experimental findings. First, in a comparison of two inflected forms of the same noun, frequency is (at one level) equal since they are the same word (e.g., RUKA and RUCI both mean hand). But the occurrence of the various grammatical cases differs such that nominative singulars (e.g., RUKA) are at least ten times more frequent than dative singulars (e.g, RUCI). When both forms are unique letter strings, the latency for nominatives is (about 80 ms) shorter. When the nominative singular is ambiguous and the dative singular is unambiguous (i.e., has one unique character), latency for datives is (about 185 ms) shorter (Feldman et al., 1983). Phonological ambiguity overrides the frequency advantage.

The second rejoinder to frequency arguments comes from a comparison of words that are ambiguous in one alphabetic transcription but unique in the other. For example, the Cyrillic version of "hawk"--KOбAч--is unique (pronounceable only as /kobats/) while its Roman version--KOBAC--is ambiguous (pronounced /kobats/ if read as Roman but /kovas/ if read as Cyrillic. With such pairs, a word can be used as its own control: Frequency, meaning, length, number of syllables are identical. Only the number of morphophonological representations is different but that is sufficient to produce a 350 ms difference in decision time (Feldman, 1981).

6 Word-pseudoword Comparisons

As indicated in Section 4.0, the three general models of word processing agree that words should be named faster than pseudowords. Their reasons are quite different, however, as are the particulars of how lexicality might interact with phonological ambiguity. A model of visual analysis suggests that words and pseudowords are read aloud by a common analogical process. Very roughly, a word finds a perfect analogy in the lexicon, with a singularly defined code for pronunciation; a pseudoword finds several analogies in the lexicon, defining several alternative pronunciations. The competition among lexical entries induced in the case of pseudowords would account for their slower naming relative to words (e.g., Glushko, 1979; Kay & Marcel, 1981). The effects of such competition ought to be especially (perhaps exclusively) apparent in experiments that compare phonologically ambiguous letter strings.

A model of phonological analysis holds that words and pseudowords are read aloud by a common phonological strategy that uses spelling-to-sound rules (based on the same principle as, though not necessarily identical to, the grapheme-to-phoneme correspondences identified by Venezky [1970]). Very roughly, the more regular the letter string the more rapid the recoding. As a rule, pseudowords will be less phonologically regular than words, resulting in slower naming latencies (e.g., Parkin, 1982; Parkin & Underwood, 1983). This residual difference should not change when both types of letter strings are chosen to be purposely ambiguous.

Finally, a dual process view asserts that words are read aloud by a visually based look-up of a word's lexical representation where the word's pronunciation can be retrieved. In contrast, pseudowords are read aloud by assembling a pronunciation on the basis of grapheme-phoneme correspondences. It is hypothesized that visual access is faster than rule-based assembly; consequently, words are named more rapidly than pseudowords (e.g., Coltheart, 1978; Coltheart et al., 1979). Phonological ambiguity should affect only pseudowords since their names alone are derived phonologically.

In Serbo-Croatian, at least, it appears that the difference in naming latencies between words and pseudowords does not change when phonological ambiguity is manipulated (Feldman, 1981). Both are slowed by about 450 ms when the letter strings can be read in two ways, suggesting that phonological involvement is the same for words and pseudowords. Certainly, this strategy is encouraged by the fairly direct correspondence to speech that the Serbo-Croatian orthographies exhibit. One might expect a different pattern with English, where the correspondence between orthography and speech is abstract. While English and Serbo-Croatian have not been compared directly (i.e., in the same experiment with the same controls) on the lexicality-ambiguity interaction, the direct comparisons that have been performed reveal differences between the languages that are germane to this issue. Since these involve a manipulation--semantic priming--that we have not yet discussed, we'll take a moment to describe its logic before summarizing the results.

It is commonly found that lexical decision and naming are facilitated when the target word is preceded by a semantically related priming word (Becker & Killion, 1977; Massaro, Jones, Lipscomb, & Scholz, 1978; Meyer, Schvaneveldt, & Ruddy, 1975). The general assumption is that when the prime activates its own lexical representation, that activation spreads to semantically related items, thereby speeding their subsequent lexical processing. Tasks that are lexically mediated ought to be facilitated; tasks that are not facilitated are unlikely to be lexically mediated.

Semantic priming of lexical decision is, in fact, found in both English and Serbo-Croatian (Katz & Feldman, 1983). For naming, however, facilitation is found only for English, suggesting that naming in the phonologically shallow Serbo-Croatian orthography need not involve the lexicon. This point is underscored by the correlations between lexical decision and naming (which may be taken as an index of processing similarity). In English, performance on semantically primed lexical decision correlates with naming, whether the latter is semantically primed or not; lexical decision without semantic priming also correlates with naming, whether primed or not. In Serbo-Croatian, the only significant correlation occurred when neither task was semantically primed. "The similarity between tasks is strongest when there is least involvement of the internal lexicon" (Katz & Feldman, 1983, p. 163).

7 Conclusion

The case for phonological analysis as the primary, not optional reading strategy in Serbo-Croatian is quite strong. It is not yet clear, however, whether or not this strategy is peculiar to Serbo-Croatian (or writing systems with similar properties): Does phonological analysis result from experience with a shallow orthography (i.e., does orthography influence processing) or is it simply easier to demonstrate in the sorts of experiments that the Serbo-Croatian orthography allows?

As strongly as we argue for a phonologically analytic strategy in Serbo-Croatian, others have claimed that Chinese characters can only be read via the visual route. Indeed, lexical decision is slowed by a visual manipulation wherein the internal components of two-character words (and nonwords) are distorted disproportionately, for example, 内 becomes  ; 功 becomes  (Hung, Tzeng, Salzman, & Dreher, 1984). This parallels the result for mixing upper and lower case letters in English (e.g., Coltheart & Freeman, 1974) but is in contrast to mixing Cyrillic and Roman letters in Serbo-Croatian. The latter slows neither lexical decision nor naming (Feldman & Kostić, 1981; Katz & Feldman, 1981). Interestingly, however, visual distortion in both Chinese and English affects poor readers more than good readers (Hung et al., 1984). This is puzzling if one assumes that the manipulation interferes with the putatively optimal strategy on which better readers ought to be more reliant. Serbo-Croatian, at least, follows the expected logic for a phonologically analytic strategy--good readers are hurt more by phonological ambiguity (Feldman et al., 1985).

We do not know if fluent readers of Chinese rely on some strategy other than visual analysis or if they can resort to some other strategy if the visual route is hindered. We do know that there are hints of some phonological analysis of Chinese characters. Detection of graphemic components (e.g., 台 /tai/) is more successful when the component carries a phonetic clue (as in 抬 /tai/) than when it does not (as in 怡 /yi/ [Hung & Tzeng, 1981]). Inconsistent characters take longer to name than consistent characters (where consistency is defined by the ratio of exemplars pronounced the same as the target to the total number of characters with that phonetic, regardless of how they are pronounced [Fang & Horng, this volume]). And a comparison of Japanese kanji (the logographic script borrowed from Chinese) with kana (a syllabary that depicts the phonetic value of its characters) reveals that colors are named faster when written in kana even though color names appear more frequently in kanji in Japanese literature (Feldman & Turvey, 1980; cf. Saito, 1981). This last finding, especially, seems troublesome for those models that restrict the role of phonological analysis. Phonological involvement is demonstrated for words (not just pseudowords) and it appears to facilitate, rather than slow, naming. One might argue that if phonological analysis is optional, then it is an option readily (eagerly?) exploited when available--even in writing systems that are biased, by design and practice, in favor of visual analysis (cf. Brooks, 1977).

References

Barron, R. W. (1978). Reading skill and phonological coding in lexical access. In M. M. Gruneberg, R. N. Sykes, & D. E. Morris (Eds.), *Practical aspects of memory*. London: Academic Press.

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.

Brooks, L. (1977). Visual patterns in fluent word identification. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a theory of reading: The proceedings of the CUNY conference*. Hillsdale, NJ: Erlbaum.

Coltheart, M. (1978). Lexical access in simple reading task. In G. Underwood (Ed.), *Strategies of information processing*. London: Academic Press.

Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E. (1979). Phonological encoding in the lexical decision task. *Quarterly Journal of Experimental Psychology*, *31*, 489-502.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI*. London: Academic Press.

Coltheart, M., & Freeman, R. (1974). Case alternation impairs word identification. *Bulletin of the Psychonomic Society*, *3*, 102-104.

Fang, S. P. & Horng, R. Y. Consistency effects in naming real and pseudo Chinese characters. (This volume)

Feldman, L. B. (1981). Visual word recognition in Serbo-Croatian is necessarily phonological. *Haskins Laboratories Status Report on Speech Research*, *SR-66*, 167-201.

Feldman, L. B., & Kostić, A. (1981). Word recognition with mixed-alphabet forms. *Haskins Laboratories Status Report on Speech Research*, *SR-66*, 203-211.

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., Lukatela, G., & Turvey, M. T. (1985). Effects of phonological ambiguity on beginning readers of Serbo-Croatian. *Journal of Experimental Child Psychology*, *39*, 492-510.

Feldman, L. B., & Turvey, M. T. (1980). Words written in Kana are named faster than the same words written in Kanji. *Language and Speech*, *23*, 141-147.

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.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 674-691.

Hung, D. L., & Tzeng, O. J. L. (1981). Orthographic variations and visual information processing. *Psychological Bulletin*, *90*, 377-414.

Hung, D. L., Tzeng, O. J. L., Salzman, B., & Dreher, J. (1984). A critical evaluation of the horse-racing model of skilled reading. In H. S. R. Kao & R. Hoosain (Eds.), *Psychological studies of the Chinese language*. Hong Kong: Chinese Language Society of Hong Kong.

Katz, L., & Feldman, L. B. (1981). Linguistic coding in word recognition: Comparisons between a deep and a shallow orthography. In A. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading*. Hillsdale, NJ: Erlbaum.

Katz, L., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 157-166.

Kay, J., & Marcel, A. (1981). One process, not two, in reading aloud: Lexical analogies do the work of non-lexical rules. *Quarterly Journal of Experimental Psychology*, *33A*, 397-413.

Klima, E. S. (1972). How alphabets might reflect language. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye*. Cambridge, MA: MIT Press.

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.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. (1980). Lexical decision in a phonologically shallow orthography. *Memory & Cognition*, *8*, 124-132.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. (1978). Bi-alphabetical lexical decision. *Language and Speech*, *21*, 142-165.

Martin, R. C. (1982). The pseudohomophone effect: The role of visual similarity in non-word decision. *Quarterly Journal of Experimental Psychology*, *34A*, 395-409.

Massaro, D. W., Jones, R. D., Lipscomb, C., & Scholz, R. (1978). Role of prior knowledge on naming and lexical decision with good and poor stimulus information. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 498-512.

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 relationship between speech and reading*. Cambridge, MA: MIT Press.

Mattingly, I. G. (1984). Reading, linguistic awareness, and language acquisition. In J. Downing & R. Valtin (Eds.), *Language awareness and learning to read*. New York: Springer-Verlag.

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* (pp. 98-118). London: Academic Press.

Parkin, A. (1982). Phonological recoding in lexical decision: Effects of spelling-to-sound regularity depend on how regularity is defined. *Memory & Cognition*, *10*, 43-53.

Parkin, A., & Underwood, G. (1983). Orthographic vs. phonological irregularity in lexical decision. *Memory & Cognition*, *11*, 351-355.

Saito, H. (1981). Use of graphemic and phonemic encoding in reading Kanji and Kana. *The Japanese Journal of Psychology*, *52*, 266-273.

Tzeng, O. J. L., & Hung, D. L. (1980). Reading in a nonalphabetic writing system: Some experimental studies. In J. F. Kavanagh & R. L. Venezky (Eds.), *Orthography, reading, and dyslexia*. Baltimore, MD: University Park Press.

Venezky, R. L. (1970). *The structure of English orthography*. The Hague: Mouton.

Wang, W. S-Y. (1973). The Chinese language. *Scientific American*, *228*, 50-60.

Footnotes

---

[1] This is akin to Klima's (1972) third convention.

[2] We find these parallels to be pedagogically useful but they may be idiosyncratic and should not be taken as representative of how linguistic demand is characterized typically. For example, Mattingly (1984) has recently revised his distinction of phonological maturity and linguistic awareness as entailing grammatical knowledge and access to such knowledge, respectively. We are less able to use this distinction for our present purpose of classifying orthographies.

THE RELATIONSHIP BETWEEN KNOWLEDGE OF DERIVATIONAL MORPHOLOGY AND SPELLING ABILITY IN FOURTH, SIXTH, AND EIGHTH GRADERS

Joanne F. Carlisle†

Abstract. This study investigated young students' knowledge of derivational morphology and the relationship between this knowledge and their ability to spell derived words. The subjects (fourth, sixth, and eighth graders) were given the Wide Range Achievement Test, Spelling subtest, and several experimental tasks--1) a test of their ability to generate base and derived forms orally; 2) a dictated spelling test of the same base and derived words; and 3) a test of their ability to apply suffix addition rules. The results indicate strong developmental trends in both the mastery of derivational morphology and the spelling of derived forms; however, spelling performances lagged significantly behind the ability to generate the same words. Success generating and spelling derived words depended on the complexity of the transformations between base and derived forms. Further, mastery of phonological and orthographic transformations most strongly distinguished the three grades in both spelling and generating derived forms. Other indications that the older students were using knowledge of morphemic structure in spelling derived forms were found in analysis of the spelling of base and derived word pairs and the application of suffix addition rules. However, incomplete mastery of the phonological and orthographic transformations suggests that students might benefit from explicit instruction in morphemic structure in order to improve their spelling of derived words.

Introduction

It is commonly acknowledged that learning to spell English words requires an understanding of the relationships between phonemes and graphemes and a memory for those words or parts of words that are "irregular." However, since our orthography is morphophonemic, it seems reasonable to believe that a knowledge of the morphemic structure of words would be helpful, perhaps even necessary, to spell accurately the many words of more than one morpheme that we use in writing. Although we know that understanding morphology develops gradually from childhood to adulthood, little is known about the extent to

---

†Also American International College.

Acknowledgment. This manuscript is based on portions of a doctoral dissertation submitted to the University of Connecticut. I would like to express my gratitude to Isabelle Y. Liberman for her guidance on this and other projects. I would also like to thank Donald Shankweiler and Ignatius G. Mattingly for their thoughtful recommendations. I am indebted to Jerome Spears, Director of Pupil Personnel Services, and the school system of Mansfield, Connecticut, for their assistance. The research reported here was sponsored by NICHD grant HD-01994 to Haskins Laboratories and by a dissertation fellowship from the University of Connecticut Research Foundation.

which this knowledge helps an individual acquire proficiency in spelling. The present study is concerned with the spelling of derived forms and addresses the question, is there a relationship between knowledge of derivational morphology and spelling ability?

Although the relationship between morphological knowledge and the acquisition of spelling skill would seem to have educational relevance, there have been very few investigations of the matter. The paucity of research studies is surprising since quite a few theorists have suggested that sensitivity to morphemic structure should enhance the ability to spell English words (Frith, 1980; Henderson, 1982; Liberman, Liberman, Mattingly, & Shankweiler, 1980; Mattingly, 1980; Venezky, 1970) and that explicit instruction in the morphemic structure of words could have benefits for the student learning to spell derived words (Chomsky, 1970; Russell, 1972).

The learning of derivational morphology is a complex matter. Although not a necessary part of the grammar of the language, the affixes allow us to express a concept (e.g., *love*) in a number of different grammatical forms, usually while retaining the basic identity of the base form (e.g., *lovable*, *lovely*, *loveliness*). While having familiar morphemes in many different words offers ease and efficiency in conveying meaning, this benefit accrues only if we are able to appreciate the morphological relationship between different words in the same word family. Unfortunately, the distance between base and derived form in phonology and semantics can sometimes be a formidable barrier. As Klima (1972) suggests, it is questionable whether most adult speakers of English recognize the many relatively obscure morphological relationships that exist in the English language. How many, for example, are aware that *crux* and *crucial* are members of the same word family?

Both the range and the complexity of the phonological transformations from base to derived forms may make derivational relationships hard to appreciate. While Chomsky and Halle (1968) have proposed that the phonological changes from base to derived forms are orderly and ruleful, a number of researchers have questioned the psychological reality of the underlying phonological rule system (Barganz, 1971: Jaeger, 1984; Moskowitz, 1973; Steinberg, 1973; Templeton, 1980). Collectively, these suggest that children and adults have varied degrees of understanding of the underlying phonological rule system.

Several characteristics of derivational morphology make productive knowledge problematic. First, the construction of derived forms does not follow consistent patterns. For example, two quite similar words such as *terror* and *horror* have only some of the same derived forms (Richardson, 1977). They have in common *terrible* and *horrible*, *terrify* and *horrify*; on the other hand, there is *terrorize*, but not *horrorize* and *horrid* but not *terrid*. Second, the range of syntactic options makes learning the proper derived forms complex. Derived nouns, for example, can end in *-ity*, *-ment*, *-ness*, *-ence*, and *-th*, just to name a few variations. In some cases, a base word occasionally has several derived forms of the same part of speech, such as *honestness* and *honesty* or *bountiful* and *bounteous*. Third, differences in the meanings of the suffixes are often subtle or nonexistent. In fact, the same suffix can have different meanings, depending on the word it is attached to (Thorndike, 1941). For example, the suffix *-ful* has different meanings in the words *cupful* and *helpful*. Finally, derived forms sometimes undergo semantic shifts that make their relationship to their base forms seem remote. This is the case with *apply* and appliance. When similarity in meaning is absent, the realization that the base and derived forms are related requires more linguistic sophistication than many individuals have.

In view of such complexities, it is possible that the learning of word-specific patterns may play an important role in the learning of derivational morphology. Awareness of the morphological relatedness of words and the ability to analyze morphemic structure may depend on combined features of phonological and semantic similarities and associations, on linguistic sophistication and even on the specific characteristics of the language tasks used to assess this ability (Derwing & Baker, 1979; Smith & Sterling, 1982).

The Development of Knowledge of Derivational Morphology

Children learn inflected forms of words rulefully. Their knowledge of most inflectional rules, evident from the ability to supply the correct forms of nonsense words in sentences, is generally complete by the time they are seven years old (Berko, 1958). Derivational rules, however, are learned more slowly and less systematically than inflectional rules. Children's vocabulary growth during the years 7 to 12 includes many words of complex morphological structure, particularly derived forms (Ingram, 1976). To some extent, morphophonemic rules appear to be learned during this time (Moskowitz, 1973). However, the productive knowledge of even basic derivational forms may not be complete even for teenagers (Selby, 1972). In fact, since derivational morphology is an open system, learning derived forms can take place throughout adulthood for individuals who have some curiosity about words (Klima, 1972).

Although derivational morphology cannot be said to be mastered within a particular developmental period, certain developmental trends in ruleful learning of derived forms have been found. Using a task modeled after Berko's, Derwing (1976) found a consistent trend among children (ages 8 to 12), adolescents, and adults toward productive knowledge of five of the six derivational patterns he selected for investigation. These were the agentive *-er*, the *-y* adjective, noun compound, instrumental *-er*, and the *-ly* adverb. (The sixth pattern was the diminutive, which did not become productive.) The developmental trend toward mastery found in this study suggests that the learning of derived forms begins soon after age seven when the inflected forms have usually been mastered, a phenomenon also evident from Moskowitz's study (1973).

The constructions that Derwing found to be productive are regular and quite transparent. The base word remains intact in the derived form and does so without requiring a change in the phonology of the base word. Not all derivational relationships are so regular in construction or so closely related in phonology and orthography (Berko, 1958). There is less evidence to suggest that children have productive knowledge of those forms with complex phonological and semantic relationships.

Morphological Knowledge and Spelling Ability

English orthography maps onto the morphophonology of the language. Chomsky and Halle (1968) note that where changes in pronunciation from a base to a derived word are predicted by the regular sound pattern of the language, the orthography does not need to reflect the change (e.g., *race* to *racial* and *reduce* to *reduction*). A number of studies have shown that the orthographic regularities seem to provide the reader with clearer clues to morphological relationships than the underlying phonological rules (Barganz, 1971; Jaeger, 1984; Jarvella & Snodgrass, 1974; LaSorte, 1980; Moskowitz, 1973; Steinberg, 1973; Templeton, 1980). The reader who can discover from the regularity of the spelling that two words are morphologically related can use this knowledge to good advantage through efficient processing of words and through appreciation of semantic relationships and syntactic variations. It is not surprising, therefore, that there appears to be quite a strong relationship between morphological knowledge and reading or vocabulary development (Barganz, 1971; Freyd & Baron, 1982; LaSorte, 1980).

The issue we are addressing here, however, is not whether orthographic regularities help the reader, but whether they are useful to the speller--whether knowledge of the morphemic structure of words, which may be more apparent from the orthography than the phonology, is drawn upon by the speller of derived words. Reading and spelling, though closely related, are quite different tasks (Frith, 1980). C. Chomsky (1970) argues that the use of orthographic knowledge to spell derived words correctly is a natural development, at least for the good speller who can recall the orthographic similarities of related words, even when the pronunciations are dissimilar. She suggests that the spellers' knowledge of word families can help disambiguate such troublesome elements as the spelling of an unstressed vowel, as in democracy (where knowing democrat helps) or a silent consonant, as in muscle (where knowing muscular helps). Russell (1972) believes that the phonological and orthographic regularities, apparent from reading words, can be emphasized in instruction in spelling. However, neither Chomsky nor Russell offers direct evidence to support the position that knowledge of morphological structure helps the speller spell derived words correctly.

While studies of the spelling of young children give some indication of a growing awareness of morphemic structure (Marino, 1979; Rubin, 1984; Schwartz & Doehring, 1977), we do not know if an awareness of simple morphemic structure carries over to the spelling of derived forms, particularly those that undergo phonological or orthographic shifts. How well an individual speller can apply morphological knowledge to the task of spelling may depend on how explicit as well as how extensive this knowledge is. It may also depend on the speller's mastery of the orthographic conventions that govern the addition of suffixes to base words.

Two studies have looked at the spellers' ability to use morphological knowledge. One is an investigation of the use of phonological knowledge and orthographic knowledge in a dictated spelling task (real and nonsense words) involving good spellers at the sixth-, eighth-, and tenth-grade levels (Templeton, 1980). The results of this study suggest that seeing a base word prompted better recall of the phonological rules governing the spelling of derived forms than hearing the base word. In addition, the students could spell the nonsense derived words better than they could pronounce them. Templeton suggested that learning about the orthographic structure of derived words might bring about a more comprehensive and productive awareness of the underlying phonological rules.

The second study of spellers' sensitivity to morphemic structure was of good and poor spellers at the college level (Fischer, Shankweiler, & Liberman, 1985). Good spellers were much better than poor spellers at spelling morphophonemically complex words. This discrepancy was particularly striking because the two groups differed less in their ability to spell words that are orthographically transparent (adverb) or orthographically deviant (Fahrenheit). Performances on additional tasks suggested that differences in spelling morphophonemically complex words were attributable to differences in linguistic knowledge, specifically knowledge of morphological structure. The good spellers were superior to the poor spellers on nonsense tasks of prefixation and suffixation, suggesting that they were not simply better spellers of real words.

While these two studies seem to indicate that good spellers between the sixth grade and college level can use morphological knowledge to help them spell derived words, this pattern may not hold for poor spellers at these levels in school or for younger students. Spelling errors made by junior high school students have been observed to indicate lack of awareness of morphemic structure (e.g., easally for easily) (Carlisle, 1984). Similarly, in an analysis of spelling errors on compositions, Sterling (1983) found that 12-year-old students treated derived words as if they were monomorphemic words. His analysis of the students' spelling errors indicated that inflected forms were spelled by ruleful system, but derived forms were spelled as unanalyzed wholes. He suggested that access to the knowledge of morphological relationships may be obscured by the complex nature of derivational morphology.

Experiment

The general purpose of the present study was to investigate the early stages of acquisition of knowledge of derivational morphology and of the ability to spell derived words. Several different considerations guided the formulation of the questions and the design of the study. First, on the basis of investigations by Berko (1958), Derwing and Baker (1979), and Selby (1972), it was expected that learning derivational morphology would begin in the third or fourth grades, following the mastery of the inflected forms. Accordingly, students in the fourth, sixth, and eighth grades were chosen as subjects in order to provide insight into the developmental mastery of derivational morphology. Second, the study was based on the hypothesis that students do acquire ruleful knowledge of the derivational morphology and that they do not simply learn to spell derived forms as unanalyzed whole words.

The research questions were as follows: First, are there developmental trends between the fourth and eighth grades in the acquisition of morphological knowledge and knowledge of the spelling of derivatives? Second, is there a relationship between the knowledge of derivational morphology and the ability to spell derived forms in the fourth, sixth, and eighth grades? Third, is there evidence that the learning of derivational morphology and the spelling of derived forms is ruleful in nature, taking into account both phonological and orthographic transformations?

In order to investigate these issues, two tasks were devised to allow for direct comparisons of the two skills--an oral test of the ability to generate derived forms and a dictated spelling test using the same words. The words were chosen to include four possible relationships between base forms and derived forms, on the assumption that these would engender errors that would reflect different levels of mastery of phonological and orthographic rules. Included were (a) word pairs in which there is NO CHANGE in the phonology or orthography (e.g., enjoy and enjoyment), (b) pairs in which there is a PHONOLOGICAL CHANGE but no orthographic change (e.g., major and majority), (c) pairs in which there is an ORTHOGRAPHIC CHANGE but no phonological change (e.g., rely and reliable), and (d) pairs in which phonology and orthography BOTH CHANGE (e.g., reduce and reduction).

In developing these tests to address the research questions, we anticipated two particular patterns of results. First, on the question of the developmental trends of morphological knowledge and spelling ability, we expected performance on the dictated spelling test to lag behind performance on the test of oral generation (called the Test of Morphological Structure), since the development of morphological knowledge most likely precedes the ability to use this knowledge in spelling. Second, on the question of the ruleful nature of learning the morphology and spelling of derived words, we expected that the words undergoing phonological and both phonological and orthographic changes would present more difficulty than the words with more transparent relationships (those undergoing no change or just orthographic change). This expectation was based on the finding of various research studies that the more remote the relationship between base and derived forms, the more difficult it is to learn the relationship rulefully (Berko, 1958; Derwing, 1976; Derwing & Baker, 1979: Moskowitz, 1973; Templeton, 1980).

A final consideration reflected the nature of orthographic rules and the derived words whose spelling is governed by these rules. The spelling of such words draws on a somewhat different kind of "ruleful" learning--the conventions of our spelling system. While a knowledge of the morphological components of words such as "sunny" would make the task of spelling easier, specific knowledge of the conventions of spelling words with suffixes (such as the rules governing the doubling of consonants) would also seem to be helpful, if not necessary. An exploratory study of the mastery of suffix addition rules between the seventh and ninth grades showed that words with suffixes made up more than half the errors in the students' compositions (Carlisle, 1984). Therefore, a test was devised that would help determine whether the students were able to apply the suffix addition rules consciously. Since the orthographic changes could be memorized as word-specific spellings, this test required the addition of suffixes to nonsense words. On the premise that mastery of the suffix addition rules is dependent on knowledge of morphemic structure and knowledge of abstract generalizations, the students' ability to apply these spelling rules was expected to develop later than their knowledge of morphological structure.

Method

Subjects

The subjects were 65 students from the fourth, sixth, and eighth grades of a rural Connecticut school system. The 22 sixth graders and the 21 eighth graders came from classes studying language arts and literature. These were selected by the teachers on the basis of class size and availability of time. The fourth-grade group was made up of 22 students from two elementary classrooms. All subjects were judged by their teachers as having at least average intelligence.

Procedures

The Spelling subtest of the Wide Range Achievement Test (WRAT) was administered to each grade level group (Jastak & Jastak, 1978). Within a week the Derived Forms subtest of the Spelling Test was administered to each grade-level group. One week later the Base Forms subtest of the Spelling Test and the

Test of Suffix Addition were administered. Two weeks after the administration of the Derived Forms subtest of the Spelling Test, the Test of Morphological Structure was administered to each student individually.

Materials

1. Wide Range Achievement Test, Spelling Subtest (Jastak & Jastak, 1978). This dictated spelling test was administered to determine the spelling capabilities of the subjects and to evaluate the validity of the experimental spelling tests. Level I was administered to the fourth-grade group, and Level II to the sixth- and eighth-grade groups. Level II, the appropriate form to use with youngsters aged 12.0 and over, was given to all sixth graders even though some of them were not yet 12 in order to permit group administration of the test and to insure an accurate comparison of spelling abilities within the sixth-grade group. The test was administered in accordance with the directions for group administration. The students' performance on the Wide Range Achievement Test (WRAT) Spelling subtest yielded the following grade-equivalent scores: fourth grade 5.9 (standard deviation of 1.0): sixth grade, 6.7 (standard deviation of 1.4); and eighth grade, 9.4 (standard deviation of 1.3). The correlation between the students' performances on the WRAT Spelling subtest and on the Derived Forms subtest of the Spelling Test (described hereafter) was .64 ($p < .001$).

2. Test of Morphological Structure. This experimental test was designed to assess knowledge of derivational morphology. The test has two subtests. For the Derived Forms subtest, the student's task was to state a specific derived form, once the examiner had given the base word and a sentence that needed the derived form as the final word to complete the sentence. (The first item on this subtest was: "Warm. He chose the jacket for its __." The target response was "warmth.") For the Base Forms subtest, the student's task was to state the base form, once the examiner had given the derived form and an appropriate sentence, designed to end with the base form. (The first item on this subtest: "Growth. She wanted her plant to __." The target response was "grow.")

The words on the test (see the Appendix) are based on four types of linguistic relationship between the base word and derived form:

a. NO CHANGE--Neither the phonology nor the orthography of the base changes in the derived form (e.g., enjoy and enjoyment).

b. ORTHOGRAPHIC CHANGE--The spelling but not the phonology of the base word changes in the derived form. Three types of changes were included in the word list: the doubling of a final consonant before the suffix (i.e., sun to sunny), the transformation of the y to i (e.g., rely to reliable), and the omission of a final e before a suffix beginning with a vowel (e.g., endure to endurance).

c. PHONOLOGICAL CHANGE--The pronunciation changes in the shift from the base word to the derived form without an accompanying change in spelling. Four kinds of phonological change were included: (a) tense to lax vowel (e.g., heal to health), (b) vowel reduction (e.g., original and originality), (c) shift in the pronunciation of a consonant (e.g., magic and magician), and (d) shifts in both a vowel and a consonant pronunciation (e.g., sign and signal).

d. BOTH CHANGE--Changes in both the orthography and the phonology occur in the shift from base word to derived form. Among the words of this group are representatives of different types of phonological shifts, including vowel shifts, consonant shifts, and shifts in the pronunciation of both vowel and consonant. Examples of BOTH CHANGE word pairs are deep and depth, decide and decision, and reduce and reduction.

Words for the two subtests, Derived Forms and Base Forms, were selected to be as similar as possible in length, frequency, affixation, and similarity in meaning of the root word and its derived form. First, words on the two subtests, type by type, do not differ in word length, as determined by number of letters. The average length for the base words is 5.6 letters for Derived Forms and 5.7 letters for Base Forms; the derived forms of both subtests average 8.5 letters.

Second, an effort was made to ensure the familiarity of the words for students in grades four through eight. As a measure of the familiarity of the written forms, only words with a Standard Frequency Index rating of 40 or above were used (Carroll, Davies, & Richman, 1971). (A Standard Frequency Index of 40 indicates a word that has an estimated frequency of one in a million words.) The words were equated for frequency by word type (NO CHANGE, ORTHOGRAPHIC CHANGE, and so on) on the two subtests, Base Forms and Derived Forms. The mean frequencies are as follows: for the base words, 55.1 (SD 1.8) on the Base Forms subtest and 55.2 (SD 0.9) on the Derived Forms subtest; for the derived words 49.6 (SD 2.4) on the Base Forms subtest and 50.6 (SD 1.8) on the Derived Forms subtest.

Third, attempts were made to control for semantic distance (i.e., the similarity of the meanings of base and derived forms), semantic variations, and syntactic options, all factors that can affect the difficulty of generating morphological forms. An effort was made to select base and derived forms with familiar and similar meanings. The sentences were written in such a way as to constrain possible choices in meaning and form. Pilot testing was used to eliminate items that did not meet these criteria.

The order of items on each subtest was determined by creating ten sets of four items, each set made up of one word of each word type (NO CHANGE, ORTHOGRAPHIC CHANGE, PHONOLOGICAL CHANGE, and BOTH CHANGE). The four word types were randomly ordered within each set, and the ten sets were randomly ordered on the test.

The test was administered by means of a tape-recording in standard English spoken by a native American male speaker. Directions and practice items were given by the examiner. The directions indicated that the student was to give the form of the word that correctly completed the sentence. Three practice items were given to all the students; the first, for example, was: "Farm. My uncle is a __." The correct response was farmer.

If the student completed the first practice item incorrectly, the correct answer was provided. The item was then repeated so that the student could give the correct answer. Once the tape was started, the administration continued without further assistance. If a student gave no response to a test item in the allotted time (5 s between the end of one item and the beginning of the next), the tape was stopped, and the student was asked if he or she could give a form of the word that completed the sentence. After this answer was recorded, the student was asked to try to give a prompt response to each item and was reminded that extra time would not be given for other items.

3. The Spelling Test. This dictated spelling test was used to determine whether the students could spell the same base words and their derived forms that make up the Derived Forms subtest of the Test of Morphological Structure. The first subtest (Derived Forms) consists of the 40 derived forms, given by dictation. The second part (Base Forms), consists of the 40 base words, also presented by dictation. The words appear in random order in each subtest. The Derived Forms subtest was administered a week before the Base Forms subtest so that the subjects would not be sensitized to the relationship between the root and derived forms in spelling the derived forms.

The test was administered by means of a tape-recording in standard English spoken by a native American male speaker. Each word was presented first alone, then in a sentence, and finally alone. There was a 10-s lapse between the last pronunciation of the spelling word and the start of the next item. The directions and two sample items were given by the examiner orally. The directions explained the nature of the test and the student's task, including giving the procedure for writing the words. The students were told that they could not pick up their pencils to write the dictated word until the test item had been completed. The students were directed "To listen carefully to each word and the way it is used in the sentence." The same directions were used for the two subtests.

4. Test of Suffix Addition. This test was designed to determine the extent to which students were able to apply the rules that govern the addition of suffixes to base words. Nonsense words were used as the base words so that the correct execution of this task could not be accomplished on the basis of familiarity. The test consists of 30 nonsense words, each followed by an addition sign (+), a suffix, an equal sign and a blank line. (The first item, for example, is "dun + er = __.") The nonsense words were constructed from real words by substituting one consonant for another (dun for run) or one consonant blend for a consonant or another consonant blend (drim for swim or prad for sad). In no case was the substituted consonant the final consonant in the original word.

Each item on the Test of Suffix Addition requires the use of one of the three suffix rules that form the basis of the ORTHOGRAPHIC CHANGE word type on the Test of Morphological Structure. There are ten items for each spelling rule--the rule governing the doubling of a final consonant (called the "doubling" rule), the final -y rule and the final -e rule. In addition, since sometimes no change is made in the base word when the suffix is added, about half the words required an orthographic change for correct suffix addition and the other half did not. The test items assess the knowledge of some fairly refined aspects of the conventions for suffix addition. For example, "leace + able = __" requires the knowledge that the e must be retained to indicate the "soft" sound of the c in leace (i.e., leaceable). Such items were included to probe the breadth of the students' knowledge of the rules that govern suffix addition.

Directions for this test were given aloud by the examiner, and two examples were completed on the blackboard to illustrate what was expected of the students. The directions indicated that the base words were not real words, but that the students were to put the two parts (base and suffix) together as if they were real English words.

The student wrote each word on a long blank line following the test item. There was no time limit. Most students took approximately 10 minutes to complete the test.

Results

Developmental Trends in the Learning of Derivational Morphology

Performances on the Test of Morphological Structure (TMS) were initially scored by tabulating the number of correct responses for each subtest, Base Forms and Derived Forms. The mean scores for each subtest of the TMS, given in Table 1, show that there was an increase in the knowledge of derivational relationships by grade level.

Table 1

Mean number correct (and SDs) on the Test of Morphological Structure, the Spelling Test, and the Test of Suffix Addition by grade level

	Test of Morphological Structure		Spelling test		Test of Suffix Addition
Grade	Base forms	Derived forms	Base forms	Derived forms	
4	30.8 (6.9)	26.3 (5.4)	24.9 (9.3)	14.6 (9.8)	16.0 (4.0)
6	35.2 (4.1)	31.9 (3.8)	34.2 (4.1)	26.0 (7.5)	17.9 (3.3)
8	39.4 (0.7)	35.7 (2.4)	38.2 (3.0)	34.4 (5.3)	21.0 (3.7)

Note: Maximum score for TMS and ST = 40.
Maximum score for TSA = 30.

The performances at the three grade levels were found to be significantly different for both the Base Forms, $F(2,62) = 18.99$, $p < .001$, and the Derived Forms, $F(2,62) = 26.37$, $p < .001$. Paired comparisons (Scheffé, $p < .05$) showed that for both the Base Forms and Derived Forms subtests the fourth grade was significantly different from the sixth grade and from the eighth grade, and the sixth grade was significantly different from the eighth grade. In the eighth grade the students' performance on both subtests was close to the ceiling level of the test.

Differences in the students' ability to generate the word forms on the TMS cannot be attributed to differences in word frequency or word length. The correlations between errors on the TMS and word frequency (taken from the norms of Carroll et al., 1971) were very low for the Base Forms, $r = .14$, $p = .20$, and for the Derived Forms, $r = -.08$, $p = .43$. Correlations between word length and errors were also very low both for Base Forms, $r = -.26$, $p = .02$, and for Derived Forms, $r = .01$, $p = .90$.

The Spelling Test (ST) was subjected to a similar analysis. The students' performances were scored on the basis of the number of words spelled correctly on each subtest, Base Forms and Derived Forms. Letters incorrectly or ambiguously formed were counted wrong. Where the legibility of a letter or word was questionable, one additional judge scored the word independently. This procedure effectively removed the few instances of uncertainty.

The increase in mean number of correct spellings on the ST, as shown in Table 1, is significant for both the Base Forms, $F(2,62) = 26.69$, $p < .001$, and the Derived Forms, $F(2,62) = 34.88$, $p < .001$. Paired comparisons of the group means (Scheffé, $p < .05$) indicate that on the Base Forms subtest the fourth graders differed significantly from the eighth graders, and the sixth graders differed significantly from the eighth graders. On the spelling of the Derived Forms the fourth graders differed significantly from the sixth and eighth graders, but the sixth graders were not significantly different from the eighth graders. The eighth graders' spelling of the base forms was practically at a ceiling level, although their spelling of the derived words was somewhat less proficient.

As would be expected from other investigations of spelling skills (see Cahen, Craun, & Johnson, 1971), the correlation between word length and spelling errors and the correlation between word frequency and spelling errors were low to moderate. For both the base words and the derived words, the correlation of word length with spelling errors was .49 ($p<.01$). The correlation of the frequency of base words with errors on base words was -.34, and the correlation of frequency of derived forms with errors on derived forms was -.37 ($p<.05$).

Developmental trends based on the relative difficulty of the TMS and ST subtests were also found. On both tests, the performances on the Base Forms subtests were significantly better than the performances on the Derived Forms subtests: for the ST, $t(64) = 13.23$, $p < .001$; for the TMS, $t(64) = 3.90$, $p < .001$. The superior performance on the Base Forms subtests suggests that the ability to extract the base word from its derived form is developed before the ability to generate the derived form from the base form. Similarly, the spelling of the base words appears to be mastered before the spelling of their derived counterparts.

Relationship Between Knowledge of Derivational Morphology and Spelling Ability

The second research question concerned the relationship between learning derivational morphology and learning to spell derived words. In order to determine the extent to which performance on the Base Forms and Derived Forms subtests of the TMS and ST accounted for variance in the performance at the three grade levels, a discriminant function analysis was carried out. This analysis generated one function that accounted for 94.8% of the variance (Wilks' Lambda 0.3680109 at a significance level of 0.0000). (A second function was not significant.) The standardized canonical coefficients of this function are as follows: TMS Derived Forms 0.82173; TMS, Base Forms -0.59577; ST, Derived Forms 0.89484; ST, Base Forms 0.01003. The particularly high loadings of this function are on the Derived Forms subtests of the TMS and ST, suggesting that knowledge of derived forms more strongly distinguished the three grade levels than knowledge of base forms. This function correctly predicted the grade level of 69.23% of the group.

A second method was used to investigate the sensitivity to morphological structure in spelling derived words. The students' spelling of each word pair, the base form and its derived counterpart, was tabulated. Performance on each pair was figured according to four possible patterns: both base and derived forms incorrect (e.g., equl and equlty), base correct but derived incorrect (e.g., begin but begginer for beginner), base incorrect but derived correct (e.g., expens for expense but expensive), and both base and derived correct (e.g., explain and explanation).

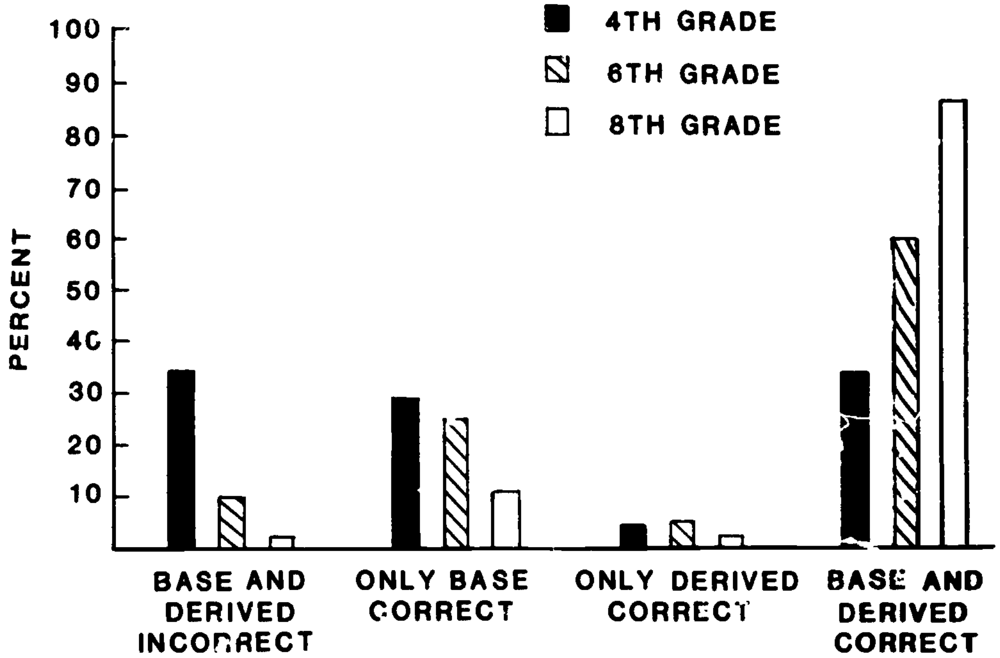


Figure 1. Comparison of correct and incorrect spellings of word pairs, base and derived forms, by grade level.

The results of this analysis, shown in Figure 1, give performance on pairs of words as a percentage of the total possible. One-way analysis of variance showed that the instances in which the students were able to spell both the base and derived words correctly increased significantly by grade level, F(2,62) = 34.51, $p < .001$. Paired comparisons of the group means

(Scheffé, $p < .05$) indicate that the fourth grade was significantly different from the sixth and eighth grades and that the sixth grade was significantly different from the eighth grade.

Furthermore, as is evident from an examination of Figure 1, generally speaking, correct spelling of the base word is a precondition for correct spelling of the derived word. While fourth- and sixth-grade students quite commonly misspelled the derived word but spelled the base word correctly, the reverse pattern was extremely uncommon--these students very seldom misspelled the base word and yet spelled the derived word correctly. The instances in which the base word was correct but the derived form was incorrect diminish markedly by the eighth grade--an indication of rapid learning of the spelling of derived forms by this grade level.

Performance on TMS and ST as a Reflection of Word Type

The third research question concerned the ruleful learning of derivational morphology and the extent to which such knowledge appears to be used in spelling derived words. To investigate this question, the experimental tests included four types of word relationships reflecting the kinds of transformations commonly found between base and derived words. As described earlier, these word types are No Change (NC), Orthographic Change (OC), Phonological change (PC) and Both Change (BC). The premise was that the more complex relationships, involving mastery of phonological and orthographic rules, would generate more errors than the more transparent relationships and would be mastered somewhat later. For the TMS, performances on both the Base Forms and Derived Forms subtests showed a pattern of performance by word type, in general reflecting more difficulty with the relationships that required phonological and/or both orthographic and phonological changes, as can be see in Figure 2.

In order to determine the extent to which the four word types (No Change, Orthographic Change, and so on) of the two TMS subtests (Base Forms and Derived Forms) accounted for variance in the performance at the three grade levels, a discriminant function analysis was carried out. This analysis generated one function that accounted for 89.23% of the variance (Wilks' Lambda 0.4420459 at a significance level of 0.0001). (The second function was not significant.) The standardized canonical coefficients of this function, shown in Table 2, indicate that the highest loading is on the Phonological Change word type of the Base Forms subtest, with moderate loadings on most of the remaining word types (the exceptions being the No Change and Orthographic Change word types of the Base Forms test). This function correctly predicted the grade level of 73.85% of the group.

As on the TMS, the students' spelling performance of the Derived Forms subtest of the ST was analyzed by word type. The question is whether students' success in spelling derived words is a reflection of the type of transformation between the base and derived form. (The words on the Base Forms subtest cannot be analyzed in the same way, since the dictated word is a single base morpheme, and there was nothing in the task to encourage the speller to consider morphological relationships.) An examination of the spelling errors on the four word types of the Derived Forms subtest of the ST (shown in Figure 2) indicated that the mean number of errors differed significantly by grade level: for NC, $F(2,62) = 24.30$, $p < .001$; for OC, $F(2,62) = 19.36$, $p < .001$; for PC, $F(2,62) = 28.30$, $p < .001$; for BC, $F(2,62) = 50.47$, p

Table 2

Standardized canonical discriminant function coefficients for the word types on the Base Forms and Derived Forms subtests, TMS

Base Forms	No Change	0.04983
	Orthographic Change	0.05298
	Phonological Change	0.63323
	Both Change	0.38875
Derived Forms	No Change	0.40498
	Orthographic Change	-0.26030
	Phonological Change	-0.19539
	Both Change	0.20406

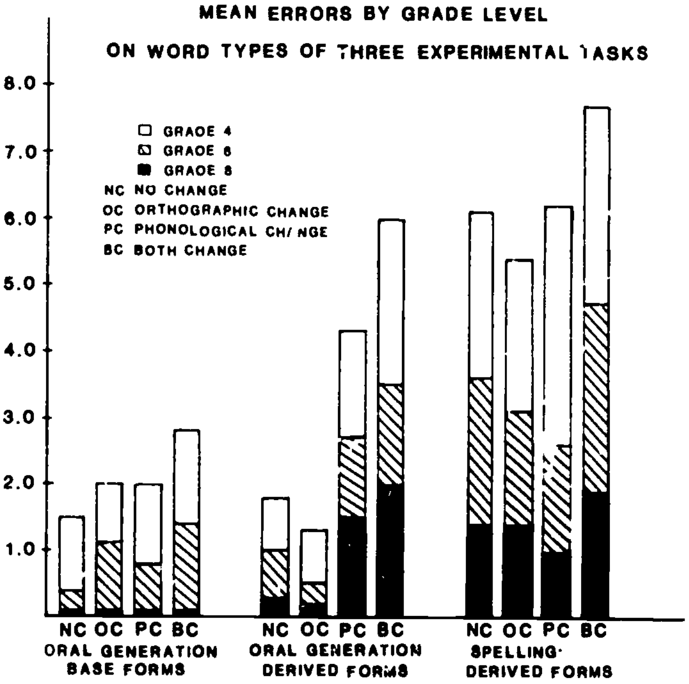


Figure 2. Mean errors by grade level on word types of three experimental tasks.

< .001. Paired comparisons (Scheffé, $p < .05$) indicated that for each word type the fourth grade differed significantly from the sixth and eighth grades, and the sixth grade differed significantly from the eighth grade.

The two Derived Forms subtests most directly assess knowledge of transformations between base and derived forms. Consequently, the two Derived Forms Subtests, TMS and ST, were analyzed in order to determine the extent to which the four word types (No Change, Orthographic Change, and so on) on the two Derived Forms subtests accounted for variance in performance at the three grade levels. A discriminant function analysis generated one function that accounted for 93.10% of the variance (Wilks' Lambda 0.2867654 at a significance level of 0.0000). (A second function was not significant.) The standardized canonical coefficients of this function, shown in Table 3, indicated high loadings on the Phonological Change word type of the TMS and the Both Change word type of the ST, suggesting that these were particularly important in accounting for the differences in performance by grade level. Both draw on knowledge of phonological rules, whether for generating or spelling derived forms. This function correctly predicted the grade level of 76.92% of the group.

Table 3

Standardized canonical discriminant function coefficients for the word types on Derived Forms subtests, TMS and ST

TMS	No Change	-0.07509
	Orthographic Change	0.09502
	Phonological Change	0.43693
	Both Change	0.04723
ST	No Change	-0.25244
	Orthographic Change	-0.14684
	Phonological Change	0.17305
	Both Change	0.92650

Analysis of Types of Errors on the TMS

Analysis of errors on the TMS provided further insight into the mastery of the rulefulness of derivational morphology. The decision to analyze the types of errors on the TMS (Derived Forms subtest) arose from the observation of patterns among the students' incorrect responses. The errors fell naturally into four categories: BASE ONLY for no response other than repetition of the base word (e.g., *sign* for *sign*); RULEFUL for ruleful but nonexistent words (e.g., *revisement* for *revision*); UNUSUAL for unusual but possible answers (e.g., *healing* instead of *health* in response to the item, "Heal. His sister was worried about his __."); and INAPPROPRIATE for nonruleful, nonexistent words (e.g., *consumeration* for *consumption*) or for existing words that were inappropriate answers (e.g., *glorify* instead of *glorious* in response to the item, "Glory. The view from the hill top was __."). Table 4 shows the average number of errors in each category made at the three grade levels.

Table 4

Mean errors (and SDs) on error types of the Derived Forms subtest, Test of Morphological Structure, by grade level

	Error types			
Grade	Base Only	Ruleful	Unusual	Inappropriate
4	4.7 (3.3)	2.5 (3.3)	2.3 (1.6)	3.5 (4.4)
6	3.4 (2.9)	1.0 (1.2)	1.5 (1.2)	1.8 (1.5)
8	1.0 (1.2)	0.7 (0.8)	1.5 (1.1)	0.8 (0.7)

Analysis of variance showed that the errors in three of the categories decreased significantly by grade level--the BASE ONLY errors, $F(2,62) = 10.66$, $p < .001$, the RULEFUL errors, $F(2,62) = 4.46$, $p < .05$, and the INAPPROPRIATE errors, $F(2,62) = 6.03$, $p < .01$. The UNUSUAL errors were not significantly different by grade level. Further examination was made of two of the errors types that seemed to be of particular interest--the RULEFUL errors and the UNUSUAL errors. Ninety-one RULEFUL errors (17% of the total) were made altogether--60.4% by fourth graders, 24.2 % by sixth graders, and 15.4% by eighth graders. Perhaps more revealing than the number of errors is the nature of the RULEFUL errors. Eighty-two percent of the errors were made on words that undergo a phonological change (with or without an accompanying orthographic change) in their derived forms (e.g., *revise* to *revision*). For 97% of these errors, the version given preserved the phonological identity of the base word (e.g., *revisement* instead of the target word, *revision*).

In addition, the students seemed to show a preference for certain suffixes in creating their RULEFUL errors. Most popular was *-ment* (accounting for 58% of the errors), followed by *-ance*, *-tion*, *-ness*, and *-less*. All of these suffixes were used to create a derived form without a phonological change in the base word. There is no reason to believe that the students were biased toward the use of any particular suffix by the other words on the test. For instance, the only test item with *-ment* as a suffix is the word *enjoyment*.

Unlike the other error categories, the UNUSUAL errors did not diminish significantly between the fourth and eighth grades. Analysis of the words on which such errors were made, as well as the kinds of responses given, indicate that the UNUSUAL errors occurred with the presentation of specific based words and sentences. Most (80%) of these errors occurred in generating the derived words from the following base words: *warm*, *deep*, *equal*, *active*, *consume*, and *heal*. Four of these six undergo a phonological change from the base to the target derived form, and yet most of the responses retained the sound of the base word (e.g., healing instead of the target word, health). The responses are unusual in that they suit the sentence but were not anticipated as likely answers, given the structure of the sentence. In this respect, the UNUSUAL answers are acceptable if not ideal and may reflect an inability to associate the base and derived word forms. We cannot infer that the students did not know the words health or consumption.

Performance on the Test of Suffix Addition

The students' mastery of orthographic "rules" was examined by means of the Test of Suffix Addition (TSA). The students' scores on the TSA consisted of the number of correct responses. In several instances, responses were written with a letter omitted, substituted, or placed in the wrong order in a part of the base word that was not essential to the suffixation. Such answers were not counted as incorrect if the suffix was correctly attached (e.g., beindish for biendish). However, where the miscopying of a base word in any way affected the addition of a suffix or where the suffix itself was misspelled, the answer was counted as wrong (e.g., pludding for pludying).

The students' performance on the TSA, shown in Table 1, indicates improvement in the ability to add suffixes to nonsense words, following the -y, -e, and doubling rules. The scores also show that even at the eighth-grade level, the students have not fully mastered the suffix addition rules. The difference between grade levels was significant, $F(2,62) = 10.25$, $p < .001$. Paired comparisons of the group means (Scheffé, $p < .05$) show that the fourth and sixth grades were significantly different from the eighth grade, but that the fourth grade was not significantly different from the sixth grade. The more pronounced growth appears to take place between the sixth and eighth grades.

Discussion

This study set out to investigate the knowledge of derivational morphology at the fourth-, sixth-, and eighth-grade levels and to investigate the extent to which this knowledge is reflected in the students' spelling of derived words. The results of the study have shown that students appear to learn a great deal about derivational morphology between the fourth and eighth grades. Their knowledge reflects varied levels of understanding of the underlying phonological rules and the orthographic rules that govern the transformations from base to derived forms. In addition, there are some indications that students learn to spell derived forms by reference to morphemic structure. Still, the spelling of derived forms lags behind the knowledge of these forms. Even by the eighth grades, students do not have a full mastery of the more complex transformations between base and derived forms or of the suffix addition rules.

Developmental Trends in the Learning of Derivational Morphology

Significant growth toward mastery was found on each of the three tasks that assessed morphological knowledge--the generation of base and derived forms and the spelling of the derived forms. The test results yield some indication of the order in which different skills are acquired. First, the ability to extract base forms from derived forms was mastered before the ability to generate derived forms from base forms. Second, the ability to spell base words was mastered before the ability to spell their derived counterparts. Third, the ability to produce the correct base and derived forms orally was generally superior at each grade level to the ability to spell the base and derived forms. Finally, application of the suffix addition rules was not fully mastered by the eighth grade, although the students' ability to apply these rules improved significantly between the sixth and eighth grade.

The task of extracting the base word (given the derived form and an appropriate sentence context) requires the ability to analyze morphemic structure, while the task of generating the correct derived form (given the base form and an appropriate sentence context) involves an awareness of the syntactic and semantic form suitable for a particular sentence context. This awareness, in turn, depends on a knowledge of the available and acceptable forms of a given word (such as equality instead of equalness). The students differed significantly in their proficiency on these two tasks--the Base Forms and Derived Forms subtests of the Test of Morphological Structure (TMS). It is evidently easier to analyze the morphemic structure of derived forms than it is to produce an appropriate derived form. While the two tasks differ in difficulty, the mean scores on both subtests increase significantly by grade level. Improvement on the Derived Forms subtest was particularly dramatic, as the fourth graders had a mean score of 26.3 correct (the maximum possible being 40), while the eighth graders had a mean score of 35.7 correct (see Table 1). The eighth graders approached the ceiling level on both subtests of the TMS, which gives an indication of the point at which students become competent at analyzing the morphemic structure of derived words and knowing the proper word forms, given words of this level of difficulty.

Since the words on the two subtests were chosen to be equally familiar, we can surmise that the particular source of difficulty in learning derivational morphology is less learning to analyze morphemic structure of derived words than learning appropriate derived word forms. One important aspect of this contrast may be the different demands each of the tasks makes on an individual. It is likely that production of a word form is more taxing than analysis of the structure of a given word. However, this general observation needs to be examined in regard to individual differences in performance. For individuals who have trouble understanding the morphemic structure of words (Wiig, Semel, & Crouse, 1973), the two tasks might be equally challenging.

Learning to Spell Derived Forms

Comparisons of the students' performances on the TMS and the Spelling Test (ST) confirm our expectation that spelling is a more difficult task than orally generating word forms. It is not surprising that skill in spelling derived forms appears to develop later than skill in generating derived forms.

Performances on the ST suggest that spelling derived forms draws on a knowledge of morphological relationships. When spelling performances on each word pair (the base and its derived form) were analyzed, mastery of the spelling of derived forms seemed to depend on initial learning of the spelling of the base forms. As Figure 1 shows, students very seldom spelled a derived form correctly when they spelled the base form incorrectly, whereas they quite commonly misspelled the derived form when they had spelled the base form correctly. It is unlikely that this pattern would be so pronounced if derived words were learned as unanalyzed whole words. In addition, there is a relatively small decrease in the percentage of the instances in which the base is correct but the derived form is incorrect (29% to 11%); in contrast, there is a very large increase in the percentage of instances in which both the base and derived forms are spelled correctly (33% to 87%). This suggests a rapid improvement in the ability to manage the "derived" part of the derived forms (including orthographic and phonological transformations), along with an improvement in the ability to spell the base forms.

Another indication of the use of morphemic analysis in spelling of derived forms comes from the students' performance on the Test of Suffix Addition. Since this test involves adding suffixes correctly to nonsense words, it requires explicit knowledge of specific suffix conventions (those governing the addition of suffixes to words ending in a silent e, in y, and in a single consonant). These suffix rules, which differ from the linguistic rules that govern phonological and orthographic transformations, are appropriately viewed as conventions of writing that govern the correct spelling of both inflected and derived forms. They are most likely learned by observation of the patterns of suffix addition in the orthography or by direct instruction in school. (The alternative would be memorization of the sequence of letters used to spell each derived word, an unwieldy system, given the large number of derived words the students are learning and can use in their writing.) The students' performances on this test show an improvement in the mastery of the three suffix rules, particularly between the sixth and eighth grades. (The fourth graders' performance was not significantly different from that of the sixth graders.) Still, even the eighth graders had not mastered the rules completely. As was anticipated, the learning of these suffix addition rules seems to take place later than the mastery of the morphological structure of words.

Ruleful Learning of Derivational Morphology

The words on the TMS represent four types of transformations between base and derived word forms. These are: NO CHANGE in the orthography and phonology (e.g., enjoy to enjoyment), ORTHOGRAPHIC CHANGE only (e.g., rely to reliable), PHONOLOGICAL CHANGE only (e.g., major to majority), and BOTH CHANGE, the orthography and the phonology (e.g., reduce to reduction). The NO CHANGE word type represents the most transparent relationship, while the BOTH CHANGE word type represents the most obscure relationship. Analysis of the test results suggests that the nature of the transformation between base and derived forms affected the accessibility of knowledge of morphological relatedness, as was expected (see Figure 2). On the Base Forms part of the TMS the NO CHANGE words had the fewest errors and the BOTH CHANGE words had the most, while PHONOLOGICAL CHANGE words and ORTHOGRAPHIC CHANGE words fall between these two extremes. On the Derived Forms subtest the PHONOLOGICAL CHANGE and BOTH CHANGE words gave much more difficulty than the NO CHANGE and ORTHOGRAPHIC CHANGE words.

A discriminant function analysis of the four word types on the two subtests of the TMS (Base Forms and Derived Forms) yielded one significant function that accounted for over 89% of the variance, indicating the power of these variables in distinguishing the students at the three grade levels. Contributing to the power of this function were all of the word types except the NO CHANGE and ORTHOGRAPHIC CHANGE types on the Base Forms subtest, possibly indicating that general mastery of the system of transformations distinguished the three grade levels.

Performance by word type was considered a particularly important indication of ruleful learning on the two Derived Forms subtests--the oral generation task (Derived Forms subtest of the TMS) and spelling (Derived Forms subtest of the ST). The four word types on these subtests were included in a discriminant function analysis. This analysis yielded one significant function that accounted for 93% of the variance. Of the standardized canonical coefficients, the heaviest loading was on the BOTH CHANGE word types of the ST, the second strongest contributor being the PHONOLOGICAL CHANGE word type of the TMS. These results suggest that mastery of both phonological and orthographic rules in spelling most strongly distinguishes the grade levels. Knowledge of the underlying phonological rule system also discriminates the three grade levels in performance on derived words, whether the task be oral generation or spelling.

Despite these findings, performance on the Derived Forms subtest of the Spelling Test shows that the distribution of errors by word type is relatively even, a pattern evident at all three grade levels (see Figure 2). There are several possible reasons for the modest effect by word type in spelling. First, the two tasks (oral generation and dictated spelling) are very different in one important respect. In generating derived forms, the student had no choice but to work with the morphemic structure of the word. However, in spelling the derived forms, the students were given the derived word by dictation, and so the task did not require them to deal with the word's morphemic structure. The fact that there is any consistency in the effects of word type by grade level suggests that some knowledge of orthographic and phonological transformations, at least, plays a role in the process of spelling derived forms. Second, spelling is a complex skill, offering many opportunities for error. Clearly, the difficulty of spelling a derived word is not simply a reflection of its word type.

Finally, analysis of the kinds of errors students made on the Derived Forms of the TMS gives additional support to the argument that the nature of the transformations between base and derived forms affects the ease of mastering morphological relationships. The two error types selected for detailed analysis (the RULEFUL errors and the UNUSUAL errors) were found to fall primarily on those words that belonged to the PHONOLOGICAL CHANGE or BOTH CHANGE word types. The students' most common error was a form of the word that retained the sound of the base word, whether the response was an actual word or a ruleful invention. This pattern suggests that the younger students know something about the system of forming derivatives but have not yet learned all of the appropriate phonological changes. In fact, a large proportion of their errors showed a resistance to making phonological changes in giving derived forms. The students often simply added one of the more common and familiar suffixes ("all-purpose" suffixes such as _-ment_) to the base word. For example, a number of students spontaneously invented the form _producement_, not knowing or not recognizing the morphological relationship of the correct response, _production_.

The RULEFUL and the UNUSUAL errors were in many respects quite similar; the UNUSUAL errors were differentiated primarily because they were existing English words, while the responses that made up the RULEFUL errors could be English words but for whatever reason are not--such is the complexity of derivational morphology. Thus, even the students who do not have a complete understanding of the complex transformations still understand some basic principles about how the system of derivational morphology works.

Instructional Implications

This study provides some evidence that sixth- and eighth-grade students draw on their understanding of the morphemic structure of the words to guide their spellings of derived words. First, there was a strong relationship between correct spelling of the base words and correct spelling of their derived counterparts. We might surmise that along with learning how to spell the base words, the students are acquiring morphological awareness--a sensitivity to word relationships and an inclination to use knowledge of morphological relationships in spelling. Second, the students demonstrated improved ability to apply the orthographic rules that govern suffix addition, indicating that general principles are learned and applied to the spelling of derivatives. Nonetheless, the spelling of derived words lagged behind mastery of the system of the transformations between base and derived forms. The test results suggest that although a student may demonstrate an understanding of morphemic structure when asked to analyze words, he or she may not put this knowledge to use on a dictated spelling test of derived words, particularly where there are phonological and orthographic transformations.

Since the students demonstrate some productive knowledge of derivational morphology, they have the potential, given suitable instruction, to develop an explicit awareness of the relationship between the word forms and their spellings. However, even the eighth graders still have not mastered the spelling of PHONOLOGICAL CHANGE and BOTH CHANGE derived words and the suffix addition rules. It seems likely that students in the fourth through eighth grades might benefit by spelling instruction that explicitly emphasizes morphological relationships and the principles that govern the addition of suffixes. One training study has been done that suggests the particular benefits of a morphemically-based spelling program (Robinson & Hesse, 1981). The seventh-grade students who received training in the morphemic structure of words showed more improvement than a control group in general spelling performance and in specific performance on morphemically complex words.

Poor spellers and learning-disabled students, who have been found to be deficient in their understanding of morphological rules (Wiig et al., 1973), might benefit particularly from intensive and explicit instruction in the morphemic structure of words. In a school system whose spelling program includes instruction in morphemic analysis and spelling rules, both good and poor spellers showed gradual improvement in their spelling of words with suffixes between the seventh, eighth, and ninth grades (Carlisle, 1984). However, the poor spellers continued to lag well behind their peers. They seem to need more intensive instruction over a longer period of time to make significant improvement in their ability to spell words with suffixes.

Explicit instruction in morphological relationships, including phonological and orthographic transformations, might enhance both the students' understanding of the structure of the language and their ability to spell derived words. Such instruction could commence at the fourth-grade level.

References

Barganz, R. A. (1971). The morphophonemic performance of good and poor readers (Technical Report #182). Madison, WI: Research and Development Center for Cognitive Learning.

Berko, J. (1958). The child's learning of English morphology. Word, 14, 150-177.

Cahen, L. S., Craun, M. J., & Johnson, S. K. (1971). Spelling difficulty--A survey of research. *Review of Educational Research*, *41*, 281-301.

Carlisle, J. F. (1984, March). *Mastery of the conventions of suffixation by good and poor spellers in grades seven through nine*. Paper presented at the New York Orton Dyslexia Society Conference, New York, NY.

Carroll, J. B., Davies, P., & Richman, B. (1971). *Word frequency book*. New York: American Heritage Publishing Company, Inc.

Chomsky, C. (1970). Reading, writing and phonology. *Harvard Educational Review*, *40*, 287-309.

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

Derwing, B. L. (1976). Morpheme recognition and the learning of rules for derivational morphology. *Canadian Journal of Linguistics*, *21*, 38-66.

Derwing, B. L., & Baker, W. J. (1979). Recent research on the acquisition of English morphology. In P. Fletcher & M. Garman (Eds.), *Language acquisition*. Cambridge: Cambridge University Press.

Fischer, F. W, Shankweiler, D., & Liberman, I. Y. (1985). Spelling proficiency and sensitivity to word structure. *Journal of Memory and Language*, *24*, 423-441.

Freyd, P., & Baron, J. (1982). Individual differences in acquisition of derivational morphology. *Journal of Verbal Learning and Verbal Behavior*, *21*, 282-295.

Frith, U. (1980). Unexpected spelling problems. In U. Frith (Ed.), *Cognitive processes in spelling*. New York: Academic Press.

Henderson, L. (1982). *Orthography and word recognition in reading*. New York: Academic Press.

Ingram, D. (1976). Aspects of phonological acquisition. In *Phonological disability in children*. London: Edward Arnold.

Jaeger, J. J. (1984). Assessing the psychological status of the vowel shift rule. *Journal of Psycholinguistic Research*, *13*, 13-36.

Jarvella, R. J., & Snodgrass, J. G. (1974). Seeing ring in rang and retain in retention: On recognizing stem morphemes in printed words. *Journal of Verbal Learning and Verbal Behavior*, *13*, 590-598.

Jastak, J. F., & Jastak, S. R. (1978). *Wide Range Achievement Test*. Wilmington, DE: Jastak Associates, Inc.

Klima, E. S. (1972). How alphabets might reflect language. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye*. Cambridge, MA: MIT Press.

LaSorte, D. M. (1980). *The differential effects of visual and auditory information in determining meaning of derived words at third and fifth grade levels*. Paper presented at the Annual Meeting of the College Reading Association, Baltimore, MD.

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.

Marino, J. L. (1979). *Children's use of phonetic, graphemic and morphophonemic cues in a spelling task*. Educational Resources Information Center, Document #ED188235.

Mattingly, I. G. (1980). Reading, linguistic awareness and language acquisition. *Haskins Laboratories Status Report on Speech Research*, *SR-61*, 135-160.

Moskowitz, B. A. (1973). On the status of vowel shift in English. In J. T. Moore (Ed.), *Cognitive development and acquisition of language*. New York: Academic Press.

Richardson, J. T. E. (1977). Lexical derivation. *Journal of Psycholinguistic Research, 6*, 319-335.

Robinson, J. W., & Hesse, K. D. (1981). A morphemically based spelling program's effect on spelling skill and spelling performance of seventh grade students. *Journal of Educational Research, 75*, 56-62.

Rubin, H. (1984). *An investigation of the development of morphological knowledge and its relationship to early spelling ability*. Unpublished doctoral dissertation, University of Connecticut.

Russell, P. (1972). *Morphology and English spelling*. Los Alamitos, CA: Southwest Regional Laboratory. Educational Resources Information Center, Document #ED108-228.

Schwartz, S., & Doehring, D. (1977). A developmental study of children's ability to acquire knowledge of spelling patterns. *Developmental Psychology, 13*, 419-420.

Selby, S. (1972). The development of morphological rules in children. *British Journal of Educational Psychology, 42*, 293-299.

Smith, P. T., & Sterling, C. M. (1982). Factors affecting the perceived morphemic structure of written words. *Journal of Verbal Learning and Verbal Behavior, 21*, 704-721.

Steinberg, D. D. (1973). Phonology, reading and Chomsky's optimal orthography. *Journal of Psycholinguistic Research, 2*, 239-258.

Sterling, C. M. (1983). Spelling errors in context. *British Journal of Psychology, 74*, 353-364.

Templeton, S. (1980). Spelling, phonology and the older student. In E. H. Henderson & J. W. Beers (Eds.), *Developmental and cognitive aspects of learning to spell: A reflection of word knowledge*. Newark, DE: International Reading Association.

Thorndike, E. L. (1941). *The teaching of English suffixes*. New York: Teachers College Press.

Venezky, R. L. (1970). *The structure of English orthography*. The Hague. Mouton.

Wiig, E. H., Semel, E. M., & Crouse, M. A. B. (1973). The use of English morphology by high-risk and learning disabled children. *Journal of Learning Disabilities, 7*, 59-67.

Appendix

Test of Morphological Structure

		Derived Forms Subtest:		Base Forms Subtest:	
		Given	(Target Response)	Given	(Target Response)
No Change	1.	warm	(warmth)	growth	(grow)
	2.	enjoy	(enjoyment)	employment	(employ)
	3.	appear	(appearance)	difference	(differ)
	4.	care	(careful)	fearful	(fear)
	5.	final	(finally)	usually	(usual)
	6.	profit	(profitable)	remarkable	(remark)
	7.	perform	(performance)	assistance	(assist)
	8.	humor	(humorous)	dangerous	(danger)
	9.	honest	(honesty)	royalty	(royal)
	10.	precise	(precisely)	extremely	(extreme)
Orthographic Change	1.	sun	(sunny)	foggy	(fog)
	2.	swim	(swimmer)	runner	(run)
	3.	begin	(beginner)	propeller	(propel)
	4.	endure	(endurance)	guidance	(guide)
	5.	active	(activity)	density	(dense)
	6.	adventure	(adventurous)	continuous	(continue)
	7.	expense	(expensive)	sensitive	(sense)
	8.	happy	(happiness)	emptiness	(empty)
	9.	glory	(glorious)	furious	(fury)
	10.	rely	(reliable)	variable	(vary)
Phonological Change	1.	equal	(equality)	humanity	(human)
	2.	original	(originality)	personality	(personal)
	3.	drama	(dramatic)	periodic	(period)
	4.	magic	(magician)	musician	(music)
	5.	protect	(protection)	election	(elect)
	6.	express	(expression)	discussion	(discuss)
	7.	electric	(electricity)	publicity	(public)
	8.	sign	(signal)	national	(nation)
	9.	major	(majority)	popularity	(popular)
	10.	heal	(health)	cleanly	(clean)
Both Change	1.	deep	(depth)	width	(wide)
	2.	type	(typical)	athletic	(athlete)
	3.	explain	(explanation)	combination	(combine)
	4.	produce	(production)	reduction	(reduce)
	5.	permit	(permission)	admission	(admit)
	6.	expand	(expansion)	extension	(extend)
	7.	absorb	(absorption)	description	(describe)
	8.	revise	(revision)	recognition	(recognize)
	9.	decide	(decision)	division	(divide)
	10.	consume	(consumption)	assumption	(assume)

RELATIONS AMONG REGULAR AND IRREGULAR, MORPHOLOGICALLY-RELATED WORDS IN THE LEXICON AS REVEALED BY REPETITION PRIMING*

Carol A. Fowler,† Shirley E. Napps,†† and Laurie B. Feldman†††

Abstract. Several experiments examined repetition priming among morphologically related words as a tool to study lexical organization. The first experiment replicated a finding by Stanners, Neiser, Hernon, and Hall (1979) that whereas inflected words prime their unaffixed morphological relatives as effectively as do the unaffixed forms themselves, derived words are effective, but weaker, primes. The experiment also suggested, however, that this difference in priming may have an episodic origin relating to the less formal similarity of derived than of inflected words to unaffixed morphological relatives. A second experiment reduced episodic contributions to priming and found equally effective priming of unaffixed words by themselves, by inflected relatives, and by derived relatives. Two additional experiments found strong priming among relatives sharing the spelling and pronunciation of the unaffixed stem morpheme, sharing spelling alone or sharing neither formal property exactly. Overall, results were similar with auditory and visual presentations. Interpretations that repetition priming reflects either repeated access to a common lexical entry or associative semantic priming are both rejected in favor of a lexical organization in which components of a word (e.g., a stem morpheme) may be shared among distinct words without the words themselves, in any sense, sharing a "lexical entry."

Words presented for lexical decision are more rapidly classified, and words presented under poor viewing or listening conditions are more readily reported, if they have been presented previously in the experimental setting than if they have not (e.g., Forbach, Stanners, & Hochhaus, 1974; Murrell & Morton, 1974; Scarborough, Cortese, & Scarborough, 1977). We will refer to this general outcome as "repetition priming." Morton (e.g., 1981) and Stanners, Neisser, Hernon, and Hall (1979) have interpreted repetition priming as a consequence of repeated access to a lexical entry. Other research has identified both episodic (Feustel, Shiffrin, & Salasoo, 1983; Jacoby & Dallas, 1981) and strategic (Forster & Davis, 1984; Oliphant, 1983) components to the priming effect as well.

---

*Memory & Cognition, 13, 241-255.

†Also Dartmouth College.

††Dartmouth College.

†††Also University of Delaware.

Acknowledgment. This research was supported by NIH Grant HD-01994 to Haskins Laboratories. We thank Vicki Hanson and George Wolford for their comments on the manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-82/83 (1985)]

The lexical interpretation is of particular interest in light of patterns of priming that are observed among morphologically-related words. Priming may occur in two forms that we refer to as "full" and "partial." Full priming is priming of one word by another that is as large, statistically, as priming of a word by itself. Partial priming is priming of one word by another that is present, statistically, but is significantly less than priming of a word by itself. Generally, the findings are that priming of a base word by regularly inflected morphological relatives is full, while priming by derived forms is partial (Stanners et al., 1979). Priming by irregularly affixed words may be partial (Stanners et al., 1979) or absent (Kempley & Morton, 1982).

Stanners et al. interpret full priming as evidence that stem forms and inflected relatives share a lexical entry; they interpret partial priming as evidence that stem forms and derived words are neighbors in the lexicon. This pattern of priming and its interpretation are appealing in supporting plausible roles for lexical entries in language use. One role has repetition priming as a by-product; a second role gives repetition priming its patterning.

In Morton's theory of the lexicon (1969, 1981), lexical entries are "logogens" which collect evidence for the occurrence in stimulation of the words they represent. Sufficient evidence, exceeding a logogen's threshold, causes the logogen to "fire." As one consequence of firing, the threshold is lowered temporarily so that less evidence is necessary for firing if the word is presented a second time. The threshold rises very slowly over time. Thresholds of frequent words are kept permanently lowered by the frequent recurrence of the words in stimulation. The frequency-sensitive thresholds of logogens explain repetition priming, but more usefully for language users, they prepare language users for perception of words most likely to occur in the environment. In this role, repetition priming is a by-product of the normal operation of the logogen system.

Arguably, this mechanism would work well if, as the repetition priming data suggest, the lexicon counted a stem morpheme and its regularly inflected, but not derived, forms as the same word. Unaffixed words and their inflected relatives are the same part of speech with essentially the same core meaning; in a sense they are the same word with the difference between them determined by the grammatical context in which the word appears. Consequently a common frequency-based expectancy is meaningful for classes of words differing only in inflectional affix. In contrast, unaffixed words and their derived relatives often are not the same part of speech, they need not be close in meaning (cf. Aronoff, 1976), and, consequently, a common frequency-based expectancy for unaffixed words and their derived relatives would not be meaningful.

The second role for a lexical entry may be in providing appropriate input to regular and productive phonological rules of the language. In generative phonology (Chomsky & Halle, 1968), a lexical entry includes just that phonological information about a word that is not predictable by rule, and hence that uniquely identifies a word. The phonological rules that are most productive and regular in English (and thus, perhaps, that are most likely to be learned by language users [cf.Berko, 1956; Ohala, 1974; Steinberg, 1973]) are rules of inflection. The finding that inflected words prime their stems fully, then, is consistent with a lexicon in which inflected words have no independent representation. Certain speech errors (for example, morpheme shifts and strandings [Garrett, 1980a, 1980b]) have been interpreted as supporting a similar conclusion; so have the speech patterns of some Broca's and jargon aphasics (see Butterworth, 1983, for a review of the relevant evidence).

Despite the consistent and plausible view of the lexicon provided by repetition-priming findings, we decided to investigate priming patterns further for two reasons. The first reason is that repetition effects are found in the memory literature (e.g., Light & Carter-Sobell, 1970) in which they are ascribed to episodic, not to lexical memory, and, episodic sources of priming are found using paradigms very similar to the repetition priming paradigms themselves (Feustel et al., 1983).

Moreover, it is not difficult to imagine how episodic influences might contribute to priming using the procedures of Morton or of Stanners et al. Subjects may explicitly recall having seen a word (or morphological relative) previously in the experiment, and in the procedure of Stanners et al., they may recall the response they made to it. This recollection may facilitate responding to a primed word. These episodic sources of priming are unlikely to exhaust the repetition priming that occurs (cf. Jacoby & Dallas, 1981); however, added to lexical sources of priming, they are likely to exaggerate the apparent loss in priming of an unaffixed form by a derived form as compared to its priming by an inflected form or by itself. This exaggerated difference would occur because derived forms generally are less formally or semantically similar to stem forms than are inflected forms (and, of course, than is a stem word to itself). Consequently, memory for a derived prime may be less likely to be cued during later presentation of the unaffixed word than memory for an uninflected prime or for the target word itself. Accordingly, full priming between a word and itself or between a word and an inflected variant may include both lexical and episodic sources of priming, whereas partial priming as between a derived prime and unaffixed target may include only lexical sources of priming.

Our second reason to explore further the patterning of repetition priming derives from questions raised about any repetition priming having a lexical rather than an episodic origin. The main question is whether repetition priming originating in the lexicon always reflects repeated access to a common lexical entry. In a recent review of the literature on word recognition, speech errors, and the speech and reading patterns of various language-disabled populations, Butterworth (1983) disputes the conclusion that lexical entries are common to unaffixed words and their affixed relatives in English. Instead, in his view, the bulk of evidence supports separate but associated entries for all words. If this interpretation is correct, then repetition priming may occur between separate entries in the lexicon. Our research investigates the distinction between shared lexical entries for morphological relatives and associated, but separate entries.

Our first experiment was designed to test for episodic sources of influence on repetition priming. Having found it, we take steps in later experiments to reduce or eliminate it and to reexamine the pattern of repetition priming among stems and regularly and irregularly inflected and derived morphological relatives. This patterning suggests hypotheses concerning the organization of morphologically related words in the lexicon.

Experiment 1

As an index of episodic priming, we chose to look at repetition priming on nonwords--both regular and irregular. The literature does not offer a clear indication of whether nonword repetition priming should be found using a lexical-decision paradigm. Forbach et al. (1974) report essentially no repetition priming among nonwords; however, Scarborough et al. (1977) report some priming of this type. Stanners et al. (1979) do not report their findings on nonwords.

In the present experiment, we examine repetition priming among words and nonwords under conditions replicating those in which Stanners et al. found full repetition priming of base forms by inflected morphological relatives and partial priming by derived forms.

Method

Subjects. Subjects were 25 Dartmouth College undergraduates who participated in the experiment for course credit. All were native speakers of English and had normal or corrected vision.

Stimulus materials. The stimuli used in the experiment were 48 English words and 48 nonwords. The words formed two groups. One group (Inflections Only) was presented both without suffixes, called "base" stimuli, and with inflectional suffixes, "s" and "ed." The second group (Derivations and Inflections) appeared as base stimuli, with the inflectional suffixes "s" and "ed," and with two derivational suffixes ("ment" and one of "er"/"or" or "able"/"ible"). Thus, within the second group, the effects of inflectional and derivational forms of the same word can be compared with each other. Words were chosen so that suffixation did not change the spelling or pronunciation of the base.

Nonwords formed three groups. Items in the first group (Nonword, Inflections Only) were created from real words having the same characteristics as the real words in the Inflections Only group. To form the nonwords, one or two letters in the real words were changed. The resulting nonwords were orthographically regular. These were presented both in a base form and with inflectional suffixes. Thus, they are the nonword counterparts of the first group of real words. The second group (Irregular, Inflections Only), consisted of ten irregular four-letter constructions and these were also presented both as base forms and with inflectional suffixes, "s" and "ed." Irregular nonwords were included in the study to provide an index of episodic priming in nonwords presumed not to have any form of representation in the lexicon. The third group of nonwords (Nonword Derivations and Inflections) were analogous to the second group of real words. They were orthographically regular and were presented as base forms, with inflectional suffixes, "s" and "ed," and with the derivational suffixes, "ment," "er"/"or" or "able"/"ible." The words used in the experiment and the words from which the 38 regular nonwords were formed were equated on average length and on mean and median frequency (Kučera & Francis, 1967). Real-word base forms are listed in Appendix A.

Five test orders were created, each one including the following priming conditions in equal numbers: (1) base as target with no prime (henceforth B1), (2) base as prime and base as target (BB; e.g., "manage"-"manage"), (3) inflection as prime and base as target (IB; e.g., "manages"-"manage"), and (4) derivation as prime and base as target (DB; e.g., "management"-"manage"). Across test orders, items appeared in identical serial positions, but the sequences differed in which version of each item served as a prime. For example, for the base word "manage," the forms "manage," "manages," "managed," "management," and "manager" served as primes in the different test sequences. In all sequences, the target was "manage". For items occurring only as base forms and inflections in the experiment, each inflected form (i.e., "s," "ed") occurred in two test sequences and the base form in one as primes. Inflections, derivations and base first occurrences were distributed proportionately over the five test sequences.

Subjects saw each morpheme only twice: once as a prime and once as a target. The average lag between the occurrence of a prime and the occurrence of its target was nine intervening trials; lags ranged from 6 to 12 and each lag was equally frequent among words and nonwords. Filler items were used as necessary to maintain appropriate lags. Each subject completed five blocks of 56 trials each, the first of which was a block of practice trials. Primes and targets were presented within one block.

Design. Five subjects were assigned to each of the five test orders. The independent variables were Priming Condition and Lexical Status (word, nonword). The main dependent variable was response time.

Procedure. Subjects were run individually. The experiment was run on a time-sharing computer interfaced with a Polytronics response timer. The stimuli were presented in upper case on a cathode ray tube. On each trial the following sequence of events occurred: (1) a fixation string of plus signs (++++++++) came on; (2) the terminal bell sounded 500 ms before the fixation mark went off; (3) a letter-string appeared as soon as the fixation mark disappeared, and remained on until the subject responded; (4) once the subject responded and the stimulus disappeared, the fixation mark returned and another trial began.

For each subject, the "K" key of the computer terminal was pressed with the right index finger for a word stimulus and the "D" key with the left index finger for a nonword stimulus. The keys were labeled with the symbols "W" and "NW" for "word" and "nonword" respectively. Subjects were informed that both accuracy and speed of responding were important, and that accuracy should be kept above 90% correct on each block of trials.

Between blocks of trials subjects were informed of their mean reaction times and proportions correct for the preceding block of trials. Blocks were initiated by the subject.

Results[1]

Errors and extreme reaction times (greater than 2000 ms or more than 2.5 standard deviations from the individual subject's or item's mean) were excluded from the analysis. This procedure excluded less than one percent of the responses. When a subject responded incorrectly to one member of a prime-target pair, both responses were excluded from the analyses. Table 1 presents mean response times and errors to base targets.

In all experiments, error rates will be reported in the appropriate tables. Analyses on the error rates will be reported only if they are significant.

Table 1

Mean Reaction Times for Words and Nonwords for the Various Prime-Target Conditions of Experiment 1

	B1	BB	IB	DB
	Words			
Inflections only	602	516(.10)	513(.12)	
Derivations and Inflections	552	499(.01)	506(.04)	525(.07)
	Nonwords			
Inflections only	689	654(.09)	648(.18)	
Irregular Inflections only	625	551(0)	585(.06)	
Derivations and Inflections	691	615(.16)	653(.13)	675(.13)

Note—Error rates are in parentheses.

One-way subject and item analyses were performed on response times to base words (conditions B1, BB, IB, and DB). Separate analyses were done on the 32 items appearing only in inflected and base forms (Inflections Only), and on the 16 items appearing in derived, inflected, and base forms (Derivations and Inflections). For the Inflections Only group of words, the effect of priming condition was significant (subjects: $F(2,40)=17.90$, $p<.001$; items: $F(2,62)=20.09$, $p<.001$). Scheffé's tests revealed that the significant main effect was due to the B1 condition differing from the BB and IB conditions (subjects: $F(2,40)=13.8$, $p<.001$; items: $F(2,62)=15.5$, $p<.001$). The difference between the BB and IB conditions was not significant.

An analogous analysis on the remaining 16 words revealed a similar outcome for inflections, but only a partial repetition effect for derivations. The main effect of priming condition was significant (subjects: $F(3,60)=6.17$, $p=.001$; items: $F(3,45)=4.87$, $p=.005$). Scheffé's tests showed that this effect was again due to the B1 condition differing from the BB and IB conditions (subjects: $F(3,60)=3.77$, $p=.015$; items: $F(3,45)=3.62$, $p=.02$). The BB and IB conditions did not differ from each other. In the DB condition, the mean response time did not differ from either B1 response time or BB and IB response times. These results are very similar in pattern to those of Experiments 1 and 3 of Stanners et al. (1979).

Similar analyses were performed on nonwords. Separate analyses were done on response times to the regular nonwords appearing only in inflected and base forms, the 16 regular nonwords appearing as derivations, inflections, and bases, and the 10 irregular nonwords. The effect of priming condition was marginally significant for the Nonword Inflections Only group in the subject analysis only (subjects: $F(2,40)=2.96$, $p=.06$; items: $F(2,42)=1.29$, $p=.28$).

Priming was significant for nonwords in the Nonword Derivations and Inflections in the subject analysis, $F(3,60)=5.55$, $p=.002$, and marginally significant in the item analysis, $F(3,45)=2.53$, $p=.06$. Scheffé's tests showed the significance of the former effect to be due to the difference between the B1 and BB conditions, $F(3,60)=4.95$, $p=.004$. Irregular Inflections Only reached significance in both analyses (subjects: $F(2,40)=7.24$, $p=.002$; items: $F(2,18)=4.25$, $p=.03$). These effects were also attributable, as shown by Scheffe's tests, to the difference between the B1 and BB conditions (subjects: $F(2,40)=7.17$, $p=.002$; items: $F(2,18)=4.16$, $p=.03$).

Discussion

The real-word results of Experiment 1 replicate the results of Stanners et al. (1979). Significant repetition priming of targets occurred for both base and inflection primes; derivations also primed their bases, but marginally. Stanners et al. interpreted the corresponding partial repetition effect they found to signify that derivations (and irregular inflections) have lexical entries separate from their base forms.

The nonword results obtained in the present experiment weaken this explanation. Presumably, nonword repetition effects, particularly those among irregular nonwords, are largely episodic rather than lexical in origin. That is, they occur because subjects remember explicitly having seen the letter strings previously in the experiment and, perhaps, having made a particular response to them. If episodic priming affects response time to nonwords, it may also contribute to repetition priming in words.[2] If it does, then partial repetition effects may reflect decreased episodic priming; the less the target in a prime-target pair looks like the prime, the less it reminds the subject of the prime.

Considerations such as these led us to repeat this study with an attempt to reduce the effects of episodic memory on subject responses.

Experiment 2

In an effort to reduce episodic contributions to the repetition effect, we extended the lag between primes and targets of a base morpheme from an average of 9 items in Experiment 1 to 48 items in Experiments 2a and 2b.

In addition, we instituted a control for unequal practice on primes and targets. Necessarily, the prime of a morpheme appears earlier in the test sequence than its target. Consequently, subjects are less practiced on the average when they respond to primes than when they respond to targets. Possibly, such an effect, too, contributes to priming.

Any asymmetrical practice of this sort can be eliminated by a procedure first used by Forbach et al. (1974) but not used subsequently by Stanners et al. (1979). In the control procedure, the test sequence of words is partitioned into blocks. In the first block of test trials, only fillers and primes of morphemes are presented. In the second block, primes from the first block are repeated as targets interleaved with a new set of primes. In subsequent blocks except the last, new primes are interleaved with repetitions of primes (now targets) from the previous block. In the final block, targets are interleaved with fillers. For most analyses, data from the first and last blocks are eliminated. In this way, analyses are restricted to comparisons of responses to primes and targets made at comparable levels of practice. Across subjects, words are counterbalanced so that every morpheme occurs equally often in each block as a prime and target.

Two experiments were run using these changes in procedure. In Experiment 2a, primes were inflections and base forms. In Experiment 2b, they were derivations and base forms.

Method

Subjects. Subjects were 72 students from the same pool as in Experiment 1. Thirty-six subjects participated in each of Experiments 2a and b. This gave three replications of all of the test-order conditions in each experiment.

Stimulus materials: Experiment 2a. Stimuli were 48 words and 48 nonwords matched in length to the words. Each word, a verb, appeared as a prime in each of three forms: uninflected (base), inflected with "s," and inflected with "ed." An individual subject saw each morpheme only twice: once as a prime and once as a target. In every instance, inflected forms preserved both the spelling and the pronunciation of the base. Targets were invariably base forms. Real-word base forms appear in Appendix B.

Nonwords were 24 orthographically regular and 24 irregular nonwords. Each nonword appeared as a prime in three forms: uninflected, inflected with "s," and inflected with "ed." As for the words, targets of nonwords were invariably "base" forms.

Experiment 2b. Stimuli were 48 words and 48 nonwords matched to the words in length. Each word and nonword appeared as a prime in each of three forms: unaffixed, and affixed with two of several derivational affixes (two of "ment," "less," "er," "ly," "ness," "able," "ful"). As in Experiment 2a, each subject saw a given morpheme only twice. All nonwords were orthographically regular. Real-word bases are listed in Appendix B.

Test orders. The test sequences consisted of one practice block and five test blocks each 48 trials in length. For purposes of counterbalancing, the 96 letter strings in the test list were partitioned into four sets. Each set included 12 words and 12 nonwords (four bases, eight affixed items). A Latin Square was used to order the sets into four different sequences. For example, the Latin Square ordering 1-2-4-3 created a test sequence in which items in the first set constituted the primes of the first block of the test sequence and the target repetitions of the second block. Primes in block one were interleaved with filler items. Items in set 2 provided the primes in block 2 of the test sequence and the target repetitions in block 3. Items in set 4 provided the primes in the third block and the target repetitions in the fourth block. Finally, items in set 3 provided the primes of block 4 and the target repetitions in the final block. In the last block, set 3 items were interleaved with fillers. The ordering procedure created a lag of 48 items between the prime and target of a morpheme.

The four test orders, each based on one row of the Latin Square, appeared in three versions. The versions were identical except for the affixes on their first occurring morphemes. For example, matched to a test order in which say, "pushes" appeared as a prime in block 2, were two test orders in which "push" and "pushed" respectively appeared as prime in block 2. In each experiment, one third of the priming items were bases; one third were words affixed with "s" in Experiment 2a and one of the derivational affixes in Experiment 2b; the remaining third included words affixed with "ed" in Experiment 2a and words with other derivational affixes in Experiment 2b. This gave 12 different test orders for each of Experiments 2a and b.

Design. Subjects experienced all levels of the independent variable, priming condition. The primary dependent measure was response time.

Procedure. The procedure was identical to that in Experiment 1.

Results[1]

Response times and errors were analyzed as in Experiment 1. Table 2 presents response times and errors to base words and nonwords in blocks 2-4 from Experiments 2a and 2b.

Response times to base words in Experiment 2a differ as a function of their priming condition (subject analysis: $F(2,70) = 54.73$, $p<.001$; item analysis: $F(2,94) = 46.59$, $p<.001$). Scheffé's tests reveal no significant difference on the subjects analysis in response times to BB and IB words, $F(2,70) = 2.56$, $p = .08$). However, the difference does reach significance on the item analysis, $F(2,94) = 3.32$, $p = .04$). The 78 ms difference between conditions B1 and IB is significant (subjects: $F(2,70) = 29.45$, $p < .001$; items: $F(2,94) = 72.9$, $p < .001$). Statistically, then, the repetition effects of inflected words on bases are full.

Table 2

Response Times to Words and Nonwords in Experiments 2A (Left) and 2B (Right)

B1	BB	IB	B1	BB	DB
			Words		
611	510(.07)	533(.07)	585	543(.05)	538(.03)
			Nonwords		
643	627(.14)	645(.10)	715	717(.17)	730(.16)

Note—Error rates are in parentheses.

Analysis of the response times to base and derived forms of Experiment 2b gives a similar picture (subjects analysis: $F(2,70) = 9.03$, $p < .001$; items analysis: $F(2,94) = 8.24$, $p < .001$).

Table 2 also provides the comparable findings on nonwords. Repetition priming among nonwords was statistically absent in both studies (Experiment 2a: Subjects analysis: $F(2,70) = 2.02$, $p = .14$; item analysis: $F(2,94) =$

1.22. In Experiment 2b both F values are less than 1.) Thus, there is no apparent episodic repetition priming on nonwords in these experiments in which a 48-item lag is used and in which the control procedure for practice is implemented.[3] (In all subsequent experiments, nonword effects will be reported in tables, but not described in the text unless they involve statistically significant effects.)

Discussion

Having significantly reduced evidence of episodic priming in nonword stimuli, we obtain a somewhat different picture of repetition priming in derived and inflected words than we obtained in Experiment 1 and than Stanners et al. (1979) report. In particular, we find that repetition priming of a base form by a derivational relative is as strong as priming by an inflectional relative. Moreover the priming is statistically and, in Experiment 2b, numerically, full.

These findings invite one of two salient interpretations. One, compatible with Butterworth's assessment of the lexicon (1983) is that repetition priming occurs among separate lexical entries in the lexicon; it is not a consequence (except in the case of exact repetitions) of repeated access to a common lexical entry. A second is that it does reflect repeated access, but a lexical entry is more inclusive than had previously been suggested by repetition-priming findings. As we will suggest in the General Discussion, the substantive differences between these views are smaller, in light of constraints on their realizations imposed by our findings, than the statements of them suggest.

In the next experiment, we further examine the kinds of morphologically-related words that are strongly associated, or that share a lexical entry. We do so by examining priming of an unaffixed form by affixed morphological relatives that do not necessarily preserve the spelling or pronunciation of the stem morpheme in the unaffixed form. In addition, we examine two types of derivationally affixed words.

Possibly, the derived words we used in Experiment 2 were special and gave rise to unrepresentatively strong priming. Chomsky and Halle (1968) identify two types of suffix in English. One, neutral affixes, includes inflections and some derivations; these affixes do not affect pronunciation of the stem morphemes to which they are attached. In contrast, nonneutral (derivational) affixes do affect the stem morpheme's pronunciation (e.g., "sign"-"signal"). In Chomsky and Halle's theory, neutral affixes are separated from the stem morpheme by a word boundary, which prevents application of phonological rules over extents spanning stem and affix. Nonneutral affixes are separated from the stem by lesser, morpheme boundaries that do not prohibit application of phonological rules over the whole domain of stem plus affix. In our Experiment 2b, affixes were neutral derivational affixes. Perhaps it is not surprising that neutrally-affixed derivations were as effective primes as inflected words.

In Experiment 3, we compare priming of unaffixed words by morphological relatives that do or do not share pronunciation or spelling of the stem morpheme with the unaffixed form. This allows us to compare priming by irregular inflected words and regular morphological relatives (cf. Kempley & Morton, 1982). In addition, in a post hoc analysis, we look specifically at neutrally and nonneutrally affixed derivations and compare their priming effectiveness.

Experiment 3

In the present experiment, we examine priming by morphologically-related forms in which either the pronunciation or the spelling and pronunciation of the common morpheme is not shared by prime and target forms. The experiment had two purposes in addition to the one just described of examining priming by derived forms with nonneutral affixes. A related purpose was to reexamine effects of decreases in formal overlap (and hence, for English, in regularity) between morphologically-related primes and targets on repetition priming. Stanners et al. had found that priming of a base by an affixed form decreases as formal overlap between the affixed and unaffixed words decreases. Kempley and Morton (1982) found no priming between irregular and regular forms when the words were presented auditorily. The present study was designed to reexamine these priming effects under the conditions we have developed which reduce episodic priming effects. Possibly, in the earlier studies, the differences in priming across conditions was episodic in origin; targets following formally identical or similar primes cued memory for the primes while dissimilar targets did not. A final purpose of the experiment was to separate effects of orthographic and phonological overlap between prime and target on the magnitude of priming.

Method.

Subjects. Thirty-six students participated in Experiment 3a and 24 different students in Experiment 3b. All came from the same subject pool used previously.

Stimulus materials. Two sets of twenty-four word triads were devised. In one set, the "Sound Only" set, each triad included one base form and two affixed forms; one affixed form preserved the spelling and pronunciation of the unaffixed form (henceforth the "NC" or "no change" form) and one preserved only the spelling (henceforth the "C" or "changed" form). A sample triad is "heal," "healer," and "health." (In six items, a silent "e" in the base morpheme was deleted in an affixed form.) The second set, the "Sound and Spelling" set, also consisted of triads including an unaffixed form and two affixed words. In this set, one affixed word shared both spelling and pronunciation of the base morpheme with the unaffixed word (the "NC" form for this set) while the other affixed word shared neither spelling nor pronunciation with the unaffixed word (the "C" form). An example is "clear," "clearly," "clarity." In both sets, words in the third category were, with few exceptions, irregular forms.

Because Experiment 2 showed no difference in priming by inflected and derived forms that shared spelling and pronunciation with the unaffixed form, we felt justified in mixing the two types of affixed forms in our new lists. However, approximately equal numbers of derived forms and equal numbers of inflected forms occurred in the Sound Only and Sound and Spelling triads, and there were sufficient numbers of pairs of neutrally-affixed forms and nonneutrally-affixed forms that they could be examined separately in a post-hoc analysis.

Phonological overlap between unaffixed and affixed words was matched across Sound Only and Sound and Spelling lists by counting each vowel, consonant, or stress change as one change and matching number of changes across the two lists. In addition, an effort was made to match type of change (vowel, consonant or stress) as closely as possible. Our final experiment (Experiment 4b) is an auditory lexical decision experiment using these materials, which shows that our matching efforts were successful.

Unaffixed words in the Sound Only and Sound and Spelling lists were matched in length and frequency (Kučera & Francis, 1967). Similarly, the two different types of affixed forms were matched in length and frequency within and across the two lists. Appendix C lists the word triads in the two stimulus sets.

We created triads of nonwords from triads of words that might have appeared as word stimuli in the experiment. They were made into nonwords by changing one or two letters, while preserving their orthographic regularity. Forty-eight nonword triads were created in this way.

From the sets of words and nonwords, three basic stimulus lists were created. Each base morpheme appeared twice in each list, once as a prime and once as a target. The lists differed in respect to which version of the morpheme (unaffixed, affixed with no sound or spelling change, affixed with a change) appeared as the prime. The target was always the unaffixed form. In each list there were sixteen of each type of prime. Half of each set of 16 items was from each of the two sets of stimulus words. There were sixteen of each type of nonword prime.

The stimulus lists were organized exactly as in Experiments 2a and 2b. As in those experiments, four versions of each basic list were created so that, across subjects, each prime occurred equally often in the first four blocks of stimuli. Each stimulus list was preceded by a practice list of 24 words and 24 nonwords randomly ordered.

Procedure. The experiment was run twice. The second experiment (3b) was identical to the first (3a) except that the stimuli were presented under degraded viewing conditions (by turning down the contrast on the CRT screen) in an effort to slow response times and thereby, perhaps, magnify the very small departures from full repetition priming we observed in Experiment 3a. This manipulation had no effect on the pattern of reaction times we observed; therefore, we present both outcomes together.

The procedure and instructions to the subjects were identical to those used in Experiment 2.

Design. Subjects participated at all levels of the two independent variables, Stimulus Set (Sound Only, Sound and Spelling) and Priming Condition (B1, BB, NC ["no-change/base"--that is, a base primed by an affixed word in which the sound and spelling of the unaffixed base morpheme is preserved], CB ["changed-form/base"--that is a base primed by an affixed word in which the base pronunciation or spelling and pronunciation is changed from the unaffixed version]). The major dependent measure is response time.

Results[1]

Extreme response times were deleted from the data as described for Experiment 1. Results for word and nonword stimuli are presented in Table 3 collapsed over the factor Stimulus Set. Separate two-way repeated measures analyses of variance with factors Priming Condition (B1, BB, NCB, CB) and Stimulus Set (Sound Only, Sound and Spelling) were performed on the outcomes of the two experiments using subjects as a random effect. Separate items analyses were also run with one within-groups factor (Priming Condition) and one between-groups factor (Stimulus Set). In Experiment 3a, the effect of Priming Condition reached significance in both subjects and items analyses (subjects: *F*(3,105) = 20.82, *p* < .001; items: *F*(3,138) = 12.81, *p* < .001). The effect of Stimulus Set was significant in the subjects analysis, with response times faster in the Sound Only condition, but was nonsignificant in the items analysis (subjects: *F*(1,35) = 12.59, *p* < .001; items: *F*(1,46) = 2.44, *p* = .12). The interaction did not approach significance in either analysis (both *F*s < 1).

Table 3

Response Times in Experiments 3A and 3B

	B1	BB	NCB	CB
	Words			
Experiment 3A	623	558(.05)	575(.09)	584(.06)
Experiment 3B	673	590(.05)	612(.06)	621(.09)
eutral and nonneutral derivations	669	579	586	601
	Nonwords			
Experiment 3A	760	748(.14)	758(.15)	746(.13)
Experiment 3B	788	777(.18)	779(.17)	783(.18)

Note--Error rates are in parentheses.

The effect of prime type is due primarily to the difference between the response to an unaffixed prime and its occurrence as a target following any of the three primes (subjects: *F*(3,105) = 17.75, *p* < .001; items: *F*(3,138) = 10.80, *p* < .001). Among the prime conditions, the difference in the effect of an unaffixed prime (BB) as compared to the effects of the other primes (NCB, CB) reaches significance in the subjects analysis, but not in the items analysis (subjects: *F*(3,105) = 2.80, *p* = .04; items: *F*(3,138) = 1.73, *p* = .16). The additional effect of sharing or not sharing spelling or pronunciation with the base (that is, the difference between 575 and 584) is not significant.

We performed additional analyses on the data of Experiment 3a having removed the six items from the Sound Only condition in which presence and absence respectively of a silent "e" distinguished the base and affixed forms. Removing these items had no effect on the outcome of the experiment. The effect of prime condition remained highly significant; neither the effect of Stimulus Set nor the interaction was significant.

A major finding of Experiment 3a is that priming by affixed forms is nearly full. The priming by NCB forms replicates the outcome of Experiments 2a and 2b. Overall in the present experiment, 17 ms less priming occurs when the prime differs from the target in being affixed, but shares sound and spelling with the prime as compared to priming by the unaffixed form itself. CB forms reduce priming by an additional 9 ms.

We ran Experiment 3b to ask whether, by slowing response times, we could magnify the small differences we observed between the BB, NCB, and CB conditions. Our manipulation, reducing the contrast on the CRT screen, slowed response time overall by 42 ms. The slowing was significant in an items analysis ($F(1,92) = 8.23$, $p = .006$), but not in the subjects analysis (F1,58) = 2.17, $p = .14$). There were no interactions involving the factor Experiment in the overall analysis and, in the analysis of Experiment 3b, there was no increase in the magnitude of the separation of BB and NCB times on the one hand or NCB and CB times on the other. Statistical analysis of the response times in Experiment 3b provided an identical pattern of significant effects to the pattern observed in Experiment 3a.

In Experiment 3b, error proportions were .04, .06, and .09 on BB, NCB, and CB items, respectively. This was significant (subjects: $F(2,46) = 6.07$, $p = .005$; items: $F(2.92) = 6.96$, $p = .002$).

A final analysis examined neutrally- and nonneutrally-affixed derivations separately from the irregular inflected forms that were included in the stimulus sets. The purpose of the analysis was to answer the question raised by the finding in Experiment 2 that derivations as well as inflections fully primed their base forms. The question raised was whether this finding is limited to neutrally-affixed derivations, which preserve the pronunciation of the base morpheme.

Eight Sound Only and ten Sound and Spelling triads permitted a comparison of priming by NC neutrally-affixed derivations and by C nonneutrally-affixed derivations. These 18 items were subjected to a one-way analysis of variance with the single factor Prime Condition (B1, BB, NCB, and CB). The analysis collapsed over the nonsignificant factor, Stimulus Set, and across Experiments 3a and b. Only the items analysis was performed. As Table 3 reveals, the pattern of means mirrors very closely that of the overall analysis. The pattern of significant and nonsignificant differences is also the same as in the overall analysis. Thus, the overall effect of Priming Condition is significant, $F(3,51) = 10.19$, $p < .001$). Moreover, the three affixed primes differed from the B1 condition both separately and as a group (overall $F(3,51) = 9.70$, $p < .001$); they did not differ from each other (all Fs less than one). This implies no substantial difference between neutrally- and nonneutrally-affixed derivations in their ability to prime an unaffixed morphological relative.

Discussion

The major outcome of the present study is that there is essentially no loss in repetition priming when the orthographic or phonological representations of affixed primes and morphologically-related targets do not fully overlap. Because we found no effect on priming of differences in form between prime and target, we could not separate effects of spelling and sound differences as intended. Experiment 4 will address that issue once again. We did find a suggestion, significant in the subjects analysis only, of a small loss in priming when an affixed prime precedes a base target as compared to exact repetition priming, but there is no significant additional loss when the affixed form differs in sound or sound and spelling from the base. This shift, too, was a shift from regularly-affixed words to largely irregular forms. Thus, we found no loss in priming between regularly-affixed forms and their irregular morphological relatives.[4] Accordingly, we conclude that, however repetition priming effects are explained--as repeated access to a common lexical entry or as priming among strongly associated but distinct entries or in some other way--the relationships of irregular and regular, derived, inflected and unaffixed forms must be explained in fundamentally the same way.

Experiment 4

We designed the final experiment with two main purposes in mind. One was to compare priming in the auditory and visual modalities. In Morton's logogen model, each logogen has paired auditory and visual inputs (Morton, 1981). That is, a word has a logogen (in the model's most recent version, an "output logogen," but not an "input logogen") in common whether it is auditorily or visually presented. This idea is supported by findings of some cross-modal repetition priming (Kirsner, Milech, & Standen, 1983). However, whereas in Experiments 3a and 3b we found strong priming of visually-presented unaffixed words by irregular morphological relatives, Kempley and Morton (1982) found no priming between auditorily-presented unaffixed words and irregular, inflected morphological relatives. Kempley and Morton used different stimuli than we did and a different paradigm with longer lags between prime and target. Consequently a variety of reasons for this difference are tenable. In the present study, we use common word sets and a common paradigm to compare priming in the two modalities directly.

Our second purpose was to examine priming when affixed words appear as targets in the repetition-priming paradigm. This allows us to address two questions, one theoretical and one methodological. The first question concerns the organization of morphological relatives in the lexicon. One possibility is that all morphologically-related words are uniformly related to each other in the lexicon. Other possibilities can be imagined as well, however. One may be developed by analogy from a theory of lexical organization in Serbo-Croatian, a highly inflected language (Lukatela, Gligorijević, Kostić & Turvey, 1980). In that so-called "satellite-entries" theory, a particular inflected form, the nominative, rather than the root morpheme, is proposed as the hub of an array of associated morphologically-related words (satellites). Inflected words other than the nominative are associated to the nominative form but not (or less strongly) to each other. In this organization, the nominative should prime and be primed by other morphologically-related affixed forms more effectively than the affixed forms prime each other. In English, the unaffixed base form is the most likely counterpart to the nominative in Serbo-Croatian. If English has an analogous organization, then the unaffixed word should prime and be primed by affixed forms more effectively than affixed forms prime each other. Our experiment is designed to discriminate between these views by examining priming of affixed words by unaffixed and other affixed morphological relatives.

The methodological question concerns the possibility that the patterns of priming that we obtain using our paradigm are largely products of a paradigm-specific strategy by which subjects predict the target given the prime. Forster and Davis (1984) and Oliphant (1983) have shown that repetition priming is severely diminished (and absent in Oliphant's study) if subjects are unaware that words are repeated in the experiment. In the work of Forster and Davis, some subjects are made unaware of the repetitions because the prime is masked. Repetition priming is small, short-lived and, in at least one respect (absence or presence of a frequency-by-priming interaction), qualitatively different in pattern from priming observed when subjects are aware of the prime.

In other research (Napps, in preparation), one of us has also found a reduction in the magnitude of repetition priming when the proportion of targets in the experiment is only .06 of all stimulus items. Nonetheless, even under these conditions, significant priming is found using the Sound and Spelling stimuli of Experiment 3 out to the longest lag examined in that experiment (10 intervening items). Napps' findings in this study and in others using low proportions of repeated items suggest that the priming we obtain with a high proportion of related items does not create the appearance of relations among morphological relatives that are unrelated in the lexicon. Rather, they enhance effects of existing relations.

To further address the question whether our priming reflects lexical organization, or instead reflects predictability of the target given the prime, we designed Experiment 4 to reduce the subjects' ability to make useful predictions. In Experiments 1-3, targets were always unaffixed words. Accordingly, given a prime, subjects could guess the identity of the target word that would appear some 50 items later in the next block of stimuli. In Experiment 4, targets were less predictable than in earlier experiments because they were one of several possible affixed morphological relatives of primes.

As a second assessment of the role of prediction, we provide a separate analysis of repetition priming effects on the very first block of the experiment in which repetitions occur, and thus before subjects have an opportunity to develop a strategy of guessing targets from primes.[5]

<u>Methods</u>

<u>Subjects</u>. Subjects were 72 students from the same subject pool used previously. Thirty-six students participated in each of Experiments 4a and 4b. All subjects had normal hearing in Experiment 4b.

<u>Stimulus materials</u>. The materials were those used in Experiment 3, with one exception. In the test lists, the NC affixed form replaced the unaffixed form in all positions in which it occurred as a target. This yielded priming conditions NC1 (first occurring affixed item), NCNC (affixed word primed by itself), BNC (affixed item primed by the unaffixed form), CNC (affixed item primed by an affixed morphological item that does not preserve the pronunciation or the spelling and pronunciation of the unaffixed morpheme).

For Experiment 4b, stimulus items were recorded onto audio tape by a female native speaker of English (CAF). These productions were sampled by computer at 10 kHz. This enabled the same token of each NC prime or target item to be used in all conditions. The test orders were recorded on one channel of an audio tape. Tone bursts were recorded on the second channel of the tape for purposes of collecting response times. The tone bursts were synchronized to the onsets of acoustic energy of each stimulus item in the test order. Therefore response times include word duration (or as much of the word as occurred before the subject made his or her button-press response). That stimulus words have different durations is unimportant in the repetition priming procedure because critical comparisons involve response times made to the same items across different priming conditions. Stimulus items were recorded onto audio tape with a three-second inter-stimulus interval.

Only three test lists were used in Experiment 4b as compared to the 12 used in Experiments 2, 3 and 4a. The three lists had the same order of stimulus items but differed in respect to which of the three prime types occurred with each target item. It was infeasible to include the additional test orders needed to counterbalance the block in which each stimulus item appeared as prime and target.

Procedure. The procedure for Experiment 4a was identical to that for the previous experiments.

In Experiment 4b, subjects listened over headphones to binaural presentations of the test list. A New England Digital Able 40 minicomputer monitored the second tape channel for the tone bursts and started a millisecond clock when one was detected. The clock was read and a response and response time were stored when subjects pressed the labeled "word" or "nonword" button on the computer-terminal keyboard. If a response was not made within 2.5 seconds following stimulus presentation, the computer stopped the tape recorder and printed, "Please make a response" on a CRT screen facing the subject. Receipt of the button-press response restarted the tape recorder. The tape recorder was also stopped between blocks as subjects received feedback on their mean response times and accuracies for the block. Subjects initiated successive blocks by hitting a key on the terminal keyboard.

Design. In both experiments, subjects participated at all levels of the independent variables, Priming Condition (NC1, NCNC, BNC, CNC) and Stimulus Set (Sound Only, Sound and Spelling). The major dependent measure was response time.

Results[1]

Errors and extreme response times were eliminated from the analysis as in the earlier experiments. Table 4 provides the mean response times and errors for Experiments 4a and 4b.

Separate two-way repeated-measures analyses of variance were performed on the response times of Experiment 4a using subjects and items as random factors. The independent variables were Prime Condition (NC1, NCNC, BNC, CNC) and Stimulus Set (Sound Only, Sound and Spelling). In both analyses, the effects of Prime Condition (subjects: $F(3,105) = 14.79$, $p < .001$; items: $F(3,138) = 16.46$, $p < .001$) and the interaction (subjects: $F(3,105) = 4.29$, $p = .007$; items: $F(3,138) = 3.00$, $p = .03$) were significant. Scheffé's tests performed on the two stimulus sets separately show that, for the Sound Only condition, all three primed conditions differ from the unprimed condition and

Table 4

Mean Response Times in Experiments 4A (Visual) and 4B (Auditory)

	NC1	NCNC	BNC	CNC
	Words			
Experiment 4A				
Sound only	633	571(.05)	585(.04)	580(.07)
Sound and spelling	687	574(.07)	591(.07)	646(.08)
Experiment 4B				
Sound only	796	734(.09)	770(.07)	780(.11)
Sound and spelling	807	734(.05)	762(.06)	772(.10)
	Nonwords			
Experiment 4A	761	757(.13)	771(.14)	768(.11)
Experiment 4B	861	868(.15)	862(.16)	863(.18)

Note—Error rates are in parentheses.

do not differ from each other. For the Sound and Spelling condition, however, whereas the B and NC primes were effective, the C prime did not lead to response times significantly faster than the no-prime condition.

With two exceptions, the outcome of Experment 4b was very similar to that of Experiment 4a. In the analysis of response times to auditorily presented targets, only the effect of Prime Condition was significant (subjects: $F(3,105) = 34.90$, $p < .001$; items: $F(3,138) = 13.46$, $p < .001$). Neither the main effect of Stimulus Set nor the interaction approacned significance. The nonsignificant interaction contrasts with the outcome of Experiment 4a. The absence of an interaction between Stimulus Set and Priming Condition with auditory presentation is not surprising in view of the fact that in Experiment 4a the interaction could be ascribed to the presence or absence of spelling differences between prime and target. The loss of the interaction indicates that we succeeded in matching the Stimulus Sets along other relevant dimensions.

Scheffé's tests on the effect of prime condition showed that all three primed conditions had shorter response times than the unprimed condition. In addition, however, the exact repetition condition differed significantly from the other priming conditions on both subjects and items analyses. This statistically partial priming is the second contrast with the outcome of Experiment 4a.

In view of the apparent effect of changing spelling between affixed primes and targets in Experiment 4a only, we compared the outcomes of the visual and auditory experiments explicitly. We transformed response times to difference scores by subtracting response times in the BNC condition from those in the CNC condition separately for the Sound Only and Sound and Spelling stimulus sets. This provides an estimate of the effects of changing pronunciation alone (Sound Only words) or of changing both pronunciation and spelling (Sound and Spelling words) between prime and target with visual and auditory presentation. We performed analyses of variance on the difference scores with factors Experiment and Stimulus Set. The effect of Stimulus Set (subjects: $F(1,70) = 4.67$, $p = .03$; items: $F(1,92) = 3.49$, $p = .06$) and the interaction (subjects: $F(1,70) = 4.12$, $p = .03$; items: $F(1,92) = 3.18$, $p = .07$) were significant in the subjects analysis and marginally significant in the items analyses. Planned comparisons on the interaction in the subjects analysis showed that the effect of a spelling difference was greater with visual than auditory presentation ($F(1,70) = 5.10$, $p = .02$); the difference between the modalities of presentation on the effect of pronunciation alone (Sound Only) was nonsignificant ($F < 1$).

One more analysis of the data from each experiment was performed. To ask whether a subject's ability to guess the target from the prime accounts for priming effects, we examined primes in the first test block and their repeated targets or morphologically-related targets in the second block in Experiments 4a and 4b.

In Experiment 4a, across subjects, all items appeared as primes in the first block and as targets in the second. In Experiment 4b, this counterbalancing was infeasible; therefore, just one fourth of the items in each condition appeared as primes and targets in the first two blocks.

Restricting our analysis to the primes in the first test block and their targets in the second, in Experiment 4a, the effects of Priming Condition are highly significant in both subjects and items analyses (subject: $F(3,105) = 8.89$, $p < .001$; item: $F(3,138) = 9.44$, $p < .001$). The effect of Stimulus Set (Sound Only, Sound and Spelling) was significant in the items analysis only; the interaction did not approach significance in either analysis. Means in the four priming conditions, NC1, NCNC, BNC, and CNC were 684, 567, 607, and 622 collapsed over stimulus sets. These times conform closely to means computed over all blocks presented in Table 4. A planned comparison of means in the NC1 (unprimed) and CNC (primed by an irregular form) conditions was significant (subject: $F(1,105) = 7.23$, $p = .008$; item: $F(1,138) = 7.41$, $p = .007$), confirming that priming among regular and irregular affixed forms is present even when subjects are not aware that primes or their morphological relatives will be presented later in the experiment.

The same analysis performed on the first two blocks of trials in Experiment 4b gave essentially the same outcome. In that set of analyses, the effect of Priming Condition was significant (subject: $F(3,105) = 11.82$, $p < .001$; item: $F(3,138) = 6.76$, $p = .001$). No other factors were significant. Means were 7[illegible]7, 695, 753, and 740 for NC1, NCNC, BNC, and CNC priming conditions, respectively. A planned comparison of the conditions NC1 and CNC was significant (subject: $F(1,105) = 9.16$, $p < .001$; item: $F(1,138) = 7.98$, $p = .001$).

The reaction-time means and the pattern of significant effects in these restricted analyses conform closely to those obtained in the overall analyses. Thus, they confirm that repetition priming in the lexical decision paradigm does not require a strategy of predicting targets from primes as the primes are presented.

Discussion

We designed Experiments 4a and 4b to address three questions. The first was whether the logogen model, with its paired acoustic and visual input logogens, was tenable, particularly in light of our findings in Experiment 3 as compared to those of Kempley and Morton (1982). In Experiment 3, we found that visually-presented irregular words do prime their unaffixed relatives fully. In contrast, Kempley and Morton (1982) found that auditorily-presented unaffixed words and their irregular inflected relatives do not prime each other. In the present study, we found very similar priming in the two modalities.

A second question was whether we would find evidence of asymmetrical relations among morphological relatives as researchers have found for Serbo-Croatian (Lukatela et al., 1980). The experiment failed to support an idea that morphological relatives have a satellite organization, with the unaffixed base word as the center of the satellite. Instead, with one exception, all relationships among morphological relatives appeared strong.

We did obtain one outcome suggesting both a difference between auditorily- and visually-presented words in the lexicon and suggestive of a satellite organization among orthographically-represented words. We found that, with visual presentation, whereas base words are primed essentially fully by affixed morphological relatives not sharing either the spelling or the pronunciation of the shared morpheme (Experiment 3), affixed targets that preserve the spelling and pronunciation of the unaffixed morpheme are not (Experiment 4a). This loss in priming apparently can be ascribed to the spelling difference between the affixed forms since an analogous effect was not obtained in the auditory version of the experiment (Experiment 4b). Further evidence will be needed to determine whether this single outcome suggestive of different organizations for phonetic and orthographic forms of words is found reliably.

A final question addressed by the experiments was whether our procedure creates priming effects by inviting subjects to generate candidate targets when primes are presented. We answered this question in the negative based on two sources of evidence. First, priming occurs over lags of nearly 50 items even when the target is not highly predictable from the prime. More convincing, perhaps, is the significant priming in the first two blocks of test trials in which subjects would have no reason to adopt a guessing strategy. These analyses yielded mean response times and patterns of significant effects remarkably similar to those of the overall analyses. In particular, priming even by irregular forms remained strong in analyses of both visually- and auditorily-presented words. Therefore, we ascribe the difference in outcome between our studies and that of Kempley and Morton either to differences in the items used or to a longer time lag between prime and target in the experiment by Kempley and Morton (1982). The latter appears more likely. Kempley and Morton used inflected forms only, and, if there is a difference in strength of priming at all between inflected and derived forms, priming by inflected forms should be stronger.

General Discussion

Our major findings can be summarized as follows. We found that losses in priming from full to partial or less, when exact repetition priming is compared with priming by morphological relatives, may be ascribed at least in part to episodic contributions to repetition priming that are larger the more similar the prime and target. By reducing the contribution of these sources of repetition priming, we find strong priming--statistically full in most cases--among inflected, derived and unaffixed words, and between regular and irregular words, with either auditory or visual presentation. Accordingly, if repetition priming is interpreted as reflecting lexical organization as we assume, then our findings eliminate a theory of lexical organization in which regular inflected forms, but not derived forms or irregular inflections, share a lexical entry with the base. Correspondingly, they eliminate a theory in which the domain of a lexical entry is just those words that can be generated by productive, grammatical rules of affixation (see Butterworth, 1983, for a similar conclusion).

Our findings invite either of two extreme interpretations previously contrasted in the literature (e.g., Butterworth, 1983). One is that full repetition priming (after Stanners et al., 1979) or full and partial priming (after Murrell and Morton, 1974) reflect a lexical entry shared by primes and targets. Therefore, they signal that inflected, derived, regular and irregular morphological relatives share a lexical entry. This interpretation offers a way of capturing the large differences in longevity that have been found between repetition priming and semantic priming in the literature (cf. Henderson, 1984). Whereas we have found priming even when nearly 50 items intervene between prime and target, in studies of semantic priming, priming is absent by a lag of 1 or 2 items (Dannenbring & Briand, 1982; Davelaar & Coltheart, 1975; Gough, Alford, & Holley-Wilcox, 1981; Meyer, Schvaneveldt, & Ruddy, 1972; see also Henderson, 1984, for a direct comparison of semantic and repetition priming).

An unappealing consequence of adopting this interpretation, however, is that the concept of lexical entry is severely weakened. Entries that are as encompassing as our findings imply lack any obvious utility for the language user. The entries cannot serve as input to regular rules of affixation. Indeed, rather than consisting of the stem morpheme, affixed by rule, each entry perhaps must be considered a cluster of tightly associated affixed and unaffixed morphological relatives--a conceptualization not very distinct from the second interpretation we will consider. A second unattractive property of the present interpretation is that each entry cannot be associated necessarily with any semantic information at all that is common to words within the domain of the entry (cf. Aronoff, 1976) or to any one syntactic class. Moreover, if the entries are logogens, they do not keep an accurate frequency-based expectancy for all words within the domain of the entry.

An alternative interpretation questions whether semantic and repetition priming are, in fact, qualitatively distinct. Possibly, morphologically-related words that prime each other over very long lags are distinct words in the lexicon that are strongly related semantically. If so, then, there are no grounds for using the priming effects as a basis for inferring sharing of lexical entries. One advantage of this hypothesis is that just one mechanism, not two, is required to account for priming. A second advantage is that language users are not presumed to have lexical entries that encompass syntactically and semantically diverse morphological relatives.

Along with other researchers (e.g., Henderson, 1984; Morton, 1981), however, we are skeptical that morphological priming is exhaustively semantic. For one thing, researchers attempt to use words with the strongest associations or the maximum semantic relatedness when they test for semantic priming; nevertheless semantic priming does not approach the longevity of repetition priming under comparable conditions. Second, derived words tend to drift semantically after they are coined so that their meaning is not a simple compositional function of the meaning of the stem plus that of the affix (Aronoff, 1976); therefore, derived words tend to be less semantically related to morphological relatives than are inflected words. However, we obtain equally strong priming from words of both types.

In any case, it may not be necessary to choose between a view that repetition priming reflects repeated access to an entry and one that it reflects associations among words in the lexicon. A third perspective on the lexicon may capture the best features of both of these views. The perspective that we propose is derived from recent network models of the lexicon (e.g., Dell, 1980, 1984; McClelland & Rumelhart, 1981; Stemberger, 1982), in particular Dell's model, which is designed to produce speech and, in so doing, to generate natural slips of the tongue. Dell's model provides a more useful source than the more obviously related model by McClelland and Rumelhart (1981), designed to generate aspects of word-recognition behavior, because Dell's model includes a required representation of morphological structure. His model has not been extended to orthographic representations of words, but there are no principled barriers to doing so.

In Dell's network model, the lexicon is a hierarchy of levels of representation including words, morphemes, syllables, syllable constituents, phonemes, and phonetic features. Words such as "swimmer" and "swimming" have distinct word representations (called "nodes") but connect to a common stem-morpheme node and from there to common syllable and phoneme nodes for the shared stem morpheme. Word nodes also have connections to semantic memory, where, presumably, "swimmer" and "swimming" connect to common and to distinct concepts. A word such as "swift" has distinct word, morpheme and syllable nodes from "swimmer," but some common phonemes. Finally, a word such as "drown" is unconnected to "swimmer" and its constituents at any level in the lexicon, but shares concepts with it in semantic memory.

The structure of the model is well-suited, in general, to explain our pattern of findings. It gives morphological relatives closer ties to each other (other things equal) than to other words in the lexicon; yet it does so without either requiring morphological relatives to share a common word node or treating morphological relations as semantic. Moreover, it can explain why we and others (Kempley & Morton, 1982; Murrell & Morton, 1974; Stanners et al., 1979) consistently find numerically or even statistically weaker priming when prime and target are not exactly the same word as when they are.

One difficulty with the model, however, is that it does not allow irregular words such as "heal" and "health" to share a morpheme node as it must to explain our priming in Experiments 3 and 4. It is prevented from doing so because the syllable structure and phonemic constituents of a word are elaborated at hierarchical levels leading from the morpheme nodes, thereby requiring that morphemes sharing a node have the same pronunciation. The model could be adjusted by having the syllable level and the levels below it connect directly to the word nodes and not to the morpheme level. Morphological structure, then, would be a hierarchical level independent of levels of phonological structure. This kind of separation may have independent motivation from theories of metrical structure in linguistics (e.g., Selkirk, 1980). However, it remains to determine whether Dell's model, so modified, would produce natural patterns of speech errors involving morphological structure.

Although the structure of the network model just outlined provides an interesting alternative to both views of the lexicon usually contrasted in the repetition priming literature, the processing assumptions of a network model cannot handle repetition priming at the lags over which we observe it. In Dell's model, nodes at each hierarchical level are connected by bidirectional excitatory lines of association. Activation of a node is progressively incremented as activation spreads from it to its associated nodes and back again. To prevent every node in the lexicon from being activated eventually, activation of a node is shut down once the relevant unit has been output by the system (in Dell's model, once a phoneme or word has been spoken). For a variety of reasons, activation does tend to rebound after a node's activation has been shut down; this promotes perseveration errors in speech (for example [from Dell, 1980]: "to the bank to pick up some money"--"to the bank to pick up some bank"), and it may explain repetition priming of the magnitude and longevity observed by Forster and Davis (1984) and by Napps (in preparation) when subjects are unaware of repetitions in the experiment. However, activation lasting for 48 subsequent items (or two days as Scarborough et al., 1977, have observed) would have disastrous consequences for the model's normal operations. Evidently, priming of the longevity we observe is strategic; possibly, it can be seen in the context of the model as strategic maintenance of activation of a node previously activated by stimulus input. This strategic activation would play no role in ordinary speech and reading, but can be exploited as we have done to strengthen repetition priming processes that reveal the organization of words in the lexicon.

References

Aronoff, M. (1976). Word formation in generative grammar. Cambridge, MA: MIT Press.

Berko, J. (1956). The child's learning of English morphology. Word, 14, 50-77.

Butterworth, B. (1983). Lexical representations. In B. Butterworth (Ed.), Language production, II. London: Academic Press.

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

Dannenbring, G., & Briand, K. (1982). Semantic priming and the word repetition effect in a lexical decision task. Canadian Journal of Psychology, 36, 435-444.

Davelaar, E., & Coltheart, M. (1975). Effects of interpolated items on the association effect in lexical decision tasks. Bulletin of the Psychonomic Society, 6, 269-272.

Dell, G. (1980). Phonological and lexical encoding in sentence production: An analysis of naturally occurring and experimentally elicited speech errors. Doctoral dissertation, University of Toronto.

Dell, G. (1984). A spreading activation theory of retrieval in sentence production. Cognitive Science Technical Report, University of Rochester.

Feustel, T., Shiffrin, R., & 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.

Garrett, M. (1980a). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production I*. London: Academic Press.

Garrett, M. (1980b). The limits of accommodation: Arguments for independent processing levels in sentence production. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand*. New York: Academic Press.

Gough, P., Alford, J., & Holley-Wilcox, P. (1981). Words and context. In O. Tzeng & H. Singer (Eds.), *Perception of print*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Henderson, L. (1984). Morphemic structure and lexical access. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance, X*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Jacoby, L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 306-340.

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., Milech, D., & Standen, P. (1983). Common and modality-specific processes in the mental lexicon. *Memory & Cognition*, *11*, 621-630.

Kučera, H.,& Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Light, L., & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *9*, 1-11.

Lukatela, G., Gligorijević, B., Kostić, A. & Turvey, M.T. (1980). Representation of inflected nouns in the internal lexicon. *Memory & Cognition*, *8*, 415-423.

McClelland, J., & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: Part 1. Account of basic findings. *Psychological Review*, *88*, 375-407.

Meyer, D., Schvaneveldt, R., & Ruddy, M. (1972). *Activation of lexical memory*. Paper presented at the meeting of the Psychonomics Society.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, *76*, 165-178.

Morton, J. (1981). The status of information processing models of language. In H. Longuet-Higgins, J. Lyons, & D. Broadbent (Eds.), *The psychological mechanisms of language*. London: The Royal Society and the British Association.

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology*, *102*, 963-968.

Napps, S. (in preparation). *Morphological, semantic and formal relations among words and the organization of the mental lexicon*. Doctoral dissertation, Dartmouth College.

Napps, S., & Fowler, C. (1983). *Orthographic organization of lexical forms*. Paper presented at the meeting of the Eastern Psychological Association.

Napps, S., Fowler, C., & Feldman, L. (1984). *Effects of orthography and phonology on morphological organization of the lexicon*. Paper presented at the meeting of the Eastern Psychological Association.

Ohala, J. (1974). Experimental historical phonology. In J. Anderson and C. Jones (Eds.), Historical linguistics, II: Theoretical description in phonology. Amsterdam: North Holland.

Oliphant, G. (1983). Repetition and recency effects in word recognition. Australian Journal of Psychology, 35, 393-403.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance, 3, 1-17.

Selkirk, L. (1980). The role of prosodic categories in English word stress. Linguistic Inquiry, 11, 563-695.

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.

Steinberg, D. (1973). Phonology, reading and Chomsky's optimal orthography. Journal of Psycholinguistic Research, 2, 239-258.

Stemberger, J. (1982). The lexicon in a model of language production. Doctoral dissertation, Department of Linguistics, University of California at San Diego.

Footnotes

[1] The response times we report here differ in absolute value from times reported in Napps and Fowler (1983) and Napps, Fowler, and Feldman (1984). The procedures we use to present stimuli to the computer-terminal screen and to collect response times create constant errors. The present response times have been adjusted for those constant errors. The times in the earlier presentations were unadjusted. The adjustments do not affect the size in ms of priming effects.

[2] One outcome in Experiment 1 is at apparent odds with the conclusion that some of the priming on words is episodic. We would expect the IB condition to give rise to slightly longer response times than the exact-repetition (BB) condition. A small difference (7 ms) in the appropriate direction does occur in the Inflections and Derivations stimuli, but it is reversed (-3 ms) in the Inflections Only stimuli. However, looking across experiments of our own and of others in the literature in which a comparison can be made, in six of eight comparisons IB exceeds BB. The differences are always small and usually nonsignificant. That they are small is not surprising, however. Inflections and base forms are orthographically and phonologically very similar. Moreover, it is possible that more lexical information than simply word forms constitutes an episodic trace in our experiments. Much of that additional information will be the same for inflections and base forms.

[3] Another assessment of episodic priming in the present experiment may be obtained by comparing response times to words in Experiments 2a and b with corresponding times in Experiment 1. Although the mean response times may differ across the experiments due to differences in lag, in subjects, and, in Experiment 2b, stimulus materials, there will be no loss in episodic priming in the B1 condition of Experiments 2a and b as compared to Experiment 1 and therefore B1 response times should be closest across the experiments. For the same reason, DB conditions should show little change when episodic priming is eliminated. The BB and IB conditions should show a relative increase in response time, however. With just one notable exception, the outcomes of Experiments 2a and b are consistent with the predictions. Conditions B1 in

Experiment 2a and B1 and DB in Experiment 2b show less change from their corresponding times in Experiment 1 than (respectively) conditions IB in Experiment 2a and BB in Experiment 2b. The exceptional point is the response time to the BB condition of Experiment 2a, which is 6 ms faster than in Experiment 1 rather than being slower as it should be. In light of the supportive evidence provided by the other conditions and, particularly, by the outcome on nonwords, we ascribe the one inconsistency to sampling error or perhaps to a floor on response times in the BB condition of Experiment 1.

[4] We should acknowledge, however, that although the difference does not approach significance, irregular forms prime base forms numerically less than do regular forms. More generally in our research using repetition priming, in nearly all instances in which the prime and target are not identical and repetition priming is statistically full, it is numerically less than full. This is the case in most comparisons in Experiments 1-4; similar trends can be seen in the findings of Stanners et al. (1979) and Morton (Morton, 1981; Murrell & Morton, 1974).

[5] This analysis assesses priming when subjects have no reason to attempt to predict a future target from a prime. It remains true, however, that by the time the targets are first presented, subjects have been exposed to a large number of morphologically-complex words. Possibly, this promotes a tendency to think of morphological relatives of primes. If it does, and thus if the set of activated relatives can remain activated over lags of 50 items or more, this finding in itself would be interesting. Moreover, it would require an explanation in terms of activation within the lexicon, most probably. Both the capacity and the temporal span of any temporary buffer would be exceeded by the memory demands required to activate a set of morphological relatives for each of the two-dozen primes presented within a 48 item span. In any case, research by Napps (in preparation) showing repetition priming with very low proportions of morphological relatives, however, suggests that this cannot be a major source of repetition-priming effects.

Appendix A

Experiment 1 Base Words

enlarge*
yell
knead
call
settle*
discern*
sign
retain
weld
gash
walk
push
punish*
agree*
toss
talk
enchant
spell
roll
command*
disagree*
invent
amend
pronounce*
replace*
gather
pick
adjust*
attain*
laugh
mow
rest
list
govern*
equip
pull
paw
wander
develop*
deploy
wait
enjoy*
latch
manage*
blink
paint
cook
detach*

*Used with both inflectional and derivational affixes.

Appendix B

Experiment 2A and 2B Base Words

Experiment 2A Base Words

enlarge
yell
knead
call
settle
discern
sigh
retain
weld
gash
walk
push
punish
agree
toss
talk
enchant
spell
roll
command
disagree
invent
amend
pronounce

replace
gather
pick
adjust
attain
laugh
mow
rest
list
govern
equip
pull
paw
wander
develop
deploy
wait
enjoy
latch
manage
blink
paint
cook
detach

Experiment 2B Base Words

develop
manage
govern
assess
announce
employ
enjoy
punish
detach
disagree
move
enforce
though
fruit
help
power
harm
care
rest
color
fear
use
hope
thank

bright
soft
eager
dark
weak
stiff
vague
complete
direct
appropriate
close
glad
bold
blind
fond
hard
awkward
fresh
rich
like
separate
vivid
fair
polite

Appendix C

Word Trials Used in Experiments 3 and 4

		Sound Only	Sound and Spelling		
Base	No Change	Change	Base	No Change	Change
heal	healer	health	creep	creepy	crept
sign	signing	signal	defend	defendant	defensive
dream	dreamer	dreamt	sleep	sleepy	slept
edit	editor	edition	repel	repellent	repulsive
deal	dealing	dealt	speak	speaker	spoke
reside	resided	residence	decide	decided	decisive
produce	producible	productive	assume	assumed	assumption
confide	confided	confidence	sweep	sweeping	swept
inhibit	inhibiting	inhibition	invade	invader	invasion
electric	electrical	electrician	persuade	persuade	persuasive
bomb	bomber	bombard	space	spaced	spatial
mean	meaning	meant	forget	forgetful	forgotten
grade	grading	graduate	sing	singer	sang
medic	medical	medicine	fall	falling	fell
compare	comparative	comparable	induce	inducement	induction
extreme	extremist	extremely	collide	collided	collision
create	creative	creature	describe	described	description
drive	driver	driven	concede	conceded	concession
rise	riser	risen	deep	deeply	depth
revise	revising	revision	picture	picturesque	pictorial
music	musical	musician	propel	propeller	propulsion
lyric	lyrical	lyricism	wise	wisely	wisdom
critic	critical	criticize	clear	clearly	clarify
clean	cleaner	cleanse	forgive	forgiveness	forgave

GRAMMATICAL PRIMING OF INFLECTED NOUNS BY THE GENDER OF POSSESSIVE ADJECTIVES*

M. Gurjanov,† G. Lukatela,† Katerina Lukatela,† M. Savić,† and M. T. Turvey††

Abstract. Two experiments examined the effect on lexical decision times for inflected Serbo-Croatian nouns when the nouns were preceded by possessive adjectives (my, your, our). For any given pairing the possessive adjective and the noun always agreed in number (singular) and case (nominative) but only agreed half of the time in gender (masculine or feminine). Lexical decisions were faster when the noun targets were of the same gender as their primes. This gender congruency/incongruency effect was shown to hold whether the inflections of the adjective and noun were the same (as is the case for typical Serbo-Croatian nouns) or different (as is the case for atypical Serbo-Croatian nouns). The results are discussed in terms of a post-lexical influence of grammatical processing on the recognition of individual words.

"Priming" is a term referring to the influence of one stimulus upon the processing of another. Most experiments on "priming" with word stimuli have considered words that are associatively related. Where lexical decision latency is the measure of processing time it has been shown that processing is more rapid when a word is preceded by an associate compared to when it is preceded by a nonassociate (Lupker, 1984). Recently other relations between and among words have come under examination. Goodman, McClelland, and Gibbs (1981) asked whether lexical decision is speeded when successive words are instances of word types that ordinarily occur in succession in the language. These authors found that when two words were syntactically legal (e.g., men swear) the target word was responded to slightly but significantly faster than when the two words were syntactically illegal (e.g., whose swear). Wright and Garrett (1984) used fragments of sentences as the priming context. They found that the grammatical structure of the incomplete sentence affected the lexical decision time for a target word that followed it. For example, modal verb contexts preceding main verb targets and preposition contexts preceding noun

---

**Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1985, 11, 692-701.

†University of Belgrade.

††Also University of Connecticut.

Acknowledgment. This research was supported in part by NICHD Grants HD-08495 and HD-01994 to the University of Belgrade and Haskins Laboratories, respectively. Dr. M. Gurjanov died in August, 1983. "Gura" has been a central figure in this research program from its inception in 1973. Without his contributions in design and analysis of experiments, computer programming, and apparatus building, very little would have been achieved. Without his consistently good-humored nature to buoy up those around him, doing research would have been a much less pleasant task. His presence will be sadly missed.

targets yielded shorter decision latencies than the contrary pairings (that is, modal/noun and preposition/verb).

English uses word order as its major syntactical device. A language like Serbo-Croatian exploits inflection as its primary means of conveying grammatical information. Experiments on syntactic or grammatical priming in Serbo-Croatian have preserved the ordinary word-type adjacencies of the language. The grammatical violations have been introduced at the level of inflected morphemes. For example, Gurjanov, Lukatela, Moskovljević, Savić, and Turvey (1985) paired adjectives and nouns in a lexical decision task. Grammatical agreement requires that the two words be of the same number, case, and gender. This agreement is to be found at the level of the inflectional morphemes that are suffixed to the adjective and noun stems. Gurjanov et al. (1985) violated case agreement and found that lexical decision times for the noun targets were slower than when the paired words were in full agreement. In another experiment with nouns, Lukatela, Kostić, Feldman, and Turvey (1983) observed slower decision times when the noun's inflection was appropriate for a preceding preposition than when it was inappropriate. And in an experiment with verb targets by Lukatela, Moraca, Stojnov, Savić, Katz, and Turvey (1982), lexical decisions were found to be faster when the preceding personal pronoun agreed in person than when it disagreed in person.

How are these various instances of syntactic influences on lexical decision to be understood? Where the context for a target word in the lexical decision task is an associate, expediting lexical decision is often described as due to an automatic, intralexical process. This process is not consciously directed. It is simply a consequence of the way in which the lexical memory is organized (Collins & Loftus, 1975; Forster, 1979). The context mechanically increases the activation level of the target's location in memory prior to the processing of the target. This fast mechanical priming is generally said to be accompanied by a slower, attentional priming. Here the idea is that the context can induce a directing of the focus of attention to a particular region of the internal lexicon (Neely, 1977; Posner & Snyder, 1975). Following a distinction suggested by Seidenberg, Tanenhaus, Leiman, and Bienkowski (1982), contexts that include an associate or semantic relative and that allow, in principle, the foregoing priming processes are termed "priming contexts." A priming context contrasts with the context under investigation in the present paper, namely, a minimal grammatical context. A context of this latter type, referred to as "nonpriming" by Seidenberg et al. (1982), does not appear to precipitate automatic spreading activation (Lukatela et al., 1982). The difference in lexical decision times that accompanies the syntactic congruency/syntactic incongruency contrast seems to be due to post-lexical processes rather than lexical processes (Seidenberg, Waters, Sanders, & Langer, 1984). The important point to be underscored is that lexical decision is a complex operation. The accessing of the context's and of the target's representations in the internal lexicon is but one component process. Other processes might include (1) recognizing the grammatical relation between context and target and (2) assigning a meaning to the context-target structure (cf. deGroot, Thomassen, & Hudson, 1982; Forster, 1979, 1982; West & Stanovich, 1982). If these post-lexical processes are completed before the internal deadline for emitting a lexical decision, they may influence positively (to shorten) or negatively (to lengthen) the response latency (West & Stanovich, 1982).

The present experiments extend the abovementioned studies on the grammatical priming of nouns. They examine the situation in which nouns agree or disagree in gender with the preceding word, a possessive adjective (in English, my, your, our, etc.). They also examine the sensitivity of the nominative singular case to priming. The preposition priming study of Lukatela et al. (1983) did not address this issue directly because the nominative singular case of Serbo-Croatian noun is not governed by a preposition. The study by Gurjanov et al. (1985) did address this issue directly and yielded a negative result: decision times for nouns in the nominative singular case were unaffected by case agreement with preceding adjectives. This issue of the priming sensitivity of the nominative singular case of nouns is important given the demonstration that this case plays a central role in the organization of the inflected forms of a noun in the internal lexicon (Lukatela, Gligorijević, Kostić, & Turvey, 1980). Although the various cases occur with different frequencies, the evidence suggests that speed of lexical access is indifferent to case frequency. The nominative singular is accessed fastest with the different oblique cases accessed at roughly the same speed.

The question posed is whether the privileged lexical status of the nominative singular is associated with a general insensitivity to grammatic context. Is it possible that case agreement and gender agreement are not of equal significance? If they are not then failure to find an effect of agreement in case (Gurjanov et al., 1985) may not extend to agreement in gender. To anticipate, the experimental outcome is that gender agreement does affect the processing of nouns in the nominative singular.

Experiment 1

The lexical decision time for any given target noun in the nominative singular form was measured in two contexts--one in which it was preceded by a possessive adjective in the nominative singular form and one in which it was preceded by a visually similar pseudopossessive adjective. For one half of the noun targets the possessive adjective agreed in gender. It was expected that if gender agreement influenced the processing of nominative singular noun forms, then gender agreement would result in faster decisions than gender disagreement.

The majority of Serbo-Croatian masculine nouns in the nominative singular case end in a consonant. In comparison, the majority of feminine nouns in the nominative singular end in A and the majority of neuter nouns end in either O or E. Some masculine nouns in the nominative singular, however, end in A. There are some feminine nouns in the nominative singular that end in a consonant. In the first experiment only typical masculine and feminine nouns were used. (In the second experiment both the typical and atypical types are examined.)

Method

Subjects. Nineteen students from the Department of Psychology, University of Belgrade, received academic credit for participation in the experiment.

Materials. Letter strings of upper case letters were typed with an IBM Selectric Typewriter. The letter strings were used to prepare black on white slides.

Two types of slides were constructed. In one type, the letter string was arranged horizontally in the upper half of a 35 mm slide and, in the other type, letters of the same kind were arranged horizontally in the lower half of a 35 mm slide. Letter strings in the first type of slides were always possessive adjectives in nominative singular form (or their pseudo-word analogues), and letter strings in the second type of slide were always ordinary nouns in nominative singular form (or their pseudoword analogues). Altogether, there were 144 "possessive adjective" stimuli and 144 "noun" stimuli with each set evenly divided into words and pseudowords.

The 36 nouns were selected from the middle frequency range of a corpus of one million Serbo-Croatian words (Kostić, 1965). Half of the nouns were masculine and half of the nouns were feminine. A different set of 36 nouns (18 masculine and 18 feminine) of the same frequency was used to generate the pseudonouns. This was done by simply changing one letter in the root morpheme. The replacement was an orthotactically and phonotactically legal letter. Importantly, all "nouns" (words and pseudowords) were five letters in length and consisted of two syllables. Thirty-six possessive adjective stimuli were possessive adjectives in the nominative singular form of the masculine gender: twelve were the first person singular (MOJ = my); twelve were the second person singular (TVOJ = thy); and twelve were the first person plural (NAS = our). The other 36 possessive adjective stimuli were the same possessive adjectives in the same case and in the same proportion but of the feminine gender (MOJA, TVOJA, and NASA). In addition to these 72 possessive adjective stimuli another 72 "possessive adjective" stimuli were constructed with the pseudoword analogues of the three masculine and feminine possessive adjectives, namely, MEJ, TLOJ, LAS, MEJA, TLOJA, LASA.

In total, a subject was presented 144 pairs of stimuli in the experimental session. Sixteen other different pairs of stimuli were used for the preliminary training of subjects.

Design. Each noun was presented two times to a given subject. On the two occasions a noun was presented, it was preceded by a possessive adjective on one occasion and by a pseudopossessive adjective on the other occasion. Importantly, between the first and second presentation of a given noun there were always 71 presentations of other pairs. This constraint on the design of the experiment meant that the 36 nouns and the 36 pseudonouns that were exposed in a pseudorandom order in the first half of each experimental session were exposed in the same order in the second half of the session. However, the priming stimuli in the first and second half of the session were mutually interchanged. Those nouns and pseudonouns, which in the first half of the session were preceded by possessive adjectives, were preceded in the second half by the corresponding pseudopossessive adjectives, and vice versa. Hence, a given subject never experienced a given pair of stimuli more than once.

As noted, for any given subject a target noun appeared only twice with one appearance preceded by a pseudopossessive adjective. The other appearance was preceded by a possessive adjective. The possessive adjective context could either agree or disagree in gender with the noun. That is, if the noun were masculine, then the preceding possessive adjective could be either masculine or feminine. Consequently, for a given subject, the nouns that occurred in an appropriate possessive adjective context were different from the nouns that occurred in an inappropriate possessive adjective context. In summarizing the data in Table 1 the fact that different word sets comprised the appropriate and inappropriate pairings is marked by the use of two exemplary masculine nouns, LONAC and SAPUN, and two exemplary feminine nouns, TABLA and PTICA. There is a further feature of the design to be remarked upon. If a target noun, say, LONAC, was preceded by a possessive adjective of the proper gender, say, MOJ, on one of its appearances, then it was preceded by a visually similar pseudopossessive adjective, say, MEJ, on the other appearance. Similarly if LONAC was preceded by the inappropriate context MOJA on one appearance, it was preceded by the pseudopossessive adjective MEJA on the other. The design therefore permitted the direct comparison within a subject of lexical decision times to the same word in two different contexts--one in which the prime agreed or disagreed grammatically and one in which the prime was a pseudoword.

To reiterate, a given subject saw 144 different pairs of stimuli: one quarter of the 144 trials consisted of possessive adjective-noun pairs (half of which agreed and half of which disagreed in gender), one quarter consisted of pseudo possessive adjective-noun pairs, one quarter consisted of possessive adjective-pseudonoun pairs, and one quarter consisted of pseudopossessive adjectives-pseudonoun pairs. The presentation order was pseudorandom.

Procedure. On each trial, two slides were presented. The subjects' task was to decide as rapidly as possible whether the letter string contained in a slide was a word. Each slide was exposed in one channel of a three-channel tachistoscope (Scientific prototype model GB) illuminated at 10.3 cd/m². Both hands were used in responding to the stimuli. Both thumbs were placed on a telegraph key close to the subject and both forefingers on another telegraph key two inches further away. The closer key was depressed for a "no" response (the string of letters was not a word); and the farther key was depressed for a "yes" response (the string of letters was a word).

Latency was measured from the onset of a slide. The subject's response to the first slide terminated its duration and initiated the second slide (at effectively a delay of 0 ms) unless the latency exceeded 1300 ms in which case the second slide was initiated automatically. The duration of the second slide, unlike that of the first, was fixed at 1300 ms.

Results

A mean reaction time was computed for each subject on each type of word pair. Latencies shorter than 300 ms and longer than 1300 ms were excluded as were latencies associated with incorrect responses. The total exclusions did not exceed 1.4 percent of all responses. The mean latencies for the primes, namely, masculine possessive adjective (e.g., MOJ), feminine possessive adjective (e.g., MOJA), pseudo masculine possessive adjective (e.g., MEJ), and pseudo feminine possessive adjective (e.g., MEJA) were: 542 ms, 543 ms, 638 ms, and 637 ms, respectively.

Because of the design of the experiment, a subject saw any given masculine noun in the nominative singular, for example, LONAC, preceded once by a masculine possessive adjective in nominative singular, for example, MOJ, and preceded once by a mutated version of that same masculine possessive adjective, viz., MEJ. Likewise, the subject saw any given feminine noun in the nominative singular, for example, PTICA, preceded once by MOJA and once by MEJA. The same arrangement was true for the incongruent pairings: MOJA SAPUN with MEJA SAPUN for a masculine noun, and MOJ TABLA with MEJ TABLA for a feminine noun. These relations and comparisons are displayed in Table 1.

Table 1

Lexical Decision Times for Examples of Masculine and Feminine Nouns Primed by Real and Pseudopossessive Adjectives

		Noun gender	
Type of prime	Prime inflection	masculine	feminine
possessive adjective	Masculine (Ø)	608±41[a] (LONAC)[b]	665±39 (PTICA)
	Feminine (A	672±27 (SAPUN)	593±36 (TABLA)
pseudoadjective possessive	Masculine (Ø)	653±40 (LONAC)	640±36 (PTICA)
	Feminine (A)	623±42 (SAPUN)	614±27 (TABLA)

[a] mean reaction time and standard deviation

[b] example of noun

Only effects that were significant by both the analysis based on subject means and the analysis based on item means are reported. The question of major interest is whether lexical decision times were affected by the grammatical relation between the prime and the target. This effect, if it exists, should be found in the two-way interaction between target gender and prime inflection and the three-way interaction among target gender, prime inflection, and lexicality. Both interactions proved to be significant: $F(1,18) = 74.93$, MSe = 641, $p < .001$ and $F(1,18) = 52.43$, MSe = 877, $p < .001$ by the subject analysis; and $F(1,32) = 17.18$, MSe = 19220, $p < .001$ and $F(1,32) = 15.68$, MSe = 21794, $p < .001$ by the item analysis. Also significant was the main effect of prime inflection: $F(1,18) = 19.79$, MSe = 291, $p < .001$ and $F(1,32) = 4.10$, MSe = 4591, $p < .05$ by the subjects and items analyses, respectively. On the average, lexical decisions following the uninflected primes (e.g., MOJ, MEJ) were slower than those following the inflected primes (e.g., MOJA, MEJA): 642 ms versus 625 ms.

The analysis supports the hypothesis that lexical decision on a noun in the minimal grammatical context provided by a possessive adjective depends on whether or not the noun and possessive adjective agree in gender. For masculine nouns the difference between the inappropriate pairing and the appropriate pairing was 64 ms; for feminine nouns it was 72 ms. These magnitudes are considerably larger than the inappropriate-appropriate difference reported by Goodman et al. (1983). Comparisons of English word sequences such as "men swears" (appropriate) and "whose swears" (inappropriate) yielded small differences of 19 ms (Experiment 1) and 13 ms (Experiment 2).

The grammatical congruent-grammatical incongruent contrast is a reliable measure of grammatical priming. Less reliable but of larger theoretical importance is the measure of grammatical priming that divides the congruency effect into facilitative and inhibitory components. This division rests on the availability of a suitable baseline. In the present experiment nouns following pseudowords provide the baseline. What is missing, however, is an independent evaluation of the effect of pseudowords on lexical decision. Another weakness of the current baseline is that a pseudopossessive adjective-noun sequence involves a negative response followed by a positive response, raising the possibility of an inhibitory influence on the noun decision-making process. The analysis that follows should be interpreted with these caveats in mind.

As noted above, because of the design of the experiment it is possible to make a within-subject comparison of a noun with itself in two different contexts, namely, those of possessive adjective and pseudopossessive adjective. Facilitation of lexical decision is here defined operationally by a significant positive difference between pairs of type MOJ LONAC (congruent prime) and MEJ LONAC (nonsense prime) or MOJA PTICA (congruent prime) and MEJA PTICA (nonsense prime), and inhibition of lexical decision is defined by a significant negative difference between pairs of type MOJA SAPUN (incongruent prime) and MEJA SAPUN (nonsense prime) or MOJ TABLE (incongruent prime) and MEJ TABLA (nonsense prime). Protected t-tests (Cohen & Cohen, 1975; the error term from the ANOVA is used as the estimate of the variance) on subject means revealed that there was facilitation: $\underline{t}(18) = 4.79$, $\underline{p} < .001$ and $\underline{t}(18) = 2.29$, $\underline{p} < .05$ for the masculine (LONAC) and feminine (PTICA) situations respectively; and that there was inhibition: $\underline{t}(18) = 4.49$, $\underline{p} < .001$ and $\underline{t}(18) = 2.50$, $\underline{p} < .05$ for the masculine (SAPUN) and feminine (TABLA) situations, respectively. These outcomes were nearly corroborated in full by protected t-tests on item means: $\underline{t}(32) = 3.49$, $\underline{p} < .001$ and $\underline{t}(32) = 1.75$, $\underline{p} < .05$ for the masculine (LONAC) and feminine (PTICA) situations, respectively; $\underline{t}(32) = 3.72$, $\underline{p} < .001$ and $\underline{t}(32) = 1.59$, $\underline{p} > .05$ for the masculine (SAPUN) and feminine (TABLA) situations, respectively.

An ANOVA conducted on the pseudonoun data revealed no main effects or interactions.

Experiment 2

The inflectional morphemes of a masculine possessive adjective in nominative singular and a typical masculine noun in nominative singular are identical, viz., Ø. Similarly, the inflectional morphemes of a feminine possessive adjective in nominative singular and a typical feminine noun in nominative singular are identical, viz., A. The second experiment examines the contribution of this identity in inflectional morphemes to the gender congruency/incongruency effect observed in Experiment 1.

As noted above, there are (very few) masculine nouns that end in A in the nominative singular and (relatively more) feminine nouns that end in Ø in the nominative singular. It is possible, therefore, to have a possessive adjective and noun that agree in nominative singular case and in gender but that do not share the same inflected ending, for example, MOJ DEDA (my grandfather), where both words are masculine nominative singular, and MOJA MATER (my mother), where both words are feminine nominative singular. The second experiment exploits pairs of the preceding kind along with pairs constructed, as before, from typical masculine and feminine nouns, for example, MOJ LONAC and MOJA PTICA. If the gender congruency/incongruency effect is not tied to the visual or linguistic identity of the prime and target suffixes, then the effect should hold for possessive adjective-noun pairs constructed with atypical nouns as it does for such pairs constructed with typical nouns. If MOJ LONAC is faster than MOJA LONAC, then MOJ DEDA should be faster than MOJA DEDA. The latter observation would rule out the hypothesis that the effect obtained in the first experiment was due to dimensions of visual similarity rather than grammatical similarity.

The design of the second experiment differed from that of the first. In the second experiment, unlike the first, no noun or pseudonoun target was repeated in the sequence of prime-target pairs seen by a subject. In the second experiment, unlike the first, the nouns preceded by congruent possessive adjectives were also the nouns preceded by incongruent possessive adjectives. This was achieved by a between-subjects manipulation. Where one group of subjects saw a given noun preceded by a grammatically appropriate prime, another group of subjects saw the same noun preceded by a grammatically inappropriate prime. The analysis of the experiment focuses on the grammatical congruency/grammatical incongruency effect. What few merits the analysis into facilitation and inhibition effects might have had in the first experiment, given its within-subject comparison of a target noun preceded by a word prime and a pseudoword prime, were reduced further by the between-subject design of the second experiment. Consequently, no attempts were made in the second experiment to quantify facilitation and inhibition.

Method

Subjects. Fifty-two students from the Department of Psychology, University of Belgrade, received academic credit for participation in the experiment. A subject was assigned to one of four subgroups according to the subjects' appearance at the Laboratory, for a total of thirteen subjects per subgroup. None of the subjects had participated in Experiment 1.

Materials. The stimuli were of the same physical appearance as in Experiment 1. Altogether, 128 "possessive adjective" stimuli and 128 "noun" stimuli were constructed, with each set evenly divided into words and pseudowords. The 64 real possessive adjective stimuli represented the possessive adjectives MOJ, MOJA, (my) and TVOJ, TVOJA (your). The 64 pseudopossessive adjective stimuli were derived from the possessive adjectives by replacement of a consonant or a vowel (MEJ, MEJA, MOS, MOSA, FOJ, FOJA, KVOJ, KVOJA, TVOK, TVOKA, TVEJ, TVEJA).

Thirty-two of the nouns in Experiment 2 were similar to those used in Experiment 1--there were 16 typical masculine nouns and 16 typical feminine nouns. In comparison to Experiment 1 an additional set of 32 atypical nouns was also used: 16 masculine nouns ending in the vowel A and 16 feminine nouns ending in a consonant. The 64 pseudonouns were generated from these typical and atypical nouns by replacing the initial or middle consonant by another consonant of same phonemic class. Consequently, 32 pseudonouns ended in a consonant and 32 pseudonouns ended in A.

In total, there were 512 different pairs of stimuli of which a given subject saw 128 pairs. Thirty-two other pairs of stimuli were used for the preliminary training of subjects.

Design. The constraint of the design of the experiment was that a given subject never experienced a given noun or pseudonoun more than once.

As mentioned, a given subject saw 128 different pairs of stimuli. Each subject saw the same nouns and pseudonouns as every other subject but not preceded by the same possessive adjective or pseudopossessive adjective type. Consider, for example, the masculine noun LONAC. In one group of subjects this noun was preceded by a possessive adjective in the same case, number, and gender (e.g., MOJ); in a second group it was preceded by a possessive adjective of the same case and number but of a different gender (e.g., MOJA); in a third group it was preceded by a pseudoword visually similar to the congruent prime (e.g., MEJ or MOJ or FOJ); and in a fourth group it was preceded by a pseudword visually similar to the incongruent prime (e.g., MEJA or MOJA or FOJA). In one half of the 128 trials the second stimulus in a pair was a noun, and in the other half the second stimulus was a pseudonoun. In one half of the 32 possessive adjective-noun trials a given subject saw 8 typical masculine and 8 typical feminine nouns. There was a similar division for the 32 pseudopossessive adjective-noun trials, the 32 possessive adjective-pseudonoun trials, and the 32 pseudopossessive adjective-pseudonoun trials. Within each combination gender-congruent possessive adjectives and gender-incongruent possessive adjectives appeared equally often.

Procedure. The procedure was the same as in Experiment 1.

Results

A mean reaction time was computed for each subject in each of the four groups. The criteria for excluding responses were the same as in Experiment 1. Approximately 3.5 percent of all responses were excluded from the analyses by these criteria.

The first question to be addressed is whether the results of the first experiment which were obtained with typical masculine and feminine nouns were replicated in the second experiment. Table 2 presents the data for typical masculine and feminine nouns as a function of prime lexicality and prime inflection. A group x prime lexicality x target gender x prime inflection analysis of variance suggests that the outcome of Experiment 2 was very similar to that of Experiment 1: Target gender was significant, $F(1,48) = 15.69$, $MSe = 2610$, $p < .001$; target gender by prime inflection was significant, $F(1,48) = 20.53$, $MSe = 4534$, $p < .001$; and target gender by prime inflection by prime lexicality was significant, $F(1,48) = 30.47$, $MSe = 2232$, $p < .001$. Although the main effect of groups was not significant, there were significant interactions involving groups: group by prime inflection, $F(3,48) = 13.66$, $MSe = 2222$, $p < .001$; group by prime lexicality, $F(3,48) = 5.57$, $MSe = 5670$, $p < .01$; group by prime inflection by prime lexicality, $F(3,48) = 11.30$, $MSe = 1958$, $p < .001$; and the four way interaction. These interactions identify the differences in the pairs of stimuli assigned to the groups.

As with Experiment 1 it can be claimed that lexical decision times for target nouns of the typical type depended on whether the inflected ending of the prime was consistent with the gender of the noun. This dependency is

---

Table 2

Lexical Decision Times and Error Rates for Typical Masculine and Feminine Nouns as a Function of Prime Lexicality and Prime Inflection

		(typical) Noun gender	
Type of prime	Prime inflection	masculine (0)	feminine (A)
possessive adjective	Masculine (0)	657±93[a] 1.4[b]	687±92 2.4
	Feminine (A)	717±112 4.8	636±79 0.50
pseudo adjective possessive	Masculine (0)	670±84 5.8	661±91 1.4
	Feminine (A)	666±80 4.3	647±73 1.9

[a] mean reaction time and standard deviation
[b] percentage of responses that were incorrect

Table 3

Lexical Decision Times and Error Rates for Atypical Masculine and Feminine Nouns as a Function of Prime Lexicality and Prime Inflection

		(atypical) Noun gender	
Type of prime	Prime inflection	masculine (A)	feminine (0)
possessive adjective	Masculine (0)	712±107[a] 5.8[b]	692±108 4.8
	Feminine (A)	734±102 7.7	647±86 1.4
pseudo possessive adjective	Masculine (0)	723±104 8.2	675±75 5.3
	Feminine (A)	730±99 8.7	652±69 2.4

[a] mean reaction time and standard deviation
[b] percentage of responses that were incorrect

---

greater for word-word pairs than for pseudoword-word pairs. Protected t-tests confirmed the difference between congruent word-word pairs and incongruent word-word pairs for the masculine nouns, $t(48) = 6.49$, $p < .001$ and between congruent word-word pairs and incongruent word-word pairs for the feminine nouns, $t(48) = 5.52$, $p < .001$. However, neither the masculine nor the feminine comparison was significant for the pseudoword-word pairs.

Is the gender congruency/incongruency effect exhibited by possessive adjective-noun pairs constructed with atypical nouns? Table 3 presents the data for the atypical masculine and feminine nouns as a function of prime lexicality and prime inflection. Comparison of Table 3 with Table 2 suggests a similar, though not identical, pattern of results. An analysis of variance conducted over the combinations of groups, prime lexicality, target gender, and prime inflection yielded significant effects for target gender, $F(1,48) = 99.87$, MSe = 3495, $p < .001$ and for the interaction of target gender with prime inflection, $F(1,48) = 21.68$, MSe = 2869, $p < .001$. There was no main effect of groups but all the interactions with group were significant, as above. Like typical nouns, atypical nouns exhibit a gender congruency/incongruency effect but, unlike typical nouns, the magnitude of the effect is less dependent on the lexicality of the prime.

It is noteworthy that there was a large difference in errors between atypical masculine nouns (more) and atypical feminine nouns (less), $F(1,48) = 11.92$, $p < .001$ and that the errors committed on these two noun types depended differently on the inflection of the preceding prime, $F(1,48) = 4.44$, $p < .05$. The same analysis on the typical nouns revealed that the masculine nouns were again the source of most errors, $F(1,48) = 7.65$, $p < .01$, but that there was no interaction of target gender with prime inflection. Overall, the errors for both analyses follow the pattern of the decision latencies (compare Tables 2 and 3) but it is not obvious why, in all analyses (Experiment 1 and Experiment 2), latencies are longer on average and errors are greater on average for masculine nouns.

The third question is whether the gender congruency/incongruency effect differs between typical and atypical masculine nouns. The number of masculine nouns that end in A is very small, as noted, and the number of nouns in this category used in the experiment almost exhausts the category. By and large, masculine nouns inflected with A in the nominative singular occur less frequently than masculine nouns inflected with ∅ in the nominative singular. A group x prime lexicality x prime inflection x target inflection (typical vs. atypical type) analysis of variance was conducted. The main effect of prime inflection was significant, $F(1,48) = 4.99$, MSe = 9249, $p < .05$--∅-inflected primes were associated with faster lexical decisions (691 ms) than A-inflected primes (711 ms). The difference between typical and atypical nouns was significant, $F(1,48) = 83.39$, MSe = 2768, $p < .001$; the atypical nouns were responded to more slowly (723 ms) than the typical nouns (680 ms) probably because of their lower frequency of occurrence. The interaction of prime lexicality and prime inflection was significant, $F(1,48) = 4.28$, MSe = 9822, $p < .05$ as was the interaction of prime lexicality and target inflection, $F(1,48) = 5.97$, MSe = 2145, $p < .01$. There was no two-way interaction between inflection of the prime and the typicality of the inflection of the noun. Lexical decision times for typical masculine nouns preceded by the congruent ∅-inflected primes (real and pseudo) were 33 ms shorter, on the average, than lexical decision times for typical masculine nouns preceded by incongruent A-inflected primes (real and pseudo). This average difference for atypical masculine nouns was 15 ms. There was, however, a significant three-way interaction among prime lexicality, prime inflection, and target inflection (typical vs. atypical), *F*(1,15) = 5.06, MSe = 3193, *p* < .05. Inspection of Tables 2 and 3 reveals that the inflection of the pseudoadjective prime did not matter for either typical or atypical nouns. The congruency-incongruency difference was -4 ms and -7 ms, respectively. In contrast, the inflection of the adjective prime did matter for both typical nouns and atypical nouns and it mattered more for the typical nouns than the atypical nouns. The congruency-incongruency difference was 60 ms and 22 ms, respectively. In sum, the data suggest that the magnitude of the gender congruency/incongruency effect differed between typical and atypical masculine nouns.

The fourth question addressed parallels the third. Does the gender congruency/incongruency effect differ between typical and atypical feminine nouns? The answer in this case is negative. A group x prime lexicality x prime inflection x target inflection (typical vs. atypical) revealed only one significant effect, namely, the main effect of prime inflection, *F*(1,48) = 17.30, MSe = 6675, *p* < .001; A-inflected primes were associated with faster lexical decision (648 ms) than 0-inflected primes (678 ms) as ought to be the case for feminine noun targets.

Finally, with respect to the pseudonoun data, separate analyses of variance revealed that for both the typical and atypical cases there was a significant effect of target inflection (0 vs. A): *F*(1,51) = 6.54, MSe = 3050, *p* < .01 and *F*(1,51) = 4.77, MSe = 4290, *p* < .05, respectively. Pseudonouns ending in A were rejected more slowly. A further significant effect was observed in the atypical analysis, namely, the interaction of prime lexicality and target inflection, *F*(1,51) = 18.90, MSe = 2827, *p* < .001. Where 0-inflected atypical pseudonouns were responded to faster when preceded by a pseudo-possessive adjective, A-inflected atypical pseudonouns were responded to faster when preceded by a possessive adjective. The data equivocate on whether or not rejecting pseudonouns was made more difficult by a grammatically and lexically proper context.

Discussion

In the present experiments, possessive adjectives provide a minimal grammatical context for nouns in the nominative singular. With case and number held constant it is shown that when the two words agree in gender, lexical decision on the target noun is faster than when the two words disagree in gender. A previous experiment (Gurjanov et al., 1985) found no effect of *case* congruency on the processing of nouns in the nominative singular. That *gender* congruency does affect the processing of nominative singulars may have implications for the representation of inflected nouns in the internal lexicon (Lukatela et al., 1980).

The lesson learned from Experiment 2 is that the gender congruency/incongruency effect is not mediated by visual identity or phonemic identity of the morphemes that inflect the possessive adjective and the noun. This latter observation implies that the gender congruency/incongruency effect must involve the recognition of the genders of the possessive adjective and the noun, which implies, in turn, that gender is part of a word's representation in the lexicon. It is not presumptuous to assume that one's knowledge of words includes a knowledge of the grammatical arrangements into which they may enter.

To know that the feminine possessive adjective MOJA cannot be entered into a grammatical arrangement with the masculine nouns LONAC or DEDA is to know that MOJA and LONAC or MOJA and DEDA are of unlike gender. On the other hand, to know that the masculine possessive adjective MOJ can be linked to the masculine nouns LONAC and DEDA is to know that these words are alike in case, number, and gender.

The argument that there is a syntactical/grammatical processor is an argument for a device separate from the device that accesses lexical representations and separate from the device that assigns meaning to an arrangement of words (cf. Forster, 1979). The syntactic/grammatical processor assigns a syntactical structure or a grammatical relation to a context-target arrangement. It obviously has a degree of autonomy; there are many celebrated examples of English syntactical structure being assignable to a list of nonsense letter strings. However, with respect to the question of the information with which the syntactic or grammatical process works, it must be supposed that that information is derived in large part by the lexical processor. Seidenberg et al. (1982) showed that in English lexical priming contexts, facilitation effects are not indifferent to the grammatical function of words and argue for a model of the internal lexicon enriched by syntactical details--an argument consonant with the suggestions of Kaplan and Bresnan (1982) and Gazdar (1982) in theoretical linguistics and continuous with the experimental efforts of Huttenlocher and Lui (1979) and Miller and Johnson-Laird (1976) and others to distinguish the mental representations of different word classes.

Given the notions of lexical processor, grammatical processor and message processor (Forster, 1979) as three relatively independent systems underlying lexical decision, an account of the gender congruency/incongruency effect takes the following form (after West & Stanovich, 1982). When a grammatically congruent pair (e.g., MOJ LANAC, MOJ DEDA, MOJA PTICA, or MOJA MATER) is presented, the outputs from the lexical processor, grammatical processor and message processor are all positive--the ideal situation for a subsequent decision-making mechanism that must arrive at the appropriate response "yes." However, when a grammatically incongruent pair (e.g., MOJA LONAC, MOJA DEDA, MOJ PTICA, or MOJ MATER) is presented, the output from the lexical processor is positive and so, perhaps, is the output from the message processor, but the output from the grammatical processor is negative. The information made available to the grammatical processor from the lexical processor is that the context is one gender and the target is another gender. Consequently, the situation for the decision-making system is less than ideal; there are discrepancies in the outputs and the no bias from the grammatical processor must be overcome (West & Stanovich, 1982). As a result, lexical decision to a grammatically incongruent pair (e.g., MOJA LONAC) is slower than lexical decision to a grammatically congruent pair (e.g., MOJ LONAC).

The foregoing account is sufficiently general to accommodate the syntactic or grammatical priming effects found with English language materials (Goodman et al., 1981; Wright & Garrett, 1984) and those found with Serbo-Croatian language materials. Where the account is weak is in its failure to distinguish those components of grammatical processing that are automatic or reflexive (Fodor, 1983; Wright & Garrett, 1984) from those that are merely strategic, that is, those that are "conscious-attentive" and shaped by the conditions of the experiment. This failure is due in part to the lack of data relevant to the contrast. It has been established empirically that associative priming involves components of both kinds and the theory of associative priming ably recognizes the distinction (Neely, 1977). If syntactic or grammatical priming proves to depend similarly on a fast-acting automatic process and a slow-acting conscious-attentive process, then this much seems certain: In syntactic or grammatical priming *both* of these processes are post-lexical (Gurjanov et al., 1985; Seidenberg et al., 1984; West & Stanovich, 1982).

References

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavior sciences.* Hillsdale, NJ: Erlbaum.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407-428.

deGroot, A. M. B., Thomassen, A. J. W. M., & Hudson, P. T. (1982). Associative facilitation of word recognition as measured from a neutral prime. *Memory & Cognition*, *10*, 358-370.

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

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. Walker (Eds.), *Sentence processing*. Hillsdale, NJ: Erlbaum.

Gazdar, G. (1982). Phrase structure grammar. In P. Jacobson & G. K. Pullum (Eds.), *The nature of syntactic representation*. Dordrecht: Reidel.

Goodman, G C., McClelland, J. L., & Gibbs, R. W. (1981). The role of syntactic context in word recognition. *Memory & Cognition*, *9*, 580-586.

Gurjanov, M., Lukatela, G., Moskovljević, J., Savić, M., & Turvey, M. T. (1985). Grammatical priming of inflected nouns by inflected adjectives. *Cognition*, *19*, 55-71.

Huttenlocher, J., & Lui, F. (1979). The semantic organization of some simple nouns and verbs. *Journal of Verbal Learning and Verbal Behavior*, *18*, 141-162.

Kaplan, R. M., & Bresnan, J. W. (1982). Lexical-functional grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The mental representation of grammatical relations*. Cambridge, MA: MIT Press.

Kostić, Dj. (1965). *Frequency of occurrence of nouns 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., Kostić, A., Feldman, L. B., & Turvey, M. T. (1983). Grammatical priming of inflected nouns. *Memory & Cognition*, *11*, 59-63.

Lukatela, G., Moraca, J., Stojnov, D., Savić, M., Katz, L., & Turvey, M. T. (1982). Grammatical priming effects between pronouns and inflected verb forms. *Psychological Research (Psychologische Forschung)*, *44*, 297-311.

Lupker, S. J. (1984). Semantic priming without association: A second look. *Journal of Verbal Learning and Verbal Behavior*, *23*, 709-733.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226-254.

Posner, M. I., & Snyder, C. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V*. London: Academic Press.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. Cognitive Psychology, 14, 489-537.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and post-lexical loci of contextual effects on word recognition. Memory & Cognition, 12, 315-328.

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, 5, 385-399.

Wright, B., & Garrett, M. (1984) Lexical decision in sentences: Effects of syntactic structure. Memory & Cognition, 12, 31-45.

GRAMMATICAL PRIMING OF INFLECTED NOUNS BY INFLECTED ADJECTIVES*

M. Gurjanov,† G. Lukatela,† Jasmina Moskovljević,† M. Savić,†
and M. T. Turvey††

Abstract. Two experiments are reported in which subjects made rapid lexical decisions about inflected nouns preceded by inflected adjectives or pseudoadjectives that did or did not agree grammatically. Both adjectives and pseudoadjectives were shown to affect lexical decision times for nouns, suggesting that the priming of inflected nouns by inflected adjectives occurred at the level of the inflections. Inflected pseudonouns, however, were not affected similarly, suggesting that lexical factors were contributing to the priming in addition to grammatical factors. This instance of grammatical priming is described as an effect that arises post-lexically, based on the outcomes of relatively independent lexical and syntactical processors.

Two broad questions may be raised with regard to the processing of nouns in an inflected language: (1) How are the cases of a noun organized with regard to each other in the internal lexicon?; and (2) How are inflected nouns linked to other lexical types such as prepositions and inflected adjectives? Serbo-Croatian is an inflected language in which the noun takes a gender (masculine, feminine, or neuter) and is declined in seven forms (nominative, accusative, instrumental, genitive, dative, locative, vocative), both in the singular and the plural. The fourteen inflected forms of a Serbo-Croatian noun can be viewed as forming a noun system (Lukatela, Gligorijević, Kostić, & Turvey, 1980). Ordinarily an inflected Serbo-Croatian noun in a sentence is grammatically related to a preposition and to one or more adjectives. Although they are not declined, prepositions are specific to inflected noun endings. A given preposition goes with at least one noun case, sometimes several cases but never with all noun cases. Adjectives are declined but not necessarily with the same inflected endings as nouns. When qualifying a noun, however, the inflection of the adjective and the inflection of the noun must

---

*Cognition, 1985, 19, 55-71.

†University of Belgrade
††Also University of Connecticut

Acknowledgment. This research was supported in part by NICHD Grants HD-08495 and HD-01994 to the University of Belgrade and Haskins Laboratories, respectively. Dr. M. Gurjanov died in August, 1983. "Gura" has been a central figure in this research program from its inception in 1973. Without his contributions in design and analysis of experiments, computer programming, and apparatus building, very little would have been achieved. Without his consistently good-numored nature to buoy up those around him, doing research would have been a much less pleasant task. His presence will be sadly missed.

agree grammatically (for example, if the noun is masculine and in the singular accusative form, the adjective must be masculine and in the singular accusative form).

With respect to the first question raised above on the organization of the cases there is evidence to suggest that frequency--precisely, the frequencies with which the various inflected noun forms occur in ordinary language usage--is not a major determinant of a Serbo-Croatian noun system's organization. In a lexical decision task nouns in the nominative singular form were accepted as words faster than nouns in the oblique forms. Among the oblique forms, however, decision times did not differ despite marked differences among the oblique forms in their respective frequencies of occurrence (Lukatela et al., 1978; Lukatela et al., 1980). Apparently, the nominative and oblique forms are qualitatively distinguished in the organization of a noun system with the nominative assuming a pivotal role. However, in either an oblique or nominative form a noun appears to be represented in the lexicon as a single unit corresponding to the complete word rather than as a combination of distinct units corresponding to morphemic constituents. The stems and suffixes of Serbo-Croatian nouns do not appear to be stored separately. An observation of the unitary representation of nouns, however, does not rule out the possibility that noun representations <u>indicate</u> their stem/suffix structure (Stanners, Neiser, & Painton, 1979; Taft & Forster, 1975).

With respect to the second question raised above (on the processing relation of nouns to other lexical types), it has been shown that with Serbo-Croatian words a preposition preceding a noun case with which it is grammatically consistent speeds up lexical decision on the noun. However, lexical decision on an inflected noun form that is grammatically inconsistent with the preceding preposition is not appreciably slowed (Lukatela, Kostić, Feldman, & Turvey, 1983). Facilitation (and inhibition) effects among words are often explained (but not always, see Discussion) by a notion of activation spreading out from one excited region of the lexicon to neighboring regions and/or by a notion of a directing of attention to a specified region of the lexicon. The first of these mechanisms may be suited to semantic relations among lexical entries but it is not easily generalized to grammatical relations such as between members of a closed class like prepositions and an open class like nouns (and it is not easily generalized to semantic relations in natural discourse, as Foss [1982] has noted). The notion of an automatic spread of activation refers to a specific linkage between particular representations of particular words (see Collins & Loftus, 1975)--(direct) stimulation of one lexical representation leads mechanically and inevitably to the (indirect) stimulation of other lexical representations. The relation of prepositions to nouns, however, is not sensibly portrayed as linkages among particular internal representations of complete words. (What would rationalize the linkage of <u>above</u> and <u>elephant</u>?) If there are linkages one might expect them to be defined over the small set of prepositions and the small set of morphemes that comprise the inflected endings of nouns. By such an account, prepositions would not be linked to the very many noun systems but to the few sets of inflected endings that the very many noun systems share. The problem with this account is that the inflected endings of (Serbo-Croatian) nouns do not appear to be stored as sets separately from their stems.

The present experiments extend the inquiry into Serbo-Croatian nouns and their processing relation to other word types. Here the focus is the relation of nouns to adjectives. Two related questions are raised. First, can adjectives affect the time to lexically evaluate nouns with which they are grammatically consistent? And second, if adjectives can affect lexical decisions on nouns do they do so at the morphological level, that is, the level of stems and affixes (rather than, say, the whole word level)? Support for the view that adjectival influences on nouns can be mediated by processes at the level of inflected endings would be provided by the demonstration that both adjective contexts and pseudoadjective contexts (letter strings derived from adjectives by changing the initial or middle consonant) expedite lexical decisions on noun targets when the inflection of the contextual item and the target are in grammatical agreement.

The selection of nouns used in the experiments was guided by the following considerations. With a few exceptions Serbo-Croatian nouns fall into three declensional classes according to the inflected ending of the genitive singular case. These three classes are designated (after Bidwell, 1970) as Class A (where the genitive singular ending in /e/, for example, ZENE), Class O (where the genitive singular ending is /a/, for example, COVEKA), and Class C (where the genitive singular ending is /i/, for example, STVARI). The dominating gender for Class A nouns is feminine. The nouns in Class C are almost exclusively feminine but Class C occurs less frequently than Class A. Class O nouns are mostly masculine and neuter nouns. From a consideration of nouns in the ordinary, written language, Kostić (1965) reported that the masculine gender accounts for 52 percent, the feminine gender for 36 percent and the neuter gender for 12 percent. Consequently, the nouns in the corpus of words from which the stimuli of the present experiment were drawn occurred in the three genders in approximately the proportions identified by Kostić, with the masculine and neuter nouns drawn from the declension Class O and the feminine nouns drawn from the declension Class A.

The adjectives in the corpus of words from which the stimuli were drawn were common adjectives all declined as indefinite adjectives. Common adjectives are those that can be declined both definitely and indefinitely. The indefinite declension of an adjective applies when the function is either predication or attribution. In the latter role the indefinite adjective is not accompanied by a deictic such as "this," "that," etc., and is referentially vague. Definite adjectives are restricted to the attributive function and are always conjuncted with a deictic. When an adjective qualifies an inanimate noun in the masculine gender, the indefinite and definite declensions are distinguished by the inflected endings of the nominative singular and accusative singular. There are, however, no such written distinctions for the definite and indefinite adjectival declensions when the word being qualified is an inanimate noun in the feminine gender (although such distinctions can be found in the spoken language in the form of stress variations). The choice of the referentially less precise indefinite declension was motivated, in part, by the desire to keep to a minimum the semantic relation between the adjectival and nominal forms paired in the experiments.

Experiment 1

The first experiment was directed at the effect of grammatical consistency between adjectives (real and pseudo) and nouns in the nominative singular and genitive singular cases. These two cases are the most frequently occurring noun cases--the nominative singular accounting for approximately 25 percent, and the genitive singular accounting for approximately 20 percent, of all instances of the noun (Kostić, 1965; Lukatela et al., 1980). The inflections of these two cases for adjectives and nouns of all three genders are shown in Table 1. Only for the feminine gender are the adjectival and nominal inflections identical.

Table 1

Nominative Singular and Genitive Singular Inflections of Serbo-Croatian Adjectives and Nouns as a Function of Gender

	MASCULINE		FEMININE		NEUTER	
	ADJECTIVE	NOUN	ADJECTIVE	NOUN	ADJECTIVE	NOUN
NOMINATIVE SINGULAR	∅	∅	A	A	∅	E or O
GENITIVE SINGULAR	OG	A	E	E	OG	A

∅ = null morpheme

There is some reason to believe that the effect of a preceding grammatically consistent adjective on lexical decision will not be of the same magnitude for nouns in the nominative singular case and nouns in the genitive singular case. As noted above, the nominative singular of a noun is qualitatively distinguished from the oblique cases of a noun and appears to play a pivotal role in the organization of a noun's case system (Lukatela et al., 1980). Moreover, the nominative singular is less dependent on grammatical factors for its interpretation than are the oblique cases (see Lukatela et al., 1983). It was expected, therefore, that for nouns in the genitive singular lexical decision would be fastest when the prime was grammatically consistent but for nouns in the nominative singular lexical decision times would be less partial to the grammatical consistency of prime and target.

Method

Subjects. Fifty-six undergraduate students from the Department of Psychology at the University of Belgrade participated in the experiment. All subjects had previously participated in reaction time experiments.

Materials. A list of 150 adjective-noun pairs was constructed with all adjectives and nouns (1) drawn from the mid-frequency range of the Kostić table, (2) in the nominative singular form and (3) comprising pairs that were congruent in gender. This list was presented to 70 students (from the Department of Linguistics) who judged the associative strength of each pair--that is, the degree to which the adjective and the noun in a pair were related. The twenty-eight adjective-noun pairs that were judged to be most weakly associated were used to generate four groups of 28 word-word pairs: nominative singular-nominative singular pairs, nominative singular-genitive singular pairs, genitive singular-nominative singular pairs and genitive singular-genitive singular pairs. (In each of the foregoing pair types, the first case is that of the adjective and the second case is that of the noun.)

A different set of adjective-noun pairs, drawn from the original list of 150 pairs, was used to generate four corresponding groups of 28 pseudoadjective-pseudonoun pairs (by changing either the initial or middle letter of both the adjective and the noun). Another set of 28 pairs from the original list of 150 pairs was used to generate four corresponding groups of 28 pseudoadjective-noun pairs (by changing either the initial or middle letter of the adjective). Finally, one further, different set of 28 pairs was transformed into four corresponding groups of 28 adjective-pseudonoun pairs (by changing either the initial or middle letter of the noun). Throughout the generation of these different groups--that paired pseudowords or paired a pseudoword with a word--the pseudoword version of a noun or adjective in nominative singular or genitive singular preserved the case ending.

The adjectives and pseudoadjectives were presented as Roman letter strings (IBM Gothic) arranged horizontally in the upper half of 35 mm slides. In contrast, nouns and pseudonouns were arranged horizontally in the lower half of 35 mm slides. The "adjective" slides and the "noun" slides were grouped into pairs as determined above to yield a total of 448 pairs of slides (28x4x4) of which a given subject saw 112 pairs.

Design. The major constraint on the design of the experiment was that a given subject never encountered a given word or pseudoword in any of the pairs more than once. This was achieved by dividing subjects into four groups with 14 subjects in each group and by dividing each set of 28 pairs into four subgroups of 7 pairs. In sum, a subject saw 7 pairs of stimuli from each of the 16 groups of pairs. Put differently, each subject saw the same adjectives and nouns as every other subject but not necessarily in the same grammatical case nor necessarily in the same type of nominative-genitive permutation.

Procedure. On each trial, two slides were presented. The subject's task was to decide as rapidly as possible whether the letter string contained in a slide was a word. Each slide was exposed in one channel of a three-channel tachistoscope (Scientific Prototype, Model GB) illuminated at 10.3 cd/m2. Both hands were used in responding to the stimuli. Both thumbs were placed on a telegraph key button close to the subject and both forefingers on another telegraph key button two inches further away. The closer button was depressed for a "No" response (the string of letters was not a word); and the further button was depressed for a "Yes" response (the string of letters was a word).

Latency was measured from the onset of a slide. The subject's response to the first slide terminated its duration and initiated the second slide unless the latency exceeded 1300 ms, in which case the second slide was initiated automatically. The duration of the second slide, unlike that of the first, was fixed at 1300 ms.

Results and Discussion

A mean reaction time was computed for each subject by averaging over the seven nouns or seven pseudonouns in each group of prime-target pairs. Reaction times less than 300 ms and longer than 1300 ms were exluded as were the times associated with erroneous responses. The total number of responses excluded by the preceding criteria did not exceed 1.5 percent.

Table 2

Lexical Decision and Percentage Error for Pseudonouns in Experiment 1 as a Function of Type and Grammatical Case of Adjectival Prime

Type of prime	Grammatical case of prime	Grammatical case of target pseudonoun	
		NOMINATIVE	GENITIVE
ADJECTIVE	NOMINATIVE	822[a] 3.3[b]	845 5.1
	GENITIVE	834 3.3	833 2.8
PSEUDOADJECTIVE	NOMINATIVE	821 3.1	822 3.3
	GENITIVE	824 3.1	833 2.6

[a] reaction time (ms)
[b] error

Table 3

Lexical Decision Latencies and Percentage Error for Nouns in Experiment 1 as a Function of Type and Grammatical Case of Adjectival Prime

Type of prime	Grammatical case of prime	Grammatical case of target noun	
		NOMINATIVE	GENITIVE
ADJECTIVE	NOMINATIVE	727[a] 4.3[b]	781 4.8
	GENITIVE	720 4.6	744 4.8
PSEUDOADJECTIVE	NOMINATIVE	713 7.9	795 6.6
	GENITIVE	694 4.3	773 2.8

[a] reaction time (ms)
[b] error

Table 2 reports the pseudonoun data. As can be seen, there were no differences due to the type of prime, the grammatical case of the prime, or the grammatical case--inflected ending--of the pseudonoun. The mean reaction times to the primes themselves were 706 ms and 726 ms, respectively, for adjectives in the nominative singular and genitive singular forms, and 841 ms and 870 ms, respectively, for pseudoadjectives inflected in the fashion of the nominative singular and genitive singular. Table 3 reports the noun data. The only effects that were significant according to the analysis of variance on both subject and item means were: grammatical case of the adjectival prime ($F(1,52) = 24.31$, MSe = 2082, $p < .001$ and $F(1,27) = 4.46$, MSe = 5676, $p < .05$) and grammatical case of the noun target ($F(1,52) = 145.26$, MSe = 2733, $p < .001$ and $F(1,27) = 26.36$, MSe = 7532, $p < .001$).

The failure to observe a significant priming effect by either adjectives or pseudoadjectives might have been expected. Approximately half of the words used in the experiment were feminine. The genitive singular form of feminine nouns (and adjectives) are identical to the nominative plural form of feminine nouns (and adjectives) (see Table 1). As noted, the nominative singular case of nouns has not proven to be sensitive to priming. If the nominative plural is similarly indifferent to priming and if the feminine "genitive singular" noun forms of the present experiment were interpreted as nominative plural forms, then the adjectival and pseudoadjectival priming of nouns would be thwarted. Table 4 distinguishes the mean decision times for the masculine/neuter items from those for the feminine items. Inspection of Table 4 suggests that (1) adjectival and pseudoadjectival effects were present for the masculine/neuter genitive singular forms (corroborated by a subject analysis, $F(1,52) = 6.22$, MSe = 21961, $p < .02$, but not by an item analysis) and absent for the feminine genitive singular forms (the prime case by target case interaction was not significant by either subjects or items analysis); and (2) the commonly obtained (e.g., Lukatela et al., 1978; Lukatela et al., 1980; Lukatela et al., 1983) faster decision times for nominative singular forms relative to oblique forms was not found with the feminine noun data, implying

Table 4

Lexical Decision Latencies of Experiment 1 as a Function of Noun Gender

		gender and case of target noun			
Type of prime	grammatical case of prime	masculine/neuter		feminine	
		NOMINATIVE	GENITIVE	NOMINATIVE	GENITIVE
ADJECTIVE	NOMINATIVE	721	810	739	719
	GENITIVE	696	746	731	727
PSEUDOADJECTIVE	NOMINATIVE	708	833	719	731
	GENITIVE	697	803	704	728

that the feminine nouns in the "genitive singular" were not being interpreted as such. It should be noted that a similar but less pronounced confounding of cases is also true for the neuter genitive singular (which is written identically to the nominative plural and genitive plural). However, whereas for the feminine gender both nouns and adjectives assume identical forms in the genitive singular and nominative plural, for the neuter gender identity of forms holds only for nouns.

Experiment 2

The second experiment used the same design, the same procedure, and the same adjective-nouns pairs as those of the first experiment but replaced the genitive singular case by the dative-locative singular case and with a new group of 56 subjects from the same subject pool. In the declension of adjectives and nouns the dative singular and the locative singular are identical in each of the three genders. The characteristic inflections common to dative singular and locative singular are shown in Table 5. With respect to the noun case confoundings identified above, the dative singular-locative singular inflection across the three genders is not shared with the nominative plural and, in fact, is shared with no other case. Thus, in comparison to Experiment 1, grammatical priming between feminine gender words should be observed in Experiment 2 if, indeed, the failure to obtain such priming in Experiment 1 was due to case confounding.

Table 5

Inflections of Dative Singular and Locative Singular Adjectives and Nouns as a Function of Gender

	Masculine	Feminine	Neuter
ADJECTIVE	OM	OJ	OM
NOUN	U	I	U

Results and Discussion

The mean lexical decison latencies were computed in the manner described in Experiment 1. The positive and negative responses to the adjectival primes were similar in pattern to those reported for Experiment 1. Negative responses to the pseudonoun targets are given in Table 6. No main effects or interactions were significant. Table 7 reports the noun data for all three genders taken together. The analysis of variance on subject means and item means (reported in parentheses) revealed significant effects for the grammatical case of the adjective prime, $F(1,52) = 40.59$, $MSe = 1372$, $p < .001$ ($F(1,27) = 8.05$, $MSe = 3460$, $p < .01$); for the grammatical case of the noun target, $F(1,52) = 61.27$, $MSe = 2508$, $p < .001$ ($F(1,27) = 22.98$, $MSe = 3343$, $p < .001$), for the type of adjectival prime, $F(1,52) = 7.88$, $MSe = 4104$, $p < .01$ ($F(1,27) = 8.85$, $MSe = 1827$, $p < .01$), and for the interaction between the

Table 6

Lexical Decision Latencies and Percentage Error for Pseudonouns in Experiment 2 as a Function of Type and Grammatical Case of Adjectival Prime

Type of prime	Grammatical case of prime	Grammatical case of target noun	
		NOMINATIVE	DATIVE/LOCATIVE
ADJECTIVE	NOMINATIVE	758[a] 4.6[b]	758 5.6
	DATIVE/LOCATIVE	757 3.6	774 2.0
PSEUDOADJECTIVE	NOMINATIVE	763 1.8	767 4.8
	DATIVE/LOCATIVE	750 3.3	760 4.8

[a] reaction time (ms)
[b] error

Table 7

Lexical Decision Latencies and Percentage Error for Nouns in Experiment 2 as a Function of Type and Grammatical Case of Adjectival Prime

Type of prime	Grammatical case of prime	Grammatical case of target noun	
		NOMINATIVE	DATIVE/LOCATIVE
ADJECTIVE	NOMINATIVE	672[a] 3.3[b]	726 3.1
	DATIVE/LOCATIVE	668 2.6	685 4.6
PSEUDOADJECTIVE	NOMINATIVE	656 3.8	708 3.3
	DATIVE/LOCATIVE	647 0.8	673 3.3

[a] reaction time (ms)
[b] error

grammatical case of the adjectival prime and the grammatical case of the noun target, F(1,52) = 16.10, MSe = 1841, p < .001 (F(1,27) = 4.84, MSe = 3060, p < .05). All other two-way and three-way interactions were nonsignificant. The significance of the type of adjectival prime may be attributed to the difference between responding positively to two successive stimuli (in the adjective trials) and responding negatively to the first stimulus and positively to the second stimulus (in the pseudoadjective trials). Intuitively, this interpretation suggests slower decision times for targets following pseudoadjectives. Inspection of Table 7 (and of Table 3) shows, to the contrary, that pseudoadjective primes were associated with overall faster decisions. One is tempted to say that the effect of word primes is predominantly "inhibitory."

Table 8 reports the mean lexical decision times for the nouns partitioned according to the masculine/neuter gender and feminine gender categories. Inspection of Table 8 and comparisons with the pattern of results in Table 4 suggest that grammatical priming occurred in both categories in the second experiment in contrast to the first and lends credence to the interpretation given of the feminine gender data of the first experiment.

Table 8

Lexical Decision Latencies of Experiment 2 as a Function of Noun Gender

		gender and case of target noun			
		masculine/neuter		feminine	
Type of prime	grammatical case of prime	NOMINATIVE	DATIVE/ LOCATIVE	NOMINATIVE	DATIVE/ LOCATIVE
ADJECTIVE	NOMINATIVE	663	717	682	741
	DATIVE	652	672	680	685
PSEUDOADJECTIVE	NOMINATIVE	651	708	662	708
	DATIVE	643	669	649	682

Discussion

The theoretically important descriptors "facilitation" and "inhibition" are not applicable to the data of Experiments 1 and 2. In neither experiment is there a neutral context to provide a baseline. The results are more prudently summarized in terms of an inequality and an equality:

(1) The lexical decision time for a noun in a grammatically congruent adjective or pseudoadjective context is less than the lexical decision time for a noun in a grammatically incongruent adjective or pseudoadjective context; and

(2) The lexical decision time for a pseudonoun in a grammatically congruent adjective or pseudoadjective context is equal to the lexical decision time for a pseudonoun in a grammatically incongruent adjective or pseudoadjective context.

An adjective or pseudoadjective defines a minimal grammatical context (cf. Kroll & Schwieckert, 1978) for a target noun. In terms of a distinction suggested by Seidenberg, Tannehaus, Leiman, and Bienkowski (1982), this minimal grammatical context is "nonpriming," meaning that it contains no lexical items that are semantic relatives or associates of the target item. By argument, a nonpriming context cannot have a selective influence on the lexicon; a selective influence is solely a consequence of intralexical processing. It is suggested that intralexical processing reflects the interconnections of entities in semantic memory but it does not reflect grammatical structure and pragmatic knowledge (Forster, 1979). The context that gives rise to intralexical processing--one that contains items associatively and/or semantically related to the target--is termed "lexical priming" by Seidenberg et al. (1982). In the introduction and elsewhere (Lukatela, Moraca, Stojnov, Savić, Katz, & Turvey, 1982) it has been argued that the effect on lexical decision of minimal grammatical contexts (e.g., a preposition for a noun, a pronoun for a verb) does not lend itself to the notion of processing based upon interconnections among individual lexical representations. Consequently, as Lukatela et al. (1982) remark "...semantic faciliation and grammatical facilitation are probably best understood not as expressions of a single mechanism but rather as an expression of different mechanisms that stand in a complementary relation...." (p. 299)

The sentiment of the preceding quotation is given expression in the language-processing system proposed by Forster (1979). Forster's system is composed of three sub-systems: (1) a lexical processor that accesses the representations in the lexicon of the target word and the context words (or word); (2) a syntactic processor that assigns a syntactic structure to the sentence constituted by the target word and its context; and (3) a message processor that assigns meaning to the syntactic structure. All three subsystems feed into a mechanism that, in the context of experiments, functions simply as a decision-maker (e.g., is it a word?). Differences in positive lexical decision times for target items associated with different contexts may originate in the decision making process, that is, post lexically (West & Stanovich, 1982). Consider a grammatically congruent adjective-noun pair in the present experiments. The output from the lexical processor and the output from the syntactic processor will both be positive. Because of the weak association between the words in the present experiments the output from the message processor might be negative or arise too slowly to contribute to the decision making (cf. de Groot, Thomassen, & Hudson, 1982). In contrast, for a grammatically incongruent adjective-noun pair the output from the lexical processor will be positive but the output from the syntactic processor will be negative. In order for the decision-making mechanism to arrive at an appropriate response in the situation of an incongruent adjective-noun pair it must overcome the bias toward a <u>no</u> decision engendered by the syntactic processor. Overcoming this bias will take time and consequently the lexical decision latency will be slowed relative to the situation in which the adjective and noun are in grammatical agreement.

A similar account can be given of the differences between grammatically congruent and grammatically incongruent pseudoadjective-noun pairs. Here, however, it must be assumed that the syntactic processor responds positively when there is an agreement of inflection despite the fact that the contextual item is nonsense. Thus, for the lexical decision on the second member of a grammatically congruent pseudoadjective-noun pair, the lexical processor and the syntactic processor will both feed positively to the decision maker--only the message processor's output will be negative. This is in contrast to the situation in which the inflection of the pseudoadjective and noun do not agree grammatically, for in this situation only the lexical processor's output will be positive. Consequently, the decision making will have to overcome more negative biasing and be slowed proportionately greater relative to the situation in which the pseudoadjective and noun are grammatically suited. For pseudoadjective-noun pairs lexical decision is faster when the inflections agree than when they do not agree.

Arguing from the perspective of Forster's (1979) language-processing system, it might be expected that the rejection of pseudonouns should be retarded by grammatical consistency. The negative outputs from the lexical processor and message processor will contrast with the positive output from the syntactic processor when the pseudonoun target and its context are in grammatical agreement. To arrive at the appropriate no response the decision maker will have to resolve the inconsistency of outputs and the bias to respond yes. In two previous experiments examining the effects of minimal grammatical contexts on lexical decision it was observed that pseudonouns were rejected more slowly when the preceding item was a grammatically congruent preposition (Lukatela et al., 1983) and pseudoverbs were rejected more slowly when the preceding item was a grammatically congruent personal pronoun (Lukatela et al., 1982). In the present experiments, however, there is no statistically significant evidence for the slowing of negative decisions by grammatical agreement.

To account for the indifference of rejection responses to grammatical congruency requires making explicit a process that is implicit in the above account of acceptance responses, namely, suffix stripping. According to the view of Taft and Forster (1975), perceiving an inflected adjective or noun involves decomposing the item into its stem and suffix (see also Taft, 1981; Stanners et al., 1979). In performing lexical decision, the representation of the stem morpheme is accessed by the lexical processor and the appropriateness of the inflected ending is determined on the basis of the information stored with the stem's representation. A similiar decomposition must occur for pseudoadjectives and pseudonouns except that for these items there would be no specific representation of the stem morpheme to be accessed, only close approximations.

It might be supposed that where the lexical processor focusses on the word stem, the syntactic processor focusses on the bearers of grammatical information, i.e., roughly, the suffixes of open-class words and the free morphemes of closed-class words. Whatever the bearers in any given context-target situation, assessing a grammatical fit takes time. Indeed, the difference between the present results and previous results with regard to negative responses might suggest that discovering the grammatical consistency in an adjective-pseudonoun or pseudoadjective-pseudonoun pair is slower than discovering the grammatical consistency in, say, a preposition-pseudonoun pair. The idea is that the longer the time taken by the syntactic processor to arrive at an output the less the likelihood that the activity of the syntactic processor will influence the time course of the lexical decision; an internally defined deadline on response selection must be assumed. For the preceding suggestion to be realizable it might have to be the case that (1) the grammatical link between closed-class, function words (e.g., prepositions, pronouns) and open-class, content words (e.g., nouns, verbs, adjectives) is "stronger" and more rapidly assessed than the grammatical link between open-class content words (e.g., the link between adjectives and nouns); and

(2) the syntactic processor can be influenced by the lexical processor. Recall that in the present experiments, although the lexical decision on a pseudonoun in the context of a pseudoadjective was not affected by grammatical consistency, the lexical decision on a noun in the same context was markedly affected. In short, the lexical status of the target made a difference--and that status is determined by the lexical processor.

In conclusion, evidence has been presented for the influencing of lexical decisions about inflected nouns by weakly associated inflected adjectives that are grammatically consistent or inconsistent with their target nouns. This effect seems to be mediated by a process that evaluates the grammar of a noun and its adjectival context primarily on the basis of the inflected morphemes. Although this effect demonstrated in "nonpriming contexts" (Seidenberg et al. 1982) can be referred to as grammatical priming (Lukatela et al., 1982; Lukatela et al., 1983) it appears to be a postlexical effect related to, but distinct from, the priming mechanisms of automatic spreading activation and context-induced attentional processing (Neely, 1977; Posner & Snyder, 1975) that have been identified in "lexical priming" contexts (Seidenberg et al., 1982). Lukatela et al. (1982) concluded that the grammatical priming of inflected verbs by pronouns and vice versa was automatic. Their conclusion was based in part on the observation that pronominal facilitation of verbs was virtually complete when the onsets of context and target were separated by only 300 ms. They recognized, however, that this automaticity did not refer to spreading activation. It is supposed that the present example of grammatical priming is also automatic but the kind of automaticity being referred to is closer to that suggested by de Groot et al.'s (1982) notion of an automatic checking for coherence (see also West & Stanovich, 1982) than it is to the more familiar notion of an automatic spreading of influences among connected representations in the internal lexicon.

References

Bidwell, C. E. (1970). Serbo-Croatian nominal inflection. In R. Filipovic (Ed.), _The Yugoslav Serbo-Croatian English contrastive project_. Washington, DC: Center for Applied Linguistics.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. _Psychological Review_, _82_, 407-428.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. Walker (Eds.), _Sentence processing: Psycholinguistic studies presented to Merrill Garrett_. Hillsdale, NJ: Erlbaum.

Foss, D. J. (1982). A discourse on semantic priming. _Cognitive Psychology_, _14_, 590-607.

deGroot, A. M. B., Thomassen, A. J. W. M., & Hudson, P. T. (1982). Associative facilitation of word recognition as measured from a neutral prime. _Memory & Cognition_, _10_, 358-370.

Kostić, Dj. (1965). _Frequency of occurrence of nouns in Serbo-Croatian_. Unpublished manuscript, Institute of Experimental Phonetics and Speech Pathology, University of Belgrade.

Kroll, J. F., & Schwieckert, J. M. (1978, November). _Syntactic disambiguation of homographs_. Paper presented at the Nineteenth Annual Meeting of the Psychonomic Society, San Antonio, TX.

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., Kostić, A., Feldman, L. B., & Turvey, M. T. (1983). Grammatical priming of inflected nouns. *Memory & Cognition, 11*, 59-63.

Lukatela, G., Mandić, L., Gligorijević, B., Kostić, A., Savić, M., & Turvey, M. T. (1978). Lexical decision for inflected nouns. *Language and Speech, 21*, 166-173.

Lukatela, G., Moraca, J., Stojnov, D., Savić, M., & Turvey, M. T. (1982). Grammatical priming effects between pronouns and inflected verb forms. *Psychological Research, 44*, 297-311.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226-254.

Posner, M. I., & Snyder, C. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V*. London: Academic Press.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14*, 489-537.

Stanners, R. F., Neiser, J. J., & Painton, S. (1979). Memory representation for prefixed words. *Journal of Verbal Learning and Verbal Behavior, 18*, 733-743.

Taft, M. (1981). Prefix stripping revisited. *Journal of Verbal Learning and Verbal Behavior, 20*, 289-297.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14*, 638-647.

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, 5*, 385-399.

DEAF SIGNERS AND SERIAL RECALL IN THE VISUAL MODALITY: MEMORY FOR SIGNS, FINGERSPELLING, AND PRINT*

Rena Arens Krakow† and Vicki L. Hanson

Abstract. This study investigated serial recall by congenitally, profoundly deaf signers for visually specified linguistic information presented in their primary language, American Sign Language (ASL), and in printed or fingerspelled English. There were three main findings. First, differences in the serial-position curves across these conditions distinguished the changing-state stimuli from the static stimuli. These differences were a recency advantage and primacy disadvantage for the ASL signs and fingerspelled English words, relative to the printed English words. Second, the deaf subjects, who were college students and graduates, used a sign-based code to recall ASL signs, but not to recall English words; this result suggests that well-educated deaf signers do not translate into their primary language when the information to be recalled is in English. Finally, mean recall of the deaf subjects for ordered lists of ASL signs and fingerspelled and printed English words was significantly less than that of hearing control subjects for the printed words; this difference may be explained by the particular efficacy of a speech-based code used by hearing individuals for retention of ordered linguistic information and by the relatively limited speech experience of congenitally, profoundly deaf individuals.

Hearing individuals have been shown to use a speech-based code in the short-term recall of linguistic information, whether spoken or printed (Conrad, 1964; Wickelgren, 1965). Their recall performance is similar in the two cases except for a recency advantage favoring spoken over printed items in the last serial positions (Corballis, 1966; Murray, 1966). Because the orthography of English is a secondary representation derived from the primary or basic

---

*Memory & Cognition, 1985, 13, 265-272.

†Also Department of Linguistics, Yale University.

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. Some of the results were reported at the 91st Meeting of the American Psychological Association in Anaheim, California, 1983. This study was made possible by the cooperation of Gallaudet College, and especially the members of the Gallaudet Linguistics Laboratory. We are also grateful to Nancy Fishbein and Carol Padden whose expertise in ASL was of invaluable assistance in this endeavor. Many thanks to Robert Crowder, Carol Fowler, Arthur Glenberg, Louis Goldstein, Harriet Magen, Sharon Manuel, Ignatius Mattingly, James Nairne, Michael Studdert-Kennedy, D. H. Whalen, and an anonymous reviewer, for valuable comments on earlier drafts of this paper.

spoken language (Mattingly, 1972), it is not surprising that orthographic representations are recoded into a speech-based code. In addition, a speech-based code may be especially useful when the memory task calls for recall of ordered information (Baddeley, 1979; Crowder, 1978; Hanson, 1982; Healy, 1975).

The relations among pri[illegible]anguage, coding strategy, and recall performance become more difficult to unravel when we consider bilingual deaf individuals who use American Sign Language (ASL) as a primary language and English as a secondary language. The term "primary language" refers to a natural language in the form in which it functions as a principal means of communication among members of a speech community. Writing systems and other invented representations that are based upon natural languages are viewed as nonprimary derived systems.

ASL is the primary visual-gestural language of the deaf community in the United States and Canada, and is acquired as a native language by children of deaf parents. Structural differences between signed and spoken languages reflect differences between auditory-vocal and visual-gestural channels of communication. For example, spoken languages are characterized by sequential forms of structuring at the abstract phonological and morphological levels. Words are composed of sequentially arranged phonemes, and morphological processes typically add one or more prefixes and/or suffixes (each composed of one or a series of phonemes) to a stem. In contrast, ASL is strikingly different from spoken languages in the extent to which it utilizes simultaneously structured units in lexical and morphological composition (Bellugi, 1980; Klima & Bellugi, 1979). Signs, the lexical items of ASL, are composed of several co-occurring formational parameters (Stokoe, Casterline, & Croneberg, 1965), and morphological relations are expressed by spatial and temporal modifications of the basic form of a sign (Bellugi, 1980).[1]

Those who use ASL as a primary means of communication also use fingerspelling for concepts lacking a sign. Fingerspelling is a manual form of English orthography that assigns a unique hand configuration to every letter of the English alphabet; as such, it is a changing-state representation of the graphic form of a spoken language. Fingerspelling is not used as a primary means of communication by members of the deaf community (Battison, 1978). Although fingerspelled words may often occur within signed sentences, this letter-by-letter sequential representation of English words differs considerably from the co-occuring formational parameters of ASL signs.

No writing system in use is based upon ASL, and educated deaf American signers read and write in English. But the use of ASL and of written or fingerspelled English by deaf bilinguals is quite different from the use of two spoken languages by hearing bilinguals. For a deaf person, learning the orthography (whether through writing or fingerspelling) of English means learning an orthographic visual system derived from a primary form to which he or she does not have normal access. In contrast, hearing bilinguals do have normal access to the primary forms of both languages that they use. Moreover, the significant structural differences between ASL and English at both the lexical and grammatical levels require the ASL-English bilingual to know two radically different forms of linguistic structuring. The bilingual who uses two spoken languages is required to know one form of linguistic structuring, that characterizing spoken languages.

The present research examined serial-order recall by deaf signers and addressed the question of how coding strategies and recall performance are affected by the requirement to remember ASL in contrast to English. Differences in performance that may stem from the presentation of English words by fingerspelling and print were also examined. The hypotheses underlying this work are discussed in the following sections on serial-position effects, coding, and accuracy of recall.

Serial-Position Effects

Although hearing subjects use a speech-based code for recall of both spoken and printed word lists, auditory presentation results in a recency advantage over visual presentation (for a review of this research, see Penney, 1975). This advantage for the more recently presented items occurs whether the experimenter or the subject reads the stimuli aloud. On the basis of such findings, the critical variable appears to be hearing the items. The "modality effect" was originally attributed to the fact that information in pre-categorical acoustic storage (PAS) has greater durability than information in an iconic sensory representation (Crowder & Morton, 1969).

However, further research provided evidence for similar effects in the visual modality in the absence of acoustic information, thus casting doubt on the PAS explanation for the recency advantage. Findings of recency advantages for ASL signs (Shand, 1980), moving hand shapes (Campbell, Dodd, & Brasher, 1983), lipread items (Campbell & Dodd, 1980; cf. Crowder, 1983), mouthed items (Nairne & Walters, 1983), and items vocalized "aloud" by deaf subjects (Engle, Spraggins, & Rush, 1982) are all incompatible with an explanation based on acoustic advantage.

Two alternative accounts to the PAS explanation have been proposed. First, the difference in recency favoring, for example, spoken, lipread, and signed information over orthographic information may reflect an advantage in recall of primary-language input over nonprimary (printed) input (Campbell & Dodd, 1980; Campbell et al., 1983; Nairne & Walters, 1983; Shand, 1980; Shand & Klima, 1981). Second, this effect may be attributed to an advantage in remembering changing-state information over remembering static information (Campbell & Dodd, 1980; Campbell et al., 1983; Nairne & Walters, 1983). Hereafter, the term "dynamic" will be used to mean "changing-state."

It is important to note that recall differences between lists of words that are heard and lists that are silently read are restricted to the recency portion of the curve, with a recency advantage for the words that are heard. Thus, there is an overall advantage for the heard lists. However, the recency advantage for lipread and for mouthed lists does not yield an overall advantage over printed (silently read) lists. This is because recall of lipread and mouthed lists is poorer than recall of printed lists at earlier serial positions. Researchers have tended to focus on the similarity in recency effects among mouthed, lipread, and spoken input conditions, without giving adequate attention to the fact that spoken input results in the best recall overall. The dynamic-presentation hypothesis and the primary-language hypothesis must therefore be examined with respect to effects that span the entire serial-position curve.

The present study was designed to separate serial position effects attributable to primary language from those attributable to dynamic presentation. Serial position functions that distinguished fingerspelled and printed

English lists from lists of ASL signs would provide support for the primary-language hypothesis. On the other hand, serial position functions that distinguished the signed and fingerspelled lists from the printed lists would provide support for the dynamic-presentation hypothesis.

Coding

Research with deaf signers can also provide insight into the question of whether a code based on one's primary language is useful when the recall task involves information whose linguistic structure is quite different from that of the primary language. Shand (1982; Shand & Klima, 1981) suggested that the primary code is the natural and most efficient code for short-term recall of linguistic information. Recoding by hearing individuals from print into a speech-based code takes advantage of the systematic relation between the spoken form and its orthography (Mattingly, 1972). However, there is no such systematic relation between ASL signs and English orthography.

Simultaneously occurring parameters of movement, place of articulation within the signing space, and hand configuration are the sublexical components of ASL signs (Stokoe et al., 1965). These formational parameters (cheremes or primes) evidently support recall of signs by deaf signers much as phonetic parameters of speech support recall of spoken information by hearing individuals (Bellugi, Klima, & Siple, 1975; Hanson, 1982; Poizner, Bellugi, & Tweney, 1981; Shand, 1982). Thus, Bellugi et al. (1975) found intrusion errors suggesting sign-based coding of ASL signs by deaf signers on a serial-recall task. The majority of the intrusion errors were signs that differed from a correct response by one formational parameter. For example, some of the subjects reported JEALOUS for CANDY. The signs for JEALOUS and CANDY are a minimal pair in that they have the same place of articulation and movement; they differ only in hand configuration. Likewise, some subjects reported NEWSPAPER for BIRD; these two signs share movement and hand configuration and differ only in place of articulation.

Evidence for both sign-based and speech-based recoding of printed words by deaf subjects has been obtained in serial-order recall tasks (Hanson, 1982; Lichtenstein, in press; Shand, 1982). Subject characteristics associated with coding preferences suggest that speech-based recoding is typically used by those prelingually, profoundly deaf adults who are better readers and who have better speech production skills (Lichtenstein, in press). A shortcoming of previous studies was that they compared the performance of different groups of subjects on the different stimulus types. Furthermore, they never included fingerspelled English. Presenting ASL signs, printed English words and fingerspelled English words to the same group of deaf signers in the present study made it possible to ascertain whether deaf individuals changed strategies as the stimuli changed or maintained a preferred strategy, such as sign-based or speech-based coding. In order to provide English words that were compatible with a sign-based code, half of the fingerspelled and printed words were chosen because they had readily available sign translations ("high-signability" words); the other half, because they did not ("low-signability" words). If deaf subjects recode into signs and recoding into one's primary language is the most natural and efficient strategy (Shand, 1982), then two outcomes might be predicted. First, high-signability words should be recalled more accurately than low-signability words. Second, recall performance on high-signability words should provide evidence of sign-intrusion errors.

Accuracy of Recall

In general, when congenitally, profoundly deaf individuals perform a task that calls for ordered recall of English words or letters, they do not perform as well as hearing subjects (Belmont & Karchmer, 1978; Belmont, Karchmer, & Pilkonis, 1976; Hanson, 1982; MacDougall, 1979; Wallace & Corballis, 1973). Belmont and Karchmer argued that the generally poorer performance of deaf individuals reflects a "mismatch" between the native language (ASL) and the language of the information to be recalled (English). However, even on serial-recall tasks involving ASL signs, deaf signers do not remember as many items as hearing subjects tested on the signs' printed (Hanson, 1982) or spoken English equivalents (Bellugi et al., 1975). Moreover, Hanson found that deaf subjects did perform as well as hearing subjects on tasks that called for free recall of printed English words. The nature of the ordered-recall task, rather than characteristics of the input, may actually favor hearing individuals.

Recent studies indicate that the speech code is particularly useful for retaining order information (Baddeley, 1979; Crowder, 1978; Hanson, 1982; Healy, 1975). For deaf subjects, accuracy of recall has been found to correlate with the use of a speech-based code; those who use this code efficiently recall more than those who use it inefficiently or not at all (Conrad, 1979; Hanson, 1982; Lichtenstein, in press). Therefore, it seems that the speech-based code may facilitate serial-order recall in a way that alternative coding mechanisms, including sign-based coding of ASL signs, do not. Furthermore, it is likely that the use of the speech-based code by deaf individuals is not as effective as it is for hearing people. The present study examined the recall performance of deaf subjects, who were highly proficient in English as well as in ASL, and asked whether accuracy of recall differs as a function of the type of linguistic input (dynamic vs. static; primary vs. nonprimary) or whether serial recall is, regardless of input characteristics, a particularly difficult task for individuals who do not have normal access to speech.

Experiment

This experiment compared the performance of congenitally, profoundly deaf signers when presented with English words and ASL signs for serial-order recall. The presentation mode of the English words was varied so that some were printed and others were fingerspelled. All the deaf subjects used ASL as their primary means of communication. The recall performance of two groups of deaf subjects was compared in order to find out whether there are performance differences between native and nonnative signers. Members of one group acquired ASL as a native language from deaf parents, and members of the other group learned ASL outside the home in the early school years. A normal-hearing control group was tested on the printed stimuli.

Method

Subjects

All subjects were tested individually and were paid for their participation.

Deaf subjects. Twenty congenitally, profoundly deaf subjects participated in the short-term memory experiment; two were eliminated because their hearing loss was less than the criterion for profound deafness (85 dB, better-ear average). Background information gathered from the subjects indicated that they all used ASL as their primary means of communication, supplemented by fingerspelling. Eight of the subjects were born to deaf parents and had acquired ASL as a native language (native signers), while 10 of the subjects had hearing parents and learned ASL outside the home in the early school years (nonnative signers). All subjects were currently attending or were recent graduates of Gallaudet College, a liberal arts college for deaf students.

Twenty congenitally deaf adults served as control subjects on a perceptual task, described below. Nine of these subjects had participated in the memory experiment several months before. Each had a hearing loss of at least 70 dB in the better ear. They were all students or graduates of Gallaudet College and reported using ASL as a primary means of communication.

Hearing Subjects. Ten hearing subjects were recruited from among Yale University students and affiliates. They were native speakers of English who reported no history of hearing impairment. Because the hearing subjects were tested on both sets of printed stimuli, 10 subjects provided sufficient data for comparison with the deaf subjects.

Stimuli

Stimulus lists were constructed from 141 high-signability (HS) English nouns and 94 low-signability (LS) English nouns. All were words considered to be commonly known by college-age adults, and were selected with the assistance of a deaf native signer. HS words were matched with LS words for frequency of occurrence in printed English (Kučera & Francis, 1967). HS words were randomly assigned to each of three presentation conditions: signs, fingerspelling, and print. LS words were randomly assigned to fingerspelling or print conditions. These assignments produced one set of stimuli. A second set of stimuli was constructed by reassigning printed items to fingerspelling or signs, reassigning fingerspelled items to signs or print, and reassigning signed items to print or fingerspelling, in order to partially counterbalance the assignment of words to presentation conditions. Thus, the following five conditions were obtained for both sets of stimuli: (1) American Sign Language signs; (2) HS fingerspelled English words; (3) LS fingerspelled English words; (4) HS printed English words; and (5) LS printed English words. Each condition contained 42 nouns, in seven lists of 6 nouns each. Previous work with deaf subjects indicated that a list containing 6 nouns could be expected to produce both primacy and recency serial position effects (Bellugi et al., 1975). An additional 5 lists of 5 nouns provided practice blocks.

Procedure

All stimulus lists were videotaped at a rate of 2 sec per trial. A native signer recorded the signed and fingerspelled lists on videotape; for maximal visibility, she was framed from forehead to waist. The signer maintained a neutral expression throughout the taping session. Printed words were videotaped directly from an Atari 400 computer and were displayed for 1.5 sec with a .5-sec interstimulus interval. Stimuli in each condition were recorded in seven continuous lists of six nouns each. One practice list preceded each of the five conditions.

The order in which stimulus conditions were presented was partially balanced across subjects as follows: There were five orders of presentation for each stimulus set and no condition ever occurred in the same ordinal position twice. Four subjects were tested on each of the orders. Order 1 was based on differences in mode of presentation: (a) Signs; (b) HS Fingerspelling; (c) LS Fingerspelling; (d) HS Print; (e) LS Print. Order 2 was also based on mode differences but it involved a rearrangement of the ordering of signs, fingerspelling, and print modes: (a) HS Print; (b) LS Print; (c) HS Fingerspelling; (d) LS Fingerspelling; (e) Signs. Order 3 arranged lists by signability differences: (a) LS Print; (b) LS Fingerspelling; (c) HS Print; (d) Signs; (e) HS Fingerspelling. Order 4 arranged lists by signability in a different ordering than order 4: (a) HS Fingerspelling; (b) HS Print; (c) Signs; (d) LS Print; (e) LS Fingerspelling. Order 5 mixed modes and signability in a random fashion: (a) LS Fingerspelling (b) Signs; (c) LS Print; (d) HS Fingerspelling; (e) HS Print.

Deaf subjects were tested on all five conditions by a native signer who provided both printed and signed instructions; nine of the subjects were tested on one set of stimuli and nine on the other. The subjects were told that they would see lists of nouns presented by various modes: ASL signs, printed English, and fingerspelled English. A message printed on the screen indicated the termination of each list. The subjects were instructed to watch the screen and to write the words they had just seen, in serial order, on the answer sheet provided. The answer sheet included the numbers 1 through 6 for each list with blank spaces for responses. The subjects were not prevented from recording words in any order. It was, however, required that words appear in their correct serial positions. Bellugi and Siple (1974) reported that deaf signers' recall performance with written report of signs was as good as their recall perfromance with signed report.

To control for possible dialectal variations on the interpretations of the signs and to ensure a fair scoring procedure, a control group of deaf subjects was tested in a perceptual task. These subjects were asked to watch the signed portions of the videotapes and to simply write down the English translation of each sign.

The hearing subjects were tested by a hearing experimenter who provided both printed and spoken directions. Stimuli for the hearing subjects, who served as partial controls in this experiment, consisted of the printed conditions only. Each hearing subject saw both sets of printed stimuli.

Scoring

All subjects' responses in the memory task were scored as follows: Items were marked correct if they appeared in the proper serial position in the current list. Dialectal differences were taken into account when scoring the answer sheets from signed trials; a response on the memory task that matched a response in the correct serial position on the perceptual task was scored as correct. Because there were seven lists in each condition, seven was the maximum score possible at each serial position for each condition.

Results

A three-way ANOVA examined the within-subjects effects of presentation condition (ASL signs, printed English, fingerspelled English), and serial position (one through six), and the between-subjects effect of group (native or nonnative signers) on the number of words the deaf subjects recalled accurately. For the purposes of this analysis, performance on high- and low-signability lists was averaged. The analysis revealed a significant main effect of serial position, $F(5, 80) = 30.01$, $p < .0001$, and no significant effect of either group or condition (both *Fs* < 1.00). These latter results indicated that native and nonnative signers could not be differentiated on the basis of their performance on these serial-recall tasks and that their recall accuracy was similar for the three presentation conditions. There was, however, a significant condition X position interaction, $F(10, 160) = 3.33$, $p < .001$, indicating differential effects on the serial-position curve as a function of condition. This interaction is shown in Figure 1, in which mean recall is plotted at each serial position for the three conditions: ASL signs, fingerspelled English words, and printed English words. In this figure, we have pooled the high- and low-signability trials and averaged across the two groups of deaf subjects.

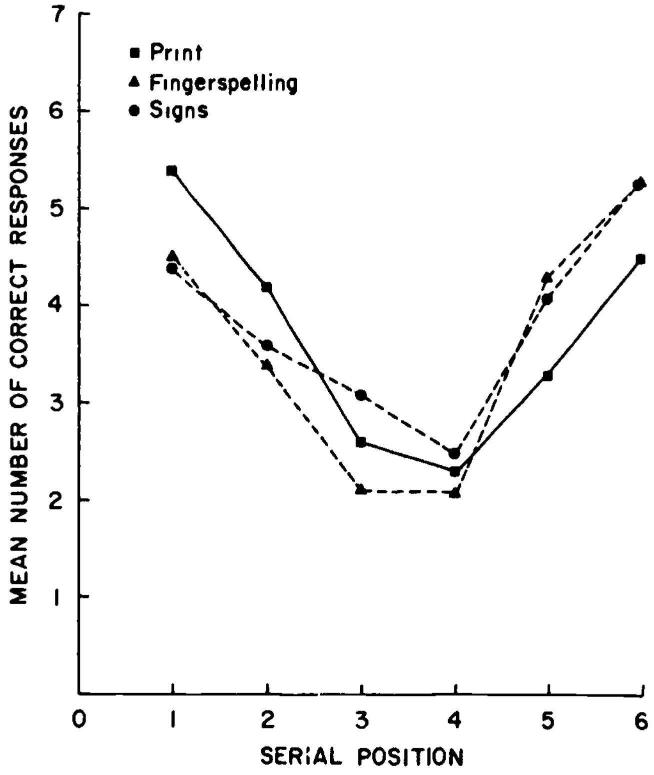


Figure 1. Mean number of printed, fingerspelled, and signed items correctly recalled by deaf subjects at each serial position.

Additional analyses were undertaken in order to understand the nature of the interaction. The competing hypotheses regarding the effects of primary language vs. those of dynamic presentation prompted examination of the differences in serial-position effects as a function of condition. To test the primary language hypothesis, one ANOVA compared performance on the print condition to that on the fingerspelling condition. This was a three-way analysis, as above, with the exclusion of the sign condition. In this comparison, the condition X position interaction was also highly significant, $F(5, 80) = 6.84$, $p < .0001$. Thus, the serial-position curves for the print and fingerspelling conditions differed. Performance on the signed and fingerspelled trials was compared in the same way. In this ANOVA, the condition X position interaction disappeared, $F(5, 80) = 1.69$, $p > .05$. This lack of a significant interaction indicates no difference in the serial-position curves for the signed and fingerspelled trials. To complete the comparison of dynamic and static conditions, an ANOVA was performed on the printed and signed trials; the results showed a significant condition X position interaction, $F(5, 80) = 3.66$, $p < .01$. As is evident in the figure, the deaf subjects were never at ceiling in their recall performance.

Taken together, these analyses indicate that the condition X position interaction in the original analysis was due to differences between the print condition on the one hand and the fingerspelling and sign conditions on the other. This is consistent with the hypothesis that recall of dynamic and static forms of linguistic information produces different serial-position curves. In order to localize the effects of dynamic vs. static input on the serial-position curve, contrasts were done at each serial position, going back to the original analysis, by comparing recall performance in the static condition (print) with that in the dynamic conditions (fingerspelling and signs). The contrast was significant at Position 1, $F(1, 34) = 10.49$, $p < .01$ and Position 2, $F(1, 34) = 4.99$, $p < .05$, with accuracy greater in the print condition than in the other two conditions. The contrast was also significant at Position 5, $F(1, 34) = 10.05$, $p < .01$, and Position 6, $F(1, 34) = 8.67$, $p < .01$, with accuracy greater in the sign and fingerspelling conditions than in the print condition. The contrast was not significant at Position 3, or Position 4 (both F's < 1.00). These results indicate that there is a recency advantage for the dynamic information (signed and fingerspelled) but a primacy advantage for the static information (printed). The existence of some recency gains in all conditions probably reflects the relatively short list length and the freedom of subjects to record the items they remembered in any order they wished.

To test specifically for the effects of signability on recall, a 3-way ANOVA was performed on the recall accuracy for the within-subjects factors of signability (HS, LS) X mode (fingerspelling, print) X serial position (1-6). Because the group factor never entered into any significant main effects or interactions, native and nonnative subjects were pooled in this and subsequent analyses. The main effect of signability was nonsignificant ($F < 1.00$); thus, the availability of a direct sign translation for an English word did not enhance its recall. Mean recall of all deaf subjects on the 6-item lists was 3.16 for the HS stimuli and 3.12 for the LS stimuli. The main effect of mode was also nonsignificant ($F < 1.0$), and the ANOVA revealed a significant mode X position interaction, $F(5, 85) = 7.42$, $p < .0001$, reflecting the differences in serial-position effects between static printed input and dynamic fingerspelled input. As in the previous analysis, the main effect of serial position was highly significant, $F(5, 85) = 30.91$, $p < .0001$.

The analysis of signability indicated that if deaf subjects were using a sign-based code to recall English words, it was not to their advantage. However, no evidence of sign-based coding of fingerspelled or printed English words was obtained in an analysis of the intrusion errors. Two deaf native signers of ASL examined each error on the sign trials and on the HS fingerspelling and print trials and judged whether or not each was formationally similar to the target item (i.e., a sign intrusion). Disagreements between the two signers were rare (occurring on only 4 of the 63 errors that did not include misorderings or blanks) and when they occurred, they were resolved by consulting a vocabulary book on ASL signs (O'Rourke, 1978). Error analysis of the sign trials showed that of the 63 errors, 30 were sign intrusions. The results of the perceptual task indicated that these sign intrusions were not due to perceptual confusions. (Many of the remaining errors consisted of words that were formationally similar to a word in another position in the same list.) Table 1 lists examples of sign intrusion errors and the corresponding target signs for the same serial positions in the recorded list of signs. In contrast, errors made on the fingerspelled and printed English conditions did not tend to be sign intrusions. The 79 errors on the HS trials (not counting misorderings and blanks) included only a single response that had a sign similar to that of the target item. This was the intrusion of "caution" for "warning," which is also semantically related. The other 78 error could not be differentiated in kind from errors on corresponding LS printed and LS fingerspelled lists. Errors made on fingerspelled and on printed lists appeared to be of the same general type, as indicated by the examples of errors on HS lists provided in Table 2. Patterns of visual resemblance of item and error pairs are obvious. Such errors could reflect either visual or phonological confusions; the present experiment was not designed to distinguish between these two possibilities. Taken together, these results suggest that well-educated deaf signers employ sign-based coding in retention of ASL signs but not in retention of English words, whether printed or fingerspelled.

Finally, recall accuracy of the deaf subjects on the printed trials was compared with that of the hearing subjects. Collapsing the data across all deaf subjects, mean recall on the six-item printed blocks was 3.14. (It should be remembered that for the deaf subjects, mean recall did not differ significantly as a function of condition: average recall on the fingerspelling and sign conditions was 3.10 and 3.17, respectively.) Mean recall of the hearing subjects on the printed blocks was 4.87, and many of them were at ceiling. An analysis comparing mean recall of the deaf subjects with that of the hearing subjects indicated that there was a significant difference in the accuracy of subjects as a function of group (deaf or hearing), $\underline{t}(26) = 6.85$, $p < .0001$. No valid tests of parallel serial position differences could b used due to the ceiling performance of so many hearing subjects.

Discussion

In the present experiment, there was no significant difference in performance between the native and nonnative signers tested. This suggests that native signers and nonnative signers who learned ASL at an early age form a homogeneous subject group; as far as these tasks are concerned, ASL functions as a primary language in the same way for both.

Serial-position effects were examined in order to test the dynamic-presentation hypothesis against the primary-language hypothesis by comparing deaf signers' recall of English print, fingerspelling, and ASL signs. The

Table 1

Item and Error Pairs in Recall of ASL Signs

Target Item	Intrusion Error	Parameter(s) of Difference
danger	algebra	movement
zero	photograph	handshape
telegram	declination	handshape
secret	patience	movement
debt	this	movement
instructions	iceskating	handshape
pope	princess	movement, location
fence	screen	handshape, location
rosary	interpreter	movement, location
sandwich	school	movement, location

Table 2

Item and Error Pairs in Recall of Fingerspelled and Printed English Words

FINGERSPELLING		PRINT	
Target	Error	Target	Error
diamond	almond	heart	horse
wrestling	recycling	concept	corn
ceremony	cemetery	leaf	leather
pipe	pope	interference	inference
bomb	bubble	rosary	rosemary
noon	noun	digit	dignity
temptation	temperature	outlaw	outline
vinegar	vineyard	cure	burn
cure	sure	antique	unique

results revealed that the serial-position curves were similar for the two types of dynamic stimuli (fingerspelling and signs) and that these curves differed from those obtained for the static stimuli (print). Recall was better for dynamic stimuli in the last two serial positions but worse in the first two serial positions.

The recency advantages found for fingerspelled English words and ASL signs add to a growing body of results indicating that "modality effects" can be obtained even in the absence of acoustic input (Campbell & Dodd, 1980; Campbell et al., 1983; Engle et al., 1982; Nairne & Walters, 1983; Shand 1980). However, the present results are inconsistent with the primary-language hypothesis, according to which differences would have been expected between the serial-position curves for the primary-language items (ASL signs) and those for the nonprimary-language items (fingerspelling and print). Rather, the present findings provide support for the hypothesis that the "modality effect" is a reflection of a recency advantage that accrues to dynamically presented information, regardless of input modality. The primacy advantage found for printed stimuli over fingerspelled and signed stimuli resembled the primacy advantage for printed over lipread and mouthed stimuli reported in previous studies (Campbell & Dodd, 1980; Nairne & Walters, 1983). As mentioned earlier, the comparison between hearing subjects' recall of spoken and of printed words reveals only a recency difference between the two conditions, and consequently, an overall advantage for the spoken words. But it appears that in spite of the recency advantage for nonacoustic dynamic stimuli (e.g., signs and lipread, mouthed, and fingerspelled words), such stimuli show no overall advantage over static stimuli (printed words). What is important to note in all of these studies is that dynamic information (whether spoken, signed, fingerspelled, etc.) and static information (printed) yield different serial-position curves.

As in previous research (Bellugi et al., 1975), analysis of the deaf subjects' intrusion errors revealed sign-based coding of the ASL signs. However, the lack of sign intrusion errors on both printed and fingerspelled English lists suggests that well-educated deaf persons do not recode English words into signs. In addition, there was no recall advantage for those English words that have direct sign translations. These results are especially noteworthy because they suggest that deaf bilinguals can change their recall strategies depending upon whether they are presented with information in English or in ASL.

The number of items recalled by deaf signers did not differ as a function of language, signability, or dynamic-static differences. But their mean recall was significantly less than that of hearing subjects when the performance of both groups on the printed trials was compared. These results are not consistent with the view that the generally poorer performance on serial-recall tasks by deaf subjects than by hearing subjects stems from the requirement to remember English. In conjunction with earlier findings that deaf signers perform as well as hearing individuals on free-recall tasks involving English stimuli (Hanson, 1982), the present study indicates a specific difficulty on the part of the deaf signers with serial-order recall.

It is important to realize that difficulties deaf individuals may have with serial-recall tasks need not interfere with their primary-language abilities in ASL because of ASL's emphasis on simultaneous production of linguistic units. But serial-recall performance may become a problem when deaf individuals learn a spoken language. English, even more than some other spoken languages, relies heavily on word order in syntactic structuring. Not surprisingly, deaf children have difficulty in learning to read and write the complex syntactic structures of English, which place a heavy load on memory for ordered units (Russell, Quigley, & Power, 1976), and deaf individuals usually do not read as well as their hearing peers (Bornstein & Roy, 1973; Karchmer, Milone, & Wolk, 1979). If we are to improve our methods for teaching deaf persons to read and write, it is crucial that we gain more insight into the strategies that deaf individuals bring to bear when remembering English letters, words, and sentences, and the ways in which deafness affects the perception of and memory for sequential flow of linguistic information.

References

Baddeley, A. D. (1979). Working memory and reading. In P. A. Kolers, M. Wrolstad, & H. Bouma (Eds.), Processing of visible language, (Vol. 1, pp. 355-370). New York: Dienem Publishing Co.

Battison, R. (1978). Lexical borrowing in American Sign Language. Silver Spring, MD: Linstok Press.

Bellugi, U. (1980). Clues from the similarity between signed and spoken language. In U. Bellugi & M. Studdert-Kennedy (Eds.), Signed and spoken language: Biological constraints on linguistic form (pp. 115-140). Weinheim: Verlag Chemie.

Bellugi, U., Klima, E. S., & Siple, P. (1975). Remembering in signs. Cognition, 3, 93-125.

Bellugi, U., & Siple, P. (1974). Remembering with and without words. In F. Bresson (Ed.), Current problems in psycholinguistics (pp. 215-235). Paris: Centre National de la Recherche Scientifique.

Belmont, J. M., & Karchmer, M. A. (1978). Deaf people's memory: There are problems testing special populations. In M. Cruneberg & R. Sykes (Eds.), Practical aspects of memory (pp. 581-588). London: Academic Press.

Belmont, J. M., Karchmer, M. A., & Pilkonis, P. A. (1976). Instructed rehearsal strategies' influence on deaf memory processing. Journal of Speech and Hearing Research, 19, 36-47.

Bornstein, H., & Roy, H. L. (1973). Comment on "linguistic deficiency and thinking: Research with deaf subjects 1964-1969." Psychological Bulletin, 79, 211-214.

Campbell, R., & Dodd, B. (1980). Hearing by eye. Quarterly Journal of Experimental Psychology, 32, 85-99.

Campbell, R., Dodd, B., & Brasher, J. (1983). The sources of visual recency: Movement and language in serial recall. Quarterly Journal of Experimental Psychology, 35A, 571-587.

Conrad, R. (1964). Acoustic confusions in immediate memory. British Journal of Psychology, 55, 75-84.

Conrad, R. (1979). The deaf schoolchild. London: Harper and Row.

Corballis, M. C. (1966). Rehearsal and decay in immediate recall of visually and aurally presented items. Canadian Journal of Psychology, 20, 43-51.

Crowder, R. G. (1978). Language and memory. In J. F. Kavanagh & W. Strange (Eds.), Speech and language in the laboratory, school, and clinic (pp. 331-376). Cambridge, MA: MIT Press.

Crowder, R. G. (1983). The purity of auditory memory. Philosophical Transactions of the Royal Society of London, B302, 251-265.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). Perception & Psychophysics, 5, 365-371.

Engle, R. W., Spraggins, A., & Rush, P. (November, 1982). *A test of sensory store theories of modality effects.* Paper presented at the meeting of the Psychonomic Society, Minneapolis, Minnesota.

Hanson, V. L. (1982). Short-term recall by deaf signers of American Sign Language: Implications of encoding strategy for order recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 572-583.

Healy, A. F. (1975). Coding of temporal-spatial patterns in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 481-495.

Karchmer, M. A., Milone, M. N., Jr., & Wolk, S. (1979). Educational significance of hearing loss at three levels of severity. *American Annals of the Deaf*, *124*, 97-109.

Klima, E. S., & Bellugi, U. (1979). *The signs of language.* Cambridge, MA: Harvard University Press.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day English.* Providence, RI: Brown University Press.

Lichtenstein, E. H. (in press). The relationships between reading processes and English skills of deaf college students: Parts I & II. *Applied Psycholinguistics.*

MacDougall, J. C. (1979). The development of visual processing and short-term memory in deaf and hearing children. *American Annals of the Deaf*, *124*, 16-22.

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

Murray, D. J. (1966). Vocalization at presentation and immediate recall with varying recall methods. *Quarterly Journal of Experimental Psychology*, *18*, 9-18.

Nairne, J. S., & Walters, V. L. (1983). Silent mouthing produces modality- and suffix-like effects. *Journal of Verbal Learning and Verbal Behavior*, *22*, 475-483.

O'Rourke, T. J. (1978). *A basic vocabulary: American Sign Language for parents and children.* Silver Spring, MD: T. J. Publishers, Inc.

Penney, C. G. (1975). Modality effects in short-term verbal memory. *Psychological Bulletin*, *82*, 68-84.

Poizner, H., Bellugi, U., & Tweney, R. (1981). Processing of formational, semantic, and iconic information in ASL. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1146-1159.

Russell, W. K., Quigley, S. P., & Power, P. J. (1976). *Linguistics and deaf children: Transformational syntax and its applications.* Washington, DC: Alexander Graham Bell Association for the Deaf.

Shand, M. A. (1980). Short-term coding processes in congenitally deaf signers of ASL: Natural language considerations. (Doctoral dissertation, University of California, San Diego, 1980) *Dissertation Abstracts International*, *41.4*, 1572A-1573A.

Shand, M. A. (1982). Sign-based short-term coding of ASL signs and printed English words by congenitally deaf signers. *Cognitive Psychology*, *14*, 1-12.

Shand, M. A., & Klima, E. S. (1981). Nonauditory suffix effects in congenitally deaf signers of ASL. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 464-474.

Stokoe, W. C., Casterline, D. C., & Croneberg, C. G. (1965). *A dictionary of American Sign Language.* Washington, DC: Gallaudet College Press.

Studdert-Kennedy, M., & Lane, H. (1980). Clues from the differences between signed and spoken language. In U. Bellugi & M. Studdert-Kennedy (Eds.), Signed and spoken language: Biological constraints on linguistic form (pp. 29-39). Weinheim: Verlag Chemie.

Wallace, G., & Corballis, M. C. (1973). Short-term memory and coding strategies in the deaf. Journal of Experimental Psychology, 99, 334-348.

Wickelgren, W. A. (1965). Acoustic similarity and intrusion errors in short-term memory. Journal of Experimental Psychology, 70, 102-108.

Footnote

[1] Sequential structuring does, of course, play a role in ASL, much as simultaneous structuring does in speech. The essential difference is in the extent to which sequential structure or parallel structure is part of the abstract organization of the language. Studdert-Kennedy and Lane (1980) suggest that speech draws on parallel organization (coarticulation, for example) to implement an abstract sequential linguistic structure, while ASL draws on sequential organization of its gestures to implement an abstract parallel linguistic structure. For example, in ASL the formation of a sign's handshape may precede the start of its movement. Clearly, there is also a sequential component in ASL syntax.

DID ORTHOGRAPHIES EVOLVE?*

Ignatius G. Mattingly†

Abstract. According to Gelb (1963), writing has "evolved" from picture writing to logography to syllabic writing to alphabetic writing. It is argued here that this widely accepted theory of orthographic evolution does not really fit the historical facts very well, and that the variety of orthographies is better explained on linguistic grounds. Orthographies have to be productive, and they can manage this only by providing devices for transcribing the possible words in the lexicon. The very limited number of different ways in which this is accomplished in different orthographies is accounted for by the structural peculiarities of the languages that the orthographies transcribe.

It is generally believed by linguists, psychologists, psycholinguists and educators that writing has "evolved." First there was picture writing, then came logographies, then syllabaries, and finally, the alphabet. At each of these stages of development, writing became more efficient, because a smaller inventory of signs was required to do the job. The alphabet is the culmination of this evolutionary process, and its nearly universal triumph over less efficient orthographies has been well deserved.

The evolutionary view of writing probably originated during the nineteenth century, when most of the decipherments that led to our present knowledge of ancient writing systems took place, and theories of cultural evolution, inspired by the theory of biological evolution, were in vogue. The evolutionary view can be found in one form or another in many of the standard accounts of the history of writing. Thus Jensen (1970):

> In the broader history of writing we can see then certain evolutionary tendencies emerging. Above all it is governed by the law of least resistance, according to which every change must in the normal way run from the more difficult to the more easy, from the more complicated to the more simple; we find, furthermore, in keeping with the general development of civilization, an increasing abstraction, a certain assimilation of the form to the self-increasing intellectuality of the content. (p. 22)

(Cf. also Pedersen, 1962, chap. VI.) And the evolutionary view has been elaborated into a theory by Gelb, whose A Study of Writing (1963) most of us

---

*To appear in RASE Topics in Learning and Learning Disorders.

†Also University of Connecticut

Acknowledgment. Support from NICHD Grant HD-01994 is gratefully acknowledged. Arthur Abramson, Alvin Liberman, Isabelle Liberman, and Harriet Magen read early drafts of this paper and offered valuable criticisms and useful examples.

who are interested in the psychology of reading turn to for enlightenment about the natural history of writing.

Gelb says that "writing had its origin in simple pictures" (p. 190), advanced to "semasiography" (that is, picture writing), and then to "phonography," which comprehends word-syllabic, syllabic, and alphabetic writing (p. 191). The development of writing is said to be "unidirectional" (p. 200):

> What this principle means in the history of writing is that in reaching its ultimate development writing, whatever its forerunners may be, must pass through the stages of logography, syllabography, and alphabetography in this, and no other, order. Therefore, no writing can start with a syllabic or alphabetic stage unless it is borrowed, directly or indirectly, from a system which has gone through all the previous stages. A system of writing can naturally stop at one stage without developing farther. Thus a number of writings stopped at the logographic or syllabic stage. (p. 201)

Thus, just as biological evolution explains the variety of natural species, orthographic evolution is said to explain the variety of orthographic species.

What I wish to do here is to reconsider the theory of orthographic evolution. I will argue that the evolution of writing has been more apparent than real, and that the variety of orthographic species is better understood from a standpoint more linguistic than Gelb adopts. The alphabet, I will suggest, is not necessarily the best way to write all languages. For the evidence that leads to these conclusions, I rely mainly on the remarkable erudition of Gelb himself.

Is this a matter of more than marginal concern for the psychology of reading and spelling? I suggest that it may be, for the evolutionary view is echoed by psychologists concerned with the reading process (Crowder, 1982, p. 148; Henderson, 1982, p. 7), and the supposed evolution of writing is sometimes taken to reflect psychological facts and even to suggest teaching strategies. Citing Gelb (1963), Gleitman and Rozin (1977) say:

> ...each orthography arose as a gradual refinement and generalization of resources already implicitly available in its predecessors, as though the early scripts formed the necessary conceptual building blocks required for further development....On these grounds, one can build a plausibility case (though only that) for organizing reading instruction in terms of a similar accumulation of conceptions: perhaps ontogeny recapitulates cultural evolution. (p. 8)

Let us begin with the claim that logography evolved from picture writing. There are seven ancient traditions of logographic writing: the Mesopotamian, Proto-Elamite, Proto-Indic, Sino-Japanese, Egyptian, Cretan, and Hittite. Decipherment has not progressed very far in the cases of Proto-Elamite, Proto-Indic, and the early Cretan writing, but in the case of the other logographic traditions there is evidence that the signs were at first iconic (Gelb, 1963, chap. II , only later becoming arbitrary and non-iconic. The obvious explanation for this development is that while iconic signs were suitable for monumental inscriptions, hieratic, commercial, and literary uses required signs that could be rapidly written rather than slowly drawn. There was thus an evolution from iconic to non-iconic writing. But regardless of their graphic form, the signs were from the beginning logograms: they stood for words (or more correctly, morphemes), not, as is sometimes said, "concepts" or "meanings." An iconic sign designated a particular word by suggesting some aspect of its meaning, but the meaning of the logographic text did not depend on these pictorial hints, but on the selection and ordering of the words, just as it does in spoken and written language in general. Non-iconic signs, arbitrarily associated with words, served the purpose equally well.

Picture writing, on the other hand, is non-linguistic. The term is a convenient cover label for a fascinating miscellany of assorted artifacts from preliterate societies: rock-drawings warning of danger nearby, pictorial "letters," narratives and proverbs, tribal and commercial identification symbols, calendar systems, and so on (Gelb, 1963, chap. II).

In what sense can logography be said to have evolved from picture writing? The claim would have some substance if it could be shown that the signs of some logography were borrowed from or paralleled those of a particular tradition of picture writing, but there appears to be no example of this sort in any of the logographic traditions. The Mesopotamian Sumerians used both cylinder seals and logographic writing on commercial identification tags, but there is no relationship between the seals and the writing (Gelb, 1963, p. 65). If cultural evolution means anything, it must imply some kind of structural development: thus the computer can reasonably be said to have evolved from the loom. But linguistic writing merely took over the communicative functions of picture writing, as the internal combustion engine took over the locomotive functions of the horse; it did not, in any interesting sense, evolve from picture writing.

The second part of Gelb's theory is that syllabaries evolved from logographies. This claim implies that within a particular orthographic tradition, there is a period of strictly logographic writing, then, perhaps, a transitional period, and then a period of strictly syllabic writing. But what we actually find, in the Mesopotamian, Hittite and Sino-Japanese traditions (Egyptian will be discussed shortly) is just the transitional period.

The writing in these traditions is what Gelb aptly calls "word-syllabic" writing, in which logograms and syllabary signs supplement each other. Thus, in Sumerian and in Japanese writing, the syllable signs are used regularly to write inflectional morphemes and can also be used to write base morphemes. Alternatively, a base morpheme can be written with a logogram, and in this case, a supplementary syllable sign is sometimes used to indicate the phonological form of the morpheme. In Chinese writing, some of the characters are simple logograms, but most of them consist of two component signs: the "radical," one of 214 signs that serve as semantic classifiers, and the "phonetic complement," a sign that in isolation has a phonological value similar or identical to that of the compound character. The compound character for /ku$_3$/, blind, for instance, is composed of the simple signs for /ku$_3$/, drum and /mu$_4$/, eye (Jensen, 1970, p. 170).[1] Since the phonetic complements have logographic values of their own, and there are in general quite a few phonetic complements for a particular syllable (10 for /li$_4$/ for example; Wieger, 1927), it might seem a bit eccentric to regard Chinese writing as systematically syllabic, rather than simply as a case of massive phonetic transfer. But the fact that a common error in the writing of Chinese is the use of an incorrect but phonologically accurate phonetic complement (H.-B. Lin, personal communication) attests to the psychological reality of the syllabary system.

In all these word-syllabic orthographies, the syllable signs clearly derive from logograms. Thus the syllable sign for /gal/ in Sumerian derives from the logogram for /gal/, *great* (Gelb, 1963, pp. 110-111); one of the phonetic complements for /ku$_3$/ in Chinese, as we have seen, derives from the logogram for /ku$_3$/, *drum*; and the Japanese kana for /mo/ derives from the character for /mo/, *hair*, borrowed from Chinese /mao$_2$/, *hair* (Jensen, 1970, p. 201). But is derivation necessarily to be equated with evolution? Gelb himself makes it quite clear that there is no period in any of these traditions during which the writing was strictly logographic; syllable signs occur in the earliest specimens (Gelb, 1963, pp. 67, 83, 85). Nor did any of these traditions lead eventually to a strict syllabary, though some of the later Mesopotamian systems came fairly close (p. 165).

The Cretan tradition is perhaps the one case that supports the claim. Whether there was a strictly logographic stage cannot be determined until the early Minoan scripts are deciphered, but the strictly syllabic Cypriote orthography appears to have developed from the earlier word-syllabic stage represented by Cretan Linear B (Gelb, 1963, p. 154).

Finally, Gelb's theory claims that alphabetic writing evolves from syllabic writing. But this part of the theory depends crucially on Gelb's particular interpretation of the structure of the Egyptian and West Semitic orthographies, and on his presumption that the latter derive from the former.

In the Afro-Asiatic family of languages, to which both Egyptian and Semitic belong, the base morphemes are, in general, simply consonantal patterns, for example, Egyptian n-f-r, *lute*; p-r, *house*; and Semitic k-t-b, *to write*; m-l-k, *to rule*. In actual words, vowels are morphologically inserted and, together with prefixes and suffixes, distinguish the various forms derived from the base. Thus the base k-t-b yields in Hebrew [ka'tav], *he wrote*; [jix'tov], *he will write*; [jik'atev], *he will be inscribed*; [mix'tav], *letter*; [ktu'ba] *marriage*, and many other forms.

Egyptian writing is a mixture, often redundant, of logograms and signs for consonants and for sequences of two consonants. These consonantal and biconsonantal signs are derived from the logograms by phonetization. Thus the sign for d̲-t, *snake*, is used for the consonant /d̲/, and the sign for /w-r/, *swallow*, is used for the consonantal sequence /w-r/ in writing /w-r-d/, *to be weary* (Jensen, 1970, p. 60). There are no obviously syllabic signs. Vowels are not ordinarily indicated, but in special cases, such as foreign proper names, the signs for the consonants /ʔ/, /j/, /w/ are used for vowels /a/, /i/, /u/, respectively. This assignment of consonantal signs to vowels is not arbitrary. /j/ is homorganic with /i/ and /w/ with /u/. While /ʔ/ is not homorganic with /a/, it is nevertheless phonologically reasonable to transcribe the low back vowel with the sign for the glottal stop, the lowest and most back consonant. As with Sumerian and Chinese, there appears to be no historical period during which the writing is strictly logographic; the consonantal signs are there from the first (Gelb, 1963, p. 74).

Ancient Semitic writing consists simply of signs that ordinarily stand for single consonants: thus Hebrew [ka'tav] is written *ktb*, and [mix'tav], *mktb*. But as with Egyptian, consonantal signs are used, when necessary, to indicate vowels: the signs for /ʔ/, /j/, /w/, aleph, yod and waw, could indicate /a/; /i/ or /e/; and /u/ or /o/, respectively. This device was used not only for proper names: [da'wid], *David*, being written *dwjd*, but also to avoid ambiguity in other words, [jix'tov] being written *jktwb* to distinguish it from [jik'atev], written *jktb*.

Pace Gleitman and Rozin, it was surely not the case that the West Semites didn't "notice" the vowels in their language (1977, p. 19): when it was important to write the vowels, they wrote them. On the contrary, what is especially significant about the Afro-Asiatic languages is that their morphological structure must have fostered awareness of segmental structure to a far greater degree than in the case of Indo-European languages. As I have argued elsewhere, such "linguistic awareness" is not automatic and is essential for alphabetic reading and writing (Liberman, Liberman, Mattingly, & Shankweiler, 1980; Mattingly, 1972).

The reason that both Egyptian and Semitic could be written without consistent indication of vowels is that, in general, the vowels carried only inflectional information. Since word-order is relatively fixed, this information is for the most part redundant. On the other hand, in Greek and in Indo-European languages generally, the base morphemes include vowels. Thus, when the Phoenician alphabet was adapted to Greek, it became a *plene* alphabet: vowels as well as consonants were regularly transcribed, aleph, yod, and waw being used for /a/, /i/, and /u/ as before, and three other Phoenician consonantal signs, he, /h/, heth, /ḥ/, and ayin, /ʕ/, for /e/, /ē/ and /o/, respectively.

To maintain his theory of orthographic evolution, Gelb has to argue, since there are no preceding West Semitic logographies or syllabaries, that the West Semitic scripts derive from the Egyptian. And since he denies the direct development of an alphabet from a logography, he has to argue that the Egyptian consonantal and biconsonantal signs are really syllabic.

In asserting the derivation of the West Semitic script from the Egyptian, Gelb very properly rejects the far-fetched attempts of other scholars to demonstrate similarities in the forms of the signs of the two scripts. His argument relies on the similarity of "inner structure" (p. 146), that is, the use of a limited set of signs to express consonants but not (ordinarily) vowels. But this argument loses what force it might have in view of the fact that it is the same peculiarity in *morphological* structure that made it possible for both languages to be written in this way. Gelb might have adduced a further similarity of inner structure: when vowels did have to be written, the signs for the same three consonants, /ʔ/, /j/, and /w/, were used to write the same three vowels, /a/, /i/, and /u/. But the similarity of Egyptian and West Semitic phonological inventories explains this. Since both had the consonants /ʔ/, /j/, /w/ phonologically related to the vowels /a/, /i/, /u/, respectively, the signs for these consonants were the obvious choices to write the corresponding vowels. Though the possibility cannot be ruled out, there is no need, in the absence of other evidence, to conclude that West Semitic script is derived from Egyptian script. The linguistic similarity of the Egyptian and Semitic languages is quite sufficient to account for the similarity of the two scripts.

As for the Egyptian consonantal signs, Gelb's proposal is that each of them represents a set of syllables or disyllables with the same consonants but varying (or zero) vowels. Thus the biconsonantal sign that other scholars transliterate as *mn* or *m-n* is transliterated by Gelb $\underline{m^{x}n^{x}}$, $\underline{m}^{\underline{x}}\underline{n}^{(\underline{x})}$, $\underline{m}^{(\underline{x})}n^{x}$, $\underline{m}^{(\underline{x})}\underline{n}^{(\underline{x})}$; $\underline{x}$ standing for whatever vowel is required in context (1963, pp. 77-78). From the reader's point of view, this might seem a distinction without a difference, but for Gelb it is crucial:

> The Egyptian phonetic, non-semantic writing cannot be consonantal, because the development from a logographic to a consonantal writing, as generally accepted by Egyptologists, is unknown and unthinkable in the history of writing, and because the only development known and attested in dozens of various systems is that from a logographic to a syllabic writing. (pp. 78-79; original in italics)

But, obviously, this argument is entirely circular; only the theory itself justifies the syllabic interpretation. One might have supposed that the West Semitic scripts, at least, could be allowed to be alphabetic without damage to the theory, but to concede this would obviously undermine the claim of inner structural similarity between them and the Egyptian script. Thus the West Semitic script must be syllabic, too, *waw*, for example, being transliterated *wa*, *wi*, *wu* (Gelb, 1963, p. 148), and the development of alphabetic writing must await the Greeks.

This claim is not only uncorroborated; it also makes it much more difficult to account for the emergence of the Greek *plene* alphabet. If the Phoenician orthography was syllabic, there is no particular reason why the Greeks, any more than other Indo-Europeans, should have become aware of the segmental character of their language when they borrowed this orthography. We should expect to find them using, at least at first, a patched-up syllabary like that of the Persians. But if it is recognized that the West Semites, thanks to the peculiar morphology of their language, had already arrived at the alphabetic principle, then the development of the Greek alphabet from the Phoenician alphabet can be seen to be simply a matter of adding two more vowel signs and using them consistently.

If we do not accept the claim for the development of West Semitic writing from Egyptian writing, and for the syllabic nature of at least the latter, then Gelb's theory is in trouble, for it would seem that, insofar as derivation can be equated with evolution, an alphabet can evolve from a logography without an intervening syllabic stage, as in the case of Egyptian; and may even, perhaps, emerge without any precursors, as in the case of West Semitic; but that no alphabets have developed from syllabic or word-syllabic systems, for apart from the Ugaritic cuneiform alphabet, of unknown origin (Gelb, 1963, p. 129), all other alphabets are derived directly or indirectly, from the West Semitic consonantal alphabets.

The theory of orthographic evolution cannot be correct, for logography cannot be shown to have evolved from picture writing in any meaningful sense; syllabaries do not generally develop from logographies; and alphabets do not develop from syllabaries. What we find instead are either logosyllabic traditions: Mesopotamian, Hittite, Cretan, and Sino-Japanese; or alphabetic traditions: Egyptian and West Semitic. We can, if we choose, regard as evolutionary the development of non-iconic logograms from iconic ones, or the development of the Greek *plene* alphabet from the Phoenician consonantal alphabet, but these are not the sorts of evolution the theory calls for.

But without the theory, how can we account for the variety of orthographies? Let us consider this question from a rather different point of view. The orthography of a language must be *productive*; that is, it must enable the user to write any of the infinite number of possible utterances of the language. Because there are many levels at which an utterance is mentally represented in production and perception, there are, in principle, many possible forms that a productive orthography might take. For example, any utterance of a particular language (in fact, any utterance of *any* language) can be written in a general system of phonetic transcription. If such a transcription were used as an orthography for all languages, any literate person could read aloud in any language. Or one could imagine an orthography that would be based on the acoustic properties of utterances (cf. the "visible speech" of Potter, Kopp, & Green, 1947, and the stylized spectrographic patterns used for speech synthesis by rule at Haskins Laboratories by Liberman, Ingemann, Lisker, Delattre, & Cooper, 1959); such an orthography would include just the information on which the listener to spoken language relies. Or one could imagine an orthography based on the semantic representations of utterances (cf. Katz & Fodor, 1963), if indeed such representations really exist (Fodor, Fodor, & Garrett, 1975); after all, it is the meaning, not the linguistic structure, that the writer really wants to convey to the reader. But it is obvious that none of these alternatives would do for a practical orthography, though it is not easy to say exactly why (see Mattingly, 1984, for some speculations).

There is in fact a very severe limitation on orthographic variety. In practical orthographies, only one basic principle has ever been used, that of transcribing utterances of a language as sequences of lexical items, that is, words. I would argue that all known orthographies are in this sense lexical, varying only in the specific ways in which they happen to transcribe the words. The lexical character of logographies seems obvious, but it might be objected that alphabetic systems are essentially transcribing the phonemes of utterances, and only incidentally the words. With a well-behaved orthography, like that of Serbo-Croatian, only the spaces between the words indicate its specifically lexical character. The point becomes clearer in the case of an eccentric orthography, like that of English, in which there is usually more than one way to write a particular sound. Thus English [ay], phonologically /ī/, can be written *-igh-*, *-y*, *-y(-)e*, *i(-)e*, *-uy*. But despite this variability, there is but one way of writing each of the words *sight*, *try*, *lye*, *dyne*, *lie*, *lime*, *buy*.

A lexical orthography can only be productive if it incorporates a system for transcribing all the words in the languages. There is, however, no principle that can specify just the actual words of a language, and provide the basis for such a system. Thus /čayf/, v., *to gather truffles on Wednesday* could perfectly well be an English word; its absence from the lexicon is accidental. Nor, since the membership of the lexicon, though finite in theory, is indefinite in practice, would it be satisfactory simply to list all the words and provide an arbitrary sign for each. Any word that was inadvertently omitted, or entered the language after the list was compiled, would be unwriteable. And the writer who could not remember the sign for a word that was on the list would be driven to paraphrase. Thus there can be no strict logographies, for a strict orthography would not be productive; and accordingly no such stage is actually found in Sumerian, Egyptian, Hittite, or Chinese.

There is, however, a way to specify all possible words in a language. The phonetics and phonotactics of a language determine the set of phonological forms that qualify for membership in its lexicon. Thus, while /čayf/ could be a word in English, and /kæt/ really is one, /λŬc/ and /stwɔyg/ could not be. By exploiting the phonological structure of the language, that is, by some form of phonetization, an orthography insures that any possible word can be transcribed. This does not mean that a writer will always know the standard way to write a particular word, or that the reader will always know what word is transcribed by a particular orthographic form. It does not preclude a particular word's being standardly transcribed in some exceptional or arbitrary way, e.g., *one*. What it does mean is that if /čayf/ should enter the English language, there will be at least one, in fact several, ways to write it; that the writer who cannot recall the standard spelling of *cat* can at least write *kat*, and that the reader confronted with a word unfamiliar in its written form will have a basis for guessing what the word is.

Although lexical items have syntactic and semantic as well as phonological properties, only the last allow the specification of the set of possible words of a language. Syntactic properties are not sufficient to specify different words uniquely, and a principled characterization of word meaning has thus far eluded the efforts of linguistic semanticists (Fodor, 1977, chap. 5). As we have seen, however, semantic properties can nonetheless play a useful auxiliary role in orthographies.

Every orthography, then, achieves productivity by incorporating some system for transcribing phonologically the possible words of the language. Since the only relevant phonological units are syllables and phonemes, there are really only two ways to do this: the syllabic way and the alphabetic way, and we have seen that all orthographies make use either of the one or the other. But why must there be even two ways? Why are not all orthographies *plene* alphabets? The answer is that, to a large extent, the morphological and phonological structure of a language defines the orthographic options. There are some languages for which a *plene* alphabet would be cumbersome and redundant, and others for which there is no really satisfactory method of phonetization. Moreover, the alphabetic option becomes an obvious one only under rather special linguistic circumstances.

A Semitic language, unless it has borrowed heavily from a non-Semitic language, has no need of a *plene* alphabet. Since lexical items are consonantal patterns, the vowels carrying only inflectional information, an extremely parsimonious system of phonetization is possible, as the West Semitic orthographies demonstrate. Under similar linguistic circumstances, Egyptian writing was able to achieve productivity in much the same way. The extensive and often redundant use of logograms does not alter the fact that the uniconsonantal and biconsonantal signs are the true basis of this orthography.

Because of their restricted syllable structure, Sumerian, Chinese and Japanese are less orthographically amenable. Japanese has only 74 phonotactically possible syllables (or more exactly, moras). Chinese has about 1200 *possible* syllables, but by no means all of them are actually used. Sumerian appears to have been similarly restricted. Restricted syllable structure surely promotes awareness of syllables, and in these cases a syllabary might seem to be the obvious phonetization device. But the morphological consequence of restricted syllable structure unfortunately, is pervasive homophony, exacerbated when, as in Chinese and Sumerian, the base morphemes are mostly monosyllabic. For example, there are 38 different Chinese words with the phonological form /li$_4$/ (Wieger, 1927). Under these circumstances, a strict syllabary is hardly practical, for it would give rise to pervasive homography, far less tolerable in writing, because of the lack of prosodic information to help specify syntactic structure, than pervasive homophony in speech. For these languages, a word-syllabic system, in which the ambiguity of syllable signs is reduced with the help of logograms, is a reasonable, if not highly efficient solution. Alphabetic writing would be no improvement. To replace the syllabic signs in Chinese and Japanese writing by alphabetic ones would do nothing to reduce homography, and to use *only* an alphabet to write these languages, convenient though it might be for printers, would be disastrous for readers.

For many other languages, a *plene* alphabet is the most efficient system of phonetization. But the alphabetic principle is not an obvious one. It did not occur to the Hittites, who used a word-syllabic system even though they did not have a homophony problem and could have used an alphabet. It occurred to the Egyptians and the West Semites only because the morphology of their peculiar character of the languages made them aware of phonological segments. It is certainly owing entirely to the West Semitic example that alphabetic writing is now so widespread.

It would, however, be pressing the point too far to say that variations in linguistic structure account for all orthographic variety. Non-linguistic factors assuredly play a role. The Akkadians, for example, spoke a Semitic language and would certainly have been well advised to use a consonantal alphabet. But being impressed by the culture of the Sumerians, they adopted the Sumerian orthography and made writing unnecessarily complicated for themselves and their Mesopotamian successors (Jensen, 1970, p. 94). Greek speakers on the island of Crete used a word-syllabic system, Linear B, no doubt influenced by the example set by the speakers of the unknown Minoan language written in Linear A (Gelb, 1963, p. 91 ff.). The bewildering complexities of the Japanese kanji, borrowed from the Chinese, have a similar historical explanation (Martin, 1972).

To summarize, Gelb's widely accepted theory of orthographic evolution must be rejected. Orthography has no relationship to picture-language, and there is no sequential development from logography to syllabary to alphabet. The forms that orthographies have taken are constrained by the requirement that they must be productive, and must transcribe lexical items. The limited variety of orthographies can be explained largely on linguistic grounds.

References

Crowder, R. C. (1982). *The psychology of reading*. New York: Oxford University Press.

Fodor, J. D. (1977). *Semantics: Theories of meaning in generative grammar*. New York: Crowell.

Fodor, J. D., Fodor, J. A., & Garrett, M. F. (1975). The psychological unreality of semantic representations. *Linguistic Inquiry*, *4*, 515-531.

Gelb, I. J. (1963). *A study of writing* (2nd ed.). Chicago: University of Chicago Press.

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: Lawrence Erlbaum.

Henderson, L. (1982). *Orthography and word recognition in reading*. London: Academic Press.

Jensen, H. (1970). *Sign, symbol and script* (3rd ed., G. Unwin, Trans.). London: Allen & Unwin.

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language*, *39*, 170-210.

Liberman, A. M., Ingemann, F., Lisker, L., Delattre, P., & Cooper, F. S. (1959). Minimal rules for synthesizing speech. *Journal of the Acoustical Society of America*, *31*, 1490-1499.

Liberman, I., Liberman, A. M., Mattingly, I., & Shankweiler, D. (1980) Orthography and the beginning reader. In J. F. Kavanagh & R. L. Venezky (Eds.), *Orthography: Reading and dyslexia*. Baltimore: University Park Press.

Martin, S. (1972). Non-alphabetic writing systems: Some observations In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye*. Cambridge, MA: MIT Press.

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.). *Language awareness and learning to read*. New York: Springer.

Pedersen, H. (1962). *The discovery of language* (J. W. Spargo, Trans.). Bloomington: Indiana University Press.

Potter, R. K., Kopp, G. A., & Green, H. (1947). *Visible speech*. New York: Van Nostrand.

Wieger, L. (1927). *Chinese characters* (4th ed., L. Davrout, Trans.). New York: Paragon, Dover.

Footnote

[1] Subscripts denote phonological tones.

THE DEVELOPMENT OF CHILDREN'S SENSITIVITY TO FACTORS INFLUENCING VOWEL READING

Danielle R. Zinna,† Isabelle Y. Liberman,†† and Donald Shankweiler††

Abstract. To disambiguate vowel assignment to a vowel digraph in a word, readers must take into account aspects of the word context beyond the vowel digraph units themselves. The present study examined the development of young readers' use of this context in two experiments. In the first experiment, first-, third-, and fifth-grade children were required to read aloud high- and low-frequency words containing vowel digraph units with variant and invariant pronunciations. Words containing vowel digraph units with variant pronunciations were further categorized by the uniformity of pronunciation of the vowel digraph-final consonant unit as it appeared in real words (i.e., the orthographic neighborhood consistency).

While word reading accuracy of all groups was enhanced by word frequency, only the third and fifth graders demonstrated sensitivity to variation in pronunciation of the vowel digraph unit. For these children, low-frequency words containing vowel digraph units with invariant pronunciations were read with accuracy comparable to that obtained for the high-frequency words. In contrast, low-frequency words containing vowel digraphs with variant pronunciations were still a significant source of error for the older readers, but chiefly when they came from inconsistent orthographic neighborhoods.

In a second experiment, pseudoword stimulus items were used to examine further the effect of the orthographic neighborhood on vowel pronunciation. The influence of the vowel digraph-final consonant unit in determining pronunciations was again indicated by limited variability in pronunciations of pseudowords ending in particular vowel digraph-final consonant units from consistent orthographic neighborhoods. Where there was variability in pronunciation, the initial consonant-vowel digraph structure appeared to be largely

---

†Bowling Green State University
††Also University of Connecticut

Acknowledgment. Paper presented at the 1984 American Educational Research Association Convention, New Orleans, April, 1984. The authors would like to thank Arnold Fassler, Director, Special Education Resource Center, Hartford, Connecticut, for enlisting the support of the Bloomfield, Connecticut, public school system; Herb Chester, Superintendent, Art Matiello, Director of Reading Services, and the reading consultants, faculty, and children of the Bloomfield, Connecticut, public school system for their cooperation in this study. This research was supported by the University of Connecticut Research Foundation to D. R. Zinna. The second and third authors were supported by NICHD Grant HD-01994 to Haskins Laboratories.

responsible. Both experiments support the hypothesis that with reading experience, children identify the systematic relationship between pronunciation and orthographic structure and utilize that knowledge in the pronunciation of unfamiliar words.

Analysis of the errors made by children as they acquire skill in word reading has provided some clues to the problems beginning readers encounter in identifying words. The well-documented finding that, in English, vowel misreadings occur with greater frequency than consonant misreadings (Fowler, Liberman, & Shankweiler, 1977; Shankweiler & Liberman, 1972; Weber, 1970) suggests that beginners in English experience particular difficulty in associating a given orthographic vowel unit with its appropriate pronunciation.

A number of explanations have been proposed to account for the difference in difficulty between vowels and consonants (Fowler, Shankweiler, & Liberman, 1979; Shankweiler & Liberman, 1976). One explanation emphasizes the differences in the linguistic properties of vowels and consonants in speech production and perception, noting that vowels are more fluid and generally less categorically defined than consonants (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Another explanation turns on the difference between vowel and consonant orthography. The preponderance of errors on vowels has been attributed to the fact that the same vowel may be spelled differently in different words. Consonants, on the other hand, have a more nearly one-to-one correspondence between orthographic unit and phonological segment. The consonant letters, with few exceptions, cue the same phonological segments wherever they occur, whereas the letters that represent vowels frequently have multiple phonological referents (Venezky, 1967). Further support for the role of the orthography, rather than the differences in vowel and consonant perception, in accounting for the vowel error pattern is reported by Lukatela and Turvey (1980). In their examination of word reading errors in Serbo-Croatian, an orthography that includes a simple vowel set but a more complex consonant set, phoneme substitutions on medial vowel segments were less frequent than substitutions on initial or final consonant segments.

In view of the complexity of the English vowel orthography, it is hardly surprising that there are more vowel errors than consonant errors in reading English words. In order to disambiguate the vowel pronunciation, readers must take into account aspects of the word contexts that are represented by the letters surrounding the vowels. Beginners' errors show that they have not yet learned to do this, but use instead grapheme-phoneme correspondences for single vowel letters (Fowler et al., 1979). With age and experience, children narrow the range of vowel renderings with greater and greater precision, taking more account of the surrounding letter context (Fowler et al., 1979).

In English, these surrounding letter contexts differ in the extent to which they constrain the selection of the appropriate vowel. The context may be tightly constrained, as in the tense or long pronunciation for orthographic vowel units appearing in the context of the silent-e marker. Or it may be loosely constrained in a vowel digraph that may have several appropriate realizations within a particular context. For example, the vowel digraph ou in the context of gh may be correctly rendered as /ɑu/ in bough, /ʌ/ in tough, /ɔ/ in thought, /u/ in through or /o/ in though. In the Fowler et al. studies, although the stimuli included a wide range of contextual constraints, the possibly differing effects among them were not considered.

Attempts to construct a model for predicting adults' pronunciations of pseudowords containing vowel digraph units (Johnson & Venezky, 1976; Ryder & Pearson, 1980) have suggested that the vowel pronunciation could be influenced either by the frequency of occurrence of that unit without regard to the context, that is, without regard to the effect of the final consonant or, alternatively, by the context provided by the final consonant. Results of those investigations support a model predicting that adult pronunciation is highly determined by frequency of orthographic patterns, but the functional unit is hypothesized to be the vowel digraph-final consonant structure.

Skilled adult readers have in fact been shown to be sensitive to the consistency or inconsistency of the pronunciation of medial vowel-final letter units (Glushko, 1979). Glushko has proposed that, in the course of reading a word, an entire neighborhood of similarly structured words and their pronunciations is automatically activated in memory. Glushko's "neighborhood" includes all monosyllabic words in the reader's lexicon that share the same medial vowel letters in combination with the same letter units in word final position. Rhyming words such as *seam*, *beam*, and *team*, sharing both the medial vowel-final letter unit and a uniform pronunciation, would thus constitute a consistent orthographic neighborhood; whereas the words *beat*, *threat*, and *great*, although sharing the medial vowel-final letter unit, fail to share a uniform pronunciation, and thus would be classified as constituting an inconsistent orthographic neighborhood. Glushko's adult readers' performance was influenced by the consistency or inconsistency in orthographic neighborhoods as evidenced by more rapid reading and more limited variation in pronunciation of words and pseudowords from consistent orthographic neighborhoods (i.e., words of similar structure sharing a uniform pronunciation). It was also indicated by a greater latency of response and significant variation in pronunciation of words from inconsistent orthographic neighborhoods (i.e., words of similar structure that fail to share a uniform pronunciation).

The vowel digraph unit in many words may be ambiguous unless the reader can exploit additional cues from the other letters in the word. The broader context, as for example, the final consonant, may supply such cues. Whether or not it does could depend on whether word items from an orthographic neighborhood for that vowel digraph-final consonant unit share a consistent pronunciation. Thus, the final consonant might be used to disambiguate the vowel digraph, but its use would involve a complex context-sensitive operation.

A study examining this skill in second-, fourth-, and sixth-grade children (Johnson, 1970) found that the factor most likely to influence children's selections was also the frequency of occurrence of a particular pronunciation for a given unit, and further, that with increasing grade level, children's responses more closely reflected the pronunciations of those units as they appear in real words. Though mention is made of some additional effects of the final consonant context and the position of the vowel digraph unit within the word, the study was not designed to investigate the development of the influence of context on children's selections as a result of reading experience. Nor did it examine the effects of the frequency of occurrence of the vowel digraph-final consonant structure and the consistency of pronunciation of that structure in real words.

To date, there has been no systematic study of the development of children's use of the final consonant context in disambiguating vowel assignment to vowel digraph units and their sensitivity to orthographic neighborhood consistency. An examination of these effects with children may provide insight into the development of children's awareness of the very complex relationship between the orthography and the phonology. In addition, it would also assist us in understanding how normally developing readers use the reading vocabulary they have mastered to develop strategies to identify unfamiliar words.

In order to explore these questions, two experiments were conducted. In the first experiment, development of children's understanding of vowel digraph pronunciation was the focus. First-, third-, and fifth-grade children were required to read aloud high- and low-frequency words containing vowel digraph units with variant and invariant pronunciations. For each grade, an examination of error rate and of the characteristics of errors was conducted to explore the effects of word frequency, of alternate pronunciations for vowel digraph units, and of consistency of orthographic neighborhood on word reading accuracy. The second experiment investigated other influences on vowel digraph reading using pseudowords containing vowel digraph units that have variant pronunciations in words. By eliminating the factor of word familiarity, pronunciation preferences for vowel digraph units, as well as factors influencing those pronunciations, could be studied and the results compared with those obtained on the real word reading task.

General Method

Subjects

The subjects in the first experiment were children from the first-, third-, and fifth-grade classes of a suburban public school system in Connecticut. Following a review of teacher ratings for reading achievement for the first and third graders, and teacher ratings and group reading achievement tests scores for the fifth graders, a pool of subjects, all average or above average readers, was identified. The final population consisted of 90 students, 30 from each grade level. The subjects participating in the second experiment were the 30 third-grade children who had participated in Experiment 1. All subjects selected were native English speakers with no known hearing or vision impairments.

Procedure

The children were tested individually in two 30-min sessions. During the first session, the experimental word reading task was presented. The words were typed in lower case primary type on 4" x 6" file cards secured in a ring binder. The stimuli were presented in random order with 20 filler words, which were single syllable items selected from the reading subtest of the Wide Range Achievement Test (Jastak, Bijou, & Jastak, 1978). These filler words were included in order that the randomization satisfy the constraint that words with the same vowel sound not precede one another, thus minimizing possible priming effects. Subjects were instructed to read each word orally and then to turn to the following card. Approximately two weeks after the initial session, a second session was held for the third-grade children during which the experimental pseudoword reading task was presented. Subjects were informed that these words were nonsense or "pretend" words and that they should not attempt to make real words out of the items. They were instructed to read each word orally and to turn to the following card after reading each word. All pronunciations were recorded on tape for later transcription and analysis.

Experiment 1

Materials

Two lists of monosyllabic real words, including 72 items in all, were developed. One list, as displayed in Table 1, included words containing vowel digraph units with invariant phonological correspondences, *ee*, *oa*, *oi*, *ai*, and *ew*. The words in the other list, as displayed in Table 2, included words containing units with variant correspondences, *ea*, *ou*, *ow*, *ie*, and *oo*. Words were selected to vary in two respects: frequency and variability of pronunciation of the vowel digraph unit. Frequency was determined by the occurrence of the words in reading material at the third-grade level as indicated in the American Heritage Word Frequency Listings (Carroll, Davies, & Richman, 1971). Classification according to variant or invariant pronunciation was based on the pronunciations reported in a thorough listing (Fischer, 1979) of monosyllabic English words containing vowel digraphs. Both word frequency and pronunciation variability were systematically controlled in both stimuli lists.

In addition, as indicated in Table 2, for each monosyllabic word containing a vowel digraph unit with a variant pronunciation, the word's orthographic neighborhood was determined from the Fischer set in the manner of Glushko (1979). This determination was made for both high- and low-frequency words. Each word with a vowel digraph-final consonant unit that is always pronounced the same way in all monosyllabic words sharing that structure, was considered to have a consistent orthographic neighborhood. In contrast, each word with a vowel digraph-final consonant unit that is pronounced differently in at least one other monosyllabic word sharing that structure was considered to have an inconsistent orthographic neighborhood.

Results and Discussion

Because the variance in performance was substantially greater for the first graders than for the third and fifth graders, a separate analysis was carried out for each grade level group. Mean percentages of correct responses, possible pronunciation responses, and error responses were calculated for each grade on each word category. These data appear in Table 3. The data for each grade were subjected to two separate factorial analysis of variance procedures. The first analysis examined factors of word frequency and pronunciation variability for the vowel digraph unit for the entire set of stimuli. In the second analysis the factors of word frequency and consistency or inconsistency of the orthographic neighborhood were examined for the variant pronunciation set of words.

Effects of Frequency and Vowel Digraph Pronunciation

First graders. The analysis of the first graders' data revealed, as expected, a significant main effect for word frequency, $F(1,29) = 45.89$, $p < .0001$. As illustrated in Table 3, these children correctly identified 65% and 63% of the high-frequency words containing vowel digraph units with variant and invariant pronunciations, respectively. Thus, it appears likely that the first graders employed a holistic word reading strategy. In contrast, identification was correct for only 50% of the low-frequency words containing vowel digraph units with invariant pronunciations and 43% of the low-frequency words containing vowel digraph units with variant pronunciations. Thus, while these

Table 1

Real-Word Stimulus Items with Invariant Pronunciations of the Vowel Digraphs (Experiment 1)

High Frequency	Low Frequency
green	sleek
street	breed
road	oat
coal	boast
soil	toil
join	joint
paint	ail
main	trait
drew	dew
flew	slew

Table 2

Real-Word Stimulus Items with Variant Pronunciations of the Vowel Digraphs (Experiment 1)

Consistent Orthographic Neighborhood		Inconsistent Orthographic Neighborhood	
High Frequency	Low Frequency	High Frequency	Low Frequency
beach	ream	read	tread
clean	dean	speak	steak
		break	teak
		head	plead
young	mount	mouth	youth
found	spout	touch	uch
group			vouch
proud			soul
tried	fried	owl	flown
piece	niece	how	tow
pie	lied	bowl	jowl
field	shield	low	pow
soon	croon	foot	loot
room	sloop	food	hood
		good	mood
		shoot	soot

first-grade readers correctly identified nearly two-thirds of the high-frequency words containing vowel digraph units with invariant pronunciations, they did not generalize that knowledge in assigning the correct pronunciation to identical vowel digraph units with invariant pronunciations embedded in the less familiar, low-frequency words.

Further analysis suggests that the first-grade readers were nonetheless beginning to acquire an awareness of alternate pronunciations for vowel digraph units with variant pronunciations. In reading high-frequency words containing vowel digraph units with variant pronunciations, first graders, as noted above, correctly identified 65% of the words; however, 58% of their error responses consisted of substitutions of possible alternate pronunciations for that vowel digraph unit. Error data obtained from their reading of low-frequency words containing vowel digraph units with variant pronunciations offer corroborative evidence for this finding. Although the overall error rate for reading low-frequency words containing vowel digraph units with variant pronunciations was substantially greater than that obtained for the high-frequency words, 53% of these errors (again greater than one-half of the total) consisted of substitutions of possible alternate pronunciations for the vowel digraph unit.

Third graders. As was the case for first graders, analysis of third-grade data again revealed a significant main effect for frequency, $F(1,29) = 55.46$, $p < .0001$. In addition, a significant main effect for pronunciation for the vowel digraph unit, not present in the analysis of the first-grade data, was obtained with the third graders, $F(1,29) = 59.93$, $p < .0001$. As illustrated on the left in Figure 1, an interaction between word frequency and pronunciation for the vowel digraph unit was obtained, $F(1,29) = 23.54$, $p < .0001$.

Like the first graders, the third-grade readers read high-frequency words containing vowel digraph units with variant and invariant pronunciations equally well, though with greater accuracy than the first graders, correctly identifying 96% and 98% of words in these categories, respectively. In contrast to the first graders, the third graders read low-frequency words with invariant pronunciations for the vowel digraph unit with accuracy comparable to that obtained for the high-frequency words. They correctly identified 92% of the low-frequency words of that orthographic type, suggesting that they had been successful in identifying the systematic relationship between pronunciation and orthographic structure among the words in their reading vocabulary. Less dependent on previous knowledge of specific words, the third graders demonstrated skill in generalizing knowledge of proper pronunciations of invariant vowel digraph units when those units appeared in the context of unfamiliar, low-frequency words.

In contrast to this performance on the invariant units, the third graders were able to read accurately only 79% of the low-frequency words containing vowel digraph units with variant pronunciations. Nonetheless, their overall error rate in this category (21%) was substantially lower than that of the first graders (57%). However, like the first-grade pattern, a majority of their errors (82%) consisted of substitutions of possible alternate pronunciations for the vowel digraph unit. As illustrated in Table 3, while the error rate declined from the first to the third grade, the ratio of substitutions of possible alternate pronuciations to errors increased. Once again, the

Table 3

Frequencies and Percentages of Correct and Incorrect Responses for Real Words Containing Variant and Invariant Vowel Digraph Units (Experiment 1)

	Grade 1		Grade 3		Grade 5	
	Freq.	%	Freq.	%	Freq.	%
Variant Unit						
High-Frequency Words						
Total Correct	509	65	751	96	766	98
Errors						
Possible Pronunciations	158	20	25	3	14	2
Impossible Pronunciations	113	15	4	1	0	0
Low-Frequency Words						
Total Correct	332	43	618	79	712	91
Errors						
Possible Pronunciations	235	30	133	17	57	7
Impossible Pronunciations	213	27	29	4	11	1
Invariant Unit						
High-Frequency Words						
Total Correct	189	63	294	98	300	100
Errors	111	37	6	2	0	0
Low-Frequency Words						
Total Correct	149	50	275	92	296	99
Errors	151	50	25	8	4	1

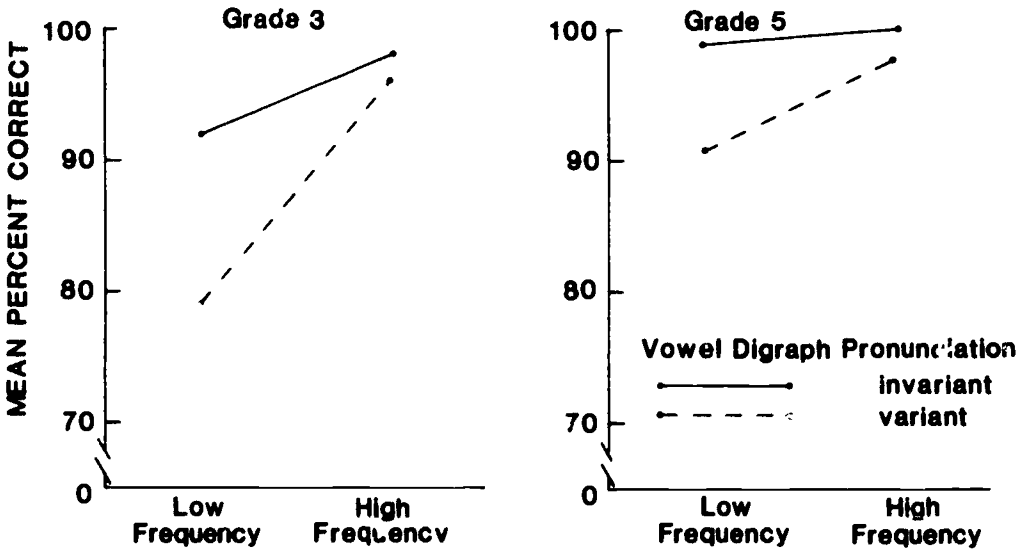


Figure 1. Performance of third and fifth graders on reading low-frequency and high-frequency words, plotted in mean percent correct.

third-grade readers demonstrated skill in generalizing knowledge of pronunciations for vowel digraph units to unfamiliar, low-frequency words.

Fifth graders. The main effects for word frequency and variant versus invariant pronunciation for the vowel digraph unit were again revealed in the analysis of the fifth-grade data, $F(1,29) = 38.40$, $p < .0001$, and $F(1,29) = 59.39$, $p < .0001$, respectively. As illustrated on the right in Figure 1, an interaction between frequency and pronunciation for the vowel digraph unit was again obtained, $F(1,29) = 26.51$, $p < .0001$. Though their performance was more accurate overall, the pattern of the fifth graders was similar in one respect to that of both earlier grades. That is, they read high-frequency words containing vowel digraph units with variant and invariant pronunciations equally well, correctly identifying 98% and 100% of the words of these categories, respectively.

As observed previously with the third graders, the fifth graders successfully identified the systematic relationship between pronunciation and orthographic structure. Thus, they were able to generalize that knowledge to the identification of words of lower frequency containing these invariant units, correctly identifying 99% of the words of this category. As was the case with the third graders, the fifth graders' reading of low-frequency words containing vowel digraph units with variant pronunciations was poorer than their reading of high-frequency words of that type: 91% of the words of this category were correctly identified. Most of their errors (87%) consisted of substitutions of possible alternate pronunciations for the vowel digraph unit embedded within these words, a slightly greater percentage of such substitutions than in the third grade (82%).

Summary. The analysis confirms the expectation that children's accuracy in word reading would be favorably enhanced by high word frequency, regardless of the number of alternate pronunciations for the vowel digraph unit contained within these words. In addition, the highly accurate performance of the third and fifth graders in reading low-frequency words containing vowel digraph units with invariant pronunciations supports the hypothesis that with reading experience, children identify the systematic relationship between pronunciation and orthographic structure and utilize that knowledge in the pronunciation of unfamiliar words. Finally, the increase in proportion of substitutions of possible alternate pronunciations among the errors, which increased with increasing grade level, provides further evidence that as children develop reading skill they identify the systematic relationship between pronunciation and orthographic structure.

Effects of Frequency and Orthographic Neighborhood Consistency

A second analysis was conducted to examine the possibility that the error rate on categories of words that contained vowel digraph units with variant pronunciations was affected by the consistency of the orthographic neighborhood of individual words. Mean percentages of correct responses were calculated for each grade level group on each word category. These data appear in Table 4.

First graders. For the first graders, the analysis revealed a significant main effect for frequency, $F(1,29) = 75.12$, $p < .0001$. As indicated in Table 4, they correctly identified 62% and 68% of the high-frequency words

Table 4

Mean Percentage of Correct Responses for High- and Low-Frequency Words Containing Variant Vowel Digraph Units from Consistent and Inconsistent Orthographic Neighborhoods (Experiment 1)

	Orthographic Neighborhoods					
	Consistent			Inconsistent		
Grade	1	3	5	1	3	5
Variant Unit						
High-Frequency Words % Correct	62	98	99	68	95	97
Low-Frequency Words % Correct	44	91	96	42	72	89

from consistent and inconsistent-orthographic neighborhoods, respectively. In contrast, they correctly identified only 44% and 42% of the low-frequency words from consistent and inconsistent orthographic neighborhoods, respectively. Once again, word frequency was the most predictive index of word reading accuracy.

Third graders. Analysis of the third grade data also revealed a significant main effect for word frequency, $F(1,29) = 76.79$, $p < .0001$. However, a significant main effect for orthographic neighborhood consistency, not found in the analysis of the first-grade data, was also obtained, $F(1,29) = 88.87$, $p < .0001$. As illustrated on the left of Figure 2, a significant interaction occurred between word frequency and orthographic neighborhood consistency, $F(1,29) = 21.12$, $p < .0001$. Like the first graders, the third-grade readers read high-frequency words from consistent and inconsistent orthographic neighborhoods equally well, though with greater accuracy than the first graders, correctly identifying 98% and 95% of words from these categories, respectively. When low-frequency words were presented, however, in contrast to the first graders' error pattern, those words from consistent orthographic neighborhoods were read with accuracy comparable to that obtained for the high-frequency words. The third graders correctly identified 91% of the low-frequency words from consistent orthographic neighborhoods, in contrast to correct identification of only 72% of the low-frequency words from inconsistent orthographic neighborhoods. This result suggests that the third graders, but not the first graders, have developed a reading vocabulary sufficient to provide a data base from which to determine the relations between orthographic structure and pronunciation.

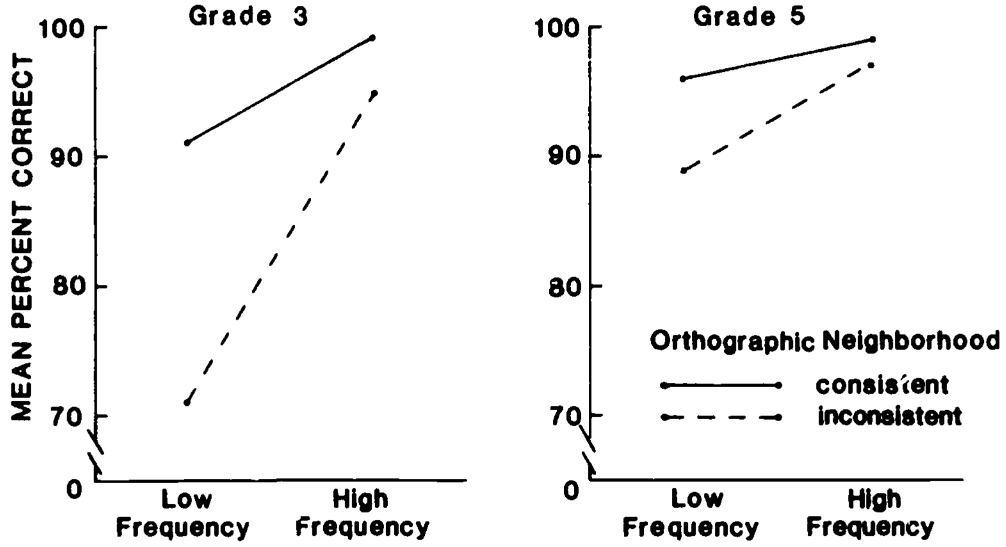


Figure 2. Performance of third and fifth graders on reading low-frequency and high-frequency words with variant vowel digraph units, plotted in mean percent correct.

Fifth graders. Main effects for word frequency and orthographic neighborhood consistency were once again found in the analysis of the fifth-grade data, F(1,29) = 37.4', p < .0001, and F(1,29) = 33.29, p < .0001, respectively. As illustrated on the right in Figure 2, a significant interaction between word frequency and orthographic neighborhood consistency was again obtained, F(1,29) = 9.64, p < .0042. Though more accurate than the first and third graders, the fifth graders also read high-frequency words from consistent and inconsistent orthographic neighborhoods equally well, correctly identifying 99% and 97% of words of these categories, respectively. Like the third graders, the fifth graders, when presented with low-frequency words from consistent and inconsistent orthographic neighborhoods, read words from consistent neighborhoods with accuracy close to that obtained for the high-frequency words. They correctly identified 96% of the low-frequency words from consistent orthographic neighborhoods, as contrasted with correct identification of 89% of the low-frequency words from inconsistent orthographic neighborhoods. Once again, support is provided for the contention that the analysis of interword relations and awareness of consistencies and inconsistencies between orthographic structure and pronunciation, in this case the vowel digraph-final consonant structure, provide the reader with the knowledge necessary to pronounce an unfamiliar word correctly.

Experiment 2

The results of Experiment 1 provide evidence that older readers' accuracy and error rate in reading real words containing vowel digraph units with variant pronunciations were influenced by the consistency of pronunciation of other words sharing the particular vowel digraph-final consonant unit. To examine this effect further and to begin exploring the effect of the initial consonant-vowel digraph unit on pronunciation selection, a second experiment was conducted. In this experiment, the third-grade children who had participated in the first experiment were asked to read monosyllabic pseudowords containing vowel digraph units with variant pronunciations. By eliminating the possibility of word familiarity, it was anticipated that factors influencing reading would be more unequivocally revealed.

Materials

A list of 60 monosyllabic pseudowords was developed that contained vowel digraph units with variant pronunciations, ea, oo; ou, ow, and ie. Each pseudoword consisted of initial and final segments that might appear in real words. The initial consonant-vowel digraph segment and the vowel digraph-final consonant segment in each of the pseudowords represented a legitimate sequence in English phonology. However, vowel digraph segments in the pseudowords might have different pronunciations in different real word contexts. For example, the ou unit in the pseudoword moung might be rendered like the ou in mouth or the ou in young. For pseudowords constructed in this manner, each item was reviewed to determine the consistency of pronunciation among monosyllabic real words sharing the vowel digraph-final consonant unit. Of the 60 items, 36 pseudowords were determined to have consistent orthographic neighborhoods, as evidenced by the uniformity of pronunciation among monosyllabic real words sharing the particular vowel digraph-final consonant structure (Fischer, 1979). The remaining 24 items were determined to have inconsistent orthographic neighborhoods, as evidenced by the lack of uniformity of pronunciation among monosyllabic real words sharing the particular vowel digraph-final consonant structure (Fischer, 1979). The final pseudoword lists are included in Tables 5 and 6.

Results and Discussion

The pronunciation preferences of the 30 third graders for reading each of the pseudowords are listed as percentages in Tables 5 and 6. Vowel digraph-final consonant units, which were determined to have consistent orthographic neighborhoods because of their uniform pronunciation in monosyllabic real words, are listed in Table 5. Items determined to have inconsistent orthographic neighborhoods, based upon the lack of such uniformity, appear in Table 6.

Influence of the Vowel Digraph-Final Consonant Unit

It is evident from Tables 5 and 6 that pronunciations for pseudowords containing the vowel digraph units oo and ea tended to vary with the designation of their orthographic neighborhood as consistent or inconsistent. Pseudoword items containing the units -ooth, -oc -oon, and -each, -ean, and -eam, all considered to have consistent orthographic neighborhoods, were usually pronounced as /u/ for the former and /i/ for the latter. These pronunciations occurred in never fewer than 90% of the cases. In contrast, the units -ool, -ood, -ook, and -ead, -eat, and -eak, all considered to have inconsistent orthographic neighborhoods, were the source of considerable variation in pronunciation. Pseudowords containing the oo unit received the /u/ pronunciation in between 50% and 97% of the cases; items containing the ea unit received the /i/ pronunciation in between 60% and 97% of the cases.

Table 5

Percentages of Total Responses to Each Item from Consistent Neighborhoods by Vowel Digraph Pronunciation (Experiment 2)

Consistent Orthographic Neighborhood

	Responses			Other Responses or Errors
	/u/	/ʊ/		
mooth	90	0		10
looth	94	3		3
troom	97	3		0
poom	94	3		3
shoon	90	3		7
smoon	100	0		0
woon	93	0		7
	/i/	/eɪ/	/ɛ/	
meach	94	3	0	3
slean	97	0	3	0
chean	97	0	0	3
team	94	3	0	3
drief	67	33		0
tiece	57	40		3
criece	60	40		0
biece	60	37		3
fiece	70	27		3

Certain units elicited the greatest variation in pronunciation. For example, the realizations for the unit oo followed by k were evenly distributed between /u/ and /ʊ/. For each of these items, the initial word segments moo- and zoo- were words likely to be in a third grade child's reading vocabulary. However, the highly frequent words book and look, also likely to be in a young child's reading vocabulary, provide the dominant pronunciation for the unit -ook as it appears in monosyllabic real words. These factors, in addition to these items' inconsistent orthographic neighborhood, may account for the pronunciation alternation.

Table 6

Percentages of Total Responses to Each Item from Inconsistent Orthographic Neighborhoods by Vowel Digraph Pronunciation (Experiment 2)

Inconsistent Orthographic Neighborhood

	Responses				Other Responses or Errors
	/u/	/ʊ/			
bool	80	7			13
smood	97	3			0
tood	73	17			10
zook	54	43			3
mook	50	47			3
	/i/	/eɪ/	/ɛ/		
stread	60	0	40		0
clead	80	0	17		3
chead	77	0	23		0
steat	97	0	3		0
preat	90	3	3		3
dreak	70	13	10		7
heak	94	0	3		[illegible]
treak	94	0	3		3
	/u/	/o/	/aʊ/	/ʌ/	
touth	30	17	43	3	7
mouch	0	7	[illegible]	13	0
fouth	7	10	[illegible]	0	23
	/aʊ/	/o/			
blowl	53	47			0
lowl	37	60			3
snowl	37	63			0
fow	57	43			0
clow	80	17			3
drow	43	47			10
cown	100	0			0
hown	97	3			0

Influence of the Initial Consonant-Vowel Digraph Unit

In contrast to the even distribution of pronunciation selections for both pseudoword items ending in *-ook* is the inconsistency in assignment of pronunciation to several other pseudoword items containing identical vowel digraph-final consonant structures. For example, similar variation in pronunciation might be expected for the three items ending in *-ead*, a unit with an inconsistent orthographic neighborhood. Instead, the *ea* unit in the pseudoword *clead* was rendered as /i/ 80 % of the time; whereas in the item *stread*, it was similarly rendered only 60% of the time. It seems likely that real words sharing the initial consonant-vowel digraph structure may be biasing the pronunciation of the pseudoword, but a final determination must await further study.

As indicated in Table 5, the consistency of pronunciation expected for the *ou* unit in pseudowords ending in *-oup*, *-oud*, and *-ound*, considered to have consistent neighborhoods and expected to be rendered as /u/, /ɑʊ/, and /ɑʊ/, respectively, was not obtained. It may be that the paucity of words ending in those structures in a third-grade child's reading vocabulary reduced the saliency of the vowel digraph-final consonant unit, allowing the initial word segment to influence pronunciation. For example, the pseudowords *proup* and *cloup* were expected to be rendered on the basis of the reader's knowledge of words such as *soup* and *group*. Instead, the *ou* unit was frequently rendered as /ɑʊ/. As an explanation of that result, we would suggest that words such as *proud* and *cloud*, which share the exact initial consonant-vowel digraph unit with *proup* and *cloup*, may have been activated and contributed to the unexpected pronunciation.

In view of that result, the apparent saliency of the /ʌ/ pronunciation for the *ou* unit in *moung* is particularly notable. Though that pronunciation occurs in English only in the single word, *young*, the *ou* unit embedded in the pseudoword *moung* received the /ʌ/ pronunciation 70% of the time, despite membership of the initial segment in a neighborhood containing *mouth* and *mountain*. In contrast, the other pseudoword item containing the *oung* unit, *groung*, received the /ʌ/ pronunciation only 37% of the time and the pronunciation /ɑʊ/ associated with the initial segment *grou-*, 50% of the time.

Mixed Influence

Additional evidence for the possibility that pronunciation selections could be influenced by the initial consonant-vowel digraph unit was revealed in the analysis of the *ow* unit in pseudowords. Any pseudoword containing the *ow* unit, whether it ended a word or was combined with "l" as in *-owl* or "n" as in *-own*, was considered to have an inconsistent orthographic neighborhood. Pronunciations of pseudowords containing the *ow* unit reflected that inconsistency, with the exception of the *ow* in the items *cown* and *hown*. The *ow* unit in these words was rendered as /ɑʊ/ in 100% and 97% of the cases, respectively. In each of these instances, the initial word segment consisted of a morpheme, the pronunciation of which was not overridden by the pronunciation inconsistency of the final unit *-own*. In addition, words likely to be present in a third-grade child's reading vocabulary, *down*, *brown*, and *town*, provide identical pronunciations for the *ow* unit and share the *-own* structure, probably accounting for the consistent rendering of these items.

Pseudowords containing the vowel digraph unit ie were all expected to reflect their consistent orthographic neighborhoods. The designation of consistency was based as always on the uniformity of rendering of the vowel digraph-final consonant unit in similarly structured real words. However, in the case of pseudowords containing the ie unit, this detection of neighborhood consistency required that the reader respond to the affixation of plural and past tense markers as a signal for the /ɑɪ/ pronunciation. The third-grade readers in this study were able to identify the ie unit in the pseudoword items kie and nie as /ɑɪ/; yet their pronunciations for similar items with the plural or past tense marker were variable. For example, the ie unit in the items bries and fied received the /ɑɪ/ pronunciation in between 50 and 70 percent of the cases only.

The ie unit in pseudowords ending in -ield, -iece, and -ief was expected to be pronounced as /i/ on the basis of knowledge of such words as field, piece, and chief. A review of the responses indicates that items ending in these units received the /i/ pronunciation in between 47% and 70% of the cases. Evidently, pronunciation preferences are being influenced by experience or instruction, but the design of the stimuli did not allow us to pinpoint the source of the variation in pronunciation of the ie unit in that context.

Summary. The results of Experiment 2 provide support for the influence of the vowel digraph-final consonant unit in determining the rendering of the vowel in English-like pseudowords. The influence of this unit could be seen in the greater uniformity of the pronunciation of pseudowords ending in particular vowel digraph-final consonant units from consistent orthographic neighborhoods. In instances where there was less uniformity in pronunciation of such items, the influence of the initial segment appears to account for most of the variability.

General Discussion

Children's acquisition of word reading skills was examined with particular emphasis on the development of young readers' response to variant vs. invariant phonologic associations for vowel digraph units, the use of the final consonant context in disambiguating vowel assignment to invariant vowel digraph units, and their sensitivity to the orthographic neighborhood consistency of that vowel digraph-final consonant structure.

The data obtained in Experiment 1 indicate that the word reading accuracy of the first-grade children was strongly affected by word frequency, but not by the variation in pronunciation of the vowel digraph unit. This finding supports the view expressed by Gough and Hillinger (1980) that initial acquisition of word reading skills may typically be accomplished through rote learning with the result that frequently encountered words are usually identified without analysis of word components.

The word reading accuracy of third and fifth graders was also affected by word frequency, but in addition, the older readers read low-frequency words containing vowel digraph units with invariant pronunciations with accuracy comparable to that obtained for the high-frequency words. This effect is consistent with results of earlier studies (Fowler et al., 1979; Venezky & Johnson, 1973; Venezky & Massaro, 1979) demonstrating children's ability to generalize knowledge of orthographic patterns beyond the words in which they were originally encountered. In contrast, low-frequency words containing vowel digraph units with variant pronunciations were a significant source of error even for the older readers.

When these low-frequency words were further categorized by consistency or inconsistency of their orthographic neighborhoods, those from consistent orthographic neighborhoods were read by the third and fifth graders with a level of accuracy close to that obtained for both high-frequency words and those of low frequency that contained invariant vowel digraph units. For the children in the higher grades, only the low-frequency words containing variant vowel digraph units with inconsistent orthographic neighborhoods were a substantial source of error. These results provide support for a model in which the final consonant predicts vowel digraph pronunciation preferences (Johnson & Venezky, 1976; Ryder & Pearson, 1980). They also support the hypothesis (Glushko, 1979) that the ability to read the vowel in words is affected by the consistency of pronunciation of words sharing a particular medial vowel-final letter unit. Despite some exceptions, these findings speak to the special salience of the vowel digraph-final consonant unit in disambiguating vowel pronunciation.

In the second experiment, pseudoword stimulus items were used to allow us to explore further the influence of the neighboring orthographic segments on vowel pronunciation. It was found that whereas the orthographic neighborhood consistency effect, as defined for medial vowel-final letter units, was obtained for many pseudoword items, the pronunciation of others was not disambiguated by the consistent pronunciation of the vowel digraph-final consonant of that item. This result was observed on the items <u>proup</u> and <u>cloup</u>, in which the <u>ou</u> unit was frequently pronounced as /ɑʊ/, despite the consistency of pronunciation evidenced by the <u>-oup</u> unit as it appears in real words. Many of these exceptions were rationalized by considering possible interference from initial consonant-vowel digraph occurrences in familiar real words. These cases suggest that in future work it will be desirable to expand the concept of neighborhood consistency to examine influences from the initial portion of the word as well as of the final.

One possible explanation for the results is the operation of a left-to-right letter string parser (Marcel, 1980). Marcel proposed that when a word or pseudoword is presented to a reader, the letter string is segmented in all possible ways. Each word segment, as it is parsed, automatically activates the pronunciations of that unit as it occurs in different words. Thus, for the young reader the pronunciation activated for the word segments <u>prou-</u> and <u>clou-</u> may result from the words <u>proud</u> and <u>cloud</u> in their reading vocabularies. Word pronunciation may result from the parsing of successive units of the letter string, during which the pronunciation of later appearing segments may override the pronunciation of prior segments (Baron & Strawson, 1976; Marcel, 1980). For the young reader, then, it may be that the strength of the association between the unit <u>ou</u> and the /ɑʊ/ pronunciation was too strong to be overridden by the pronunciation of <u>-oup</u> as it appears in the words <u>soup</u> and <u>group</u>.

The proposal put forth by Marcel (1980) also explains the pronunciation of the <u>ou</u> unit in the item <u>moung</u> as /ʌ/. According to that explanation, as a child attempts pronunciation of the pseudoword <u>moung</u>, the initial segment parsed is mou-, the <u>ou</u> unit likely to be pronounced as /ɑʊ/ on the basis of knowledge of words such as <u>mouth</u> and <u>mountain</u>. When the child parses the final segment of the letter string -oung, however, a different pronunciation for that unit is activated on the basis of the occurrence of that unit in the word young. As it happened, the pronunciation of the ou unit in the pseudoword moung was frequently /ʌ/, attesting to the strong effect the final word segment maintains over word pronunciation.

A left-to-right parser, with capacity to override and disambiguate pronunciations activated for earlier segments of a word, would require that the reader have a substantial reading vocabulary and awareness of the phonemic segmentation of the words in the lexicon. It has been well documented not only that phonemic awareness is a predictor of reading achievement (Blachman, 1983; Bryant & Bradley, 1980; Liberman, 1973; Lundberg, Olofsson, & Wall, 1980), but also that this awareness is enhanced by reading experience and instruction (Liberman, Liberman, Mattingly, & Shankweiler, 1980; Morais, Cary, Alegria, & Bertelson, 1979). We may speculate, therefore, that the limited reading vocabularies of the first graders, in combination with underdeveloped phoneme awareness and segmenting skills, effectively limit the amount of information that most first graders are able to utilize in reading new words. As a result, they were more likely to identify high-frequency words correctly than low-frequency words, regardless of the number of alternate pronunciations for the vowel digraph. Insensitive to orthographic neighborhood consistency or inconsistency, the first-grade readers were unable to use the larger vowel digraph-final consonant context to disambiguate vowel assignment to a vowel digraph.

We must ask whether this result may be an artifact of instruction. All children participating in this study have received what is best identified as an eclectic approach to reading instruction. As reported, the third graders, and, even more so, the fifth graders, had developed a sensitivity to the orthographic neighborhood consistency, taking account of the wider vowel digraph-final consonant context to disambiguate vowel assignment to vowel digraphs. Apparently by the third grade, children who are progressing normally in reading have acquired a corpus of words in their reading vocabularies adequate to meet the demands of an operation that requires phoneme awareness, segmenting skill, and prior word knowledge to determine the pronunciation of an unfamiliar word. In contrast, the first graders, as they learn new words, are just beginning to identify phoneme correspondences of individual graphemes and may depend heavily on these to identify vowel digraphs. Thus, their responses, though incorrect, include some substitutions that are possible in certain other contexts.

This difference between the performances of the first and third graders raises critical questions for future investigation. We are interested to know if, during that second year of formal reading instruction, children merely acquire a more extensive reading vocabulary in a rote manner, or if they begin then to analyze interword relations identifying consistencies between orthographic structures larger than the individual letters and their pronunciation. Moreover, we should like to know whether different methods of instruction will make a difference in the development of these skills, and even whether there may be lasting effects of such instructional differences. In addition, our attention must turn to those older children who fail to acquire automatic word reading skills. Are these older, poorer readers functioning like the first-grade readers, or are they utilizing different information to determine the pronunciation of an unfamiliar word? It is clear that the answers to these questions will further our understanding of reading and how it develops.

References

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. Journal of Experimental Psychology: Human Perception and Performance, 2, 386-393.

Blachman, B. (1983). Are we assessing the linguistic factors critical in early reading? Annals of Dyslexia, xxxiii, 91-109.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read - a causal connection. Nature, 301/3, 419-421.

Bryant, P. E., & Bradley, L. (1980). Why children sometimes write words which they cannot read. In U. Frith (Ed.), Cognitive processes in spelling. London: Academic Press.

Carroll, J. B., Davies, P., & Richman, B. (1971). Word frequency book. New York: American Heritage Publishing Company, Inc.

Fischer, P. E. (1979). Words according to the GFB sequency for teaching and testing reading. (Available from Dr. Phyllis E. Fischer, Department of Special Education, Southern Connecticut State University, New Haven, Connecticut).

Fowler, C. A., Liberman, I. Y., & Shankweiler, D. (1977). On interpreting the error pattern of the beginning reader. Language and Speech, 20, 162-173.

Fowler, C. A., Shankweiler, D., & Liberman, I. Y. (1979). Apprehending spelling patterns for vowels: A developmental study. Language and Speech, 22, 243-252.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. Journal of Experimental Psychology: Human Perception and Performance, 5, 674-691.

Gough, P. B., & Hillinger, M. L. (1980). Learning to read: An unnatural act. Bulletin of the Orton Society, 30, 179-196.

Jastak, J. F., Bijou, S. W., & Jastak, S. R. (1978). Wide range achievement test. Wilmington, DE: Jastak Associated.

Johnson, D. D. (1970). Factors related to the pronunciation of vowel clusters (Tech. Rep. No. 149). Madison, WI: Research and Development Center for Cognitive Learning.

Johnson, D. D., & Venezky, R. L. (1976). Models for predicting how adults pronounce vowel digraph spellings in unfamiliar words. Visible Language, 10, 257-268.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.

Liberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. Bulletin of the Orton Society, 23, 65-77.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. Venezky (Eds.), Orthography, reading and dyslexia. Baltimore, MD: University Park Press.

Lukatela, G., & Turvey, M. (1980). Some experiments on the Roman and Cyrillic alphabets of Serbo-Croatian. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading and dyslexia. Baltimore, MD: University Park Press.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years, predicted from phonemic awareness skills in kindergarten. Scandinavian Journal of Psychology, 21, 159-173.

Marcel, A. J. (1980). Surface dyslexia and beginning reading: A revised hypothesis of the pronunciation of print and its impairments. In M. Coltheart, K. Patterson, & J. Marshall (Eds.), *Deep dyslexia*. London: Routledge & Kegan Paul.

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

Ryder, R. J., & Pearson, D. D. (1980). Influence of type-token frequencies and final consonants on adults' internalization of vowel digraphs. *Journal of Educational Psychology*, *72*, 618-624.

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. G. (1976). Exploring the relations between reading and speech. In R. M. Knights & D. J. Bakker (Eds.), *The neuropsychology of learning disorders: Theoretical approaches*. Baltimore, MD: University Park Press.

Venezky, R. L. (1967). English orthography: Its graphical structure and its relation to sound. *Reading Research Quarterly*, *2*, 75-105.

Venezky, R. L., & Johnson, D. D. (1973). Development of two letter-sound patterns in grades one through three. *Journal of Experimental Psychology*, *64*, 109-115.

Venezky, R. L., & Massaro, D. W. (1979). The role of orthographic regularities in word recognition. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading* (Vol. 1). Hillsdale, NJ: Erlbaum.

Weber, R. (1970). A linguistic analysis of first-grade reading errors. *Reading Research Quarterly*, *5*, 427-451.

PUBLICATIONS
APPENDIX

PUBLICATIONS

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25-33). New York: Academic Press.

Best, C. T. (Ed.). (1985). *Hemispheric function and collaboration in the child*. New York: Academic Press.

Borden, G. J., Baer, T., & Kenney, M. K. (1985). Onset of voicing in stuttered and fluent utterances. *Journal of Speech and Hearing Research*, *28*, 363-372.

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 Press.

Carello, C., & Turvey, M. T. (1984). On vagueness and fictions as cornerstones of a theory of perceiving and acting: A comment on Walter (1983). *Cognition and Brain Theory*, *7*, 247-261.

Carello, C., Turvey, M. T., & Kugler, P. N. (1985). The information support for upright stance. *The Behavioral and Brain Sciences*, *8*, 151-152.

Feldman, L. B., Lukatela, G., & Turvey, M. T. (1985). Effects of phonological ambiguity on beginning readers of Serbo-Croatian. *Journal of Experimental Child Psychology*, *29*, 492-510.

Fischer, F. W., Shankweiler, D., & Liberman, I. Y. (1985). Spelling proficiency and sensitivity to word structure. *Journal of Memory and Language*, *24*, 423-441.

Fowler, C. A., Napps, S., & Feldman, L. B. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition*, *13*, 241-255.

Gurjanov, M., Lukatela, G., Lukatela, K., Savić, M, & Turvey, M. T. (1985). Grammatical priming of inflected nouns by the gender of possessive adjectives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 692-701.

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

Harris, K. S., Rubin-Spitz, J., & McGarr, N. S. (1985). The role of production variability in normal and deviant developing speech. In J. Lauter (Ed.), *ASHA Reports* *15*, 50-57.

Kelso, J. A. S. (in press). Pattern formation in multidegree of freedom speech and limb movements. *Experimental Brain Research Supplementum*.

Kelso, J. A. S., & Kay, B. (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., Saltzman, E. L., & Tuller, B. (in press). The dynamical perspective in speech production: Data and theory. *Journal of Phonetics*.

Kelso, J. A. S., & Scholz, J. P. (in press). Cooperative phenomena in biological motion. In H. Haken (Ed.), *Complex systems: Operational approaches in neurobiology, physics, and computers*. Berlin: Springer Verlag.

Krakow, R. A., & Hanson, V. L. (1985). Deaf signers and serial recall in the visual modality: Memory for signs, fingerspelling, and print. *Memory & Cognition*, *13*, 265-272.

Kugler, P. N., & Turvey, M. T. (in press). *Information, natural law, and the self-assembly of rhythmic movement*. Hillsdale, NJ: Erlbaum.

Laufer, A. (1985). Phonetic description of Hebrew consonants. *Publication of the Israeli Academy of Sciences*, *7*, 67-107.

Laufer, A. (1985). The pharyngeal consonants. *Publication of the Israeli Academy of Sciences*, *7*, 39-66.

Levitt, A. G., & Healy, A. F. (1985). The roles of phoneme frequency, similarity, and availability in the experimental elicitation of speech errors. *Journal of Memory and Language*, *24*.

Liberman, A. M., & Mattingly, I. G. (in press). The motor theory of speech perception revised. *Cognition*.

Mann, V. A. (in press). Phonological awareness: Nature and nurture both contribute. *Annual Bulletin Research Institute of Logopedics and Phoniatrics*.

Michaels, C., Prindle, S., & Turvey, M. T. (1985). A note on the natural basis of action categories: The catching distance of mantids. *Journal of Motor Behavior*, *17*, 255-264.

Munhall, K. G. (in press). An examination of intra-articulator relative timing. *Journal of the Acoustical Society of America*.

Munhall, K. G., & Ostry, D. J. (1985). Ultrasonic measurement of laryngeal kinematics. In I. R. Titze & R. C. Scherer (Eds.), *Vocal fold physiology*. Denver: Denver Center for the Performing Arts.

Repp, B. H. (in press). Perceptual coherence of speech: Stability of silence-cued stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*.

Repp, B. H., & Williams, D. R. (1985). Influence of following context on perception of the voiced-voiceless distinction in syllable-final stop consonants. *Journal of the Acoustical Society of America*, *78*, 445-457.

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

Saltzman, E. L. (1985). On the coordination and regulation of complex sensorimotor systems. [A review of H. T. A. Whiting (Ed.). (1984) *Human motor actions: Bernstein reassessed*]. *Contemporary Psychology*, *30*, 531-533.

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.

Scholz, J. P., Turvey, M. T., & Kelso, J. A. S. (in press). Naturalizing the context for SMA function. *The Behavioral and Brain Sciences*.

Turvey, M. T. (in press). Intentionality: A problem of multiple reference frames, specificational information and extraordinary boundary conditions on natural law. *The Behavioral and Brain Sciences*.

Turvey, M. T. (in press). Beyond anatomical specificity. *The Behavioral and Brain Sciences*.

Turvey, M. T., & Carello, C. (1985). The equation of information and meaning from the perspectives of situation semantics and Gibson's ecological realism. *Linguistics and Philosophy*, *8*, 195-202.

Turvey, M. T., & Carello, C. (in press). The ecological approach to perceiving-acting: A pictorial essay. *Acta Psychologica*.

Turvey, M. T., & Kay, B. (1985). Control of complex systems: An intrinsically derived or imposed process? [Review of Selfridge, O. G., Rosslard, E. L., & Arbib, M. A. (Eds.). Adaptive control of ill-defined systems: NATO conference series. II. Systems science Vol. 16.] Contemporary Psychology, 30, 687-688.

Whalen, D. H., & Samuel, A. G. (1985). Phonetic information is integrated across intervening nonlinguistic sounds. Perception & Psychophysics, 37, 579-587.

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 A099958	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	**	**

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
Computer Microfilm International
Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories **Status Report** on **Speech Research** is abstracted in Language and Language Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) Haskins Laboratories 270 Crown Street New Haven, CT 06511	2a REPORT SECURITY CLASSIFICATION Unclassified 2b GROUP N/A

3 REPORT TITLE

Haskins Laboratories Status Report on Speech Research, SR-82/83, 1985

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

Interim Scientific Report

5 AUTHOR(S) (First name, middle initial, last name)

Staff of Haskins Laboratories, Alvin M. Liberman, P.I.

6 REPORT DATE	7a. TOTAL NO OF PAGES	7b NO OF REFS
October, 1985	300	613

8a CONTRACT OR GRANT NO		9a. ORIGINATOR'S REPORT NUMBER(S)
HD-01994	NS13870	SR-82/83, 1985
HD-16591	NS13617	
N01-HD-5-2910	NS18010	9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
RR-05596	N00014-83-K-0083	None
BNS-8111470		

10 DISTRIBUTION STATEMENT

Distribution of this document is unlimited*

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY
N/A	See No. 8

13 ABSTRACT

This report (1 April-30 September) 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:

-On vagueness and fictions as cornerstones of a theory of perceiving and acting: A comment on Walter (1983)
-The informational support for upright stance
-Determining the extent of coarticulation: Effects of experimental design
-The roles of phoneme frequency, similarity, and availability in the experimental elicitation of speech errors
-On learning to speak
-The motor theory of speech perception revised
-Linguistic and acoustic correlates of the perceptual structure found in an individual differences scaling study of vowels
-Perceptual coherence of speech: Stability of silence-cued stop consonants
-Development of the speech perceptuomotor system
-Dependence of reading on orthography: Investigations in Serbo-Croatian
-The relationship between knowledge of derivational morphology and spelling ability in fourth, sixth, and eighth graders
-Relations among regular and irregular, morphologically-related words in the lexicon as revealed by repetition priming
-Grammatical priming of inflected nouns by the gender of possessive adjectives
-Grammatical priming of inflected nouns by inflected adjectives
-Deaf signers and serial recall in the visual modality: Memory for signs, fingerspelling, and print
-Did orthographies evolve?
-The development of children's sensitivity to factors influencing vowel reading

DD FORM 1 NOV 65 1473 (PAGE 1)

UNCLASSIFIED

Security Classification

S/N 0101-807-6811 *This document contains no information not freely available to the general public. It is distributed primarily for library use.

A-31409

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Perception: motor theory, learning, perceptual structures, vowels, individual differences, scaling, linguistics, perceptuomotor, system development, stop consonants, silent cues, coherence Speech Articulation: speaking acquisition, coarticulation extent, phoneme frequency effects, speech errors, experimental design Motor Control: perceiving and acting, theory, information, perception, development, upright stance Reading: development, orthography, Serbo-Croatian, morpnology, spelling ability, fourth grade, sixth grade, eighth grade, lexicons, priming, inflection, nouns, adjectives, grammar, possessive tense, sign language, serial recall, deafness, memory, fingerspelling, print, evolution, vowels						

DD FORM 1 NOV 65 1473 (BACK)

101-807-6821

UNCLASSIFIED

Security Classification

A-31409