DOCUMENT RESUME

ED 257 159 CS 504 934

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, January 1-March 31, 1985.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Institutes of Health (DHHS), Bethesda, Md.; National Inst. of Child Health and Human Development (NIH), Bethesda, Md.; National Inst. of Neurological and Communicative Disorders and Stroke (NIH), Bethesda, Md.; National Science Foundation, Washington, D.C.; Office of Naval Research, Washington, D.C.
REPORT NO	SR-81-(1985)
PUB DATE	85
CONTRACT	NICHHD-NO1-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	321p.
AVAILABLE FROM	U.S. Department of Commerce; National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22151.
PUB TYPE	Reports - Research/Technical (143) -- Information Analyses (070)
EDRS PRICE	MF01/PC13 Plus Postage.
DESCRIPTORS	Academic Aptitude; Adults; *Articulation (Speech); Cognitive Processes; *Communication Research; Elementary Secondary Education; *Language Processing; *Language Research; Linguistic Theory; Memory; Perception; Reading Difficulties; *Speech Communication; Speech Handicaps; Speech Improvement; *Speech Skills; Spelling

ABSTRACT

One of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications, this report covers the period January 1 to March 31, 1985. Studies summarized in the report cover such topics as (1) segmentation of coarticulated speech in perception (2) intrinsic time in speech production, (3) a theoretical model of phase transitions in hand movements, (4) repetitive naming and the detection of word retrieval deficits in the beginning reader, (5) linguistic abilities and spelling proficiency in kindergarten children and adult poor spellers, (6) errors in short term memory for good and poor readers, (7) longitudinal prediction and prevention of early reading difficulty, (8) temporary memory for linguistic and nonlinguistic material in relation to the acquisition of Japanese Kana and Kanji, (9) speech perception and the prelinguistic infant, (10) categorical trends in vowel imitation, (11) cognitive processes in reading, and (12) processing kinematic data. (HOD)

SR-81 (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 January – 31 March 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
Margo Carter, Technical Illustrator
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++
Daniel Holender1
Satoshi Horiguchi2
Leonard Katz*
J. A. Scott Kelso
Gary Kidd+
Andrea G. Levitt*
Alvin M. Liberman*
Isabelle Y. Liberman*
Leigh Lisker*
Kristine MacKain*
Virginia Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Kevin Munhall+++
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
Margo Carter
Philip Chagnon
Alice Dadourian
Vincent Gulisano
Donald Hailey
Raymond C. Huey*
Sabina D. Koroluk
Bruce Martin
Betty J. Myers
Nancy O'Brien
Gail K. Reynolds
William P. Scully
Richard S. Sharkany
Edward R. Wiley

Students*

Christopher Allen
Joy Armson
Dragana Barac
Sara Basson
Eric Bateson
Suzanne Boyce
Teresa Clifford
Andre Cooper
Jan Edwards
Jo Estill
Nancy Fishbein
Carole E. Gelfer
Bruce Kay
Noriko Kobayashi
Rena A. Krakow
Hwei-Bing Lin
Katrina Lukatela
Harriet Magen
Sharon Manuel
Richard McGowan
Jerry McRoberts
Susan Nittrouer
Lawrence D. Rosenblum
Richard C. Schmidt
John Scholz
Robin Seider
Suzanne Smith
Katyanee Svastikula
Daniel Weiss
David Williams

---

*Part-time
1Free University of Brussels, Brussels, Belgium
2Visiting from University of Tokyo, Japan
+NIH Research Fellow
++NRSA Training Fellow
+++Natural Sciences and Engineering Research Council of Canada Fellow

Status Report on Speech Research

Haskins Laboratories

SEGMENTATION OF COARTICULATED SPEECH IN PERCEPTION*

Carol A. Fowler†

Abstract. The research investigates how listeners segment the acoustic speech signal into phonetic segments and explores implications that the segmentation strategy may have for their perception of the (apparently) context-sensitive allophones of a phoneme. Two manners of segmentation are contrasted. In one, listeners segment the signal into temporally discrete, context-sensitive segments. In the other, which may be consistent with the talker's production of the segments, they partition the signal into separate, but overlapping, segments freed of their contextual influences. Two complementary predictions of the second hypothesis are tested. First, listeners will use anticipatory coarticulatory information for a segment as information for the forthcoming segment. Second, subjects will not hear anticipatory coarticulatory information as part of the phonetic segment with which it co-occurs in time. The first hypothesis is supported by findings on a choice-reaction time procedure; the second is supported by findings on a 4IAX discrimination test. Implications of the findings for theories of speech production, perception and of the relation between the two are considered.

Skilled listeners sometimes behave as if they have have not extracted all of the phonetic structure from an acoustic speech signal. Listeners to fluent speech more readily recognize target syllables and words than they recognize target phonetic segments (McNeill & Lindig, 1973; Savin & Bever, 1970) and, in their perceptions of a fluently-produced sequence, they are likely to "restore" phonetic segments that are overdetermined and missing (Samuel, 1981; Warren, 1970) or mispronounced (Marslen-Wilson & Welsh, 1978).

Despite this apparent inattention to the phonetic structure of speech by skilled listeners, the structure persists in languages; that is, "duality of patterning" (Hockett, 1960) is not, apparently, disappearing from them. Indeed it is universal to languages, presumably because it is required to maintain the openness of their open lexical classes. Moreover, the phonetic structure of words _is_ psychologically real, even to the skilled listeners just described when they talk. For example, speech errors commonly consist of phonetic-segment misorderings and substitutions (cf. Fromkin, 1973), and many

---

*Perception & Psychophysics, 1984, 36, 359-368.

†Also Dartmouth College.

Acknowledgment. The research was supported by NSF Grant BNS 8111470 and NICHD Grant HD-16591-01 to Haskins Laboratories. I thank George Wolford for his comments on the manuscript and advice on data analysis.

language games (including rhyming, alliteration, and Pig Latin, among others) involve operations performed on the phonetic- (or phonological-) segmental structure of words (cf. Pisori, in press).[1]

If the phonetic structure of words is to be perpetu[illegible] in languages, and if language learners are to become talkers who Spoone[illegible] play language games, the learners must be able to extract phonetic [illegible]ure from an acoustic speech signal even if they will not always do so when they become skilled users of the language. This observation implies that an acoustic speech signal must provide sufficient information for extraction of the phonetic structure of the talker's intended message. Yet the signal has provided major barriers to investigators' efforts to extract phonetic segments from it.

Two related barriers are those of segmentation and invariance. Both problems arise because speech is coarticulated--that is, because articulatory gestures for successive phonetic segments are not temporally discrete. The segmentation problem is to understand how separate phone-sized segments may be extracted from a signal in which information for the segments overlaps in time. The invariance problem is to rationalize listeners' classifications of phonetic tokens into types. It is called the "invariance" problem because the presumption has been (e.g., Stevens & Blumstein, 1981) that its solution lies in discovering acoustic invariants that exist across tokens. The search for invariance is rendered difficult by coarticulation, which ensures that the acoustic signal during a time window most closely identifiable with one phonetic segment is context-sensitive, not (wholly) invariant. Moreover, the problem of explaining listeners' classifications goes beyond the search for acoustic invariance. Certain sets of phones (for example, the [d]s in [di], [da] and [du]) are always classified as tokens of a common phonemic type even though acoustic information for the different tokens is largely (and, in some synthetic stimuli, entirely) context-sensitive, and even though listeners attend to the context-sensitive information (in the example, the second-formant transitions) more closely than to any invariant information (e.g., the shape of the release-burst spectrum; cf. Stevens & Blumstein, 1981) that may be present when they identify the phones (Walley & Carrell, 1983).

The present research contrasts two possible ways that listeners may segment the acoustic speech signal into phone-sized segments. These strategies offer different perspectives on the problem of explaining listeners' classifications of apparently context-sensitive phonetic segments into types.

Figure 1a displays an acoustic speech signal schematically, and Figures 1b and 1c illustrate the two segmentation strategies. In Figure 1a, the horizontal axis is time and the vertical axis a provisional dimension, "prominence." The prominence of a segment in an acoustic signal refers to the extent to which acoustic properties characteristic of that segment are salient in the signal. For example, in a syllable, /si/, /s/ is more prominent than /i/ during the frication noise even though production of /i/ begins before or during closure for the frication (Carney & Moll, 1971), and evidence of its production is available in the signal.

One segmentation strategy, illustrated in Figure 1b, uses the relative prominence of successive segments to establish boundaries between them. This essentially is the procedure used in the phonetics literature when measure-

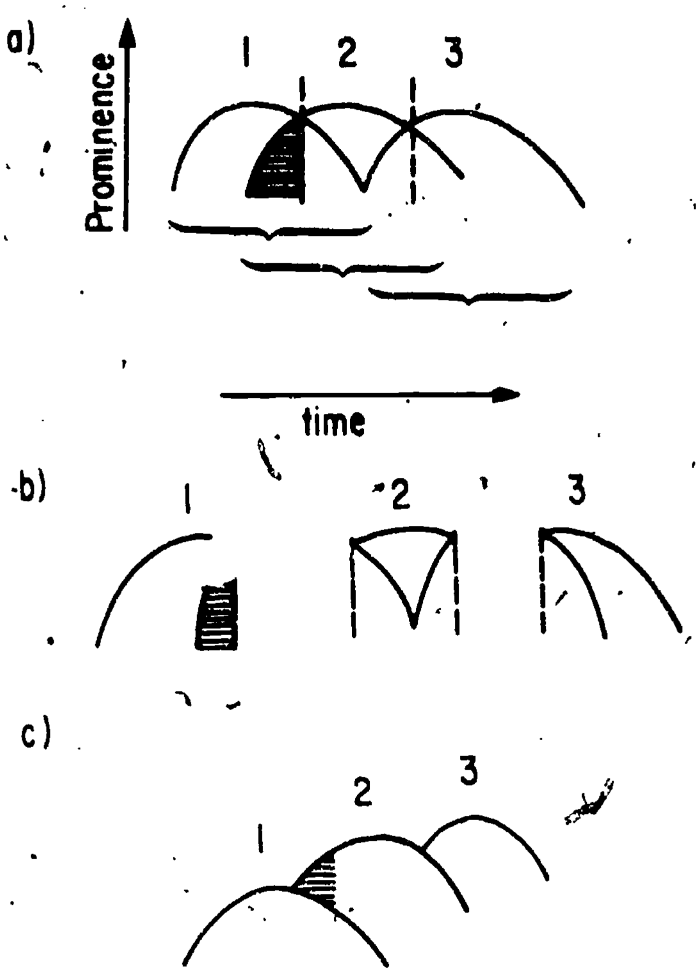


Figure 1. Schematic display of segment production. In Figure 1a, segments are produced in overlapping time frames. In Figure 1b, segmentations are made at points in time when one segment ceases to dominate in the signal and another takes over. This divides speech into discrete, context-sensitive segments. In Figure 1c, speech is segmented along coarticulatory lines into overlapping segments freed of their contextual influences.

ments are made of phonetic-segment durations, acoustically realized (e.g., Klatt, 1975; Peterson & Lehiste, 1960; and see Lisker, 1974, for other references). This procedure divides the acoustic speech signal into discrete context-sensitive phonetic segments. Disagreements concerning where to draw the segmentation lines arise when neither of two neighboring segments clearly predominates in some acoustic interval. For example, Lisker (1972) cites research in which vowel onsets sometimes include and sometimes exclude formant transitions following consonant release.

A second possible strategy is illustrated in Figure 1c. The acoustic signal is segmented along coarticulatory lines into overlapping phonetic segments, free from the contextual influences of phonetic neighbors. Thus, for example, in /si/, the onset of /i/ is identified where production of /i/ is first detectable within the /s/ frication, not where its acoustic manifestations begin to predominate in the signal.

Measurement conventions reflecting the segmentation strategy of Figure 1b are adopted in the phonetics literature to maximize reliability, not necessarily either to mimic listeners' segmentation strategies or to capture any articulatory lines of segmentation that the signal may reflect. Indeed, the literature offers a hint that the conventions do not mirror the listener's manner of segmenting the signal. The hint is provided by two independently developed, but possibly converging, lines of research.

First is evidence that listeners use anticipatory coarticulatory influences on one phonetic segment as information for the influencing segment (e.g., Alfonso & Baer, 1983; Martin & Bunnell, 1982; Ochiai & Fujimura, 1961; Whalen, 1984; but see Lehiste & Shockey, 1972). For example, Whalen (1984) cross-spliced friction noises across tokens of /sa/, /su/, /ša/, and /šu/ and asked listeners to identify the vowels of each syllable in a choice reaction-time procedure. Listeners were faster and more accurate when the frication noises provided accurate anticipatory information for the vowels than when they provided misleading information.

In itself this finding can be explained assuming the segmentation strategy of Figure 1b. Having partitioned the signal into discrete, context-sensitive allophones (Wickelgren, 1969, 1976), listeners may use the context-sensitivity of the allophone that precedes a target segment to predict the target's identity.

The second line of research shows that listeners "compensate" for coarticulatory influences on phonetic segments when they identify them (Liberman, Delattre, & Cooper, 1952; Mann, 1980; Mann & Repp, 1980). For example, /s/ and /š/ are distinguished acoustically in part by the relative locations of energy concentrations in their spectra, that for /s/ being higher than that for /š/. In the context of a following /u/, however, the spectra for both consonants are lowered by anticipatory lip rounding. Compatibly, listeners accept stimuli with concentrations of energy in lower frequencies as tokens of /s/ if the frication is followed by a rounded vowel than if it is followed by an unrounded vowel (Mann & Repp, 1980). The same frication noise may be identified as /š/ before /i/, but as /s/ before /u/.

Again, in itself, this finding can be explained assuming segmentation into discrete, context-sensitive segments. The explanation is that compensation reflects an adjustment to information for an earlier segment based on knowing the effects that anticipation of the following segment should have had on it (cf. Mann & Repp, 1980).

Considered together, however, this account and the foregoing account of listeners' use of anticipatory coarticulatory information are paradoxical. For the segmentation hypothesis illustrated in Figure 1b to account for the findings from the reaction-time procedure, it must be supposed that the coarticulatory effects on a segment are identifiable as such before the onset of the anticipated segment. Otherwise, reaction times would not be reduced when coarticulatory information is "predictive." (Nor, as Meltzer, Martin, Mills, Imhoff, & Zohar, 1976, have shown, would they be improved even further when the anticipatory information is shifted earlier in time than its natural time of occurrence.) However, for it to explain a finding of compensation, it must be supposed that the segment following a context-sensitive allophone is used to guide the identification of the contextual influence on the allophone. That is, in the one instance, the coarticulatory information facilitates later identification of the segment it anticipates; in the other, it can only be identified as a coarticulatory influence on an allophone after the segment it anticipates has itself been identified.

The alternative segmentation hypothesis under consideration satisfies both sets of findings. Listeners may segment the speech stream along its coarticulatory lines into overlapping phonetic segments (Figure 1c). There are two consequences of this segmentation strategy: anticipatory coarticulatory information is perceived as the onset of the segment it "anticipates" and the same information therefore is not integrated with concurrent information for the preceding phonetic segment. That is, "compensation" occurs as a necessary by-product of segmentation.[2] From this perspective, compensation is symptomatic of an additional consequence of segmentation. Because sources of context-sensitivity are not integrated with information for segments with which they co-occur in time, the same phonetic segment in different coarticulatory contexts is predicted to sound approximately the same to listeners. That is, listeners may perceive the tokens of a phonetic type as the same or very similar across different phonetic contexts because sources of contextual influence have not been integrated with the tokens.

The present research is designed to contrast the foregoing accounts of segmentation. In particular, the research first asks whether listeners will use coarticulatory information for a vowel within the acoustic domain of a preceding phonetic segment (in the present case, /g/) as information for the vowel (cf. Martin & Bunnell, 1982; Whalen, 1984). It next asks, if they do, whether they also show evidence of "compensation" for contextual influences of the vowel in their perceptual judgments of the preceding phonetic segment. If listeners exhibit both behaviors on the same syllables, then the segmentation strategy of Figure 1b can be ruled out on grounds previously outlined: that strategy requires that listeners identify the coarticulatory information as such before identifying the segment it anticipates in the paradigm used by Martin and his colleagues; however, it requires the reverse ordering of identification to explain apparent compensation. The segmentation strategy of Figure 1c provides a unified account of both findings.

Use of anticipatory coarticulatory information to identify a forthcoming segment will be tested using the cross-splicing, choice reaction-time procedure developed by Martin and Bunnell (1982) and used by Whalen (1984). Compensation will be assessed using a 4IAX discrimination procedure on the same stimuli (cf. Pisoni, 1971).

Listeners are said to compensate for contextual influences if their perceptual judgments of a phonetic segment suggest that the contextual influences have been eliminated or reduced (cf. Mann & Repp, 1980). A 4IAX trial such as the following will allow assessment of compensation:

$g_i i$-----$g_i u$----------------$g_i i$-----$g_u u$

The trial includes four syllables temporally organized into pairs. Members of a pair have different vowels, but the vowels are the same across the pairs. "g" refers to a stop burst originally produced either in a [gi] syllable (subscripted with "i") or in a [gu] syllable (subscripted with "u"). Subjects are asked to decide which pair has members that sound more similar. If listeners make their assessment based on the relative acoustic overlap between members of a pair, they should select the members of the first pair as more similar than members of the second because the former have identical bursts. The opposite prediction is made if listeners make their judgments with contextual influences eliminated from the different consonantal segments (that is, if they "compensate" for those influences). In that case, influences of different vowels are eliminated from identical stop bursts in the first pair, yielding different residuals. In the second pair, the different influences of /i/ and /u/ are eliminated from contextually-appropriate stop bursts yielding, by hypothesis, identical, context-free phonetic tokens. Thus, members of the second pair should be judged more alike than members of the first.

This research continues a series of studies reported elsewhere (Fowler, 1983b; Fowler & Smith, in press). The earlier research used the paired choice reaction-time and 4IAX procedures just described to test listeners' perceptions of coarticulatory influences of stressed vowels on preceding or following cross-spliced, unstressed schwa. Predictions based on the hypothesis that listeners segment speech along coarticulatory lines were partially supported for these stimuli. We found positive evidence for the segmentation strategy of Figure 1c (and correspondingly, disconfirming evidence for the strategy of Figure 1b) when the coarticulatory effects under study combined both carryover and anticipatory effects of stressed vowels (as in /ib bi/ and /ab ba/). The reaction-time procedure also provided positive evidence for contexts in which coarticulatory effects were only anticipatory (as in /b bi/ and /b ba/). However, in the 4IAX task, responses were random when coarticulatory effects were anticipatory only.

We hypothesized that the chance performance in the 4IAX study in which only anticipatory coarticulation was present was due to a lack of sensitivity of the 4IAX procedure as compared to the choice reaction-time procedure, and not to a restriction on the applicability of the segmentation hypothesis to carryover coarticulatory influences. This interpretation is plausible because anticipatory coarticulation of stressed vowels is more limited in these contexts than carryover coarticulation (Bell-Berti & Harris, 1976; Fowler, 1981a), and because, as compared to the choice reaction-time procedure, the discrimination procedure places severe memory demands on the listener and requires a difficult judgment. However, it remains to be demonstrated that segmentation occurs along coarticulatory lines whether the lines reflect anticipatory or carryover coarticulation.

The present experiment used stimuli in which anticipatory coarticulatory effects of a vowel on a preceding segment are larger than they were on the schwas of our previous study. Stimuli in the experiment are the stop-vowel syllables /gi/ and /gu/. Because the stop immediately precedes the vowel and because velar stops coarticulate extensively with vowels, I expected these stimuli to enable observation of segmentation of anticipatory coarticulatory influences of the vowel from the acoustic domain of the consonant if it occurs.

Experiment

Methods.

Subjects. Subjects were 36 students at Dartmouth College. All were native speakers of English and reported normal hearing.

Materials

Stimuli. Stimuli were two tokens each of the monosyllables /gi/ and /gu/ produced by a female talker. They were input to a New England Digital minicomputer, sampled at 20 KHz and filtered at 10KHz.

Based on criteria provided by Dorman, Studdert-Kennedy, and Raphael (1977) the release bursts of each utterance were identified and segmented from the remainder of the syllable. Release bursts ranged in duration from 16 to 20 ms and did not vary systematically in duration with the identity of the vowel. (These values compare to averages across /gid/ and /gud/ of 7.5 ms for one speaker reported by Dorman et al. and 22.5 for the second speaker.) The period of aspiration following release was removed from the vocalic portion of the syllable and was replaced by an equivalent period of silence. This was done to avoid abrupt discontinuities in the spectra when bursts and vocalic segments were cross-spliced. In the test orders, stimuli were presented in low levels of white noise, which improved the perceived quality of the stimuli by masking the temporal discontinuity. The intervals of aspiration ranged from 6 to 15 ms (as compared to averaged values of 11.5 and 12.5 for talkers in Dorman et al.). Durations of the voiced portion of each syllable were 429 and 430 for tokens of /gi/ and 359 and 361 for tokens of /gu/.

Three types of test syllables were constructed from the syllable fragments just described. The four "original" syllables consisted of release bursts and vocalic portions that had originally been produced together. They were separated by a period of silence equivalent to the original period of aspiration for the vocalic segment. "Spliced" syllables were release bursts from one token of a syllable type attached to a silent interval and vocalic portion originally associated with production of the other token of the same syllable type. (That is, for example, a burst from one token of /gi/ was spliced onto an interval of silence and a silent interval and vocalic portion of the other token of /gi/.) There were four spliced syllables. Eight "cross-spliced" syllables were created by attaching a release burst from a token of one type onto a silent interval and vocalic portion associated with a token of the other phonemic type. (That is, for example, a burst from a token of /gi/ was spliced onto the vocalic portion of a /gu/ syllable.)

Identification test. An identification test presented release bursts, vocalic portions and whole CVs for identification in that order in separate blocks of 32 trials.

The identification test was originally presented to 12 naive subjects, who had not heard the stimuli before, and to 12 subjects who had just completed the choice reaction-time and 4IAX tests to be described. These 24 subjects were given an answer sheet with alternatives "bee," "dee," "gee," "boo," "doo," and "goo" arranged in three blocks of 32 rows, one row for each trial of the identification test. In the test, both groups of subjects exceeded chance in their ability to identify the syllables' vowels from their release bursts alone. This suggested the possibility that some or all of the subjects heard diphthongal vowels in the cross-spliced syllables. To assess that, a new group of 12 subjects took the identification test preceded by the reaction-time and 4IAX procedures. The response sheet given to these subjects for the identification test allowed six new response alternatives: "bwee," "dwee," "gwee," "byoo," "dyoo," and "gyoo" in addition to the original six.

Choice reaction-time test. The choice reaction time study, modeled after the paradigm of Whalen (1984), consisted of original, spliced and cross-spliced stimuli randomly presented one at a time in four blocks of 48 trials. Predictions were that, because the release burst would provide misleading information for the vowel in cross-spliced stimuli, reaction time and accuracy to identify the vowel in those stimuli would be inferior to the same measures taken on original and spliced stimuli.

4IAX discrimination test. The 4IAX test consisted of three blocks of 64 trials. One half of the trials were of type A and one half were of type B, both illustrated by example below. Stimuli in this test were either spliced or cross-spliced; no original stimuli were presented. (As before, subscripts on the [g]s indicate the vowel with which the release burst had originally been produced.)

Trials of type A were designed to test whether listeners could distinguish the different bursts in the context of a vowel. As described in the introduction, trials of type B provided the critical test of the segmentation hypotheses.

A: $g_i i$—$g_i i$----------$g_i i$—$g_u i$

B: $g_i i$--$g_i u$----------$g_i i$—$g_u u$

In either trial type, four stimuli were presented per trial, arranged temporally in two pairs. Members of one pair of an A trial were identical. One member of the second pair was identical to the members of the first pair. The fourth syllable differed from the others in its release burst. That syllable was always cross-spliced. Trials of type B were like trials of type A except that the vocalic segments within a pair were different. In B trials, then, the members of one pair had identical release bursts. In the other pair, one item had the same release burst as the members of the first-mentioned pair; the other had a different release burst. In pairs where bursts were identical, one member of the pair was spliced and one cross-spliced. For the pair with different release bursts, both members were spliced so that the bursts were in vocalic contexts compatible with those in which they had originally been produced.

In a trial, the offset-onset time was 200 ms within a pair; between pairs it was 500 ms. If the stimuli in the sample trials above are labeled 1, 2, 3 and 4, then their ordering in the sample trials is 12-34. In addition to this ordering were equal numbers of occurrences of orders 21-43, 34-12 and 43-21. In the sample trials above, release burst [g_i] occurs three times and burst [g_u] occurs just once. There were equal numbers of trials in which [g_u] was the more frequent burst in the trial.

Procedure

Group 1. The twelve subjects in this group took only the identification test. They listened to stimuli over headphones. Stimuli were presented on-line on a New England Digital minicomputer. In this test as in the others the stimuli were mixed with a low level of white noise.

Subjects were told that stimuli on the first third of the test were the first few milliseconds of a CV syllable and their task was to guess the identity of the whole syllable from the fragment. The syllable types they might hear were pronounced for them. They were instructed to circle the response choice on the answer sheet that best represented the syllable from which the fragment had been excised. They were required to guess if necessary. In addition, they were told that they might hear all or only some of the syllables represented on the answer sheet. Therefore, they should circle their best guess based on what they heard and not attempt to distribute their responses evenly among the response alternatives. On the second block, they were told that the stimuli were the remainders of the CV syllables with the first few milliseconds excised. Instructions were the same as on the first block. Finally, on the third block, they were told that stimuli were the two types of syllable fragments they had just been listening to, but rejoined to make a whole CV syllable. Instructions were to identify the CV on each trial as one of the six listed on the answer sheet.

Trials were initiated individually by key press; therefore, subjects had unlimited time to make their responses.

Groups 2 and 3. The twenty-four subjects in these groups took the choice reaction-time test, the 4IAX test and the identification test in that order. The procedures for these groups were identical; they differed only in the response sheets they received on the identification test.

In the reaction time procedure, subjects listened over headphones to stimuli presented on-line and mixed with noise as in the identification procedure just described. They were instructed to identify the vowel in each syllable as "ee" or "oo" by hitting the appropriate labeled key on the computer terminal's keyboard as quickly and as accurately as possible. They received response-time feedback after every trial and averaged response-times and accuracy at the end of each of the four blocks of 48 trials. They were asked to keep their accuracy above .9. The first block of trials served as practice.

In the 4IAX procedure, subjects were instructed to choose the first or second pair of stimuli on each trial as having the more similar members. They signaled their selection by typing "1" or "2" into the computer, using the calculator pad on the keyboard. They followed that selection with a confidence judgment (1: guess, 2: intermediate certainty, 3: high level of confidence). Neither response was timed.

In this test there were three blocks of 64 trials, the first block serving as practice. Trials were self-paced and there was no feedback.

Last in the session, subjects took the identification test.

Results

Identification. Identifications of bursts, vocalic portions and original, spliced and cross-spliced CVs are provided in Table 1 for all three groups of subjects. Consonant and vowel identifications are displayed separately.

In identification of consonants from isolated bursts, subjects are close to chance in Group 1 (naive listeners). Performance for experienced subjects, particularly those in Group 3 is better than chance. More remarkable is performance identifying the vowel from the burst. All groups exceeded chance on this identification task.

As for the isolated vocalic portions of the syllables, the vocalic portion of [gu] led to predominantly "b" identifications in all groups. In contrast, the vocalic portion of [gi] evidently was more ambiguous leading to substantial numbers of identifications in all consonantal response categories. Vowel identifications based on vocalic portions of the syllables were accurate.

Subjects in all groups were accurate in identifying the vowels and consonants of original and spliced whole syllables. As for cross-spliced syllables, "g" was the predominant identification in [g_ui], but, in two groups, "b" was the predominant consonant identification for [g_iu]--a finding also reported by Cole and Scott (1974; for a related finding, see Liberman, Delattre, & Cooper, 1952). Subjects in Group 2 gave predominantly "d" responses for the consonant in the latter syllable.

Subjects in Group 3 did report more diphthongs in the cross-spliced syllables than elsewhere, particularly in the syllable [g_ui]. These data, as well as those for subjects reporting "b"s or "d"s in cross-spliced syllables will be used later to examine individual subject's performances in the choice reaction time and 4IAX procedures.

Overall, this test provides two pieces of information necessary to the interpretation of the next two tests. First, despite the surgery performed on the syllables, the bursts are integrated with the vocalic portions sufficiently in whole CVs that consonant identifications based on the two fragments together are different from identifications based on the separated parts. Second, the identification test provided a finding that we had expected to uncover only in the reaction-time procedure--namely, that listeners are sensitive to the information for the following vowel in the release burst. This led to the only effect that the burst appeared to have on vowel identification in the identification test. Some subjects in Group 3 identified the vowel as diphthongal.

Table 1

Identifications (Proportion of Responses) of Bursts, Vocalic Portions and Whole Syllables by Naive (Group 1) and Experienced Listeners (Group 2) with 6 Response Choices and by Experienced Listeners (Group 3) with 12 Alternatives

		b	d	g	i	wi	u	yu
				BURSTS				
Group 1	gu	20	38	42	12		88	
	gi	22	42	36	89		11	
Group 2	gu	21	30	49	24		76	
	gi	13	39	48	88		12	
Group 3	gu	17	23	60	11	05	65	19
	gi	09	19	72	73	13	06	08
				VOCALIC PORTIONS				
Group 1	gu	83	12	05	05		95	
	gi	25	50	25	97		03	
Group 2	gu	71	11	08	09		91	
	gi	44	29	27	95		05	
Group 3	gu	79	09	12		02	94	04
	gi	20	38	42	84	14	01	01
				WHOLE SYLLABLES				
				Original				
Group 1	gu			100			100	
	gi	04		96	100			
Group 2	gu			100	04		96	
	gi		02	98	98		02	
Group 3	gu		02	98			98	02
	gi	02	04	94	98	02		
				Spliced				
Group 1	gu	04		96			100	
	gi	02	02	96	98		02	
Group 2	gu	02		98			100	
	gi		04	96	100			
Group 3	gu			100			98	02
	gi		06	94	96	04		
				Cross-Spliced				
Group 1	gu	88	7	04	02		98	
	gi	18	35	47	100			
Group 2	gu	18	44	38	02		98	
	gi	08	18	74	88		12	
Group 3	gu	57	28	18	05	03	65	27
	gi	02	21	78	45	56		

Choice Reaction Time

Table 2 provides response times and accuracies in the choice reaction time procedure. In Table 2a, means are collapsed over the subjects in Groups 2 and 3. Although subjects in Group 3 responded more rapidly (by an average of 70 ms) and more accurately (by an average of 5%) response patterns and outcomes of separate ANOVAs performed on the data from each group were the same.

Table 2

Reaction Times (in ms) and Accuracy (Percent Correct) for Groups 2 and 3 in the Choice Reaction Time Study

	gu		gi	
	RT	Accuracy	RT	Accuracy
Original	455	93	441	98
Spliced	453	93	425	98
Cross-spliced	504	73	520	86

Reaction times and accuracy were subjected to separate two-way analyses of variance with factors: Syllable-type (original, spliced, cross-spliced) and vowel (/i/, /u/). In the analysis of reaction times, the main effect of syllable type was significant, $F(2,46) = 29.40$, $p < .001$, reflecting the substantially longer response times to cross-spliced as compared to spliced and original syllables. In addition, the interaction of vowel and syllable type reached significance, $F(2,46) = 4.05$, $p = .02$, because the slowing caused by cross-splicing was more marked for the syllable [g_ui] than for [g_iu].

The accuracy measure provided a compatible outcome, with performance lowest in cross-spliced as compared to original and spliced syllables, $F(2,46) = 20.87$, $p < .001$. In this analysis, the interaction did not reach significance.

The identification test had revealed that some subjects heard diphthongal vowels in cross-spliced syllables, particularly in [g_ui]. This provides an alternative account of the slowing on cross-spliced stimuli. If subjects hear diphthongs, then, as predicted, they hear the vowel information in the burst; their reaction times to cross-spliced stimuli are slowed, however, because the perceived vowels include both response alternatives and subjects have to choose just one. Subjects who do not report diphthongs may also extract vowel information from the bursts in cross-spliced syllables, yet still hear the syllable vowel as monophthongal because later vocalic information overwhelmingly contradicts information in the burst. This latter was the possibility the experiment had been designed to establish and test.

Post-hoc analyses of responses by individual subjects in Group 3 were performed to determine whether subjects responded differently depending on whether they heard the vowel as monophthongal or diphthongal. For the syllable $[g_u i]$, seven subjects consistently reported diphthongs in the identification test, three consistently reported monophthongs and two reported some of each. (Consistency in identification was defined operationally as selection of a diphthongal [monophthongal] response on at least 6 of 8 opportunities on the identification test.) For the syllable $[g_i u]$, numbers of subjects falling into the three categories were two, ten, and zero, respectively. Some subjects fell into the same category twice, because they heard the vowel in the same way on both syllables. In those instances, their data for the two syllables was pooled. Average response times and accuracy were collapsed over syllables for the seven subjects consistently reporting diphthongs in Group 3 and separately for the ten subjects reporting monophthongs. An analysis of variance comparing the two groups on the original, spliced, and cross-spliced stimuli yielded a highly significant effect of splicing condition, $F(2,30) = 39.09$, $p < .001$, but no effect of subject group and no interaction (both Fs less than one). It seems that whether or not subjects experience the anticipatory vowel information in the burst as a glide, it serves them as information for a vowel and, in cross-spliced stimuli, subjects are misled by it.

4IAX

Table 3 provides the outcome of the 4IAX test collapsed over subjects in Groups 2 and 3. The data were collapsed over the groups because analyses performed on the individual groups did not differ.

Table 3

Outcome of the 4IAX Test Collapsed Over Subjects in Groups 2 and 3

	Cross-Spliced Syllable*		Response Selection**	
Trials	$g_u i$	$g_i u$	Acoustically Identical	Acoustically Different
A	.94	.82	2.48	1.91
B	.34	.27	1.86	2.25

*Proportion of A and B trials in which listeners selected syllables having acoustically identical bursts as more similar than syllables having acoustically different bursts. **Confidence judgments.

As predicted, on A trials, listeners reliably chose syllables having acoustically identical bursts in their proper contexts as more similar than syllables having acoustically different bursts in identical contexts (/gu/: $t(23) = 18.34$, $p < .001$; /gi/: $t(23) = 11.84$, $p < .001$). This verifies that the anticipatory coarticulatory information is audible in the context of a syllable.

Of greater interest is performance on B trials. On these trials, listeners compared syllables with acoustically different bursts, each in their proper coarticulatory contexts (e.g., $[g_i i]$-$[g_u u]$) to syllables with acoustically identical bursts, one in its original context and one not (e.g., $[g_i i]$-$[g_i u]$). As predicted, on these trials in contrast to A trials, listeners reliably selected the syllables with different bursts as more similar than those with identical bursts (/gu/: $t(23) = -3.26$, $p = .004$; /gi/: $t(23) = -3.96$, $p < .001$).

As shown in Table 3, confidence judgments mirror the response selections. The confidence judgments in Table 3 are collapsed over syllable type. This was necessary because subjects occasionally had no responses either in the "acoustically different bursts" category on A trials or in the "acoustically identical bursts" category on B trials. No subject had missing data when the data were collapsed over $[g_i]$ and $[g_u]$ trials. On A trials, subjects are more confident of their (correct) judgments that syllables with identical bursts are more similar than those with different bursts. On B trials, their confidence reverses. A two-way analysis of variance (trial type [A,B] by judgment [syllables with acoustically identical bursts, those with different bursts]) was performed on the confidence judgments. In that analysis, the effect of trial type, $F(1,23) = 5.02$, $p = .03$, and the interaction, $F(1,23) = 39.33$, $p < .001$, were significant. The significant interaction reflected the effect of interest. Listeners were more confident of their correct selections of syllables having acoustically identical bursts on A trials than of their errors, $F(3,23) = 9.31$, $p < .001$; on B trials, they were less confident of their selection of those having acoustically identical bursts than of their selection of different bursts in their proper contexts, $F(3,23) = 4.28$, $p = .02$.

Response selection by individuals hearing diphthongs was examined separately from individuals hearing monophthongs. The average performance on B trials of the eight subjects reliably hearing diphthongs did not differ from that of the ten subjects hearing monophthongs, $t(16) = 1.07$, $p = .30$.

It is also of interest to look separately at subjects for whom cross splicing changed the identity of the consonant to /b/ or /d/ and those for whom it did not. For subjects of the first type, the 4IAX task confronts them with an easy between-category discrimination. For subjects in the second category, the task is one of within-category discrimination.

For these analyses, data from Groups 2 and 3 were pooled. In all, there were 15 subjects who reported the syllables with the cross-spliced burst reliably as /b/- or /d/-initial in at least one syllable. All but two of these were subjects in the condition with cross spliced $[g_i]$. Across Groups 2 and 3 there were 19 subjects reliably reporting "g" on at least one syllable. Performance differences were significant between these two groups, as expected from the general findings that between-category discrimination is easier than within-category discrimination, $t(32) = 2.92$, $p < .01$. However, subjects hearing /b/ or /d/ were not wholly responsible for the outcome on B trials. Of those 15 subjects, 13 had performance levels below .5, $t(14) = -5.76$, $p < .001$. Of the 19 subjects hearing /g/, 12 showed the predicted direction of difference, $t(18) = -2.02$, $p = .056$. We conclude, then, that although the within-category discrimination is much more difficult than the between-category discrimination, it is not qualitatively different from between-category discrimination. Overall, subjects hear syllables with acoustically different bursts in their proper coarticulatory contexts as more similar than those with acoustically identical bursts, one in its proper context and one not; making the discrimination at all is facilitated if the segmentation process leads the cross-spliced burst to fall into a different phonemic category than its original one.

Discussion

In this study, as in the earlier research reported by Fowler and Smith (in press), subjects' choice reaction-time and discrimination performances reflect the segmentation strategy of Figure 1c more closely than that of Figure 1b. Listeners use coarticulatory information as information for the influencing segment, and they do not integrate it into their perceptual experience of the segment with which it co-occurs in time. The present study extends the findings of Fowler and Smith to anticipatory coarticulatory influences and to coarticulatory relationships of consonants and vowels.

The segmentation of speech that our research supports closely resembles that achieved by a recent computer model of speech perception described by Elman and McClelland (1983). In their model (cf. McClelland & Rumelhart, 1981), features, phonemes and words are represented by "nodes" interconnected by excitatory and inhibitory links. In general, excitatory connections link nodes that are mutually consistent; inhibitory connections link nodes that are inconsistent. (For example, phoneme nodes excite words of which they are constituents; word nodes inhibit each other.) Acoustic information input to the model activates features compatible with it; in turn, the features activate phonemes consistent with them, and phonemes activate words. Of particular interest here is the segmentation of the acoustic signal that the model achieves over time as it identifies phonetic segments from an acoustic speech signal. Over time, the acoustic signal first provides stronger and then weaker evidence for the presence of a particular phonetic segment. I have called that waxing and waning of information the "prominence" pattern for a segment. In the model of Elman and McClelland, the activation pattern for a phonetic segment tracks the waxing and waning of information for the segment in the acoustic signal.

Due to coarticulation, in most time frames, the model receives featural information consistent with two phonetic segments concurrently—for example, a syllable-initial consonant and a following vowel. When that happens, two phonemes are highly activated concurrently. Eventually information for the first segment dies out leaving the highly activated second segment. The activation patterns for a sequence of phonemes, therefore, resemble the prominence curves represented in Figure 1a. Thus, in the model, although there is no explicit segmentation process separate from the process of identifying phonetic segments, nonetheless, a segmentation of the signal is achieved, and it is precisely the segmentation that I have found characteristic of human listeners. In the present study, listeners begin using acoustic information for a segment as such whenever it occurs in the speech signal. This leads to a reaction-time advantage for original and spliced over cross-spliced stimuli in the choice reaction-time study: If the information is coarticulatory, they do not integrate it with information for a segment with which it co-occurs in time. This leads to the findings in the 4IAX study.[4]

The model of Elman and McClelland would not achieve the segmentation it does if the acoustic signal did not support it. It has not been obvious that the signal does support this segmentation, however, because visible displays of the signal do not invite it; indeed, acoustic analysis guided by visible displays have not achieved it. (This is true not only of segmentations used in the phonetics literature as described in the introduction; it also appears to characterize segmentations described by naive subjects learning to read spectrograms based on a whole-word training procedure [Greene, Pisoni, & Carrell, 1984].) It will be important for future research to make explicit the relationship betweeen the listeners' and the model's segmentation of the acoustic speech signal on the one hand and the support for it that the signal provides on the other.

One step further back in the chain of communication, the acoustic speech signal could not reliably give rise to the segmentation it does without support from the talkers' articulations. That is, the gestures correspondin a given phonetic segment must, in some sense, cohere in articulation and those corresponding to different segments must, in the same sense, be separable.[5] This line of reasoning in turn suggests that Hockett's (1955) often-cited Easter-egg analogy is misleading. Hockett compared the effects of coarticulation on phonetic segments to a process of sending a row of Easter eggs through a wringer. His analogy reflected the view, still current (cf. MacNeilage & Ladefoged, 1976), that coarticulation destroys both the coherence of individual phonetic segments and their separation one from the other. Necessarily, then, the acoustic signal cannot be supposed to provide sufficient information, in itself, to support perception of the segments; rather, phonetic identifications must be interpretations imposed on the signal by a listener (cf. Studdert-Kennedy, in press).

The present findings and the behavior of Elman and McClelland's computer model render this perspective on articulation doubtful, however. In view of that, it is not surprising that research on articulation suggests a picture tidier than Hockett's analogy implies. For example, research by Ohman (1966), Carney and Moll (1971), Butcher and Weiher (1975), and Barry and Kuenzel (1976) agree in showing that vowel-to-vowel movements of the tongue-body occur before, throughout and after the production of an intervocalic consonart in a VCV production. Ohman's interpretation is that, in VCVs, consonantal gestures are superimposed on on-going diphthongal vowel-to-vowel gestures. In this type of utterance, then, coarticulation does not destroy the coherence of features of individual phonetic segments or the separation among distinct segments as the Easter-egg analogy implies. Indeed, rather than being an irrecoverable smearing of consonantal and vocalic gestures, in these utterances coarticulation is the overlapping occurrence of two distinct types of gestures—one for the vowel-to-vowel movements and one for the consonantal gestures.

This research on C-V coarticulation converges with other production research, in which segment durations are measured. In that literature, vowels are measured to shorten as consonants are added to a syllable, and in similar fashion, stressed vowels are measured to shorten as unstressed vowels are added to a word or stress foot (e.g., Fowler, 1977, 1981b; Lindblom & Rapp, 1973). Data in Fowler (1981b) suggest, however, that at least some of the measured shortening is not articulatory shortening in fact, but rather reflects the sort of articulatory overlap reported by Ohman and others (and illustrated in Figure 1). It is identified as shortening only because measurement conventions do not include, as part of a vowel's duration, those parts of its coarticulatory extent where another segment predominates in the signal. Together ~~with~~ the articulatory measures, the shortening measures further support the hypothesis that consonants and vowels (and stressed and unstressed vowels) are nondestructively overlapped in production in a way consistent with the perceptual segmentation of the acoustic signal that the present research and that of Fowler and Smith (in press) suggest.

The production research just described and our interpretation of the present findings both predict that the perceived duration of a phonetic segment should exceed its measured duration, corresponding instead approximately to its coarticulatory extent. A similar expectation can be derived from Liberman and Studdert-Kennedy's discussion (1978) of reasons why coarticulation may be necessary for perceivers.

In their view, talkers have to produce speech that meets two competing requirements. Because meanings of grammatical utterances have to be extracted from grammatically-coherent groups of words, and cannot be determined word-by-word, speech may have to be transmitted at a rapid rate. The listener has to be able to remember the beginning of a syntactic phrase at the time the end of it is produced. Second, however, the rate cannot exceed that at which listeners are no longer able to determine the order of sequences of sounds (Warren, 1976). Liberman and Studdert-Kennedy point out that coarticulation allows relatively long-duration segments to occupy relatively short intervals of time. However, this would only be a perceptual advantage to a listener who heard coarticulatory overlap as overlap rather than as context-sensitivity of discrete phonetic segments. In recent work, I have found some evidence that the perceived duration of a vowel does indeed exceed its measured duration (Fowler, 1983a).

Together, the research and theoretical considerations outlined here suggest a coherent perspective on the production and perception of speech (cf. Fowler, 1983a, 1983c). Talkers produce phonetic segments in overlapping time frames. The articulatory overlap, however, does not smear the segments; rather it preserves the coherence of the temporally extended parts of an individual phonetic segment and the separation of distinct segments. Compatibly, the acoustic signal provides information for the separation of overlapping segments and the coherence of temporally-extended parts of a segment. Finally, listeners segment the signal realistically, recovering the segments that talkers produce.

References

Abbs, J. H., & Gracco, V. L. (in press). Control of complex motor gestures: Orofacial muscle responses to load perturbations of the lip during speech. Journal of Neurophysiology.

Alfonso, P. J., & Baer, T. (1982). Dynamics of vowel articulation. Language and Speech, 25, 151-173.

Barry, W., & Kuenzel, H. (1975). Co-articulatory airflow characteristics of intervocalic voiceless plosives. Journal of Phonetics, 3, 263-282.

Bell-Berti, F., & Harris, K. S. (1976). Some aspects of coarticulation. Haskins Laboratories Status Report on Speech Research, SR-45/46, 197-204.

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

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

Cole, R., & Scott, B. (1974). The phantom in the phoneme: Invariant cues for stop consonants. *Perception & Psychophysics*, *15*, 101-107.

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.

Elman, J., & McClelland, J. (1983). *Speech perception as a cognitive process: The interactive activation model* (ICS Report No. 8302). San Diego: University of California, Institute of Cognitive Science.

Fowler, C. (1977). *Timing control in speech production*. Bloomington, IN: Indiana University Linguistics Club.

Fowler, C. A. (1981a). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research*, *46*, 127-139.

Fowler, C. A. (1981b). A relationship between coarticulation and compensatory shortening. *Phonetica*, *38*, 35-50.

Fowler, C. A. (1983a). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet. *Journal of Experimental Psychology: General*, *112*, 386-412.

Fowler, C. A. (1983b, May). *Perceiving the natural structure of coarticulated speech*. Paper presented at the 105th meeting of the Acoustical Society of America, Cincinnati, OH.

Fowler, C. A. (1983c). Realism and unrealism: A reply. *Journal of Phonetics*, *11*, 303-322.

Fowler, C. A., & Smith, M. (in press). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability of speech processes*. Hillsdale, NJ: Erlbaum.

Fromkin, V. A. (Ed.). (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

Gelb, I. (1952). *A study of writing*. Chicago: University of Chicago Press.

Greene, B., Pisoni, D., & Carrell, T. (1984). Recognition of speech spectrograms. *Journal of the Acoustical Society of America*, *76*, 32-43.

Hockett, C. (1955). *Manual of phonology*. Bloomington, IN: Indiana University Publications in Anthropology and Linguistics, No. 11.

Hockett, C. (1960). The origin of language. *Scientific American*, *203*, 89-96.

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

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, *3*, 129-140.

Lehiste, I., & Shockey, L. (1972). On the perception of coarticulatory effects in VCV syllables. *Journal of Speech and Hearing Research*, *15*, 500-506.

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., & 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, 143-178.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. *Papers from the Institute of Linguistics* (University of Stockholm), *21*, 1-59.

Lisker, L. (1974). On time and timing in speech. In T. A. Sebeok (Ed.), *Current trends in linguistics* (Vol. 12, Part 10, pp. 2387-2418). The Hague: Mouton.

MacNeilage, P., & Ladefoged, P. (1976). The production of speech and language. In E. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 7): *Language and perception*. New York: Academic Press.

Mann, V. A. (1980). Influence of preceding liquid on stop consonant perception. *Perception & Psychophysics*, *28*, 407-412.

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

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

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. (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.

McClelland, J., & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception, Part 1. *Psychological Review*, *88*, 375-407.

McNeill, D. A., & Lindig, K. (1973). The perceptual reality of phonemes, syllables, words and sentences. *Journal of Verbal Learning and Verbal Behavior*, *12*, 419-430.

Meltzer, R., Martin, J., Mills, C., Imhoff, D., & Zohar, D. (1976). Reaction time to temporally-displaced phoneme targets in continuous speech. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 277-290.

Ochiai, K., & Fujimura, O. (1971). Vowel identification and phonetic contexts, *Reports from the University of Electrocommunications* (Tokyo), *22-2*, 103-111.

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

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, *32*, 693-703.

Pisoni, D. (1971). On the nature of categorical perception of speech sounds. *Supplement to the Haskins Laboratories Status Reports on Speech Research*, *SR-27*.

Pisoni, D. (in press). Acoustic-phonetic representations in word-recognition. *Cognition*.

Samuel, A. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*, 474-494.

Savin, H. B., & Bever, T. G. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, *9*, 295-302.

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

Studdert-Kennedy, M. (in press). Sources of variability in early speech development. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability of speech processes*. Hillsdale, NJ: Erlbaum.

Walley, A. C., & Carrell, T. D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. Journal of the Acoustical Society of America, 73, 1011-1022.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. Science, 167, 392-393.

Warren, R. M. (1976). Auditory sequence and classification. In N. Lass (Ed.), Contemporary issues in experimental phonetics. New York: Academic Press.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. Perception & Psychophysics, 35, 49-64.

Wickelgren, W. A. (1969). Auditory or articulatory coding in verbal short-term memory. Psychological Review, 76, 232-235.

Wickelgren, W. A. (1976). Phonetic coding and serial order. In E. Carterette & M. Friedman (Eds.), Handbook of perception: Language and speech (Vol. 7). New York: Academic Press.

Footnotes

[1] Pisoni (in press) marshalls evidence from a variety of sources (from linguistics: synchronic and diachronic phonological regularities, systematic alternations among morphological relatives in the lexicons of languages; from psychology: language games, alphabetic writing systems and speech errors) converging on conclusions that phonetic structure is psychologically real and that it plays an important role in language use.

[2] We will use the word "compensation" to label findings that contextual influences are not integrated with information for a segment with which they co-occur in time. We do not intend to imply any active process of compensation by perceivers, however. Our hypothesis (see also, Elman & McClelland, 1983) is that compensation is a by-product of segmentation, which itself is a necessary consequence of phone identification.

[3] Subjects in Group 3 performed better than those in Group 2 on all three tests. Subjects in the two groups were recruited from the same type of population (an Introductory Psychology class) and received the same instructions. The relevant difference between them, I think, is that subjects in Group 2 were recruited at the end of one academic term and those in Group 3 at the beginning of the next one. Individuals who look for extra credit at term's onset may be more highly motivated overall than individuals who seek it at term's end.

[4] There is a possible difference in the view of segmentation depicted in Figure 1c and that achieved by the model of Elman and McClelland. In the model, segmentation is a by-product of procedures for identifying the component phonetic segments of an utterance. The strategy of Figure 1c, could be, but is not necessarily or explicitly a strategy for identifying the phones; it is essentially a strategy for keeping separate the information for different phones that is provided in overlapping time windows. The findings of the present study appear to be fully compatible with the model of Elman and McClelland, however.

[5] Research by Abbs and Gracco (in press) and by Kelso, Tuller, V.-Bateson, and Fowler (1984) provides preliminary evidence for this. In the research of Kelso et al, a subject's jaw is perturbed as it closes for the final consonant of /bæb/ or /bæz/. If the utterance is /bæb/, compensatory movement by the

upper lip achieves the lip closure necessary for the bilabial segment. No reactive activity is found in the tongue, which is not involved in /b/ production. Thus, two articulators involved in the production of an individual segment are found to be coupled; an articulator involved in production of other segments is not. A different outcome is observed when the final consonant is alveolar /z/. There, the tongue does compensate for jaw braking during closure; although excitatory lip activity is observed in this case, no lip movements occur. Again, two articulators involved in the production of a single segment are functionally coupled in articulation. They are not coupled to articulators required to produce different segments.

INTRINSIC TIME IN SPEECH PRODUCTION: THEORY, METHODOLOGY, AND PRELIMINARY OBSERVATIONS*

J. A. S. Kelso† and Betty Tuller††

Abstract. A continuing challenge to our understanding of speech production and perception is the fact that utterances with markedly different acoustic, kinematic, and electromyographic characteristics can nevertheless be perceived as the "same" word. In this paper, we discuss the importance of examining articulation relative to an intrinsic, activity-defined metric and show how such an analysis of intervocalic consonant timing across different speaking rates and stress patterns significantly reduces both interspeaker and intraspeaker variability. Next we explore whether the observed relative temporal stability can be achieved without reference to an extrinsic clocking device, but rather in terms of the dynamic topology of the system's behavior. To this end, using a phase plane description of articulatory motion, we show how the temporal analysis originally offered can be redescribed in terms of critical position-velocity states (or, in polar coordinates, phase angles) for interarticulator cooperation. Such coordination, we propose, can be captured in terms of events that are intrinsic to the system's dynamics, not in terms of conventional durational metrics.

1.0 Introduction

One primary focus of speech production research has been to understand how the many articulatory degrees of freedom are temporally organized. In traditional (and many current) theories, the problem is "explained" by invoking the notion of a program, a representation of the behavior coded in some mental or neural device that exists before the behavior is realized in the real world (see Kelso, 1981; Kelso, Tuller, & Harris, 1983; Kugler, Kelso, & Turvey, 1980, for criticisms). The function of the temporal program is to instruct the articulators when to become active and for how long. For

---

*To appear in E. Keller & M. Gopnik (Eds.), Sensory processes and language. Hillsdale, NJ: Erlbaum.

†Also University of Connecticut.

††Also Cornell University Medical College.

Acknowledgment. The research referred to here was supported by National Institutes of Neurological Communicative Diseases and Strokes Grants NS-17778 and RR-05396 to Cornell University Medical College, NS-13617 to Haskins Laboratories, and by a grant from the Stuttering Center, Department of Neurology, Baylor College of Medicine/Ariel and Benjamin Lowin Medical Research Foundation. The first author was also supported by U. S. Office of Naval Research Contract No. N0014-83-K-00. Elliot Saltzman gave us extremely helpful comments on this paper, for which we are grateful.

example, in the early, influential theories proposed by Kozhevnikov and Chistovich (1966) and Lindblom (1963), execution of discrete linguistic units was thought to be triggered by an independent "rhythm generator" or "timing program," which timed the units with respect to one another. More recently, so-called central pattern generators have been thought to be the neural embodiment of timing programs.

This class of theory can be termed indicational (cf. Reed, 1981). That is, the role of the plan is to indicate, instruct, or command the articulators how and when they should be active. The emphasis of indicational theories is placed firmly on the symbolic mode of description with little or no attention paid to the detailed dynamical processes that the symbol ode is said to indicate or direct. To use a favorite example (cf. Pattee, 1977), a stop sign indicates to a driver that the car should be stopped, but provides no detailed information about how to stop the car, i.e., how, where, and by how much to decelerate, apply the brakes, etc. Thus, the symbolic or indicational mode greatly underdetermines the information actually required to perform an activity. In the case of speech, indicational theories pay no regard to the dynamical behavior of the articulatory system, i.e., the ordered motions of the articulators in space and time. The timing program or rhythm generator concept emphasizes the symbolic, indicational mode and provides no account of how the multiple degrees of freedom of the articulatory system are actually coordinated in the course of an activity.

Indicational theories (which are pervasive in biology and psychology) not only ignore, in large part, dynamical processes but also lack a rationale for how it is and by what means one particular symbol string is created rather than another (cf. Kugler, Kelso, & Turvey, 1980; 1982; Turvey & Kugler, 1984). In speech research, different variants of an indicational theory include features, segments, or syllables as candidates for units of articulation. What is missing, then, is an account of the symbolic mode that is not arbitrary with respect to the dynamics that it instructs. The origins of the symbolic mode must, it seems, be lawfully derived from dynamics.[1]

The foregoing arguments serve to focus attention on dynamics as a source of understanding natural activities, such as speaking, whose spatiotemporal organization is the main concern of this paper. Dynamics, by definition, means simply motion and change in space/time. Maxwell (1877) described dynamics as the "simplest and most abstract description of the motion of a system." Thus, and this is important, there is no logical reason why dynamics, although rate-dependent, cannot be conceived of as abstract. One reason why dynamics has been undervalued is that it has been interpreted as local and concrete (purely biomechanics?) rather than global and abstract. Yet recent developments in the field of dynamical systems indicate otherwise: Systems possessing huge numbers of dimensions can be abstractly characterized in a low-dimensional space (cf. Abraham & Shaw, 1982; Haken, 1983). Mor ver, it is certainly possible, in principle, to characterize the global behavior of a dynamical system using symbolic representations (e.g., Crutchfield & Packard's, 1983, "symbolic dynamics"). Here again, however, the question of the non-arbitrary or privileged coupling between a symbolic representation of the dynamics and the dynamics itself remains.

In this paper, we shall try to accomplish three goals, all of which relate to a dynamic perspective on speech production with particular emphasis on its temporal aspects (see also Kelso & Tuller, 1984a, 1984b). First, we will review some of our recent research, which reveals that the relative timing among articulatory events is a significant index of the stable performance of the speech motor system across suprasegmental changes in stress and speaking rate. This ubiquity of relative timing, not only in speech production but other activities as well, raises a number of issues about the role of time in biological systems (e.g., conventional questions about how the system meters and monitors time, how duration is controlled, etc.). The whole question of time has been a subject of fascination since the dawn of modern thought. Aristotle is said to have associated time with motion, yet also advocated "a soul which counts" (cf. Prigogine, 1984). In the second main part of the paper, we will argue that the soul that counts (today's temporal program?) can be usefully replaced by a view of time that is intrinsic to a sequence of events, and whose units are defined entirely in terms of the state variables of the system.

Fowler (1977; see also Fowler, Rubin, Remez, & Turvey, 1980) has proposed an intrinsic timing account of speech production that, though appealing, has lacked an empirical methodology for its detailed evaluation. In the third part of the paper, we will ground the notion of intrinsic time empirically by recasting our relative timing data into a phase plane description of articulator trajectories, in which time itself does not appear explicitly. In our final remarks we will address the advantages--both theoretical and experimental--of this methodology in which intrinsic time is revealed by the geometry of the system's dynamical behavior. This dynamical perspective stresses different observables and motivates a simpler and more elegant account of relative timing in speech productio

2.0 Relative Timing of Articulatory Gestures

Our basic intuition is that it is probably incorrect to assume, as conventional accounts do, that speech timing is based on standard temporal units, such as milliseconds. This intuition is strengthened by well-known facts: namely, that the absolute duration of individual electromyographic, kinematic, and acoustic speech events can change dramatically as a function of speaking rate, syllable stress, and phonetic context, among other things, yet the perceptual identity of constituents is preserved (e.g. Fry, 1955; 1958; Lehiste, 1970; Lindblom, 1963). This suggests that time might usefully be measured in relative, rather than absolute, temporal units that are, in a sense, "normalized" to the activity being performed.

The notion that the relative timing of articulatory events provides a more appropriate metric than their absolute durations is mirrored by the importance of relative acoustic durations for perception of speech. For example, the duration of the interval between release of supraglottal occlusion and the onset of glottal pulsing, the so-called voice onset time, is a strong cue to the voicing category of a stop consonant (Lisker & Abramson, 1964). However, perception of the voicing category does not switch when some absolute interval duration of voice-onset-time is reached. Rather, the category boundary is perceived relative to the overall speech rate (Summerfield, 1975; see also Port, 1979). Similarly, the duration of formant transitions is a strong distinguishing cue between perception of /b/ and /w/ in word initial position (Liberman, Delattre, Gerstman, & Cooper, 1956). But again, the absolute duration of these transitions is less important than the transition duration relative to syllable duration (Miller & Liberman, 1979).

Other evidence that time is relative in motor control comes from analyses of nonspeech activities in which the timing of individual events appears to be constrained relative to a longer, activity-defined period. This relative temporal stability of muscle activities, or kinematic events, is apparent over scalar changes in rate or force of production that often result in large variations in absolute duration. Although early demonstrations of relative temporal stability were provided from activities that are qualitatively repetiti and potentially pre-wired (e.g., locomotion, respiration, and mastication; see Grillner, 1977, for review), more recent work has revealed that less repetitive activities show the same organizational features (e.g., two-handed movements, typing, handwriting, postural control, and speech-manual coordination; Hollerbach, 1981; Kelso, Southard, & Goodman, 1979a, 1979b; Kelso, Tuller, & Harris, 1983; Lestienne, 1979; Nashner, 1977; Schmidt, 1982; Shapiro, Zernicke, Gregor, & Diestel, 1981; Viviani & Terzuolo, 1980).

In recent papers we have presented data suggesting that the production of speech can be described by a similar style of organization (Harris, Tuller, & Kelso, in press; Tuller & Kelso, 1984; Tuller, Kelso, & Harris, 1982; 1983). This demonstration should be of some interest to neuroscience and neuropathology, because it supports the idea that a single set of organizational principles may underlie very different motor skills, including one as highly symbolic as human speech (e.g., see contributions by Grillner, Evarts, and Granit in Grillner, Lindblom, Lubker, & Persson, 1982; Ostry & Cooke, this volume; Ostry, Keller, & Parush, 1983). The data are also interesting, we believe, because they highlight the importance of a commensurate vocabulary for the symbolic description of an activity, and the activity itself, and offer an activity-sensitive metric for measuring timing in speech production.

In our work on speech production, we varied two suprasegmental aspects of speech that are believed to be particularly important--syllable stress and speaking rate--in an effort to discover underlying articulatory invariance across speech segments. We approached this problem by examining kinematic and acoustic recordings of speakers' productions of two-syllable nonsense utterances embedded in the carrier phrase "It's a ____ again." The utterances under discussion here were /ba#Cab/, and /bæ#Cab/, where the medial C was from the set /b, p, w/. Half of the tokens were produced with primary stress placed on the first syllable and half were produced with primary stress on the second syllable. Twelve repetitions of each utterance were produced at each of the two stress patterns and at two self-selected speaking rates, conversational and somewhat faster.

Kinematics of articulatory movements were monitored in the up-down direction using an optical tracking system that followed the movement of lightweight, infrared, light-emitting diodes (LEDs) attached to the subject's lips, jaw, and nose. In order to minimize head movements during the experiment, output of the LED on the nose was displayed on an oscilloscope placed directly in front of the subject, who was told to keep the display on the zero line. The movement records were recorded on multichannel FM tape for later computer analysis. For each token, displacement maxima and minima (corrected for head movements) and the times at which they occur, were obtained individually for the jaw, the upper lip, and the lower lip.

The spatiotemporal coordination among articulator events was analyzed to determine whether stable relative timing of kinematic events is characteristic of speech. This analysis requires demarcation of some period of articulatory activity and the latency of an articulatory event within the defined period. Although we examined relative timing of nine different articulatory events, linguistic evidence suggests that temporal stability among phonetic segments will be defined relative to the period between successive vowels. For example, when short silent intervals are inserted into a sentence, listeners notice the insertions only when they disrupt the relative timing of the stressed vowels (Huggins, 1972). Here we will discuss only the timing of consonant production relative to the interval between flanking vowels, by far the most stable of our measures. Over linguistic variations, in this case stress and rate, these intervals change in their absolute durations. The question is whether they change in a systematically related manner.

Figure 1, taken from Tuller and Kelso (1984), shows one subject's data for production of the utterances /babab/, /bapab/, and /bawab/. The data were similar for all four subjects. The x-axis represents the interval (in ms) from the onset of jaw lowering for the first vowel to the onset of jaw lowering for the second vowel. The y-axis is the interval from the onset of jaw lowering for the first vowel to the onset of upper lip lowering for the medial labial consonant. In this figure, the jaw component has been subtracted from the lower lip movement. Each point on the graph is one token of an utterance type. A Pearson product-moment correlation was calculated for each distribution. The high correlations obtained (.97, .97, and .89) signify that the relative timing of these articulatory events was maintained over the large variation apparent in their absolute durations. For utterances with /æ/ as the first vowel, we find the same results as for the utterances displayed in Figure 1; the temporal changes are highly correlated (r = .90, .89, and .84, for medial consonants b, p, and w, respectively).

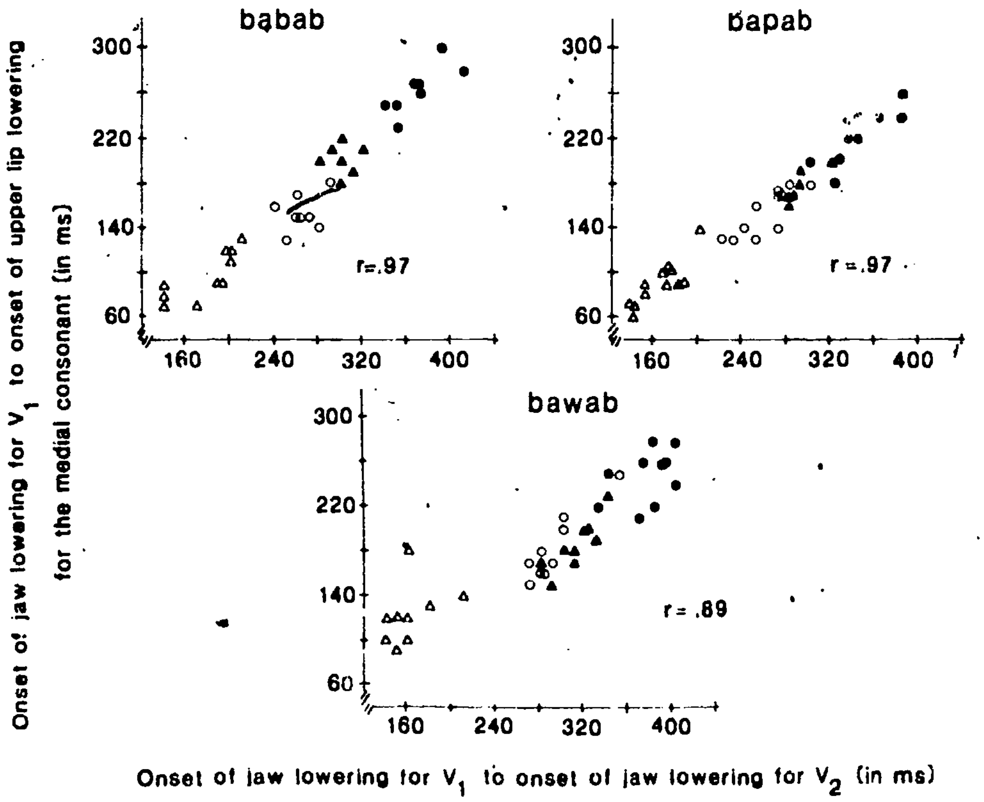


Figure 1. Timing of lower lip raising for medial consonant articulation as a function of the vowel-to-vowel period for one subject's productions of the indicated utterances. Each point represents a single token of the utterance. (●) primary stress on the first syllable, spoken at a conversional rate; (○) primary stress on the second syllable, spoken at a conversational rate; (▲) primary stress on the first syllable, spoken at a faster rate; (△) primary stress on the second syllable, spoken at a faster rate. [From Tuller & Kelso, 1984]

Although Figure 1 illustrates the data from only a single subject, the three other subjects showed essentially the same pattern. Values obtained for the other subjects by correlating the period between the onsets of successive vowel articulation ranged from .84 to .97 whether the onset of consonant articulation was defined by the raising gesture of the lower lip or by the lowering gesture of the upper lip. In all cases, the calculated correlation was significantly higher ($ps < .05$) than what would be expected solely on the basis of changes in vowel duration or part-whole correlations (cf. Barry, 1983; Munhall, submitted; Tuller & Kelso, 1984; Tuller et al., 1983). Note, however, that the relationship is not ratiomorphic--lip latency varies systematically with jaw cycle deviation, plus some intercept value that changes with speaker and phonetic context. This basic description has recently been replicated and extended to tongue gestures produced by English speakers (Harris et al., in press; Munhall, submitted), and lip and jaw gestures produced by speakers of French and Swedish (Gentil, Harris, Horiguchi, & Honda, 1984; Lubker, 1983; see also Linville, 1982). Let us underscore again that this strong relationship holds even though other aspects of the movements, such as their displacement, velocity, and absolute duration, change substantially. For example (in agreement with the acoustic/phonetic literature), the production of destressed syllables shows smaller displacement, lower velocity, and shorter duration movements than the same phonetic segment spoken with primary stress (e.g., Harris, 1971, 1978; Kent & Netsell, 1971; Lindblom, 1963; MacNeilage, Hanson, & Krones, 1970; Mermelstein, 1973; Stone, 1981; Sussman & MacNeilage, 1978; Tuller, Harris, & Kelso, 1982).

These relative timing results indicate, we believe, a functional constraint on movement--a coordinative structure (cf. Easton, 1972; Fowler, 1977; Kelso et al., 1979a, 1979b; Turvey, 1977) or unit of action (cf. Ghiselin, 1981)--in which a system possessing a large number of potential degrees of freedom is compressed into one that requires few control decisions (Bernstein, 1967). During a movement, the timing of individual elements is constrained within a particular relationship. Flexibility can then be attained by adjusting control parameters over the total unit.

3.0 Intrinsic Versus Conventional Metrics for the Analysis of Timing

The ubiquity of stable relative timing in so many different types of activities, including speech production, raises a number of fundamental questions about the underlying basis of timing regularity in biological systems. For example, is the duration of each articulatory movement controlled directly via an extrinsically-imposed timing program, or is the duration of an articulator's movement a consequence of some other parameter, as yet unspecified? If not controlled by an extrinsic clock, what is the informational basis for the observed temporal stability? How, in a complex system of articulators, does a given articulator "know" when it should be activated in relation to other active articulators? With respect to our relative timing data, for example, what information is needed for the upper lip (a remote, non-mechanically linked articulator) to move in appropriate temporal relation to the jaw? Although an intrinsic timing theory of speech production has been proposed (Fowler, 1980), a generalized methodology for evaluating "intrinsic time" has yet to be offered. Before proposing such a methodology and applying it to experimental data, let us first clarify the basic notion of intrinsic time (cf. Richardson & Rosen, 1979).

Although it is convenient to measure time intervals between events as durations, this is purely a convention since, it can be argued and shown, time itself has no unique dimension. A few examples will clarify what we mean. Consider a primitive clock, such as a candle. Time in this case corresponds to a change in a spatial variable, namely, the length of the candle that is burned. The units of time intrinsic to this particular dynamical system are inches. Similarly, in a water clock, the unit of time corresponds to number of drops. We see very quickly from these simple but intuitive examples that time itself is not, strictly speaking, a fundamental observable; rather, it is intrinsically determined by the particular system involved. Thus, paraphrasing Richardson and Rosen (1979), time is demarcated by defining some state variable appearing in the events; it is intrinsic to that sequence, that is, to a dynamical process, and takes its units from the system's state variables. This kind of intrinsic time is quite different from conventional time, which is imposed on a system regardless of its particular dynamics. Conventional or mechanical time (measured in seconds, hours, etc.) plays a role when it is necessary to determine the relationship between the intrinsic times of two devices whose dynamic parameters are very different. In such cases, a standard is introduced. For example, 1/86,400 of the earth's rotation is, by convention, called a second. Again, according to convention, all harmonic oscillators are calibrated in terms of seconds, not in terms of an event that is intrinsic to all harmonic oscillators, namely, the cycle.

We want to stress that calibration and convention, though important, are not the central issues of concern here. Our aim, rather, is to define an intrinsic metric in terms of the dynamical behavior of a biological system, in this case that of speech production. The examples of the candle and the clock indicate how a given system can generate its own intrinsic time entirely according to its constitutive parameters. We recently came across similar sentiments expressed in a rather different context, namely irreversible thermodynamics and nonequilibrium physics. There, Prigogine (1984) uses the concept of internal time. In Prigogine's words:

> "to grasp the intuitive meaning of internal time, think about a drop of ink in a glass of water. The form the drop takes gives us an idea of the interval of time that has elapsed." (p.6; italics ours).

And later:

> "The internal time T is quite different from the usual mechanical time, since it depends on the global topology of the system" (p. 7)

The drop of ink, the candle, and the water clock do not possess any knowledge of their own dynamics, nor do these systems contain an explicit representation of time. Time evolves from the "playing out" of the dynamics, but there is no programming, time control, or time representation anywhere (see also Kelso & Holt, 1980; Kelso, Tuller, & Harris, in press; Kelso, V.-Bateson, Saltzman, & Kay, 1985). These terms are simply not descriptions appropriate to the concept of intrinsic time.

4.0 Grounding Intrinsic Time in the Geometry of the Speech System's Dynamics

Let us now follow through with the claim that the notion of intrinsic time is open to measurement and evaluation. This requires that we characterize articulatory motion in terms of certain state variables (see Section 3.0 above). To do this, we introduce a phase portrait methodology developed originally by Poincaré and Liapounov for many-body problems in dynamics, and employed by us and others for analyses of movements (e.g., Kelso et al., 1985). The phase portrait captures the forms of motion (a purely geometric/kinematic description) produced by an underlying dynamic organization (cf. Abraham & Shaw, 1982). But first let us consider the standard representation of our speech timing data (see Section 2.0 above).

Consider the simple case we have described in which the latency (in ms) of onset of lower lip motion for a medial consonant is measured relative to the interval (in ms) between onsets of jaw motion for flanking vowels. As we have shown, the two events are highly correlated across rate and stress in different speakers. Although this strictly temporal description has been useful, it does not necessarily imply that the speech motor control system is keeping track of the duration of articulatory motions. In contrast, preliminary work suggests that the data can be understood without recourse to an extrinsic timer or timing metric (more extensive analyses are currently underway). Here we describe in general terms how, using a phase portrait description of articulatory trajectories on the phase plane, a very different view of articulatory "timing" emerges (see also Kelso et al., in press).

The phase plane is the space of all possible states of the system, in the plane whose axes are the articulator's position (x) and its velocity ($\dot{x}$). The position and velocity values act as coordinates of a point on the articulator in two dimensional space. As time varies, the point P ($x,\dot{x}$) describing the motion of the articulator moves along a certain path on the phase plane.

Figure 2 illustrates the mapping from time domain to phase plane trajectories. H[illegible]othetical jaw and upper lip trajectories (position as a function of time) are shown for an unstressed /bab/ (Figure 2a, left) and a stressed /bab/ (Figure 2b, left). On the right are shown the corresponding phase plane trajectories. In this figure and those following we have reversed the typical orientation of the phase plane so that position is shown on the vertical axis and velocity on the horizontal axis. Thus, downward movements of the jaw are displayed as downward movements of the phase path. The vertical crosshair indicates zero velocity and the horizontal crosshair indicates zero position (midway between minimum and maximum displacement). As the jaw moves from its highest to its lowest point (from A to C in Figure 2), velocity increases (negatively) to a local maximum (B) then decreases to zero when the jaw changes direction of movement (C). Similarly, as the jaw is raised from the low vowel /a/ into the following consonant constriction, velocity peaks approximately midway through the gesture (D) then returns to zero (A).

Note that time, although implicit and usually recoverable from the phase plane description, does not appear explicitly. For different initial conditions (such as starting position or the level of articulator stiffness) there will be different corresponding paths, and the totality of all possible trajectories constitutes the full phase portrait of the system's dynamic behavior. It is useful to transform the Cartesian $x,\dot{x}$ coordinates into equivalent polar coordinates, namely, a phase angle, $\phi = \tan^{-1}[\dot{x}/x]$, and a radial amplitude, $R = [x^2 + \dot{x}2]^{\frac{1}{2}}$. These polar coordinates are indicated on the phase planes shown in Figure 2. The phase angle will be a key concept in our re-analysis of interarticulator timing because it signifies position on a cycle of states (cf. Abraham & Shaw, 1982; Garfinkel, 1983).

The phase plane trajectory preserves some important differences between stressed and unstressed syllables. For example, maximum lowering of the jaw for the stressed vowel is greater than lowering for the unstressed vowel and

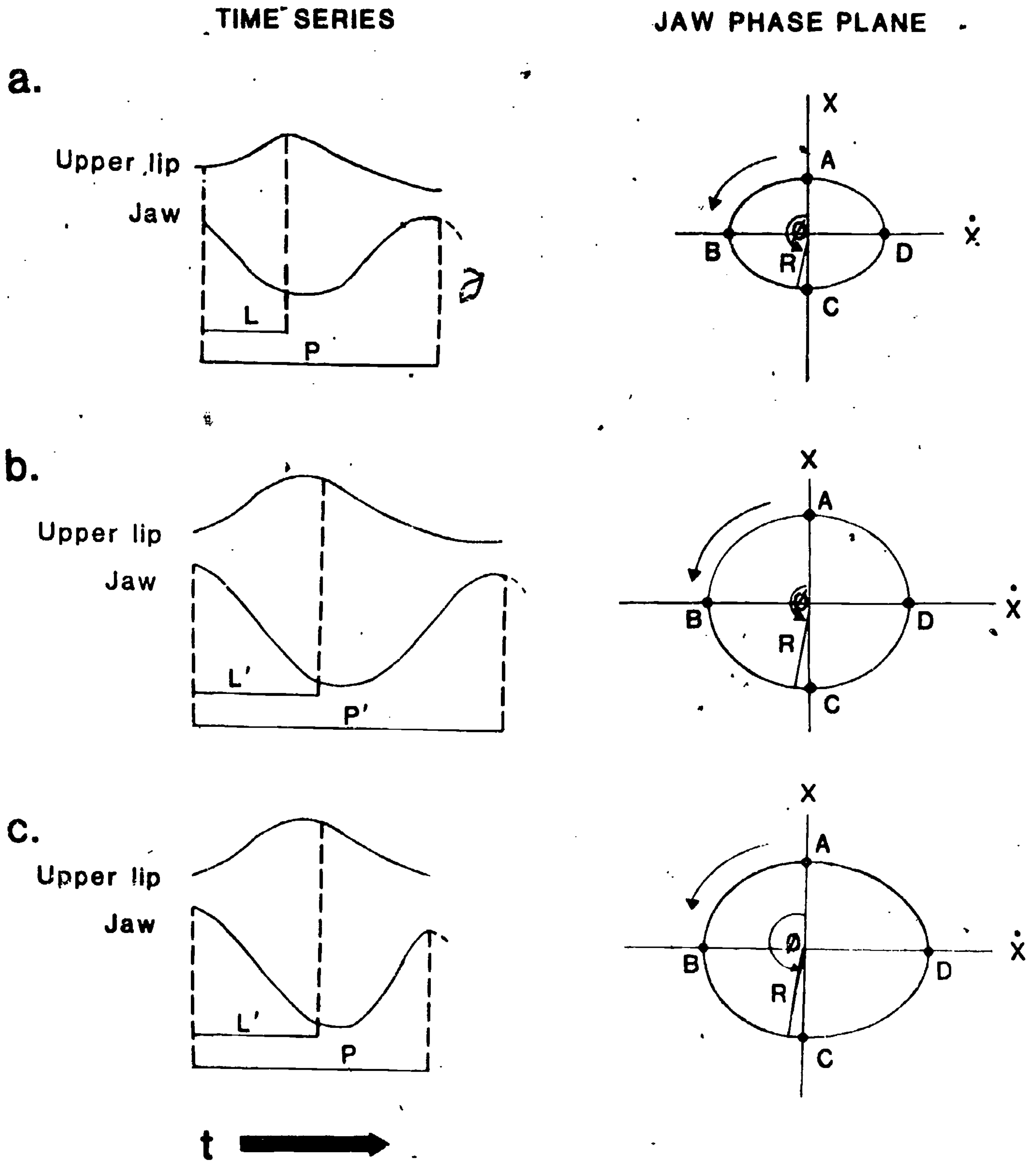


Figure 2. Left: Time series representations of idealized utterances. Right: Corresponding jaw motions, characterized as a simple mass spring and displayed on the 'functional' phase plane (i.e., position on the vertical axis and velocity on the horizontal axis). Parts a, b, and c, represent three tokens with vowel-to-vowel periods (P and P') and consonant latencies (L and L') that are not linearly related. Phase position of upper lip movement onset relative to the jaw cycle is indicated (see text).

maximum articulator velocity differs noticeably between these two orbits (e.g., Kelso et al., 1985; MacNeilage et al., 1970; Stone, 1981; Tuller, Harris, & Kelso, 1982). In contrast, the different durations taken to traverse the orbit as a function of stress are not represented in this description. To reiterate, a crucial point about this description is that duration does not appear explicitly. Jaw cycles of different durations are still characterized as a single orbit on the plane, i.e., they are topologically the same.

Now we can rephrase the question of how the lip "knows" when to begin its movement for the medial consonant by asking where on the cycle of jaw phase angles the lip motion for medial consonant production begins. One possibility is that lip motion begins at the same phase angle of the jaw across different jaw motion orbits (i.e., across rate and stress). This outcome is not necessarily entailed, or predicted by, the relative timing results. For example, Figure 2a through 2c shows three utterances whose vowel-to-vowel periods and consonant latencies do not change in a linearly related fashion. Nevertheless, the phase angle at which upper lip motion begins relative to the cycle of jaw states is identical in the three cases. Thus, the information for "timing" of a remote articulator (e.g., the upper lip) may not be time itself, nor absolute position of another articulator (e.g., the jaw), but rather a relationship defined over the position-velocity state (or, in polar coordinates, the phase angle) of the other articulator. Although this conceptualization is intriguing, we want to re-emphasize that it constitutes an alternative description of the relative timing data set. For example, Figure 3

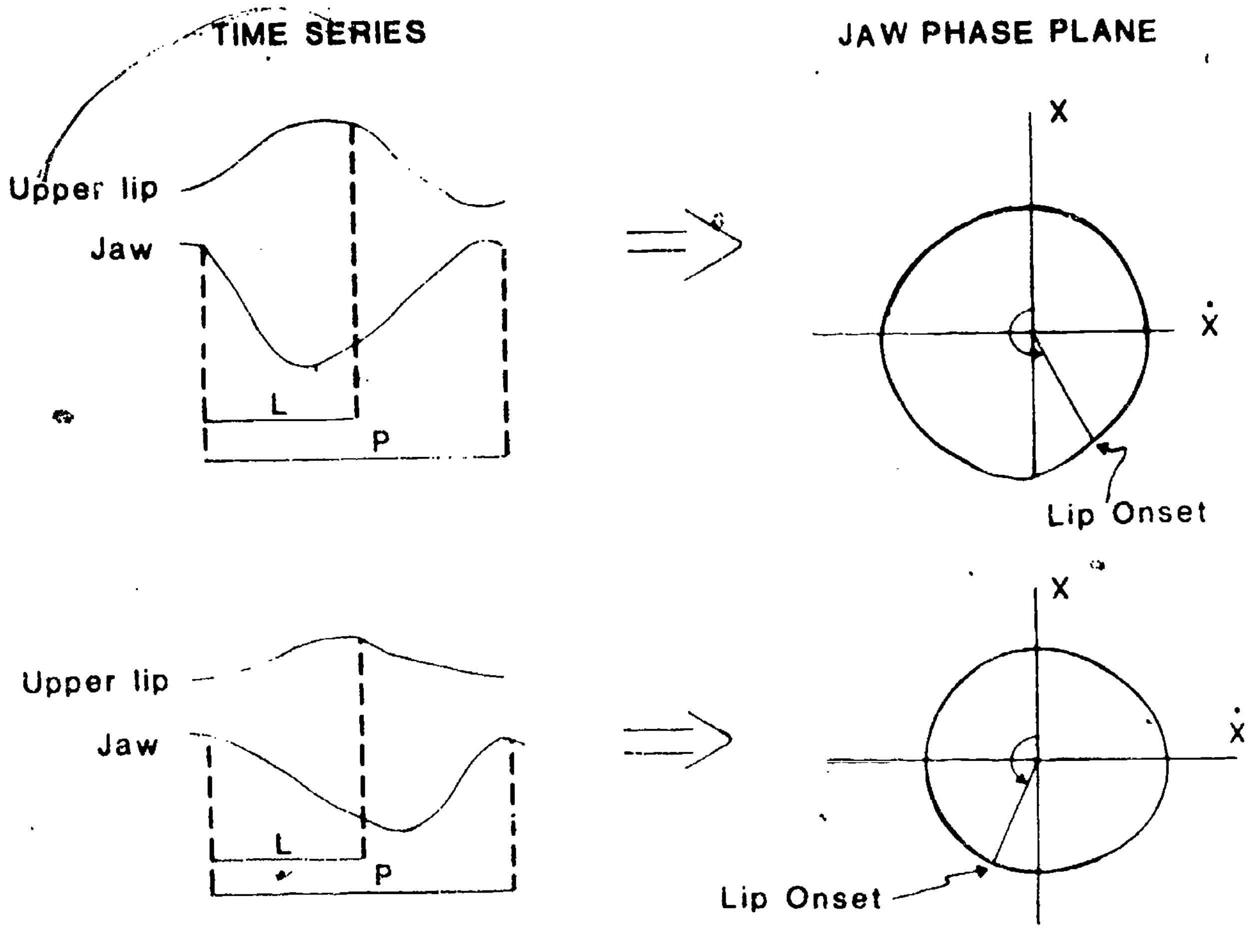


Figure 3. Two hypothetical utterances having identical vowel-to-vowel periods (P) and consonant (upper lip) latencies (L) but different phase angles of upper lip onset. [See caption Figure 2.]

illustrates the converse of Figure 2, namely, that two (hypothetical) utterances with identical vowel-to-vowel periods (P) and consonant latencies (L) can nonetheless show very different phase positions for upper lip movement onset. To be specific, the phase angle analysis incorporates the full trajectory of motion; the relative timing analysis is independent of trajectory once movement has begun and is based only on the onsets and offsets of events.

We also want to emphasize that the jaw motion need not be perfectly sinusoidal in order to apply a phase angle analysis. In fact, the motions actually observed are usually not sinusoidal; position at zero velocity is affected by the stress and rate characteristics of the surrounding vowels (see Figures 2 and 4). For this reason, each jaw cycle is normalized by determining the minimum and maximum jaw positions for the consonant-vowel gesture. The midpoint between them is used as the best approximation of the equilibrium or zero position for that syllable. Similarly, jaw velocity (from zero to peak lowering velocity) is normalized for each cycle.

Figure 4 shows jaw motion on the phase plane for the first syllable of /'ba#bab/ (top) and /ba#'bab/ (bottom) produced at a fast rate. Each token shown is the first instance produced of the utterance type. On the left is the entire jaw cycle for each stress pattern; on the right, the jaw cycle is reproduced only until the point of onset of upper lip movement downward for production of the medial bilabial consonant, as measured from the first deviation from zero velocity. The calculated phase position at which upper lip motion begins is indicated for each token. Notice that the jaw displacement and velocity are both greater for the stressed than the unstressed syllable. Nevertheless, upper lip motion begins at essentially the same phase angle for both tokens. If upper lip motion began at a phase angle of 180°, it would be synchronous with the jaw "turnaround" point.

Figure 5 shows the mean data and the standard error of the mean from the same speaker whose relative timing data were shown in Figure 1. The phase angle subtended, in degrees, is shown on the y-axis, stress-rate condition on the x-axis. A 2 X 2 ANOVA for each utterance type showed no significant main effects or interactions on the phase angle of upper lip onset for medial consonant production. For /babab/, Fs (1,27) = 2.50, 0.03, and 1.85; for /bapab/, Fs (1,30) = 2.39, 0.01, and 0.10; for /bawab/, Fs (1,29) = 0.06, 1.16, and 0.61. F-ratios are for rate, stress, and their interaction, respectively, ps > .5.

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

The phase conceptualization also has a number of theoretical advantages over our original relative timing analysis. First, once articulatory motions are represented geometrically on the phase plane, duration is normalized across speaker, stress, speaking rate, etc. Strictly speaking, the system's <u>topology</u> is unaffected by durational changes. For example, the "down-up" cy-

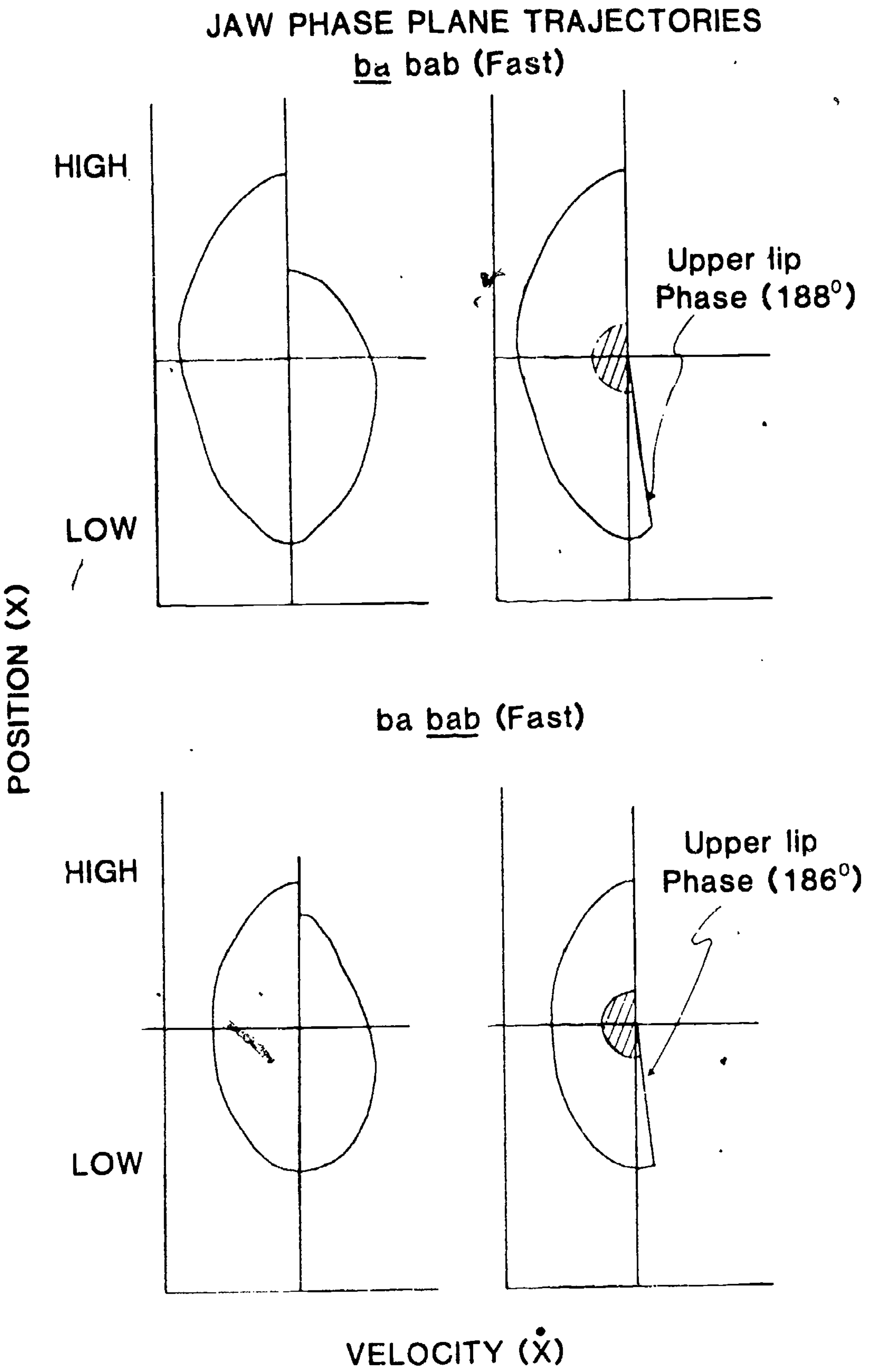


Figure 4. Left: Jaw cycle on the phase plane for the first token produced of stressed /bab/ (top) and unstressed /bab/ (bottom), spoken at a fast rate. Right: Jaw cycle until the onset of upper lip lowering for the second /b/.

39

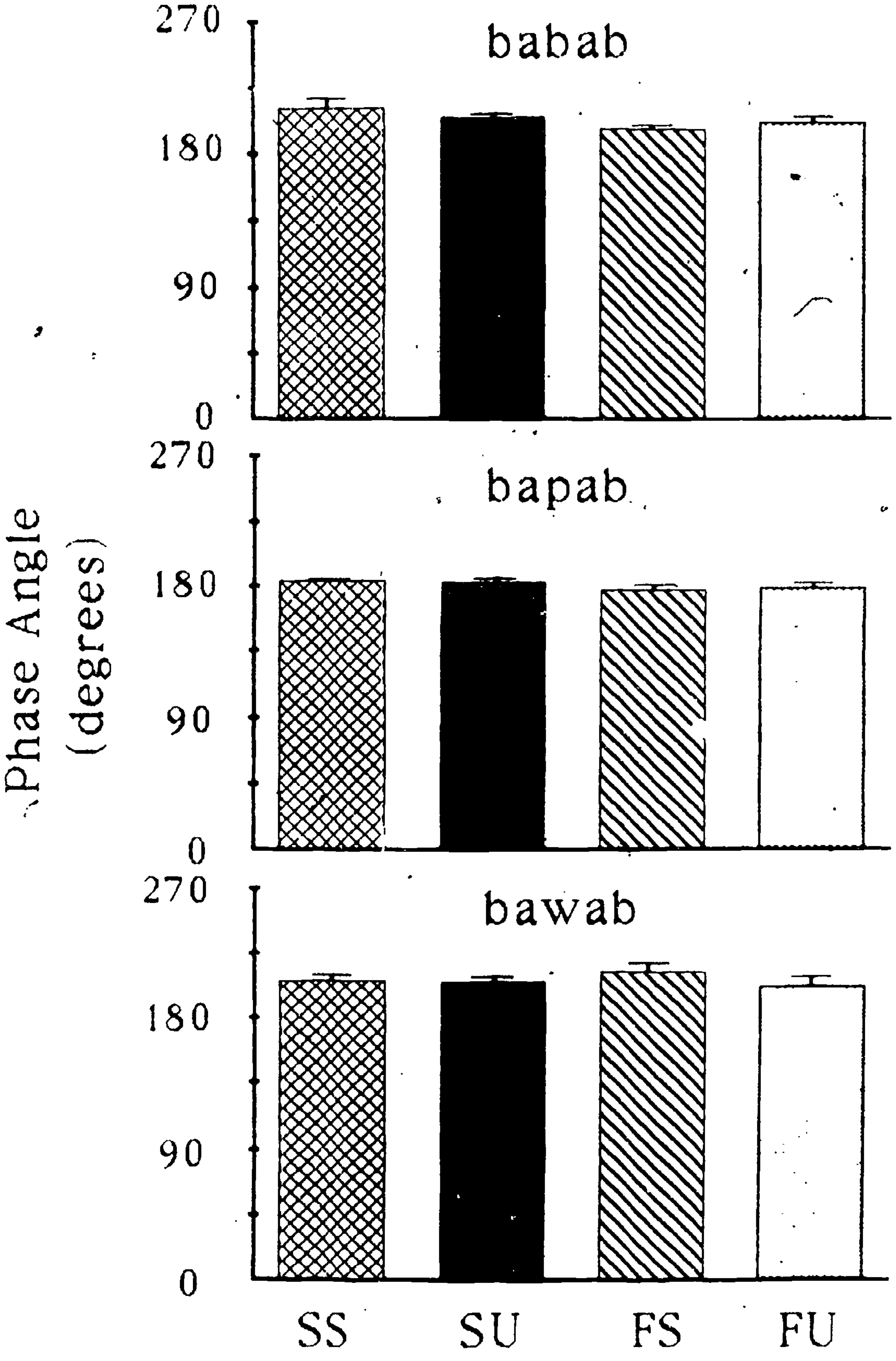


Figure 5. Mean phase angles of upper lip lowering relative to the jaw cycle for one subject's productions of /ba#Cab/. Standard error of the mean is also shown. SS: slow, first syllable stressed; SU: slow, first syllable unstressed; FS: fast, first syllable stressed: FU: fast, first syllable unstressed.

cle of the phase plane description is independent of the duration required for the gesture. This potentially provides a grounding for so-called intrinsic timing theories of speech production (e.g., Fowler, 1980; Fowler et al., 1980). Second, neither absolute nor relative durations have to be extrinsically monitored or controlled in this formulation. There is no need to posit any kind of time-keeping mechanism or time controller. As an aside, it has never been clear how the speech system could keep track of time, at least peripherally, because there is no known afferent basis (such as time receptors) for time-keeping in the articulatory structures themselves (Kelso, 1978). On the other hand, an informational basis (e.g., in position and velocity sensitivities of muscle spindle and joint structures) is a physiological given in the phase angle characterization. It might well be the case that certain critical phase angles provide information for coordination between articulators (beyond those considered here) and/or vocal tract configurations, just as phase angles of the leg joints provide coupling information for locomotory coordination (Shik & Orlovskii, 1965). Thus, the temporal orchestration of articulatory events in the speech motor system unfolds as a consequence of its dynamic parameters. By extension, it seems unlikely that the symbol structure for speech production includes a specification of durational rules defined in conventional mechanical time. As in a candle and a watch, time is not a possessed, programmed, or represented property of the speech production system.

References

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

Barry, W. J. (1983). Some problems of interarticulator phasing as an index of temporal regularity in speech. Journal of Experimental Psychology: Human Perception and Performance, 9, 826-828.

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

Crutchfield, J. P., & Packard, N. H. (1983). Symbolic dynamics of noisy chaos. Physica, 7D, 201-223.

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

Fowler, C. (1977). Timing control in speech production. Bloomington, IN: Indiana University Linguistics Club.

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

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.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. Journal of the Acoustical Society of America, 27, 765-768.

Fry, D. B. (1958). Experiments in the perception of stress. Language and Speech, 1, 126-152.

Garfinkel, A. (1983). A mathematics for physiology. American Journal of Physiology: Regulatory, Integrative, and Comparative, 14, R455-R466.

Gentil, M., Harris, K. S., Horiguchi, S., and Honda, K. (1984). Temporal organization of muscle activity in simple disyllables. Journal of the Acoustical Society of America, 75, S23. (Abstract)

Ghiselin, M. T. (1981). Categories, life, and thinking. The Behavioral and Brain Sciences, 4, 269-313.

Grillner, S. (1977). On the neural control of movement--A comparison of different basic rhythmic behaviors. In G. S. Stent (Ed.), Function and formation of neural systems (Life Sciences Research Reports, Vol. 6). Berlin: Dahlem.

Grillner, S., Lindblom, B., Lubker, J., & Persson, A. (Eds.) (1982). Speech motor control. Oxford: Pergamon Press.

Haken, H. (1983). Synergetics. Heidelberg: Springer-Verlag.

Harris, K. S. (1971). Vowel stress and articulatory reorganization. Haskins Laboratories Status Report on Speech Research, SR-28, 167-178.

Harris, K. S. (1978). Vowel duration change and its underlying physiological mechanisms. Language and Speech, 21, 354-361.

Harris, K. S., Tuller, B., & Kelso, J. A. S. (in press). Temporal invariance in the production of speech. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum.

Hollerbach, J. (1981). An oscillator theory of handwriting. Biological Cybernetics, 39, 139-156.

Huggins, A. W. F. (1972). On the perception of temporal phenomena in speech. Journal of the Acoustical Society of America, 51, 1279-1290.

Kelso, J. A. S. (1978). Joint receptors do not provide a satisfactory basis for motor timing and positioning. Psychological Review, 85, 474-481.

Kelso, J. A. S. (1981). Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance (IX). Hillsdale, NJ: Erlbaum.

Kelso, J. A. S., & Holt, K. G. (1980). Exploring a vibratory systems analysis of human movement production. Journal of Neurophysiology, 43, 1183-1196.

Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979a). On the nature of human interlimb coordination. Science, 203, 1029-1031.

Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979b). On the coordination of two-handed movements. Journal of Experimental Psychology: Human Perception and Performance, 5, 229-238.

Kelso, J. A. S., & Tuller, B. (1984a). Converging sources of evidence for common dynamical principles in speech and limb coordination. American Journal of Physiology: Regulatory, Integrative, and Comparative, 246, R928-R935.

Kelso, J. A. S., & Tuller, B. (1984b). A dynamical basis for action systems. In M. S. Gazzaniga (Ed.), Handbook of neuroscience. New York: Plenum.

Kelso, J. A. S., Tuller, B. H., & Harris, K. S. (1983). A "dynamic pattern" perspective on the control and coordination of movement. In P. MacNeilage (Ed.), The production of speech (pp. 138-173). New York: Springer-Verlag.

Kelso, J. A. S., Tuller, B. H., & Harris, K. S. (in press). A theoretical note on speech timing. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum.

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

Kent, R. D., & Netsell, R. (1971). Effects of stress contrasts on certain articulatory parameters. Phonetica, 24, 23-44.

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.)

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.

Lehiste, I. (1970). Suprasegmentals. Cambridge, MA: MIT Press.

Lestienne, F. (1979). Effects of inertial load and velocity on the braking process of voluntary limb movements. Experimental Brain Research, 35, 407-418.

Liberman, A. M., Delattre, P. C., Gerstman, L. C., & Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. Journal of Experimental Psychology, 52, 127-137.

Lindblom, B. (1963). Spectrographic study of vowel reduction. Journal of the Acoustical Society of America, 35, 1773-1781.

Lindblom, B. (1983a). Economy of speech gestures. In P. F. MacNeilage (Ed.), The production of speech (pp. 217-245). New York: Springer-Verlag.

Lindblom, B. (1983b). Phonetic universals in vowel systems. In J. Ohala (Ed.), Experimental phonology. New York: Academic Press.

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.), Universals workshop. The Hague: Mouton.

Linville, R. (1982). Temporal aspects of articulation: Some implications for speech motor control of stereotyped productions. Unpublished doctoral dissertation, University of Iowa.

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

Lubker, J. (1983, October). Comment on "Temporal invariance in the production of speech," by Harris, Tuller, and Kelso. Paper presented at Conference on invariance and variability in speech processes. MIT, Cambridge, MA.

MacNeilage, P. F., Hanson, R., & Krones, R. (1970). Control of the jaw in relation to stress in English. Journal of the Acoustical Society of America, 48, 120. (Abstract)

Maxwell (1877). Matter and motion. (Reprinted Dover Books, 1952).

Mermelstein, P. (1973). Some articulatory manifestations of vowel stress. Journal of the Acoustical Society of America, 54, 538 (L).

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. Perception & Psychophysics, 25, 457-465.

Munhall, K. (submitted). Relative timing of tongue dorsum movements. Journal of the Acoustical Society of America.

Nashner, L. M. (1977). Fixed patterns of rapid postural responses among leg muscles during stance. Experimental Brain Research, 30, 13-24.

Ostry, D. J. & Cooke, D. This volume

Ostry, D. J., Keller, E., & Parush, A. (1983). Similarities in the control of the speech articulators and the limbs: Kinematics of tongue dorsum movement in speech. Journal of Experimental Psychology: Human Perception and Performance, 9, 622-636.

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

Port, R. F. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. Journal of Phonetics, 7, 45-56.

Prigogine, I. (1984). Irreversibility and space-time structure. In W. Horsthemke & D. K. Kondepudi (Eds.), Fluctuations and sensitivity in nonequilibrium systems. Berlin: Springer-Verlag.

Reed, E. S. (1981). Indirect action. Unpublished manuscript, University of Minnesota, Center for Research in Human Learning.

Richardson, I. W., & Rosen, R. (1979). Aging and the metrics of time. Journal of Theoretical Biology, 79, 415-423.

Schmidt, R. A. (1982). Motor control and learning: A behavioral emphasis. Champaign, IL: Human Kinetics.

Shapiro, D. C., Zernicke, R. F., Gregor, R. J., & Diestel, J. D. (1981). Evidence for generalized motor programs using gait pattern analysis. Journal of Motor Behavior, 13, 33-47.

Shik, M. L., & Orlovskii, G. N. (1965). Coordination of the limbs during running of the dog. Biophysics, 10, 1148-1159.

Stone, M. (1981). Evidence for a rhythm pattern in speech production: Observations of jaw movement. Journal of Phonetics, 9, 109-120.

Summerfield, Q. (1975). Information processing analyses of perceptual adjustments to source and context variables in speech. Unpublished doctoral dissertation, Queen's University of Belfast.

Sussman, H. M., & MacNeilage, P. F. (1978). Motor unit correlates of stress: Preliminary observations. Journal of the Acoustical Society of America, 64, 338.

Tuller, B., Harris, K. S., & Kelso, J. A. 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 timing of articulatory gestures: Evidence for relational invariants. Journal of the Acoustical Society of America, 76, 1030-1035.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1982). Interarticulator phasing as an index of temporal regularity in speech. Journal of Experimental Psychology: Human Perception and Performance, 1982, 8, 460-472.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1983). Further evidence for the role of relative timing in speech. Journal of Experimental Psychology: Human Perception and Performance, 9, 829-833.

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

Turvey, M. T., & Kugler, P. N. (1984). A comment on equating information with symbol strings. American Journal of Physiology, 246, R925-R927.

Viviani, P., & Terzuolo, V. (1980). Space-time invariance in learned motor skills. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland.

Footnote

[1] We see efforts to explore the origins of language in terms of more fundamental processes, such as particular perception-production mappings, as entirely consistent with this view (e.g., Lindblom, 1983a, 1983b; Lindblom, MacNeilage, & Studdert-Kennedy, 1983).

A THEORETICAL MODEL OF PHASE TRANSITIONS IN HUMAN HAND MOVEMENTS*

H. Haken,† J. A. S. Kelso,†† and H. Bunz†

Abstract. Earlier experimental studies by one of us (Kelso, 1981a, 1984) have shown that abrupt phase transitions occur in human hand movements under the influence of scalar changes in cycling frequency. Beyond a critical frequency the originally prepared out-of-phase, antisymmetric mode is replaced by a symmetrical, in-phase mode involving simultaneous activation of homologous muscle groups. Qualitatively, these phase transitions are analogous to gait shifts in animal locomotion as well as phenomena common to other physical and biological systems in which new "modes" or spatiotemporal patterns arise when the system is parametrically scaled beyond its equilibrium state (Haken, 1983). In this paper a theoretical model, using concepts central to the interdisciplinary field of synergetics and nonlinear oscillator theory, is developed, which reproduces (among other features) the dramatic change in coordinative pattern observed between the hands.

1.0 Introduction

While researching voluntary oscillatory motions of the two index fingers, one of us (Kelso, 1981a) observed an interesting phenomenon.[1] Under instructions to increase the frequency of out-of-phase, antisymmetrical motion (involving simultaneous flexor and extensor muscle activities), the subject's finger movements shifted abruptly to an in-phase symmetrical mode that involved simultaneous activation of homologous muscle groups. This finding was not restricted to finger movements. In later work (Kelso, 1982, 1984) that employed similar experimental manipulations, modal transitions in hand motions around the wrist were also observed: the antisymmetrical phase relationship between the hands was replaced by symmetrical phasing. Moreover, although the phase transition occurred at very different frequencies of hand motion for different subjects, it was nevertheless predictable. When the transition frequency was expressed in units of preferred frequency, i.e., an independent measure of the rate at which each subject was content to cycle the hands "as if he/she were going to do it all day," the resulting dimensionless ratio or "critical value" was constant for all subjects. Introducing a frictional re-

---

*Biological Cybernetics, in press.

†Institut für Theoretische Physik, Universität Stuttgart.

††Also Departments of Psychology and Biobehavioral Sciences, University of Connecticut.

Acknowledgment. The authors wish to thank the Volkswagenwerk Foundation for financial support to the synergetics project. J. A. S. Kelso was also supported by Contract No. N0014-83-K-0083 from the U.S. Office of Naval Research, NIH Grant NS-13617, and Biomedical Support Research Grant RR-05596.

sistance to movement systematically changed both the preferred and transition frequencies for each subject, but did not change the critical value across all subjects (Kelso, 1984).

The most dramatic aspect of these simple experiments, addressed in detail in the present theoretical model, is the sudden and completely involuntary change in the ordering or phasing among muscle groups that occurs at a critical, intrinsically defined frequency (see Figure 1). In this feature, the hand movement data share a likeness to gait transitions in locomotion.[2] For example, Shik, Severin, and Orlovskii (1966) showed that a steady increase in electrical stimulation to the midbrain region of the decerebrate cat was sufficient not only to induce an increase in locomotion rate, but, above a certain value of current, gait shifts as well. Like the hand experiments in which "flipping" from one mode to another occasionally occurred at higher movement frequencies, they too noted the presence of unstable regions in which the cat shifted from trotting to galloping and back again. Though the hand

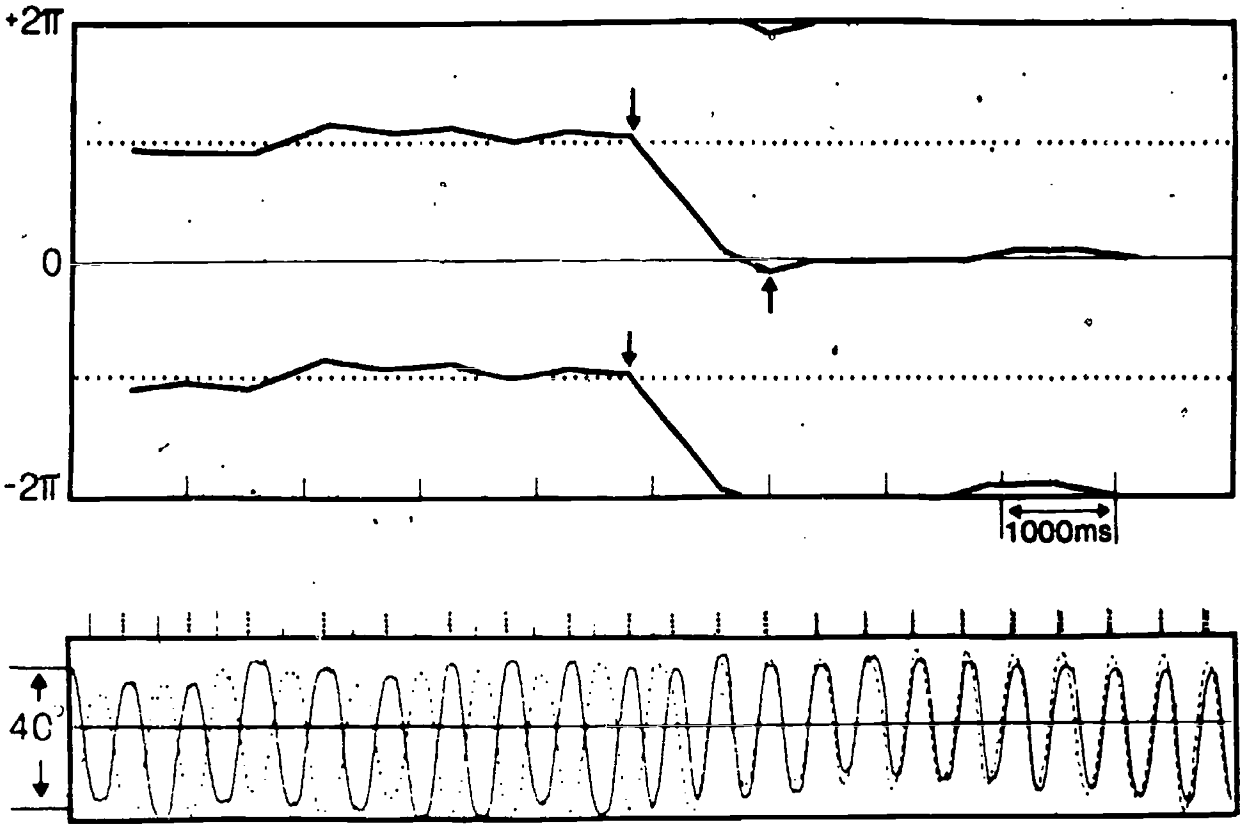


Figure 1. Bottom. Displacements over time of left (solid line) and right (dashed line) hands. The subject is simply increasing cycling frequency in an antisymmetric mode in response to a verbal cue from the experimenter. Top. Phase relationship between the two hands. The peaks of one hand movement act as a 'target' file and their phase position is calculated continuously relative to the peak-to-peak period of the other 'reference' file. The graphic display repeats the phase curve so that phase lags and leads can be noted.

data as well as these findings on quadruped gait strongly suggest that changes in coordination may be ordered by changes in a single parameter, the neural processes underlying such motoric phase transitions are still poorly understood. As Grillner (1982, p. 224) notes for the case of quadruped gait transitions, the general conception is that there is a "switch mechanism" in which "coordinating fibers" serve to switch among hindlimb neural networks. But such coordinating fibers have yet to be identified neuroanatomically (Grillner, 1982) and their exact functional role in determining locomotor pattern remains to be explained.

This problem of relating neuronal events to global patterns of behavior--in the present case abrupt macroscopic changes in the phasing of neuromuscular activities and changes in characteristic quantities such as frequency and amplitude--is somewhat reminiscent of a similar problem confronting physicists about 50 years ago. After it was discovered that matter consists of atoms and after the properties of atoms were understood theoretically, it may have seemed straightforward to derive the macroscopic properties of matter directly from the properties of the individual atoms. It turned out, however, that such a goal could not be reached immediately and it proved extremely fruitful to introduce macroscopic quantities for purposes of system description. Only later did it become possible to derive the equations governing the macroscopic quantities by means of a microscopic theory (for review and examples, see e.g., Wilson, 1979). It has been shown quite generally in the interdisciplinary field of synergetics (e.g., Haken, 1983) that in many cases the behavior of complex systems can be successfully modeled by means of a few macroscopic quantities in those situations where the behavior of the system changes qualitatively. Such macroscopic observables are called "order parameters" following a term first introduced by Landau (1936) to describe the "degree of order" (cf. Ter Haar, 1965, p. 208) of matter as it undergoes changes in phase. In synergetics, however, which deals with cooperative phenomena in non-equilibrium, open systems, the concept of order parameter has added significance: not only is it created by the cooperation of the individual components of a complex system, but the order parameter in turn governs the behavior of these components (for many examples, see Haken, 1975). Even in physical and chemical systems, finding the correct order parameter(s) is not always a simple matter. In the case of biological systems in general, and movement control and coordination in particular, the strategic approach of synergetics allows some license in selecting order parameters, an issue that we turn to next.

2.0 Initial Development of the Model:
Order Parameters and the Potential Function

To summarize, the main features of the experiments described briefly above are: (i) the presence of only two stable phase (or "attractor") states between the hands (which one is observed is a function of how the system is prepared, i.e., an instruction to move the hands in the out-of-phase or in-phase mode); (ii) the abrupt transition from one attractor state to the other at a critical cycling frequency; (iii) beyond the transition, only one mode (symmetrical in-phase) is observed; and (iv) when cycling frequency is reduced, the system stays in the symmetrical mode, i.e., it does not return to its initially prepared state--a result that suggests coexistence of the basins of attraction for the symmetrical and antisymmetrical modes and the depletion of one of them. Taken together, these results as well as other findings in the motor control literature support the hypothesis that phase is a relevant macroscopic (or "essential") parameter of certain movement patterns.[3] For example, the internal phasing structure of activities as widely varied as chewing, locomotion, handwriting, and speech remains invariant across scalar changes in force or rate (Grillner, 1982; Kelso, 1981b; Schmidt, 1982; Tuller, Kelso, & Harris, 1982). Similarly, in the experiments described above, phase is preserved constant over a wide range of frequencies, even though the magnitudes and durations of muscle activities and other kinematic variables change considerably. Only when frequency is scaled beyond a critical value does a phase shift occur.

In the present paper it seems reasonable to propose phase as an order parameter for at least two reasons. First, unlike many other possible candidates, phase is an accurate reflection of the cooperativity among the components of the system. Thus, we can say, in a manner consistent with synergetics, that the configuration of the subsystems (in the present context defined as the individual hand motions) specifies their phase relation, and conversely, that the phase variable specifies the spatiotemporal ordering of the subsystems. Second, it is phase that remains invariant across transformations in many motor activities that involve very different anatomical substrates. This highlights an important further feature of the order parameter concept, namely, that the order parameter (by hypothesis here, the relative phase) changes much more slowly than the variables describing the behavior of the individual components (e.g., velocities of each hand motion).

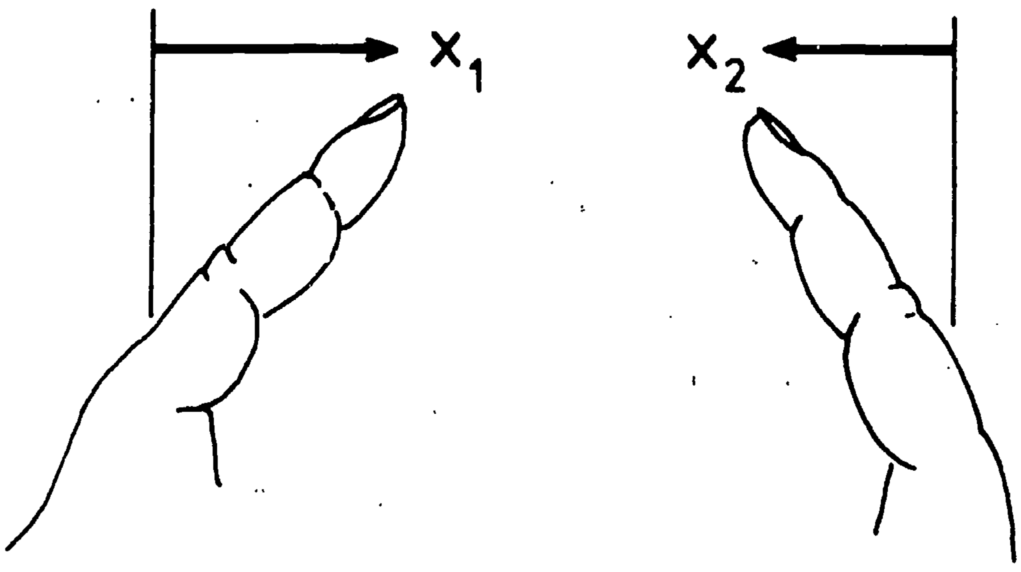


Figure 2. The displacements of x_1 and x_2 of the finger tips of the left and right hand in the symmetrical (in-phase, homologous) mode.

Our first step in the development of the present model is to provide a mathematically accurate description of the main qualitative features of the data. We therefore specify a potential function that corresponds to the layout of attractor states and show how that layout is altered as a control parameter is changed. In a following section, we employ nonlinear oscillator theory to show how the model equations describing the potential function can be derived from the equations of motion of each hand and a (nonlinear) coupling between them.

For sake of clarity we introduce the elongations of the finger tips x_1 and x_2 as shown in Figure 2. In order to define the relative phase ϕ we assume that the motion of the hands is more or less harmonic (see Figure 1) so that we put

2.1 $x_1 = r_1 \cos(\omega t + \phi_1)$,
2.2 $x_2 = r_2 \cos(\omega t + \phi_2)$,

where ω is the basic frequency of the hand movement, while the amplitudes r_1, r_2 and the phases ϕ_1, ϕ_2 are time dependent quantities whose time dependence is assumed to be much slower than that defined by the frequency ω. The relative phase is defined by

2.3 $\phi = \phi_2 - \phi_1$.

In order to describe the change of phase we adopt basic ideas from synergetics. As shown in synergetics, in many cases the equations for order parameters are of the form

2.4 $\dot{\phi} = \frac{\partial V}{\partial \phi}$,

where V is the so-called potential function. In our search for a model we make a few rather obvious assumptions about V. Since ϕ occurs under cosine or sine functions (cf. [2.1, 2.2])[4] the properties of the physical system must not change when ϕ is replaced by $\phi + 2\pi$. Consequently, we shall postulate that the potential V is periodic:

2.5 $V(\phi + 2\pi) = V(\phi)$.

We furthermore introduce the assumption that both hands play a symmetric role.[5] In such a case the behavior of the system must not depend on the way we label the right hand and the left hand. This means that V must remain unchanged when we exchange the indices 1 and 2 in Equation 2.3. This in turn means that the potential V is symmetric:

2.6 $V(\phi) = V(-\phi)$.

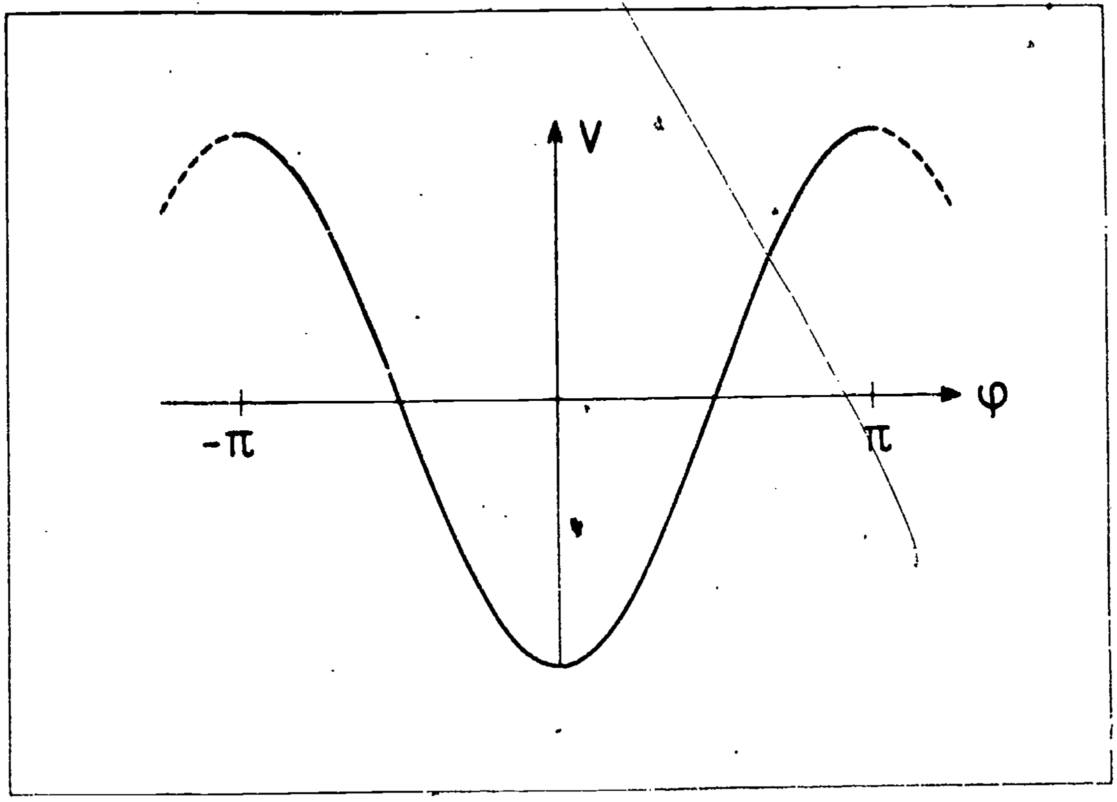


Figure 3. The potential V (2.7) for b = 0.

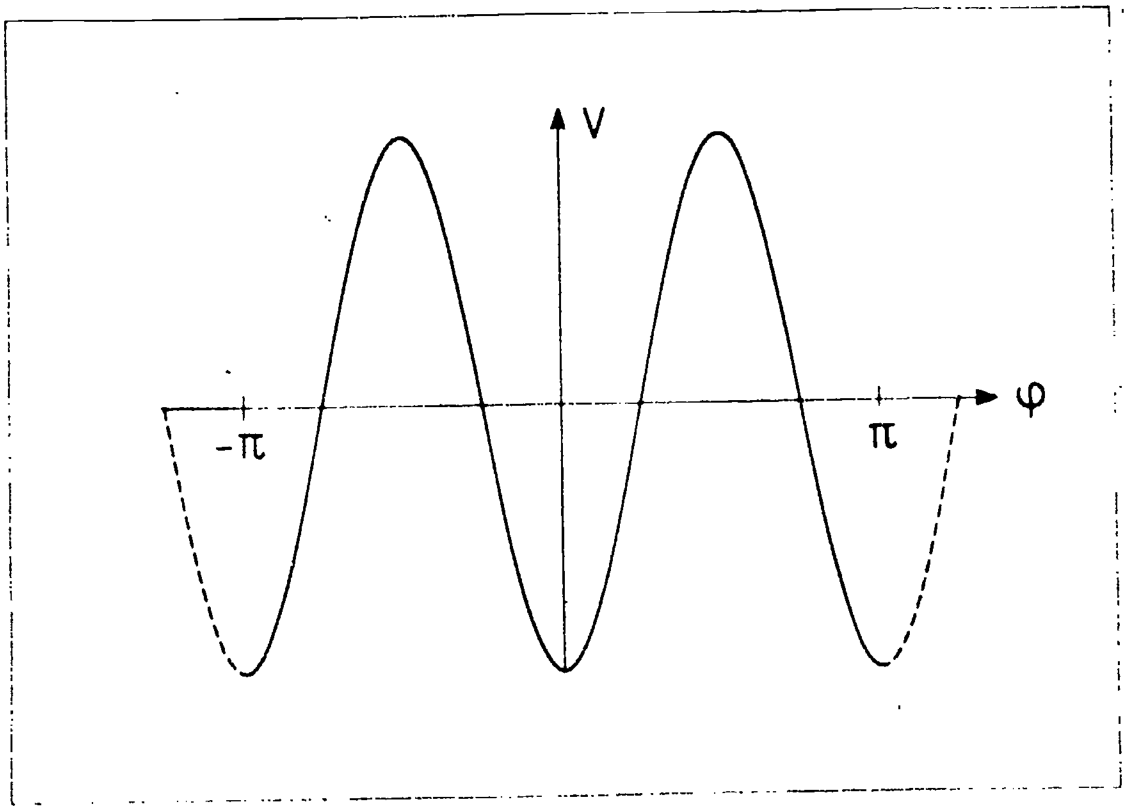


Figure 4. The potential V (2.7) for a = 0.

We assume that V obeys the conditions (2.5) and (2.6) in the simplest form that explains the above mentioned experimental results. To this end we write V as a superposition of two cosine functions:

$$V = -a\cos\phi - b\cos 2\phi. \tag{2.7}$$

As is known from synergetics, the behavior of the system obeying the equation (2.4) can be easily described by identifying ϕ with the coordinate of a particle which moves in an overdamped fashion in the potential V. To illustrate this let us consider Figure 3 where b is put equal to 0. There is only one stable equilibrium position, namely at $\phi = 0$. When we take $a = 0$, $b \neq 0$, the potential function looks like the one shown in Figure 4. Here we have two equivalent positions, namely at $\phi = 0$ and $\phi = \pi$ (which is equivalent to $\phi = -\pi$). When we take the total superposition (2.7) but change the ratio b/a we run through a series of potential fields shown in Figure 5. When we initially prepare the system in a state shown by the black ball and increase the frequency, and likewise assume that b/a decreases with increasing frequency we obtain a critical value ω_c where the ball falls to the lower minimum belonging to $\phi = 0$. This means that the hand movement made a transition from the antisymmetric ($\phi = -\pi$ state) into the symmetric state with $\phi = 0$. The hand movement stays in that state when ω is further increased. When we decrease ω starting from high values, the system remains all the time in the $\phi = 0$ state even if ω drops below ω_c. This "hysteresis" phenomenon is well known in many physical and biological systems.

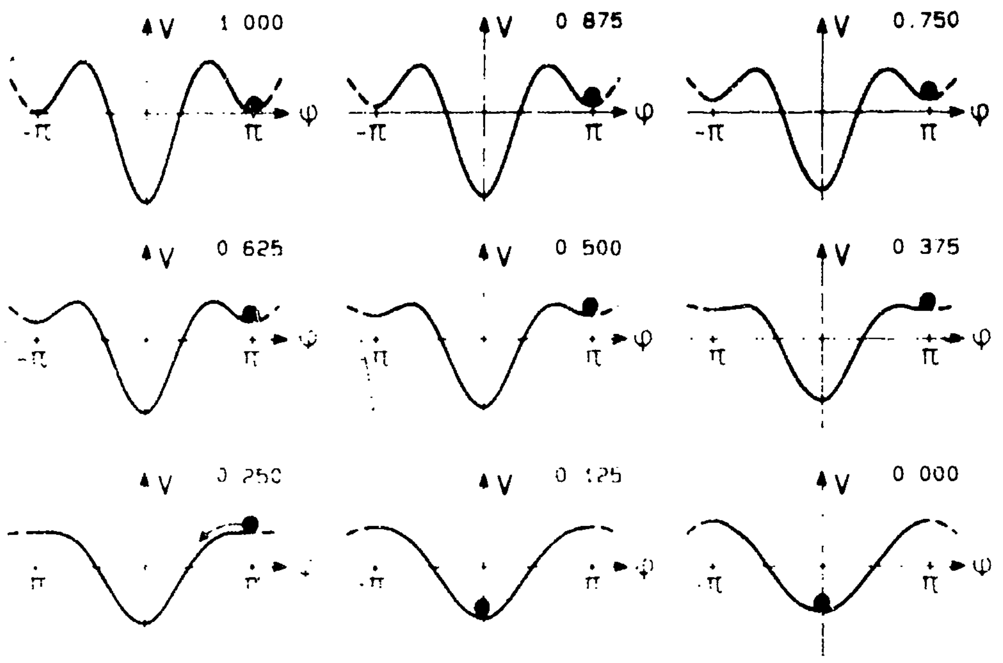


Figure 5. The potential V/a for the varying values of b/a. The numbers refer to the ratio b/a.

In order to study at which value of b/a the transition occurs, we seek the extrema of V that are defined by

2.8 $\frac{dV}{d\phi} = 0.$

Using (2.7), (2.8) reads:

2.9 $-a \sin\phi - 2b \sin 2\phi = 0.$

The second term can be transformed by means of:

2.10 $\sin 2\phi = 2\sin\phi \cos\phi$

so that (2.9) can be cast into the form:

2.11 $-a\sin\phi - 4b\sin\phi\cos\phi = 0.$

One set of roots is given by:

2.12 $\sin\phi = 0$

namely

2.13 $\phi = 0, \ \phi = \pm\pi.$

The other set of roots is given by:

2.14 $-a - 4b \cos\phi = 0$

or, when we solve for $\cos\phi$, by:

2.15 $\cos\phi = -\frac{a}{4b}.$

This value of $\cos\phi$ corresponds to the inner maxima of V. The transition occurs when these maxima vanish, which is the case if (2.15) can no more be fulfilled by a real ϕ. This happens provided:

2.16 $\left|\frac{a}{4b}\right| > 1$

or

2.17 $|b| < |a|/4$

i.e., the transition occurs if $|b|$ drops below the critical value $b_c = |a|/4$. On the other hand, we know from experiments that the amplitudes r_1, r_2 decrease with increasing ω. This suggests that b can be expressed by means of the amplitude $r = r_1 = r_2$ and a critical amplitude r_c so that we may write the potential function in the form:

2.18 $$V = -a(\cos\phi + \frac{r^2(\omega)}{4r_c^2}\cos 2\phi),$$

where r_c is defined as that value of r where the transition occurs.

3.0 Further Development of the Model

In the next step of our analysis we want to show how the model equations derived in the previous section can be derived from equations for the movements of the individual hands and a coupling between them. We write the corresponding equations in the form:

3.1 $$\ddot{x}_1 + f_1(x_1,\dot{x}_2) = I_{12}(x_1,x_2),$$

3.2 $$\ddot{x}_2 + f_2(x_2,\dot{x}_2) = I_{21}(x_1,x_2).$$

The left hand sides describe the motion of the individual hands with amplitudes x_1 and x_2, respectively, while the right hand sides describe the coupling. Of course, the coupling is achieved via the nervous system and in this way the equations (3.1) and (3.2) describe a complex system composed of the mechanical motions of the hands generated, in large part, by neuromuscular input. It is our goal to derive a minimal model for the macroscopic observables that are now the amplitudes and phases of the hand motion. Since the motion is basically oscillatory, we need at least a second order differential equation so that the terms $\ddot{x}_1$, $\ddot{x}_2$ occur. With respect to the restoring and damping forces we have a certain repertoire at hand and in all likelihood the choice of f_1 and f_2 is not unique. Since the hand movement has a more or less stable amplitude the equations must be nonlinear. We study several different examples. The first is well known from the operation of vacuum tube oscillators, but here, of course, we shall use only its mathematical properties. Let us consider the Van der Pol equation of the form:

3.3 $$\ddot{x} + \varepsilon(x^2 - r_0^2)\dot{x} + ax = 0,$$

where ε, r_0 are adjustable, but then fixed parameters, while a serves as a control parameter. In order to solve this equation for not too high amplitudes and in order to cast it into a form convenient for our later purposes we put:

3.4 $$x = Ae^{i\omega t} + A^*e^{-i\omega t},$$

where $\omega^2 = a$, and the complex amplitude A can be time dependent. It is assumed, however, that its time dependence is much slower than that of $e^{i\omega t}$. One can then perform two approximations well known in the theory of nonlinear oscillators (e.g., Haken, 1984). The "slowly varying amplitude approximation" means that we neglect terms $\dot{A}$ compared to terms ωA. The "rotating wave approximation" means that we may neglect terms containing $e^{3i\omega t}$ and $e^{-3i\omega t}$ compared to $e^{i\omega t}$ and $e^{-i\omega t}$. By means of these approximations (3.3) is transformed into:

3.5 $$e^{i\omega t}\, i\omega(2\dot{A} + \varepsilon(A|A|^2 - Ar_0^2)) = 0.$$

In the steady state the amplitude A is a constant and the only nontrivial solution reads $|A|^2 = r_0^2$. Thus the amplitude becomes frequency independent. In order to find a decrease of the amplitude with frequency we adopt a new model equation, namely:

3.6 $$\ddot{x} + \varepsilon(\dot{x}^2 - \omega_0^2 r_0^2)\dot{x} + ax = 0.$$

Making again the rotating wave approximation and the slowly varying amplitude approximation and using the hypothesis (3.4) we readily obtain for (3.6):

3.7 $$e^{i\omega t}\, i\omega(2\dot{A} + \varepsilon(3A|A|^2\omega^2 - A\omega_0^2 r_0^2)) = 0.$$

In the steady state where $\dot{A} = 0$, (3.7) has the nontrivial solution

3.8 $$|A| = \frac{r_0\omega_0}{\sqrt{3}\,\omega}.$$

Thus the amplitude indeed drops with ω^{-1}, giving us therefore a model equation that describes both the oscillatory motions of the hands and a drop in amplitude with increasing ω.

The experimental results suggest a superposition of a constant amplitude function (corresponding to ε_1 in eq. 3.9) and a function that decreases with ω (corresponding to ε_2 in eq. 3.9). That is, there is an intercept as well as a slope to the observed relationship between amplitude and frequency of hand movement. Such a behavior can be modeled by a superposition of (3.3) and (3.6):

3.9 $$\ddot{x} + [\varepsilon_1(x^2 - r_0^2) + \varepsilon_2(\dot{x}^2 - \omega_0^2 r\omega_0^2)]\,\dot{x} + ax = 0.$$

This leads in the steady state to:

3.10 $$|A|^2 = \frac{(\varepsilon_1 + \varepsilon_2\omega_0^2)r_0^2}{(\varepsilon_1 + 3\varepsilon_2\omega^2)}.$$

As we shall see, however, the main features of the phase transition can be modeled by choosing 3.6 as a basic equation.

We now come to the central problem, namely to derive a suitable coupling between the two macroscopic quantities, i.e., the amplitudes x_1 and x_2. The simplest hypothesis would be a linear coupling of the form:

3.11 $$I_{12} = \alpha(x_1 - x_2).$$

However, as we shall see below such a coupling will not lead to the required potential V for the relative phase. Rather we have to add a nonlinear coupling. Requiring that this coupling term has the same symmetry properties as (3.11) we are led to a coupling term of the following kind:

3.11a $$I_{12} = \alpha(x_1 - x_2) + \beta(x_1 - x_2)^3.$$

A detailed analysis reveals that such a coupling term still lacks an important feature, namely, it does not produce the correct phase relation between the motions of the individual hands. We have to introduce the coupling term either via a time delay or by using time derivatives. We first study the coupling by a time delay. This can be achieved by averaging over past values of (3.11a) so that we can replace (3.11a) by:

3.11b $$I_{12} = \int_{-\infty}^{t} [\alpha(x_1 - x_2)_\tau + \beta(x_1 - x_2)_\tau^3] e^{-\gamma(t-\tau)}\, d\tau.$$

An equivalent formulation is obtained by the requirement that I_{12} obeys the differential equation:

3.11c $$\dot{I}_{12} + \gamma I_{12} = \alpha(x_1 - x_2) + \beta(x_1 - x_2)^3.$$

In order to facilitate the subsequent calculation we shall assume that γ is much smaller than ω. This assumption is not all that crucial, however, since it does not change the basic structure of the equations. In order to proceed further we differentiate equations (3.1) and (3.2) with respect to time. Making use again of the slowly varying amplitude approximation and the rotating wave approximation, equation (3.1) acquires the form:

3.12 $$-\omega^2(2\dot{A}_1 + \varepsilon(3A_1|A_1|^2\omega^2 - A_1\omega_0^2 r_0^2)) = \alpha A_1 + 3\beta A_1|A_1|^2 + K_{12}.$$

The first two terms on the right hand side of (3.12) can be absorbed into the terms on the left hand side containing the factor ε and do not alter qualitatively the behavior of the system. The term K_{12} specifies the coupling influence of oscillator 1 on oscillator 2, corresponding to the motions of the two hands. In the above mentioned approximation K_{12} reads:

3.13 $$K_{12} = -\alpha A_2 - 3\beta(A_1^2 A_2^* + 2|A_1|^2 A_2) + 3\beta(2A_1|A_2|^2 + A_1^* A_2^2) - 3\beta A_2|A_2|^2.$$

We are now in a position to show how our model equations (3.1), (3.2) with the specific choice (3.11b) allow us to derive the order parameter equation (2.4). To this end we make the hypothesis:

3.14 $$A_j = r_j e^{i\phi_j},\ j = 1,2,$$

where r_j and ϕ_j may be time dependent, which transforms (3.12) into:

3.15 $$e^{i\phi_1}[-\omega^2\{2\dot{r}_1 + 2i\dot{\phi}_1 r_1 + \varepsilon(3\omega^2 r_1^3 - \omega_0^2 r_0^2 r_1)\} - \alpha r_1 - 3\beta r_1] = K_{12}.$$

K_{12} acquires the form:

3.16 $$K_{12} = -\alpha r_2 e^{i\phi_2} - 3\beta r_1^2 r_2(2e^{i\phi_2} + e^{2i\phi_1 - i\phi_2}) + 3\beta r_1 r_2^2(2e^{i\phi_1} + e^{2i\phi_2 - i\phi_1}) - 3\beta r_2^3 e^{i\phi_2}.$$

Similarly the equation for oscillator 2 contains the coupling term:

3.17 $$K_{21} = -\alpha r_1 e^{i\phi_1} - 3\beta r_2^2 r_1(2e^{i\phi_1} + e^{2i\phi_2 - i\phi_1}) + 3\beta r_2 r_1^2(2e^{i\phi_2} + e^{2i\phi_1 - i\phi_2}) - 3\beta r_1^3 e^{i\phi_1}.$$

We divide equation (3.15) by $-\omega^2 r e^{i\phi_1}$ and consider its imaginary part,

3.18 $$\dot{\phi}_1 = -\frac{1}{g(r)} \mathrm{Im}(e^{-i\phi_1}K_{12}),$$

where $g(r) = 2\omega^2 r_1$.

Applying the analogous procedure to oscillator 2 and taking the difference of the two equations for ϕ_1, ϕ_2 we obtain after a small intermediate calculation

3.19 $$\dot{\phi} = -\frac{2}{g(r)}[(\alpha r + 6\beta r^3)\sin\phi - 3\beta r^3 \sin 2\phi].$$

We have assumed that $r = r_1 = r_2$ and is well-stabilized, so that r is practically time independent. If the assumption $\gamma \ll \omega$ is dropped, then g(r) is replaced in eq. (3.19) by $\hat{g}(r) = 2(\omega^2+\gamma^2)r$. As we shall see β must have a sign opposite to α in order to obtain agreement with experimental findings. Therefore we put:

3.20 $$\beta = -\hat{\beta}.$$

Thus, we are left with our final equation:

3.21 $$\dot{\phi} = -\frac{2}{g(r)}[(\alpha r - 6\hat{\beta} r^3)\sin\phi + 3\hat{\beta} r^3 \sin 2\phi],$$

which indeed has the required structure of the order parameter equation (2.4) with (2.7). However, we are now in a position to relate the coefficients a and b to the amplitude r. The phase transition takes place for:

3.22 $$|b| = \frac{|a|}{4} \quad \begin{cases} \text{if } 4|b| > |a| & \text{bistable} \\ \text{if } 4|b| < |a| & \text{monostable} \end{cases}$$

(cf. [3.16]). Comparing the coefficients a and b with those occurring in (3.21), enables us to cast (3.22), into the form:

3.23 $$\frac{3}{2}\hat{\beta} r_c^2 = \frac{1}{4}(\alpha - 6\hat{\beta} r_c^2)$$

or, after a little algebra, into

3.24 $$r_c^2 = \frac{\alpha}{12\hat{\beta}}.$$

We thus find that bistable operation, particularly in the antisymmetric mode, occurs when r^2 fulfills (3.24). In the other case, with decreased amplitude the system becomes monostable and operates in the symmetric mode.

As mentioned above, there is still another possibility of defining I_{12}, namely by means of time derivatives. In the sense of a minimal model we choose:

3.25 $$I_{12} = (\dot{x}_1 - \dot{x}_2) \cdot (\alpha + \beta(x_1 - x_2)^2).$$

Again, to the same degree of approximation as used in eqs. (3.11)-(3.21) we obtain the equation for the phase:

3.26 $$\dot{\phi} = (\alpha + 2\beta r^2) \sin\phi - \beta r^2 \sin 2\phi.$$

The critical amplitude is then given by:

3.27 $$r_c^2 = \frac{\alpha}{4\hat{\beta}}.$$

In the transitions generally studied in synergetics, fluctuating forces play an important role. Extrapolating to the present case, a transition, say from $\phi=\pi$ to $\phi=0$ can be initiated only if fluctuating forces, F are present. To this end we enlarge the equations (3.1, 3.2) to include such forces, so that these equations now read:

3.28 $$\ddot{x}_1 + f_1(x_1,\dot{x}_1) = I_{12}(x_1,x_2) + F_1(t),$$

3.29 $$\ddot{x}_2 + f_2(x_2,\dot{x}_2) = I_{21}(x_1,x_2) + F_2(t).$$

In the context of the present paper it suffices to assume F_j, $j=1,2$ as a random small variable, which can be easily mimicked on a digital computer. At present, we cannot say much about the source of these fluctuations from existing experimental data. However, ongoing experimental work in which fine-wire electrodes are inserted into the finger muscles involved, is exploring their possible neuromuscular origin (see also Goodman & Kelso, 1983, for evidence pertaining to the relationship between physiological tremor "fluctuations" and voluntary movement).

4.0 Numerical Results

In this section we present some numerical results that correspond to the analytical treatment provided above. We solve the minimum model given by eq. (3.6) along with the coupling (3.25) on a digital computer using a fourth order Runge-Kutta method. To test the stability of a stationary solution small random fluctuations of finite amplitude are introduced. The resulting simulation shown in Figure 6 compares quite favorably with the experimental data (e.g., Figure 1). In Figure 6a the displacements x_1, x_2 are plotted over time and in Figure 6b the corresponding phase difference between the oscillators is plotted for the same motions. As in the bimanual experiments, the coupled oscillation is prepared in the state $\phi=\pi$ and the frequency ω is in-

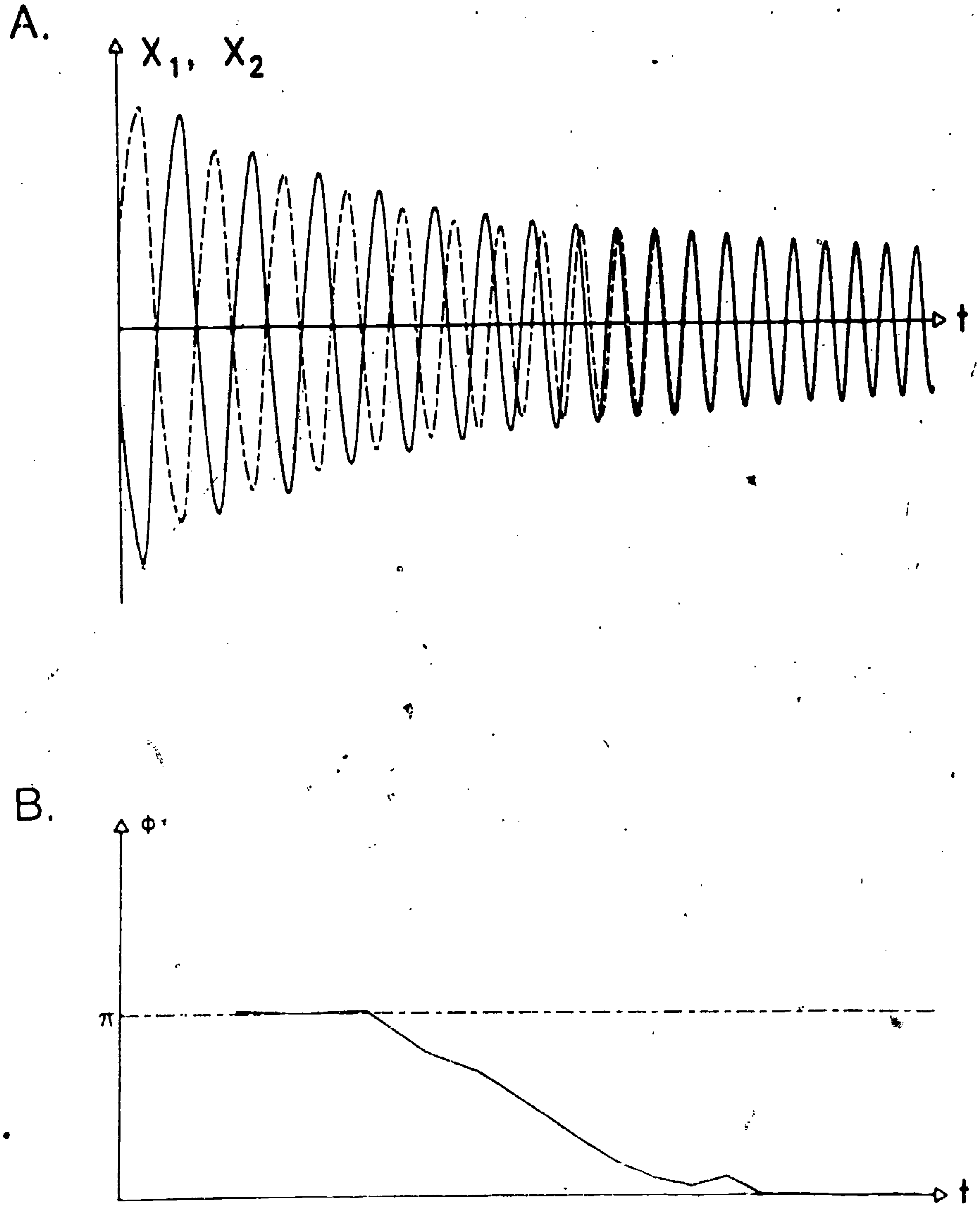


Figure 6. A numerical simulation of the phase transition in voluntary cyclical hand movement. In Figure 6a the displacements of the oscillators and in Figure 6b the corresponding phase difference between the oscillators is plotted over time. The parameters of the eq. (3.6, 3.25) were fixed at $\varepsilon=1$, $\omega_0^2 r_0^2=1$, $\alpha=-0.2$, $\beta=0.2$. From the left to the right of the displays, ω changes from $\omega=1.17$ to $\omega=3.05$.

creased monotonously. A transition from the out-of-phase mode to the in-phase mode is observed, when ω exceeds a critical value. However, the frequency of the oscillation changes rather quickly so that stationary oscillations are not reached. Thus the exact form of the curves depends strongly on the noise level and the rate of changing ω.

The steady state amplitudes for the in-phase mode and the out-of-phase mode are shown in Figure 7. The unstable branch of the out-of-phase mode is shown by dotted lines. The ω^{-1} dependence of the amplitudes is quite clear. This feature is exhibited only by the simplified model equations and will change if eq. (3.9) is used. As shown in Figure 7 for ω smaller than ω_c, the in-phase mode and the out-of-phase mode are both stable. Due to the coexistence of two basins of attraction, the particular mode observed depends on the initial conditions, i.e., which coordinative state is prepared.

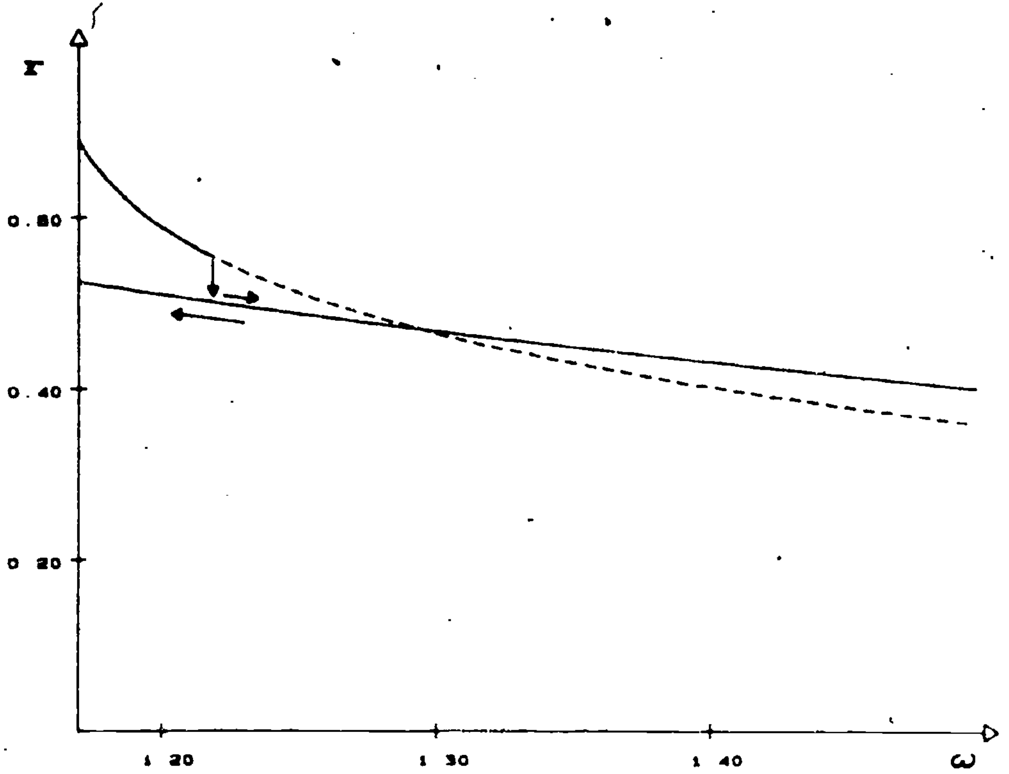


Figure 7. The steady state amplitudes of the in-phase mode (1) and the out-of-phase mode (2) are shown as a function of ω. The other parameters are fixed at the same values as in Figure 6. Stable branches of the oscillations are shown by the solid lines, and the unstable branch by the dotted line.

If one starts in the antisymmetric phase and increases ω slowly, the oscillation remains in this mode until the solution becomes unstable. At this point a jump in amplitude occurs and the only stable stationary solution revealed by the system corresponds to the in-phase mode. Such is the case when ω is increased further. On the other hand, if ω is decreased slowly the system stays within the basin of attraction of this solution even when ω drops below ω_c. As we mentioned earlier, this hysterisis phenomenon is typical for such bistable situations. To summarize, it is quite clear that the main features of the experimental data described at the beginning of Section 2.0 are captured by the present mathematical formulation as illustrated by these numerical results.

5.0 Concluding Remarks

In this paper we have introduced a minimal theoretical model that reproduces a number of the observed facts. The hand movements are described by two nonlinearly coupled oscillators that are self-sustained, i.e., not driven from the outside. The assumption of autonomous limit cycle oscillators is quite consistent with perturbation studies of two-handed cyclical movements, showing that an unexpected perturbation to one hand does not disrupt the phasing relation between the hands. The perturbed hand returns to its limit cycle almost immediately (see Kelso, Holt, Rubin, & Kugler, 1981; Yamanishi, Kawato, & Suzuki, 1980). Similar results in very different preparations (e.g., Cohen, Holmes, & Rand, 1982; Willis, 1980, for reviews) have also led to limit cycle models of neural pattern generation.

In the present model the frequency ω is defined as a control parameter via the coefficient, a, of the restoring force. The model describes not only the observed decrease in hand movement amplitudes with increasing frequency ω, but, more importantly, the phase transition, i.e., the change of qualitative behavior from antisymmetric to symmetric hand movement. A relation that automatically results from the equations is that the transition takes place at a critical frequency via the amplitudes. This prediction is now open to further experimental test. In future studies a number of phenomena known to accompany phase transitions in synergetic systems (e.g., critical slowing down; critical fluctuations) will be analyzed.

For the moment, any speculation on the origin of the coupling between the two hands is certainly premature. One coupling may be established via the corpus callosum, the well-known band of fibers that joins the two hemispheres of the brain. On the other hand, recent experiments (Tuller & Kelso, 1984) with patients whose corpus callosum has been severed, effectively cutting off communication between the left and right cerebral hemispheres, show that even in this case, control of the two index fingers in cyclical tasks is not independent. When asked to follow two pacing lights whose phase was varied between synchrony and alternation, split-brain subjects produced predominant synchrony or alternation even when paced at intermediate phase values. This bias in intermanual phase toward temporal symmetry is extremely powerful (see e.g., Kelso, Southard, & Goodman, 1979; Yamanishi et al., 1980, for evidence in normal populations) and suggests that the neural coupling for the voluntary hand movements may be established subcortically.

In conclusion, although we have shown here how a transition from one modal configuration to another is possible in our model, it remains for further theoretical and experimental research to address how it is that only two stable modes emerge in the first place from a wealth of possibilities, i.e., how these particular cooperativities arise. What is clear, however, from the present analysis, borne out by our numerical results, is the need to characterize the individual oscillators as nonlinear. But more important the coupling between oscillators must be nonlinear for the phase transition to occur. Although the present formulation clearly points to the important role of nonlinearities in certain basic motor behaviors (i.e., the frequency-amplitude relation in individual hand movements, modal transitions between the hands), the physiological underpinnings of such nonlinearities remain an open issue.

On the other hand, though their physiological basis may be obscure at present, it is entirely reasonable to inquire how a complex neuromuscular system might exploit these nonlinearities. Why are they important attributes for a neural control system to possess? What are they for? First, nonlinearity affords a stable coupling between the fundamental physical variables of space and time (i.e., the amplitude-frequency relation). In a linear system no such preferred coupling exists between these variables. Second, nonlinearity provides a means by which switching among coordinative states is possible (though other properties, e.g., fluctuations, play a key role also). In principle, there is no reason to limit this conclusion to the two phasing relations studied here. Thus, both of these attributes, we hypothesize, guarantee--in the present context--stability and flexibility of motor function.

References

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

Cohen, A. H., Holmes, P. J., & Rand, R. H. (1982). The nature of the coupling between segmental oscillators of the lamprey spinal generator for locomotion: A mathematical model. Journal of Mathematical Biology, 13, 345-369.

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

Goodman, D., & Kelso, J. A. S. (1983). Exploring the functional significance of physiological tremor: A biospectroscopic approach. Experimental Brain Research, 49, 419-431.

Grillner, S. (1982). Possible analogies in the control of innate motor acts and the production of sound in speech. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson, (Eds.), Speech motor control. Oxford: Pergamon Press.

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

Haken, H. (1983). Synergetics: An introduction. Heidelberg: Springer-Verlag.

Haken, H. (1984). Laser light dynamics. Amsterdam: North-Holland.

Kelso, J. A. S. (1981a). On the oscillatory basis of movement. Bulletin of the Psychonomic Society, 18, 63.

Kelso, J. A. S. (1981b). Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance (IX). Hillsdale, NJ: Erlbaum.

Kelso, J. A. S. (1982, March 1-5). Invited paper presented at Kroc Foundation Conference on Nonlinear Mechanisms in Brain Function, Santa Barbara, CA.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 15, R1000-R1004.

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from nonlinear limit cycle oscillatory processes: Theory and data. Journal of Motor Behavior, 13, 226-261.

Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979). On the nature of human interlimb coordination. Science, 203, 1029-1031.

Landau, L. (1936). The theory of phase transitions. Nature, 138, 840. Also in D. Ter Haar (Ed.), Collected papers of L.D. Landau. Oxford: Pergamon Press, 1965.

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

Schmidt, R. A. (1982). Motor control and learning: A behavioral emphasis. Champaign, IL: Human Kinetics.

Shapiro, D. C. (1981). Personal communication to J.A.S. Kelso.

Shik, M. L., Severin, F. V., & Orlovskii, G. N. (1966). Control of walking and running by means of electrical stimulation of the mid-brain. Biofizyka, 11, 659-666.

Ter Haar, D. (Ed.) (1965). Collected papers of L. D. Landau. Oxford: Pergamon Press.

Tuller, B., & Kelso, J. A. S. (1984, February). The kinematics of bimanual movements in commissurotomy patients. Paper presented at the International Neuropsychology Society Meeting, Houston, TX.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1982). Interarticulator phasing as an index of temporal regularity in speech. Journal of Experimental Psychology: Human Perception and Performance, 8, 460-472.

Willis, J. B. (1980). On the interaction between spinal locomotor generators in quadrupeds. Brain Research Reviews, 2, 171-204.

Wilson, K. G. (1979). Problems in physics with many scales of length. Scientific American, 241, 158-179.

Yamanishi, J., Kawato, M., & Suzuki, R. (1980). Two coupled oscillators as a model for the coordinated finger tapping by both hands. Biological Cybernetics, 37, 219-225.

Footnotes

[1] In discussions of these experiments with D. Shapiro (Shapiro, 1981, personal communication) it was learned that Cohen (1971) observed occasional involuntary shifts into the in-phase coordinative mode when out-of-phase motions at a single cycling frequency (3 Hz) were required. Cohen did not, to our knowledge, examine the phenomenon further. Similar experimental findings on bimanual finger movements have been reported by MacKenzie and Patla (1983) and by Baldissera, Cavallari, and Civaschi (1982) on ipsilateral hand and foot movements.

[2] Indeed, it was the slogan "Let your fingers do the walking" promoted by advertisers of the Yellow Pages in U.S. telephone directories, that led to the idea behind the present experiments.

[3] The reader must be warned that the word "phase" in the context of this paper has two different meanings:

1) "phase" as a temporal relationship whose precise definition is given in 2.1-2.3.

2) "phase" as a state of aggregation of matter (e.g., liquid or solid) or, more generally, different modes of behavior. Therefore, in physics, "phase transition" means transition from one state, e.g., fluid, to another one, e.g., solid. In synergetics, transitions between different dynamic states (e.g., behavioral modes) are also called phase transitions.

Since in the present paper the behavioral modes are characterized by definition (1), the notion "phase transition" is unique--in spite of the double meaning of "phase."

[4] This becomes obvious when we change the origin of time so that $\omega t + \phi_1 = \omega\tau$. In this case $x_1 = r_1 \cos(\omega\tau)$, $x_2 = r_2 \cos(\omega\tau+\phi_2-\phi_1)$.

[5] Our model can easily be generalized to include asymmetries.

REPETITIVE NAMING AND THE DETECTION OF WORD RETRIEVAL DEFICITS IN THE BEGINNING READER*

Robert B. Katz† and Donald Shankweiler††

Abstract. The claim has been advanced that children with severe reading disability are generally deficient in word retrieval compared with normal readers. Support for the claim is based largely on studies of rapid naming of repetitively presented pictured objects or other nameable stimuli, a task that is apparently more sensitive to retrieval problems than the confrontation naming of items presented singly. The purpose of this study was to examine whether there is a general relationship between word retrieval speed and reading ability in beginning readers. Although such a relationship has not been detected with confrontation naming, repetitive naming may provide a more sensitive test. Accordingly, second-grade children were required to name as rapidly as possible repeated presentations of five pictured items drawn from a single category. Separate naming tests were made for objects, colors, animals, letters, and words. The results showed that there was no relationship between reading ability and naming times when the test items were selected from sets of objects, colors, or animals, whereas on letters and words, a significant relationship was found. The less-skilled readers were not, therefore, consistently slower in all repetitive naming situations. Instead, their word retrieval deficits extended only to the orthographic materials.

It has often been claimed that many elementary school children with reading disorder experience word retrieval problems, a difficulty they share with most adult aphasics (e.g., Goodglass, 1930; Howes, 1964). Support for the claim derives largely from studies of children's performances on object-naming tasks. The most widely used procedure for testing naming is by a so-called confrontation naming test. The subject is presented with objects one at a time and is required to name each item as it appears. Generally, each pictured object is presented only once. Reading-disabled children have been

---

*Cortex, in press.

†Now at the University of Miami and the Veterans Administration Medical Center, Miami.

††Also University of Connecticut.

Acknowledgment. This investigation was supported by NICHD Grant HD-01994 and BRS Grant RR-0556 to Haskins Laboratories. We are grateful to Carol Hill, supervisor of reading, and the principals and teachers of the Manchester, Connecticut, public school system. We are also grateful to the second-grade children for their participation. Our thanks also go to Anne Fowler, Carol Fowler, and Maryanne Wolf for their comments on earlier versions of this manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-81 (1985)]

found to make a greater number of errors on this task than normal readers (Denckla & Rudel, 1976a; Jansky & deHirsch, 1973; Mattis, French, & Rapin, 1975).

Other studies, however, have failed to find evidence of naming problems in poor readers. Response times of less-skilled readers have been found not to differ from those of skilled readers on objects, colors, or digits (Perfetti, Finger, & Hogaboam, 1978; Stanovich, 1981); less skilled readers responded as quickly and as accurately even in studies using letters as stimuli (Stanovich, 1981; Wolford & Fowler, 1983). The choice of subjects may partially account for these discrepant findings. The first-mentioned studies obtained subjects from learning disability clinics, whereas the latter recruited poor readers from ordinary school classes. Thus, it is possible that the first studies tested children with more severe reading disability and more severe associated cognitive deficits than the second series. The task employed may also account for the variability in the results. It is possible that naming deficits may exist even in the more moderately impaired poor readers, but that confrontation naming tasks are not sensitive enough to detect them reliably.

This possibility is suggested by studies using another procedure. An alternative naming task involves continuous rapid naming of a small set of repeated pictured items, typically drawn from a single category. The task requires the subject to scan the display of pictures arrayed in horizontal rows and to name each picture in succession as rapidly as possible. Each item occurs several times and at various positions in the display. Such a repetitive naming test was used by Denckla and Rudel (1976b) to compare normal readers with a group of severely disabled readers selected from special school programs and clinics. The results indicated that the overall response times of these poor readers were longer than those of normal readers on every category of item tested (objects, digits, colors, and letters).

It is apparent that the two types of naming tasks are quite different in the demands they make, and that each provides us with different information. The repetitive naming task is the focus of our interest here because of the discovery that response times on this task reliably differentiate normal and poor readers even on producing response words of high frequency (Denckla & Rudel, 1976b). These findings raise the possibility that some less-skilled readers have a general problem in word retrieval that could not have been discovered using the apparently less-sensitive confrontation naming test. But to date, only one study (Blachman, 1981) has tested repetitive naming with a whole (first-grade) school class. It was found that color naming and letter naming correlated with reading ability, but object naming did not. However, the age of the subjects in the Blachman study limits the conclusions that can be drawn. Tested in the first grade, these children may not have been old enough to permit the reading problem cases to be identified.

The purpose of the present study was to examine whether there is a relationship between naming times on a repetitive naming task and reading ability among second-grade children selected from ordinary school classes. Naming ability was tested with the following categories of items: pictured objects, pictured animals, colors, letters, and words. If the less-skilled readers have consistently longer naming times than the skilled readers, then they may have general word retrieval deficits. If, on the other hand, a relationship between reading ability and naming time holds only for selected categories of items, then it may be possible to delimit possible explanations for why the less-skilled readers are slower at repetitive naming of these types of test materials and not others. In the latter case, some potential explanations, such as differential familiarity with the classes of stimulus items, should be considered. Other explanations, based solely on the mechanical aspects of repetitive naming, such as efficiency of scanning, could be ruled out.

Method

Subjects

The subjects were the 18 children from two second-grade classes in a suburban Connecticut public school, for whom parental permission for testing was granted. Of these, two were excluded from testing because English was a recent second language. The remaining 16 children were given the word identification and word attack subtests of the *Woodcock Reading Mastery Tests* (Woodcock, 1973) and the *Peabody Picture Vocabulary Test* (PPVT) (Dunn, 1959) near the end of the school year. The eight children (6 females, 2 males) with the highest combined raw scores on the two Woodcock subtests were designated the skilled readers.[1] The eight remaining children (5 females, 3 males) were designated the less-skilled readers.

The skilled readers had a mean combined Woodcock score of 140.2 (range: 129 to 164), compared with the less-skilled readers' mean of 92.6 (range: 50 to 128). Moreover, the two reading groups achieved significantly different scores on each of the two component parts. On the word identification subtest (which tests the reading of single words) alone, the skilled readers' average reading grade level was 3.9 (range: 3.3 to 5.6), whereas the less-skilled readers' average was 2.8 (range: 2.2 to 3.6), $t(14) = 4.1$, $p = .002$. On the word attack subtest (which tests the reading of pseudowords), the skilled readers' mean reading grade level was 5.8 (range: 4.1 to 12.9), compared with the less-skilled readers' mean of 2.6 (range: 1.2 to 6.1), $t(14) = 4.6$, $p < .001$. IQ scores derived from the PPVT yielded the following results: the skilled readers obtained a mean IQ of 118.6 and the less-skilled readers 94.4, $t(14) = 4.7$, $p < .001$. A difference in the mean age of the members of the two groups was evident also; the skilled readers' mean age was 7 years, 9 months, whereas the mean for the less-skilled readers was 8 years, 4 months, $t(14) = 3.9$, $p = .002$. In anticipation of the results, we should mention that the differences in IQ and age, which distinguished the two reading groups, were apparently of no consequence.

Materials

The test stimuli (printed words, letters, colored squares, and line drawings of objects) were arrayed in consecutive rows in a manner similar to that of Denckla and Rudel (1974, 1976b). Each array consisted of five items of a single category, each of which was presented a total of 10 times in a matrix of 5 rows of 10 on white cardboard. The order of the items was random with the constraints that no item immediately succeed itself and that each item appear twice in every row. The categories, and the specific stimulus items in each, were: 1) line drawings of animals (*bird*, *cow*, *dog*, *goat*, *pig*); 2) objects (*ball*, *box*, *door*, *hat*, *table*); 3) colors (*blue*, *green*, *red*, *yellow*, *black*); 4) lower-case letters (*a*, *d*, *o*, *p*, *s*); and 5) common words (*ball*, *box*, *door*, *hat*, *table*). For each chart, a practice sequence consisting of the five items was constructed.

Procedure

Each child was tested individually in one 30-min session. At the beginning of the session, the child was given the Woodcock and PPVT. Following a brief memory test, the naming tests were given. The charts were always presented in the following order: objects, colors, letters, words, animals. Before being tested on a chart, the child was asked to name the five items in the practice sequence in order to ensure that standard names could be elicited to each item. Every child was able to do this without hesitation.[2] Then the subject was instructed to name each item on the chart as rapidly as possible without making any mistakes, following along the rows from left to right. The child began the task on a signal from the experimenter. The experimenter responded to hesitations with the injunction to "Keep going." A stopwatch was used to measure the time from the child's first response to the last response.

Results

The first point to be noted is that naming errors occurred infrequently, as was expected. Accordingly, the data base consisted of the mean naming times on each class of items. These are shown in Table 1. It can be seen from the table that the naming time varied considerably with item type. The order of response times across item types is generally consistent with that of previous repetitive naming studies (Biemiller, 1977-1978; Blachman, 1981; Denckla & Rudel, 1974, 1976b). Moreover, it is a standard finding that it takes longer to identify pictured objects than to read the objects' names. This result has been obtained both with adults (Cattell, 1886; Potter & Faulconer, 1975) and with children (Ligon, 1932; Seymour & Porpodas, 1980). It is of interest to discover that it applies even to readers so near the beginning stages of skill acquisition. Examining the effect of reading ability, we note that there was only a small overall difference in mean naming time between the skilled and the less-skilled readers. This was due to the fact that the less-skilled readers tended to be faster than the skilled readers on objects and animals, but slower on letters and words.

Table 1

Mean Naming Time (sec) for Each Item Type by Reading Level

Item Type	Skilled Readers	Less-skilled Readers
Animals	53.3	50.2
Objects	56.2	52.1
Colors	44.6	46.0
Letters	26.3	31.9
Words	29.1	35.8
Mean	41.9	43.2

The data were subjected to an analysis of variance with one between-groups factor (reading ability) and one within-groups factor (item type). The analysis indicated that the main effect of reading ability was not significant, $F < 1$, whereas the main effect of item type was highly significant, $F(4,56) = 74.3$, $p < .001$. Furthermore, the interaction between reading ability and item type was significant, $F(4,56) = 3.5$, $p = .013$. Fine-grained analyses of this interaction were conducted using protected t-tests (Cohen & Cohen, 1975). These analyses indicated that for objects, animals, and colors the naming times of skilled and less-skilled readers were not significantly different: animals, $t(14) = -.6$, $p > .5$; objects, $t(14) = -.8$, $p = .409$; colors, $t(14) = .3$, $p > .5$. In contrast, significant differences were found on both letters, $t(14) = 2.5$, $p = .026$, and words, $t(14) = 2.8$, $p = .015$.[3]

It could be maintained that the analysis of variance obscures a general relationship between reading ability and naming time, since the children were divided into only two levels of reading ability, thus eliminating fine distinctions in actual reading skill. A correlational analysis was therefore carried out to assess the degree of relationship between the children's actual reading scores and their naming times. These correlations, which are shown in Table 2, form two groups of clustered variables. The first cluster consists of animals, objects, and colors; the second is formed by the reading score, letters, words, and colors. The variables in the first cluster have little relationship with reading score. Moreover, with the exception of colors, the variables in one cluster correlate relatively little with those in the other cluster. The correlations clearly indicate, as did the previous analyses, that there is a strong relationship between reading ability and the naming times for letters and words, but not for objects, animals, and colors.

Table 2

Correlations between Naming Times and Reading Score

Variable	1	2	3	4	5	6
1. Reading Score	-	.04	.12	-.07	-.58*	-.70**
2. Animals		-	.87***	.87***	.43	.31
3. Objects			-	.74**	.42	.30
4. Colors				-	.52*	.34
5. Letters					-	.62*
6. Words						-

*$p < .05$
**$p < .01$
***$p < .001$

Discussion

The purpose of the present experiment was first to examine whether word retrieval ability (independent of the reading process *per se*) is related to reading ability. Second, we wished to discover whether any difficulties that might be found are of a general nature or are circumscribed, becoming manifest in the retrieval of only certain classes of stimuli. The results of our study of rapid naming of repetitively present items indicated that the less-skilled readers were slower than the skilled readers at naming letters and words, but responded as quickly on objects, animals, or colors. Thus, the less-skilled readers were slower only on orthographic material.

The pattern of results obtained in this study raises two questions. First, why are less-skilled readers slower than skilled readers at repetitive naming of orthographic material but not on pictorial material? Because less-skilled readers responded as quickly as skilled readers on som classes of items, we can rule out those arguments that invoke reading-related differences in scanning rates, response strategies, use of peripheral information, response interference, or visual-verbal association. Similarly, although the less-skilled readers had lower IQ scores than the skilled readers, this relationship cannot account for the results. The lower IQ scores would have been expected to lengthen the naming times of the less-skilled readers on all five classes of stimuli. Although it is possible that lower IQ scores led to longer naming times on letters and words, other findings make this implausible. Other tests of repetitive letter naming using skilled and less-skilled readers with equivalent IQ scores (Biemiller, 1977-1978; Staller & Sekuler, 1975) also found an effect of reading skill on naming time.

One might also suppose that the obtained response time differences may be a function of familiarity with stimulus items. Since the less-skilled readers have read less extensively, they may well be less practiced than the skilled readers in identifying letters and printed words; this is less likely to be the case with objects, animals, or colors. Such experience-related effects could be expected to work against the less-skilled readers when tested on letter naming and word naming. It may be the case that differential familiarity with letters and words creates differences in the attentional resources needed by skilled and less-skilled readers to name items selected from these categories. Although the performance of skilled and less-skilled readers on naming letters presented singly has been found to be equivalent (Stanovich, 1981; Wolford & Fowler, 1983), the members of the two reading groups may still differ in the extent to which letter naming is automatized (LaBerge & Samuels, 1974). Thus, a less-skilled reader may have to invest a large share of processing capacity in order to name a letter as quickly as a skilled reader, who may have to devote relatively less attention to the task. Such differences in degree of automatization might be expected to become manifest only on repetitive naming tasks, since these tasks require that subjects do more than simply retrieve a single name. The additional task requirements of scanning and responding sequentially may be sufficient to expose reading-related differences in the rate at which certain types of items can be named. That is, naming letters, unlike naming pictured objects, is a somewhat fragile skill in poor beginning readers. It is easily disrupted when other factors related to sequential responding complicate the task. This possibility could account for the difference in outcome between our results on repetitive letter naming and those on naming single letters (Stanovich, 1981; Wolford & Fowler, 1983).

A second question that must be entertained concerns the source of the differences between our results and those of other studies that might be considered comparable. A comparison of our results with those of Blachman (1981) and Wolf (1981) suggests that the relationship between repetitive naming and reading scores varies with grade level; it is most robust at the very early stages of learning to read, after which it diminishes. Conceivably, this pattern of results is indicative of a developmental trend in which the less-skilled reader recovers from a general slowness in naming or acquires an increasing level of automatization for processing certain types of items. One may wonder, then, whether less-skilled readers who are slightly older than those studied here will show equal letter naming times relative to skilled readers of the same age. Previous studies (Biemiller, 1977-1978; Staller & Sekuler, 1975; Wolf, 1981) suggest that they will not, but additional research is needed to clarify the matter.

Our findings show that the less-skilled readers, compared with the skilled readers, were not characterized by a general slowness on repetitive naming. Therefore, the results indicate that less-skilled readers do not have a general problem in retrieving names. Rather, their word retrieval deficits may extend only to orthographic items.

It should be noted that neither the present study nor earlier studies that exploit the repetitive naming paradigm permit any definitive conclusion concerning possible deficiencies in the general ability of skilled and less-skilled readers to access their mental lexicons. The stimuli employed were limited to a small set of very common items. Moreover, since the items were known to the subjects prior to testing, it is likely that their names were stored in a temporary short-term buffer. Lexical memory may not have played an important role. Thus, the requirements of repetitive naming may differ from those of confrontation naming. In the latter, names must be retrieved from the lexicon itself because the test items are not made known to the subject beforehand. Research has shown that good and poor readers do indeed sometimes differ on confrontation naming (Denckla & Rudel, 1976a; Jansky & deHirsch, 1973; Katz, in press; Mattis et al., 1975; Wolf, 1981). Furthermore, data obtained on a confrontation naming task (Katz, in press), suggest that poor readers are able to access the lexicon adequately; it was hypothesized that the naming problems arise because of deficiencies in processing phonological information stored at specific lexical addresses and possibly also because of deficiencies in the quality or completeness of the phonological specifications.

References

Biemiller, A. (1977-1978). Relationships between oral reading rates for letters, words, and simple text in the development of reading achievement. Reading Research Quarterly, 13, 223-253.

Blachman, B. A. (1981). The relationship of selected language measures and the McCarthy Scales to kindergarten and first-grade reading achievement. Unpublished doctoral dissertation, University of Connecticut.

Cattell, J. M. (1886). The time it takes to see and name objects. Mind, 11, 63-65.

Cohen, J., & Cohen, P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum.

Denckla, M. B., & Rudel, R. G. (1974). Rapid "automatized" naming of pictured objects, colors, letters, and numbers by normal children. Cortex, 10, 186-202.

Denckla, M. B., & Rudel, R. G. (1976a). Naming of object-drawings by dyslexic and other learning disabled children. Brain and Language, 3, 1-15.

Denckla, M. B., & Rudel, R. G. (1976b). Rapid 'automatized' naming (R. A. N.): Dyslexia differentiated from other learning disabilities. Neuropsychologia, 14, 471-479.

Dunn, L. M. (1959). Peabody picture vocabulary test. Circle Pines, MN: American Guidance Service, Inc.

Goodglass, H. (1980). Disorders of naming following brain injury. American Scientist, 68, 647-655.

Howes, D. (1964). Application of word-frequency concept in aphasia. In A. V. S. DeReuck & M. O'Connor (Eds.), Disorders of language. Boston: Little, Brown.

Jansky, J., & deHirsch, K. (1973). Preventing reading failure. New York: Harper & Row.

Katz, R. B. (in press). Phonological deficiencies in children with reading disability: Evidence from an object naming task. Cognition.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology, 6, 293-323.

Ligon, E. M. (1932). A genetic study of color naming and word reading. American Journal of Psychology, 44, 103-122.

Mattis, S., French, J. H., & Rapin, I. (1975). Dyslexia in children and young adults: Three independent neuropsychological syndromes. Developmental Medicine and Child Neurology, 17, 150-163.

Perfetti, C. A., Finger, E., & Hogaboam, T. (1978). Sources of vocalization latency differences between skilled and less skilled young readers. Journal of Educational Psychology, 70, 730-739.

Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. Nature, 253, 437-438.

Rudel, R. G., Denckla, M. B., & Broman, M. (1978). Rapid silent response to repeated target symbols by dyslexic and nondyslexic children. Brain and Language, 6, 52-62.

Seymour, P. H. K., & Porpodas, C. D. (1980). Lexical and non-lexical processing in developmental dyslexia. In U. Frith (Ed.), Cognitive processes in spelling. London: Academic Press.

Staller, J., & Sekuler, R. (1975). Children read normal and reversed letters: A simple test of reading skill. Quarterly Journal of Experimental Psychology, 27, 539-550.

Stanovich, K. E. (1981). Relationships between word decoding speed, general name-retrieval ability, and reading progress in first-grade children. Journal of Educational Psychology, 73, 809-815.

Wolf, M. (1981). The word-retrieval process and reading in children and aphasics. In K. Nelson (Ed.), Children's language (Vol. 3). New York: Gardner Press.

Wolford, G., & Fowler, C. A. (1983). Differential use of partial information by good and poor readers. Developmental Review, 4, 16-35.

Woodcock, R. W. (1973). Woodcock reading mastery tests. Circle Pines, MN: American Guidance Service, Inc.

Footnotes

[1] The use of combined scores seemed justified in light of the high correlation between the children's scores on the two Woodcock subtests, $r(14) = .89$, $p < .001$.

[2] All the children were also very accurate at naming the items, except for one boy who consistently read the word "ball" as "bell."

[3] The skilled readers' mean naming times were affected by the extreme times of one subject. Contrary to the expectations based on previous findings, this child took several seconds longer than any other subject on the object, color, and animal charts. With her data eliminated, the mean times of the skilled readers are somewhat faster: animals, 48.9 sec; objects, 53.0;

colors, 41.1; letters, 25.1; words, 28.7. Nevertheless, the statistical effects reported here are maintained.

[4]Wolf's findings only recently came to our attention. Although the motivation for her study differed from ours, the results largely support our findings.

72

LINGUISTIC ABILITIES AND SPELLING PROFICIENCY IN KINDERGARTENERS AND ADULT POOR SPELLERS*

Isabelle Y. Liberman,† Hyla Rubin,†† Susan Duquès,††† and Joanne Carlisle††††

The research effort over the past several years by the Haskins Laboratories reading research group has bolstered our conviction that the problem of most beginners who have difficulties in acquiring literacy is basically linguistic in nature. That is, in our view, the problem is not visual or auditory or motor, as many have proposed, but lies rather in the ineffective use of phonologic strategies. We have found this linguistic deficiency in regard to two major requirements of reading proficiency--lexical access and representation in short-term memory. We have recently also begun to look more closely at spelling from the standpoint of linguistic sophistication. In this paper we will describe two recent studies we have done that are concerned with linguistic abilities and spelling--one in a group of kindergarteners and the other in an adult literacy class. But, first, since the nature of the orthography is a central consideration in spelling, we should like to prepare the way by describing our assumptions about how the alphabetic orthography represents language, assumptions that are, in effect, the guiding principles of our research.

Some Guiding Principles

Everyone would agree, we believe, that an orthography represents a language, that languages are used to convey meaning, and that words are the basic

---

*Paper presented at a Special Conference on the Biobehavioral Measures of Dyslexia sponsored by the National Institute of Child Health and Human Development, Bethesda, Maryland, September, 1984.

†Also University of Connecticut.

††Now at University of Toronto.

†††University of Connecticut.

††††Now at American International College.

Acknowledgment. The introductory section of this paper ("Some Guiding Principles") is adapted from "A Language-Oriented View of Reading and its Disabilities" by Isabelle Y. Liberman, which appeared in H. Myklebust (Ed.), Progress in learning disabilities (Vol. 5). New York: Grune and Stratton, 1983. The research of the authors was supported by NICHD grant HD-01994 to Haskins Laboratories. Special thanks for their generous cooperation in the kindergarten experiment are due to Jerome Spears, Director of Pupil Personnel of the Mansfield, CT school system, and, in the Annie E. Vinton School, to James Chilleri, principal and Anita Soriano, teacher of the kindergarten class and her pupils. For their thoughtful assistance in the adult literacy experiment, we are also indebted to Isobel Kaufman, Director, Adult and Community Education Services and to Thomas Rameaka, teacher and the students of the adult education class.

units by which languages do that. What is often forgotten, however, is that whether one receives language by eye or by ear, one must get to the word before one can get to its meaning and that a word exists apart from its typically various meanings. Moreover, a word has a uniquely linguistic, complex, phonological structure that must be somehow apprehended before one can deal with a message conveyed by language, whether written or spoken. In the primary language functions of speaking and listening, a special processor copes with that phonological structure in operations that normally function naturally, quite automatically. and below the level of awareness. One does not need to understand how it works in order to speak and listen. In contrast, both the reader and writer must have some fundamental understanding of the structure. Indeed, they must become, to some degree, linguists of sorts, sufficiently aware of the phonology to be able to divide the spoken language into the constituent segments that the orthography represents. How easy or how difficult that will be will largely depend on the nature of the linguistic segment that the orthography represents.

In the case of orthographies in which the segment represented is the word, as it is in the Chinese logography and the Japanese kanji, or the syllable, as it is in the Japanese kana,[1] we contend that the beginner's task is relatively simple. If what they need to do is separate the word or syllable from the speech stream for purposes of pairing it with its appropriate orthographic unit, they will find it to be readily isolable. This has been found to be true both here and abroad in a number of experiments with young children (Alegria & Content, 1983; Fox & Routh, 1975; Holden & MacGinitie, 1972).

The problem faced by beginners in an alphabetic orthography is much more difficult. The essence of the problem can be put in this way: though it is often said that an alphabetic orthography represents speech or ought to, it is, in fact, an abstraction from speech. Although it does bear a fairly regular relation to speech, the nature of that relation will be hard for a child, or indeed any beginner, young or old, to apprehend. To understand why that is so, we would remark briefly, first, on why it is misleading to say that the alphabetic orthography represents the sounds of speech; second, why it is also misleading to say that it does or should represent speech phonetically; and, finally, what it means to say that it is an abstraction from speech. Let us consider these remarks one at a time.

The Alphabet Does Not Represent the Sounds of Speech

The alphabetic orthography does not transcribe the sounds of speech. In the first place, the letters obviously do not portray acoustic events, as they might if they were bits of oscillograms or spectrograms. So in that rather trivial sense, the alphabet certainly does not represent the sounds of speech. However, there is a more significant sense in which the alphabet does not transcribe sounds. The point that needs to be made, and which is not trivial, is that the segmentation of the sound does not correspond to the segmentation of the letters. To take a simple example--it would be impossible to divide a recording of the spoken word, "big," into three parts, such that when played back, one part would be "buh," a second part "ih," and the last part "guh." That is because in the spoken syllable, "big," there is only one piece of sound and the three phonological segments we write with the letters B, I, and G are nearly simultaneously encoded into it.

Encoding several segments of the phonology into one segment of sound provides an important gain in efficiency for the listener, who, as we have said, has a built-in processor nicely equipped to deal with it automatically. But it has quite adverse consequences for the beginner dealing with the written language. One unfortunate consequence of the very odd relation between phonological structure and sound is that the phoneme, which is the segment represented by the alphabetic orthography, unlike the word and the syllable, is not easily separable from the speech stream. If we had not found this to be so in our research with children (Liberman, Shankweiler, Fischer, & Carter, 1974), we should have suspected it from the history of writing systems--the system using an alphabetic unit was the last to be developed, long after logographies and syllabaries. Moreover, unlike those others, the basic unit of the alphabetic orthography, the phoneme, was apparently discovered only once and all other alphabets were later adapted from that original, brilliant discovery (Diringer, 1948).

If readers and writers must be able to appreciate the relationship between the orthographic character and the linguistic unit it represents, as we believe they must, then, beginning learners of an alphabetic system are put at a disadvantage initially. They will find it difficult to see the relation between spelling and sound. And it will even be difficult for teachers to demonstrate that relationship to them. If teachers wish to do this with even a simple word like "big," they will try to isolate three sounds and in the process will unavoidably produce, not three phonemes, but three syllables: "buh," "ih," and "guh." Put together, these form a nonsense trisyllable "buhihguh" and not the monosyllable "big" that comprises the three phonological segments we spell as B-I-G (see Liberman, 1983, for a more complete discussion of this point and of the two sections that follow).

The Alphabet Does Not Represent the Phonetic Surface of Speech

If it is now evident that the alphabet is not a transcription of the sounds of speech, what about the alphabetic orthography as a phonetic transcription? It is, of course, possible to use an alphabet phonetically. Linguists do just that when they use a phonetic transcription to represent as precisely as possible what they perceive when they listen to speech. Unfortunately, the wealth of phonetic information that our natural speech-perceiving mechanisms can use creates serious problems when, as in reading and writing, we try to put all that information through the eye.

A phonetic transcription, such as the linguist uses, preserves much surface information that is not represented in any alphabetic orthography. It includes all the context-conditioned variations of speech both within words and across syllable and word boundaries. For example, in a phonetic transcription, the plural s in cats would be transcribed as s but its counterpart in "dogs" would be z to reflect its pronunciation in that context. To take another example, the final consonant in the word, "sit," would be transcribed as t, but what we ordinarily consider to be the same consonant in the related word, "sitter," would have to be changed from t to d to reflect accurately that manner change in our pronunciation. Similarly, in American English, the contraction "what's" would be transcribed differently in the context of "What's he saying?" from its rendition in the context of "What's your idea?", where, because of context-conditioned effects, it would be coarticulated with "your" to produce "Wuhchuh (idea)?"

In view of these confusing context-conditioned variations, one would suppose that it should be extremely difficult to apprehend messages conveyed by means of a strictly phonetic transcription. And so it is, in fact. To be sure, any literate adult can learn to decode a phonetic transcription more easily than s/he can decode the visual display of acoustic events in a spectrogram, but even highly trained phoneticians cannot read an unfamiliar text written phonetically as fluently as they would the same passage written in our English orthography. The representation of context-driven articulatory distinctions, to say nothing about differences in linguistic stress, emphasis, idiolect, and dialect, seriously detracts from the broader requirements of language representation. This is certainly a case in which, except for very specialized purposes, more is definitely not better.

The Alphabet Represents Phonological Structures

Given that reading the sounds of speech is hard and reading a phonetic transcription is only slightly less so, what is it that an alphabet should represent if reading is to be made as easy as possible? Presumably, in the ideal case, the representation should match the way words are organized in our heads, in what linguists refer to as our lexicons. It stands to reason that our lexicons must be organized in terms of phonological or morphophonological[2] segments that are sufficiently abstract to stand above the many variations at the auditory and phonetic surfaces. We have described the difficulties we get into when, in trying to put language in by eye, we begin with the variable auditory and phonetic forms. To get around these problems, we would want ideally to have words spelled in a way that matches the abstract (morpho)phonological structures as they must be stored in the speaker's lexicon. But that is an ideal that is not easily achieved.

The problem is that there are undoubtedly great differences among speakers of the language in the way in which their lexicons are organized and in exactly how abstractly the items are entered. To take one example, for the would-be reader who understands that such pairs as heal/health, steal/stealth, and even weal/wealth are related, the individual members of those pairs might well be entered quite differently than they would be for the reader who has never noted those relationships. The entries of the former reader would in this instance be closer to the way English spelling deals with the language.

For better or worse, English spelling happens to be quite far out on the abstractness dimension, rising considerably above the phonetic surface variations to preserve the identities of lexical cognates. The spelling is, in this sense, morphophonological in nature. As such, it necessarily must strain the linguistic sophistication of many would-be readers and spellers. The young child is especially likely to lack even the tacit knowledge, what we have elsewhere called "phonological maturity" (Liberman, Liberman, Mattingly, & Shankweiler, 1980), that is needed to rationalize so much of the spelling. For example, the use of the same alphabetic characters for phonological segments that are phonetically quite different, as in such pairs as muscle/muscular and magic/magician, preserves the morphemic relations of the words and thus may increase fluency and efficient comprehension for a mature reader but would serve only as a roadblock to the young child who is trying to figure out how the system works.

In summary, the point that should be emphasized here is that no matter how abstract it may often be and how far or how close to a given reader's lexicon, the alphabetic orthography does represent the (morpho)phonological structure of the spoken word, not its sounds or its phonetic surface. Now we can consider how this characteristic of the orthography relates to the attainment of literacy.

Linguistic Awareness and Reading

It has been our contention that in addition to the obvious need to have some command of the spoken language and the ability to discriminate the graphic symbols, the first requirement for beginning readers is to acquire a certain degree of linguistic sophistication, beyond that required for speaking and listening. One important aspect of linguistic sophistication is what Mattingly (1972) of the Haskins research group has dubbed "linguistic awareness." Though it could be taken to have a more general connotation, this term has been defined in a rather special way in the context of initial reading acquisition--to refer to the awareness of the units of speech that are represented by the orthography. In the alphabetic orthography, the phoneme, the unit of which the learner must become aware, is a constituent part of larger units, the word and the syllable, both of which have considerably more salience, as we have said. Awareness that these larger units have parts and the ability to identify those parts does not come easily and does not happen all at once. The ability to segment speech into its constituent units, of whatever size, has been found to show improvement from ages four to six or seven (Calfee, Chapman, & Venezky, 1970; Liberman, 1973; Treiman & Baron, 1981). But in this developmental sequence, awareness of the phonemic unit is always harder, develops later, and is generally found to be a more sensitive predictor of reading skills in kindergarteners and first graders than awareness of the syllable or word. There is by now a long list of studies, originating both here and abroad, which have been strongly supportive of phoneme segmentation skill as a predictor of reading ability. Among the studies that come to mind (and there are surely others as well) are Blachman (1983), Helfgott (1976), Mann and Liberman (1984), Zifcak (1981) from our research group, Bradley and Bryant (1983) in England, Lundberg and associates (1980) in Sweden, Fox and Routh (1975) in the States, and Bertelson's laboratory (Alegria & Content, 1983) in Belgium.

Most of the previous research on linguistic awareness and the acquisition of literacy has been concerned with the attainment of reading skills. Recently, we have begun to look more closely at spelling from this vantage point. We should like to report on two of these investigations in this paper, one in which we examined the invented spellings of kindergarteners and the other in which we explored the virtually uncharted territory of linguistic factors in adult illiteracy.

Linguistic Abilities and the Invented Spellings of Kindergarteners

The first study we will describe looked into the linguistic abilities of kindergarteners in relation to their skill in invented spelling. Before reporting on our findings, we should take a moment to say what we mean by invented spellings. When spelling words in their spontaneous writings, preschoolers are, of course, limited by their meager orthographic knowledge, which in the beginning may include only the knowledge of names of letters ("bee" and "dee," for example). In his seminal work in the early seventies, Charles Read (1971) demonstrated that the invented spellings of preschoolers display a predictable pattern in their choice of the letter symbols used to represent the spoken language. Relying on their apparently quite acute perception of the phonetic, surfacy features of both the utterance to be recorded and the letter names they know, young children begin by devising what amounts to a primitive phonetic transcription, rather than the phonological representation of our spelling system.

To the extent that English is written abstractly, it assumes, as we said earlier, a user who has, to a considerable degree, what we have called "phonological maturity" (Liberman et al., 1980). These younger children clearly do not have the requisite degree that our orthography demands. Given a word like "train," to borrow one of Read's examples, a preschooler might produce an H as the first letter of the word in an attempt to represent the first phone in their own spoken version of the word ("chrain") by its closest counterpart in a letter name they know ("aitch"). As the children begin to develop more sophisticated apprehensions of the phonology, the purely phonetic transcription becomes less prevalent in their spellings. They begin to assimilate the rules according to which our abstract spelling makes sense.

This growing awareness of the (morpho)phonological rule structure of words is implicit in their invented spelling productions. Therefore, if one is interested, as we are, in evaluating the level of linguistic sophistication in children's spellings, it is possible to do so by constructing a scoring system that is fashioned to reflect that awareness, rather than being limited to a consideration simply of right/wrong judgments. Louisa Cook Moats (1983), using an analytic scoring system of this kind, did a pioneering study in which she found the misspellings of dyslexic fourth through eighth graders to be quite similar linguistically to those of nondyslexic second graders.

We were curious to learn whether we could find possible precursors of linguistic deficiencies in the spelling of much younger children--those in a public school kindergarten class. Accordingly, we chose to examine the relationship between kindergarteners' proficiency in invented spelling and in their other linguistic abilities.

All the children in the kindergarten class were given a dictated, real word spelling test. A given word could receive a score from 0 (for simply random letters) to 6 (for a correct English spelling). In this scoring system, we measured the children's spelling proficiency along two dimensions--the number of phonemes that the child included in spelling the word and also the level of the orthographic representation. Thus, for the target word, "sick," increasing scores would be awarded for the following sequences of responses: one phoneme with conventional letters (s, c); more than one phoneme but not all (sk, ck); all phonemes with phonetically related letters (sec, sek); all phonemes with conventional letters (sic, sik); all with correct spelling (sick).

In addition to the spelling test, eight language-based tasks were administered to the class. We found that four of these made a difference in a multiple regression analysis, accounting for 94% of the variance in invented spelling proficiency. Of these, a phoneme segmentation task patterned after Elkonin (1973) accounted for 67% of the variance in invented spelling performance and one that measured phoneme dictation accounted for an additional 20%. A phoneme deletion task ("Say meat without the 'm'"), adapted from Rosner's

Test of Auditory Analysis Skills (Rosner, 1975), added another 6%; and a measure of expressive vocabulary, the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1976), 1%.

The four other language-based tasks in our study did not contribute significantly to the variance of the invented spelling performance. They included a test of receptive vocabulary (the Peabody Picture Vocabulary Test, Dunn & Dunn, 1981); a syllable deletion task ("Say cowboy without the 'cow'"), also adapted from Rosner's TAAS (1975); word repetition (correctly repeating words spoken by the examiner); naming letters and writing letters to dictation. Analysis of the children's performance on word repetition, letter naming, and letter writing revealed only developmentally appropriate errors such as slight infantilisms in articulation and occasional confusions of visually similar letters in writing. In regard to the latter type of error, it is of interest to note that letter reversals, though present in some protocols, were not found to be related to invented spelling ability. That is, children could be good invented spellers at the kindergarten level without having fully mastered the correct orientation of the reversible letters.

This study suggests that spelling skill develops systematically as young children master the ability to analyze words into their constituent phonemes. That conclusion is supported by the strength of phoneme segmentation ability, writing phonemes to dictation, and the deletion of phonemes as predictors of invented spelling ability. These three tasks all require a degree of explicit awareness of internal phonological word structure that is not tapped by the other language tasks. The other tasks all reflect certain aspects of language development but either do not include the analytic component at all (Peabody Picture Vocabulary, letter naming, letter writing, and word repetition) or tap it at a less abstract level, closer to the basic unit of articulation (syllable deletion).

Linguistic Abilities and Adult Poor Spellers

Among kindergarteners, then, we had found that the children who were the better spellers in the class exhibited better skills as well in analyzing the phonemic constituents of words. Recently, we examined a group of adults enrolled in a community literacy class with a view toward finding out whether their profiles would be similar to those of younger learners. The subjects in this study were nine men whose occupations ranged from lower-level management to semi-skilled labor, all of whom reported serious difficulties with spelling. Five of the men had repeated a grade in school, but only one had received remedial assistance.

Once again, as we had with the kindergarteners, we administered a number of tests of language ability as well as measures of spelling proficiency. To determine the kinds of spelling problems characteristic of adult illiterates, we used two dictated lists of spelling words. One list was taken from the spelling subtest of the Gallistel-Ellis Test of Coding Skills (1974), which includes real words of both regular and irregular orthographic construction. The other was a list of pseudowords taken from the reading subtest of the Gallistel-Ellis Test. In order to provide a comparison with their spelling proficiency in spontaneous writing, we also collected writing samples, using the stimulus pictures of the Test of Written Language (Hammill & Larsen, 1978).

Four tests of language ability were included in the adult study. As a check on the possibility that gross problems in speech perception might be at the root of their difficulties, we examined the performance of the adults on the Sound Mimicry subtest of the Goldman-Fristoe-Woodcock Auditory Skills Test Battery (1974). In this subtest subjects are required to repeat taped nonsense words, one to three syllables in length.

In view of previous findings with children that have suggested that poor readers and spellers may have difficulty analyzing other aspects of internal structure of words (Carlisle, 1984; Rubin, 1984), we wished to measure the analytic abilities of the adults at both the phonemic and morphemic levels of language. The test used for phonemic analysis was the Sound Analysis subtest of the Goldman-Fristoe-Woodcock Auditory Skills Test Battery (1974). Here, monosyllabic nonsense words are presented on tape and the subject is required to identify the first, middle, or last phoneme in the word. To determine the subjects' ability to apply basic inflectional and derivational rules of morphology, they were also given the Berry-Talbott Test of Language (1966). Normal children have been found to develop the morphological abilities tapped by this test in a systematic progression from the easier items at the beginning to the more difficult ones at the end, with mastery expected by age seven or eight.

Finally, in addition to testing them on spelling and language tasks, we also measured the oral reading ability of our adult subjects. For this purpose, both single word and passage reading measures were used. The reading subtest of the Gallistel-Ellis Test of Coding Skills was chosen to assess reading of single words. This subtest includes real words of two types--those of irregular (unpredictable by the more common orthographic rules) construction, and those of regular construction that are presented in order of increasing difficulty by syllable type. The test also includes nonsense words that are arranged by syllable type as well. The Spache Diagnostic Reading Scales (1972) were used to assess oral passage reading.

We can turn now to the results of our study of adults and look first at their spelling performance. On the dictated spelling of real words, they did somewhat better on the irregular than on the regular words--63% as against 57% correct, reflecting, perhaps, the tendency to rely on the memory of the global appearance of words often found clinically in poor readers. This possibility is supported by the large drop to 38% correct in their performance on nonsense words often found clinically in poor readers. In order to appreciate the seriousness of the drop, it should be noted that whereas there is, of course, only one correct response for real words, whether of regular or irregular construction, there can be several acceptable spellings of each nonsense word. Consider the pseudoword "lete," for example. Four spellings--lete, leet, leat, liet--would all be scored correct.

To be sure, even with this apparent advantage for the nonsense words, one might still expect some discrepancy in performance between the spelling of nonsense and real words, in favor of the real words. However, it was evident from the pattern of their results that our adult subjects had not mastered the basic phoneme-grapheme spelling patterns that would allow them consistently to produce even phonetically reasonable renditions of words they had not seen before. One striking example that occurred during the reading test comes to mind. When presented with the written word, <u>peg</u>, one of the adult subjects puzzled over the word for some time and finally said: "Pig? Well, I know it's not pig, because there's a e in the middle, but I guess I'll go with pig." (Letter discrimination was obviously not his problem.) It is relevant to remark at this point that our adults had little difficulty in the spoken repetition of the auditorily presented nonsense words, performing there with 92% success. This suggests that the problems they are having with spelling cannot be attributed to gross difficulties in speech perception (nor in articulation, for that matter, or some bias against using nonsense words.

Thus far we have reported only on spelling performances on dictated single words. The spelling of the adults on the spontaneous writing samples was somewhat better than on the dictated words--78% of the words were spelled correctly. But it was apparent that whatever improvement occurred here could probably be attributed to the tendency of the subjects to limit their productions to words that they thought they could spell correctly. A large proportion of one subject's output even included his copying of the wording of the printed test directions, for example.

An informal qualitative analysis of the errors made by the adults on the writing samples showed clearly that they had serious linguistic problems over and above their poor spelling. Approximately a third of the errors reflected grammatical weaknesses--difficulties with function words accounted for 12% of the errors, and omissions or substitutions of inflectional endings accounted for another 21%.

In the light of these findings, it is of interest to note that our adults passed only 63% of the items on the Berry-Talbott test, which measures inflectional and derivational knowledge. In contrast, in a study recently completed by one of us (Rubin, 1984), a group of 60 first graders was able to pass 57% of the same items. Moreover, the young children did well on the lower level items and less well on the higher, thus showing a systematic development of morphemic understanding. The adults, on the other hand, often performed poorly on even the simplest categories like plural and past tense inflections, though they were able to use them correctly in their spontaneous speech. It would seem that when they have to do even a moderate degree of analysis of their language, whether written or spoken, their linguistic abilities are strained to the point of breakdown.

In view of their performance on the morphemic analysis task, it was not surprising to find that language analysis at the phonemic level was especially trying for these adults. On our very simple phoneme analysis task, which is similar to those used in training kindergarteners and first graders in phonemic segmentation, only 58% of their responses were correct. Moreover, they found the task frustrating and unpleasant.

This inability of adults with literacy problems to perform well in a task requiring explicit understanding of the internal structure of words has also been found by other investigators (Byrne & Ledez, 1983; Marcel, 1980; Morais, Cary, Alegria, & Bertelson, 1979; Charles Read. personal communication, 1984). Particularly convincing in this connection is the finding by Morais and associates (1979) that the performance of first graders in the third month of school was slightly better in both phoneme deletion and addition tasks than that of the adult illiterates in their study.

We now turn to a comparison of the reading and spelling of our adult subjects. We found that their reading of single real words was better than their spelling, as would be expected in any comparison of recognition and production measures. But the pattern of performance was quite similar. Real words were read with greater accuracy than nonsense words, just as had been the case with spelling. But perhaps the most telling result was found in a direct comparison of reading and spelling as a function of word type. Whereas the reading of real words, as we have said, was generally better than the spelling, the situation was quite different in regard to the nonsense words. On nonsense words, for which a structural analytical approach is obligatory rather than optional as it may be in dealing with real words, the performance of the adults on both reading and spelling was virtually identical in quality and quite poor.

In short, the adult subjects, like the poorest invented spellers in the kindergarten study, appeared to have only the dimmest understanding of the phonemic structure of words. Qualitative analysis of their successes in reading suggested that memorizing words as global entities was a favored strategy. The analysis revealed, for example, that they often did well on words they had seen frequently before, even when the words were polysyllabic or irregular in construction. They might, for example, read correctly a ten-letter, trisyllabic word with uncommon spelling like photograph, but be at a loss to deal with a simple trigram like peg that they had never encountered before. The same kind of contrast between performance on complex practiced words and unknown but relatively simple words was also apparent in their spelling.

Their oral reading of connected text on the Spache test (1972) was clearly superior to their reading of single words, suggesting, as has been found in other studies of poor readers (Perfetti & Hogaboam, 1975), that our adult subjects were relying heavily on context to assist them in apprehending familiar words. At all events, reading, like spelling, was patently a struggle for these men, generating once more grammatical errors not present in their everyday speech. Examination of their errors in oral passage reading revealed a pattern of incorrect use of inflectional morphemes and functors much like that noted in their spontaneous writing samples.

Educational Implications

In our introductory remarks, we have advanced our reasons for expecting that the acquisition of literacy would be related to a certain degree of linguistic sophistication, that is, to the ability to deal with the structure of language in an analytic manner. In the first study reported here we found that among kindergarteners, the better spellers, like the better readers among older children in previous investigations, have developed this ability to a higher degree than those who spell more poorly. In the other study, a group of adults in a community literacy class, who showed no serious deficiencies in everyday speaking and listening, were extremely poor spellers and at the same time were also deficient in various tasks requiring analytic linguistic skills. They found phonological segmentation tasks particularly troublesome.

The ability to stand back from one's language and analyze its structure apparently does not develop naturally as a result of cognitive maturation. It must be learned or taught. But there are several ways in which it can be learned or taught. Many children, as we have pointed out elsewhere (Liberman & Shankweiler, 1977), will develop linguistic insights as a consequence of their experiences in learning to read an alphabetic orthography. But even for those children, the process might have been made easier and faster by giving them explicit instruction. For other children, including especially those who for whatever reason are at risk linguistically, more explicit instruction will be required if they are ever to attain true literacy--to be able, that is, to deal effectively with unknown or previously unlearned words, which is what an alphabetic orthography is all about. The adults in our community literacy class can still profit from explicit instruction in all aspects of the structure of language (phonological, syntactic, and morphological). But they would have been spared much grief and embarrassment if their deficits had been discovered and addressed in kindergarten instead.

Several investigations have now demonstrated that linguistic awareness can be trained (Bradley & Bryant, 1983; Olofsson & Lundberg, 1983; Vellutino, this volume, 1983) and will make a difference in reading acquisition. Recent studies make it clear that phonological (Fischer, Shankweiler, & Liberman, in press) and morphological (Carlisle, 1984; Fischer et al., in press; Rubin, 1984) knowledge make a difference in spelling proficiency. At all events, it now seems reasonable to suggest, in view of the present findings and in light of the characteristics of the alphabetic orthography and how it relates to language, that more and earlier training in all aspects of linguistic sensitivity may promote better spelling and should be encouraged. Kindergarten is not too soon to start.

References

Alegria, J., & Content, A. (1983, March). *Explicit analysis of speech and learning to read: Some experimental studies.* Paper presented at the Third Language Acquisition Conference. Gent.

Berry, M. (1966). *Berry-Talbott Language Test: Comprehension of grammar.* Rockford, IL.

Blachman, B. A. (1983). Are we assessing the linguistic factors critical in early reading? *Annals of Dyslexia*, *33*, 91-109.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature*, *301*, 419-421.

Byrne, B., & Ledez, J. (1983). Phonological awareness in reading-disabled adults. *Australian Journal of Psychology*, *35*, 185-197.

Calfee, R. C., Chapman, R. S., & Venezky, R. L. (1970). *How a child needs to think to learn to read.* Madison, WI: Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin.

Carlisle, J. F. (1984). *The relationship between derivational morphology and spelling ability in 4th, 6th, and 8th graders.* Unpublished doctoral dissertation, University of Connecticut, Storrs.

Diringer, D. (1948). *The alphabet: A key to the history of mankind.* New York: Hutchinsons.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test, Revised.* Circle Pines, MN: American Guidance Service.

Elkonin, D. B. (1973). U.S.S.R. In J. Downing (Ed.), *Comparative reading.* New York: Macmillan.

Fischer, F. W., Shankweiler, D., & Liberman, I. Y. (in press). Spelling proficiency and sensitivity to word-structure. *Journal of Verbal Learning and Verbal Behavior.*

Fox, B., & Routh, D. K. (1975). Analyzing spoken language into words, syllables, and phonemes: A developmental study. *Journal of Psycholinguistic Research*, *4*, 331-342.

Gallistel, E., & Ellis, K. (1974). Gallistel-Ellis Test of Coding Skills. Hamden, CT: Montage Press.

Goldman, R., Fristoe, M., & Woodcock, R. W. (1974). G-F-W Sound-Symbol Tests, Goldman-Fristoe-Woodcock Auditory Skills Test Battery. Circle Pines, MN: American Guidance Service.

Hammill, D. D., & Larsen, S. C. (1978). Test of Written Language. Austin, TX: Pro-Ed.

Helfgott, J. (1976). Phonemic segmentation and blending skills of kindergarten children: Implications for beginning reading acquisition. Contemporary Educational Psychology, 1, 157-169.

Holden, M. H., & MacGinitie, W. H. (1972). Children's conceptions of word boundaries in speech and print. Journal of Educational Psychology, 63, 551-557.

Kaplan, E., Goodglass, H., & Weintraub, S. (1976). Boston Naming Test. Boston, MA: Boston University School of Medicine.

Liberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. Bulletin of the Orton Society, 23, 65-77.

Liberman, I. Y. (1983). A language-oriented view of reading and its disabilities. In H. R. Myklebust (Ed.), Progress in learning disabilities (Vol. V). New York: Grune and Stratton.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading, and dyslexia. Baltimore, MD: University Park Press.

Liberman, I. Y., & Shankweiler, D. (1977). Speech, the alphabet, and teaching to read. In L. Resnick & P. Weaver (Eds.), Theory and practice of early reading. Hillsdale, NJ: Erlbaum.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. Journal of Experimental Child Psychology, 18, 201-212.

Lundberg, I., Olafsson, 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.

Mann, V. A., & Liberman, I. Y. (in press). Phonological awareness and verbal short-term memory: Can they presage early reading problems? Journal of Learning Disabilities.

Marcel, A. (1980. Phonological awareness and phonological representation: Investigation of a specific spelling problem. In U. Frith (Ed.), Cognitive processes in spelling. London: Academic Press.

Mattingly, I. G. (1972. Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly [illegible], Language by ear and by eye. Cambridge, MA: MIT Press.

Moats, L. C. (1983). A comparison of the spelling errors of older dyslexic and second-grade normal children. Annals of Dyslexia, 33, 121-140.

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

Olofsson, A., & Lundberg, I. (1983). Can phonemic awareness be trained in kindergarten? Scandinavian Journal of Psychology, 24, 35-44.

Perfetti, C. A., & Hogaboam, T. W. (1975). The relationship between single word decoding and reading comprehension skill. Journal of Educational Psychology, 67, 461-469.

Read, C. (1971). Pre-school children's knowledge of English phonology. Harvard Education Review, 41, 1-34.

Rosner, J. (1975). Helping children overcome learning disabilities. New York: Walker and Co.

Rubin, H. (1984). An investigation of the development of morphological knowledge and its relationship to early spelling ability. Unpublished doctoral dissertation, University of Connecticut, Storrs.

Spache, G. D. (1972). Diagnostic Reading Scales. Monterey, CA: California Test Bureau.

Treiman, R., & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. In G. E. McKinnon & T. G. Waller (Eds.), Reading research: Advances in theory and practice (Vol. 3). New York: Academic Press.

Vellutino, F. R. (in press). Phonological coding: Phoneme segmentation training and code acquisition in poor and normal readers. In J. Kavanagh, D. Gray, & D. Pearl (Eds.), Biobehavioral measures of dyslexia. Baltimore: York Press, Inc.

Zifcak, M. (1981). Phonological awareness and reading acquisition. Contemporary Educational Psychology, 6, 117-126.

Footnotes

[1] The ways in which the individual characters in these orthographies represent their respective languages are more complex than can be described here. It should be noted that the segments represented by the characters in the Chinese logography and the Japanese kanji are more correctly defined as morphemes rather than words; similarly, the segments in the Japanese kana are better described as moras than syllables. However, it is sufficient for our purposes, and sufficiently accurate, to speak of the word in the first case and the syllable in the second.

[2] The representation in the ideal speaker-hearer's lexicon is often morphophonological, that is, the word is represented as a sequence of systematic phonemes divided into its constituent morphemes. For example, the words heal, health, healthful, have the morphophonological representations /hēl/, /hēl+θ/, /hēl+θ+ful/, respectively (see Liberman et al., 1980, for a more complete discussion of the morphophonological nature of orthographies).

ERRORS IN SHORT-TERM MEMORY FOR GOOD AND POOR READERS

Susan Brady,† Virginia Mann,†† and Richard Schmidt†††

Abstract. Good and poor readers in the second and third grades repeated 4-item lists of consonant-vowel (CV) syllables for recall in which each consonant shared 0, 1, or 2 features with other consonants in the string. While poor readers, as in previous studies, performed less accurately than good readers, the nature of their errors was the same: Both groups revealed significant effects of phonetic similarity and adjacency on the incidence of errors. These findings suggest that poor readers employ a phonetic coding strategy in short-term memory, as do good readers, though less skillfully.

Children who have difficulty learning to read have consistently been found to perform less well than good readers on a wide variety of short-term memory (STM) tasks (Bauer, 1977; Brady, Shankweiler, & Mann, 1983; Hogaboam & Perfetti, 1978; Jorm, 1983; Katz, Healy, & Shankweiler, 1983; Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Mann, Liberman, & Shankweiler, 1980; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979; Torgesen, 1982). If, as is generally agreed, text comprehension depends on the ability to preserve temporarily the phonetic form of linguistic input in STM, then a short-term memory deficit may be central to the comprehension and reading fluency problems of poor readers.

Research in the last decade suggests that the deficits poor readers display on STM tasks are related to less efficient phonetic coding processes. This conclusion is supported by three lines of evidence. First, the contrast between reading groups for performance on STM tasks seems to be restricted to procedures with "phonetically recodable" stimuli. When visual stimuli are selected that do not lend themselves to phonetic coding, the performances of good and poor readers are the same. For example, for stimuli such as photographs of strangers, nonsense doodle drawings, or symbols from an unfamiliar writing system, recall by good and poor readers is comparable (Katz, Shankweiler, & Liberman, 1981; Liberman, Mann, Shankweiler & Werfelman, 1982; Vellutino, Pruzek, Steger, & Meshoulam, 1973). Similarly, in memory for auditory stimuli such as tones that are not readily recoded phonetically, poor readers perform equally well on STM tasks (Holmes & McKeever, 1979). The difference between reading groups for recall of "phonetically recodable" stimuli and the lack of differences in performance for stimuli that are "phoneti-

---

†Also University of Rhode Island.

††Also Bryn Mawr College.

†††Also University of Connecticut.

Acknowledgment. We wish to thank Isabelle Liberman and Donald Shankweiler for suggestions and insightful criticisms of earlier versions of the paper. The research and the preparation of the manuscript were supported in part by a grant to Haskins Laboratories from the National Institute of Child Health and Human Development (HD-01994).

[HASKINS LABORATORIES: Status Report on Speech Research SR-81 (1985)]

cally unrecodable" highlight the poor readers' difficulty with the use of a phonetic code.

A second line of evidence finds that manipulations of certain phonetic properties of stimuli in an STM task generally have less effect on the performance of poor young readers than on that of the good readers (Brady et al., 1983; Liberman et al., 1977; Mann et al., 1980; Mark, Shankweiler, Liberman, & Fowler, 1977; Olson, Davidson, Kliegl, & Davies, 1984; Shankweiler et al., 1977). With stimuli in which there is a low density of phonetic confusability (i.e., nonrhyming items), good readers' recall is superior. However, if a high density of confusability is present, the performance of good readers is impaired much more than that of poor readers. It has been supposed that this interaction, also found for adults (Baddeley, 1966; Conrad, 1972), results from the fact that the good reader is better able to form a sufficient phonetic code for the maintenance of information in STM. Stimuli that introduce confusion among the phonetic representations in STM, such as strings of rhyming words, therefore tend to have a greater effect on the performance of the more skilled readers.

Third, and of primary concern to us here, is evidence that poor readers may make use of a phonetic code, and not some other coding strategy, but may do so less accurately or efficiently than do good readers (Katz, 1982). Earlier, Conrad (1971) had reported that children 6 years of age or older produced the same pattern of results on STM tasks as do adults. These children had better recall for nonrhyming sets of pictures than for rhyming sets. In contrast, children younger than 6 years old did not show a difference in recall for the nonrhyming set. This study raised the question of whether young children might initially be using some other, non-phonetic, coding strategy in short-term memory. However, more recent research with even younger children (4 yrs) has found the adult pattern, suggestive of phonetic coding (Alegria & Pignot, 1979). That is to say, in the Alegria and Pignot experiments 4-year-old children recalled nonrhyming items better than rhyming items, leading the authors to conclude that by 4 years of age children are already using a phonetic code to store and organize information in short-term memory. Thus, at the present, questions must be raised as to whether there are developmental changes in the type of code employed in short-term memory and, in extension, whether poor readers are indeed using a nonphonetic strategy.

Several findings compel us to reconsider this issue. Using a paradigm that tested recall for time periods longer than the assumed limits of STM, one investigator (Byrne & Shea, 1979) did obtain evidence that poor readers were able to use a phonetic code (albeit poorly) when forced to do so by pseudoword stimuli, but otherwise tended to favor a semantic code (though this has not been replicated [Winbury, 1984]). However, when errors made by poor readers were examined for a standard short-term memory task, there was no indication that poor readers were using a semantic strategy (Brady et al., 1983). Instead, their errors, like those of good readers, indicated that the stimuli were being processed phonetically. Out of 437 intrusion errors (items that were not in the original list) by both reading groups, only one appeared to have been a possible semantic error ("station" for "train"); the vast majority of the remainder could be accounted for in terms of the phonetic units present in the particular string and the preceding string. What was noteworthy was that both reading groups appeared to be using a phonetic coding strategy, although the incidence of recombinations (transpositions) of the phonetic information was significantly more frequent for the poor readers.

In sum, present evidence is consistent with a hypothesis that the deficit of poor readers on STM tasks has its origin in deficiencies in creating and maintaining a phonetic code. The poor readers' specific difficulty with "phonetically recodable" stimuli, reduced sensitivity to rhyme, and greater frequency of phonetic errors of transposition all support this conclusion.

Most of the studies of short-term memory for good and poor readers have assessed performance by calculating the number of items correctly reported. Analyzing the nature of errors, rather than just the incidence of errors, offers a way to determine how good and poor readers are actually functioning in STM tasks. In the last study mentioned above (Brady et al., 1983), a preliminary effort was made to do this and the results proved interesting. The occurrence of transposition errors for good and poor readers pointed to the use of a phonetic code by both reading groups rather than revealing an alternate strategy for the poor readers. Yet the higher frequency of these errors for poor readers was suggestive of reading group differences in phonetic processing skills. However, the lists in that study had not been designed to allow the rigorous analysis of errors carried out in recent studies of STM in adults (cf. Drewnowski, 1980; Ellis, 1980), so the results are tentative.

Given the importance of short-term memory for language processing and the evidence of STM deficits in poor readers, we thought it worthwhile to conduct a more analytic study of errors in STM for good readers and poor readers. Accordingly, two experiments were conducted in which the nature of the items in the memory lists was controlled. We manipulated the phonetic similarity of the initial consonants as suggested by Ellis (1980). Nonsense syllables were selected as stimuli for two reasons: 1) The small number of semantic errors in a study of children's memory (Brady et al., 1983) and the complete lack of such errors by adults (Drewnowski & Murdock, 1980) suggests that semantic information is not the critical dimension. This conclusion is also supported by the finding that recall level by good and poor readers for sentences is not influenced by whether the sentences are meaningful or anomalous (Mann et al., 1980). 2) With nonsense syllables it is easier to control the distinctive features of the stimuli. Following Ellis's design, lists of nonsense syllables were constructed such that items in the strings shared 0, 1, or 2 features of the initial consonant. This permitted us to determine the effects of the phonetic structure of the materials on the pattern of errors, including that of phonetic similarity on the incidence of transposition errors between adjacent items. Analysis of such effects as they relate to reading ability and the accuracy of recall can provide information on the use of phonetic coding by good and poor readers.

Experiment 1

Methods

Subjects. In Experiment 1, the subjects were second-grade children from two elementary schools in a suburban school district in Rhode Island. A school reading specialist, a principal, and the classroom teachers helped to pre-select the poorest readers and the best readers from the second-grade classes. In a supplementary screening procedure, the Word Attack and Word Recognition subtests of the Woodcock Reading Mastery Tests, Form A (Woodcock, 1973), and a test of receptive vocabulary, the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn 1981), were administered to the children.

Inclusion in the study was determined by the following criteria: 1) To insure valid classification as a good or poor reader, the scores on the two Woodcock subtests had to be consistent. 2) In order to restrict the range of IQ scores, only subtests with scores from 90 to 135 on the PPVT were eligible for further testing. 3) Given the evidence that STM span increases with age (cf. Dempster, 1981), subjects were selected whose ages fell within the limited range of 88 to 100 months.

Twenty-eight children satisfied the requirements for participation in the study. Based on the scores that were obtained on the reading tests, two groups were formed that were non-overlapping in reading level. The 14 children who qualified as good readers were well beyond the end of second-grade reading performance (testing was done in the spring) with a mean reading grade level of 5.0. The 14 children labeled poor readers had an average reading grade level of 2.1, and lagged considerably behind their peers. The IQ scores, as determined by the PPVT, did not differ significantly. The mean IQ score for the good readers was 114, for the poor readers 111. The reading groups also did not significantly differ in age. The good readers had a mean age of 96.1 months, the poor readers had a mean age of 96.4 months.

Materials and procedure. The materials comprise lists of four nonsense syllables presented auditorily in three practice trials and twelve test sequences. In all sequences the stimuli were the consonant-vowel (CV) syllables /Sə/, /Zə/, /Gə/, and /Kə/. In these syllables, the vowel is held constant, and the initial consonants share 0, 1, or 2 phonetic features. The four syllables can be combined into 6 possible pairs, two of which share 0 features, two which share 1 feature, and two which share 2 features (as detailed in Table 1). The trials consisted of randomizations of these four syllables in which each consonant occurred only once per list and three times at each of the serial positions 1 to 4. For the six stimulus pairs, each occurred twice (once in each order, e.g., /Sə/, /Zə/, and /Zə/, /Sə/) at each of the serial positions 1 and 2, 2 and 3, and 3 and 4.

Table 1

Experiment 1: The consonant pairs described in terms of shared distinctive features

Consonant Pairs	Distinctive Feature			Number of Shared Features
	Voicing	Place	Consonant	
SZ	−	+	+	2
GK	−	+	+	2
SK	+	−	−	1
ZG	+	−	−	1
SG		−	−	0
ZK	−	−	−	0

The practice trials and test sequences were read with a neutral intonation by a phonetically-trained male speaker and recorded on magnetic tape. The materials were presented to subjects through headphones. Within each list the syllables were spoken with a neutral prosody at the rate of one per second.

Each child was tested individually for two sessions in a small room provided by the school. The first session consisted of the screening procedure, the second the memory task. In the second session the practice trials were presented (repeated if necessary), followed by the test sequences. For each trial, subjects were instructed to repeat the items in the order they had been presented, as soon as the list ended.

Results and Discussion

In a preliminary analysis of the data, the number of correct responses was tabulated in terms of item order and serial position, as is customarily done in studies of the short-term recall in good and poor readers. The innovation was to further analyze the errors qualitatively in relation to the syllables that were adjacent to the target syllable in the test sequence, and in relation to the phonetic features of the target syllable.

Analysis of correct responses. Since the same items were presented on each trial, varying only in terms of order, an order-correct scoring procedure was adopted. A response was considered correct only if it had been assigned to the appropriate serial position. Figure 1 shows the mean number correct at each serial position for each group of subjects. Consistent with earlier studies the good readers were notably more accurate overall than the poor readers, $F(1,26) = 9.99$, $p = .004$. Both groups showed a significant effect of serial position, $F(3,78) = 36.46$, $p < .001$.

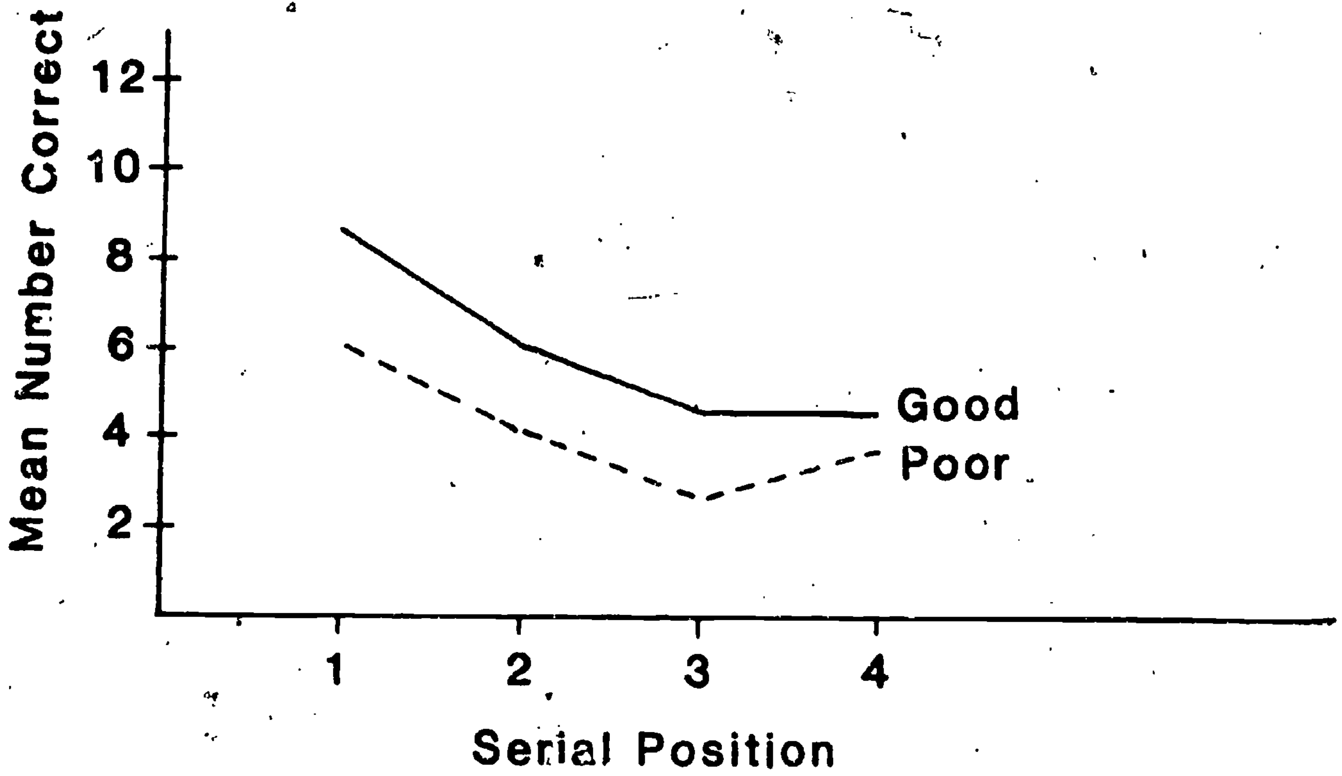


Figure 1. Experiment 1: The mean number of items correctly reported plotted by serial position.

Analyses of covariance using IQ and age as the covariates were conducted to evaluate whether the obtained differences in performance might be attributed to differences in age or intelligence between good and poor readers in our sample. Neither analysis altered the pattern of results:[1] with IQ as the covariate the good readers were still superior in recall, F(1,25) = 8.72, p = .007; likewise with age as the covariate, F(1,25) = 9.47, p = .005.

Thus, it was clear that the characteristic differences in STM recall between good and poor readers were obtained with the present task. Having replicated this pattern, we turned our attention to the nature of the errors for both reading groups.

Error analysis. The errors were first categorized as either misplacements of phonetic information that had been in the string, or as errors of omission (no response) or substitution errors (phonemes that were not in the original list). For both good and poor readers, the majority of errors were the first category, phonemes that had occurred elsewhere in the string, F(1,26) = 56.01, $p < .001$. Since the vowel was the same for all items, it is not possible to determine whether these order errors are phoneme transpositions or entire syllable order errors.

However, the construction of the present experiment, using phonemes that differ systematically in shared features, does allow us to examine the conditions under which these order errors occurred. Two parameters were measured. First, an error was evaluated as having been, in the original string, adjacent to the target or nonadjacent (e.g., target string "Gə, Zə, Kə, Sə," reported string "Gə, Sə, Kə, Zə": original location of misreported items was nonadjacent to error position). Second, an error was scored in terms of the number of features shared between the substitute response and the target item (e.g., error /G/, target /K/: two features in common).

As shown in Figure 2, the source (i.e., original location) of substituted stimuli was a significant factor both for good and poor readers. Errors were significantly more likely to involve a syllable adjacent to the target than a nonadjacent syllable, F(1,26) = 11.44, p = .002. This suggests that the subjects had retained some information about the relative position of items in the original string even when they were unable to make a fully accurate report.

The second qualitative scoring procedure, which evaluated the phonetic similarity between errors and target stimuli, is more central to our interest in the coding skills of good and poor readers. Would the incidence of order errors be greater for stimuli that shared two features than for those with one or none in common? A significant effect of phonetic similarity was obtained, F(2,52) = 7.93, $p < .001$, underscoring the phonetic basis for storage in STM. Yet the effect did not differ for good and poor readers, suggesting that both reading groups were relying on phonetic representation.

In Figure 3, the effects of feature similarity on the occurrence of errors is plotted for the order errors involving both adjacent and nonadjacent target stimuli. Here it can be seen that there is a tendency for nonadjacent errors to be more influenced by phonetic similarity than for adjacent errors (adjacency x similarity: F(2,52) = 2.72, p = .075). Another way of viewing this tendency is in terms of the distinction between memory for item identity and memory for order (Healy, 1975). In this analysis, adjacent errors are

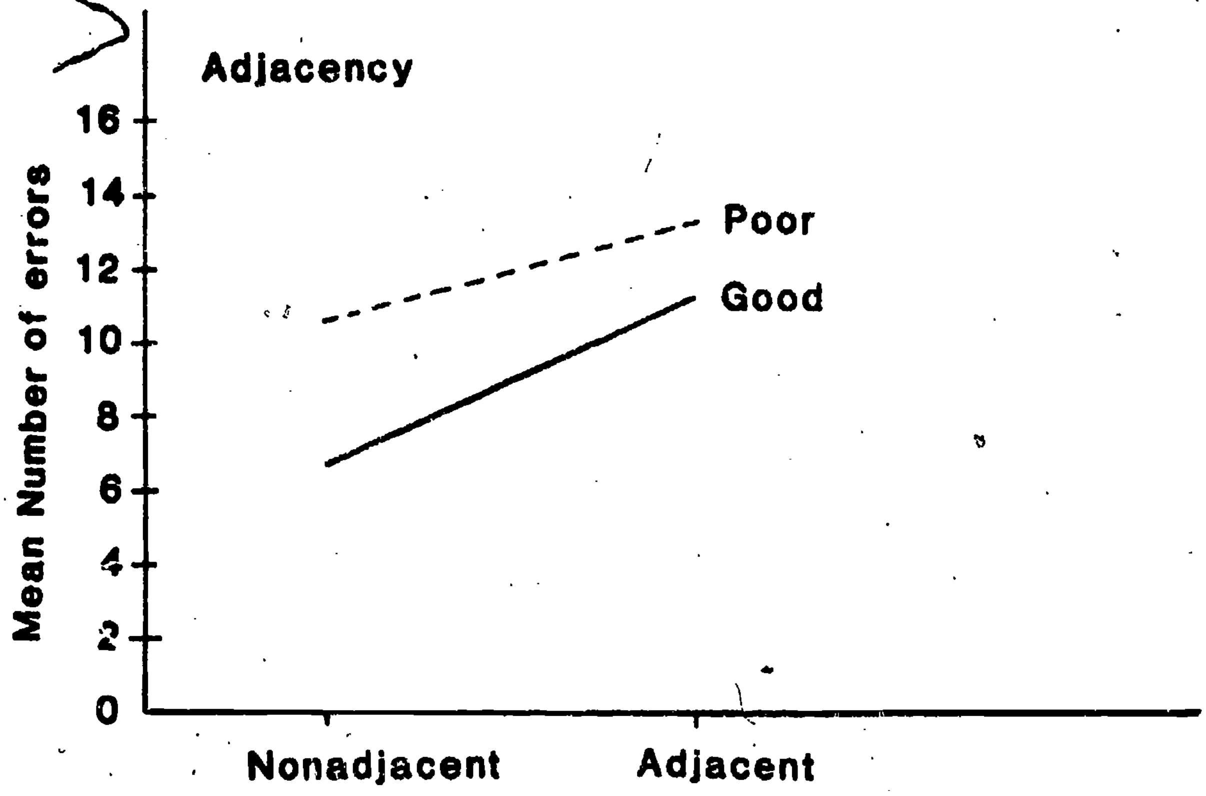


Figure 2. Experiment 1: The mean number of times the error consisted of an item from a nonadjacent or an adjacent position.

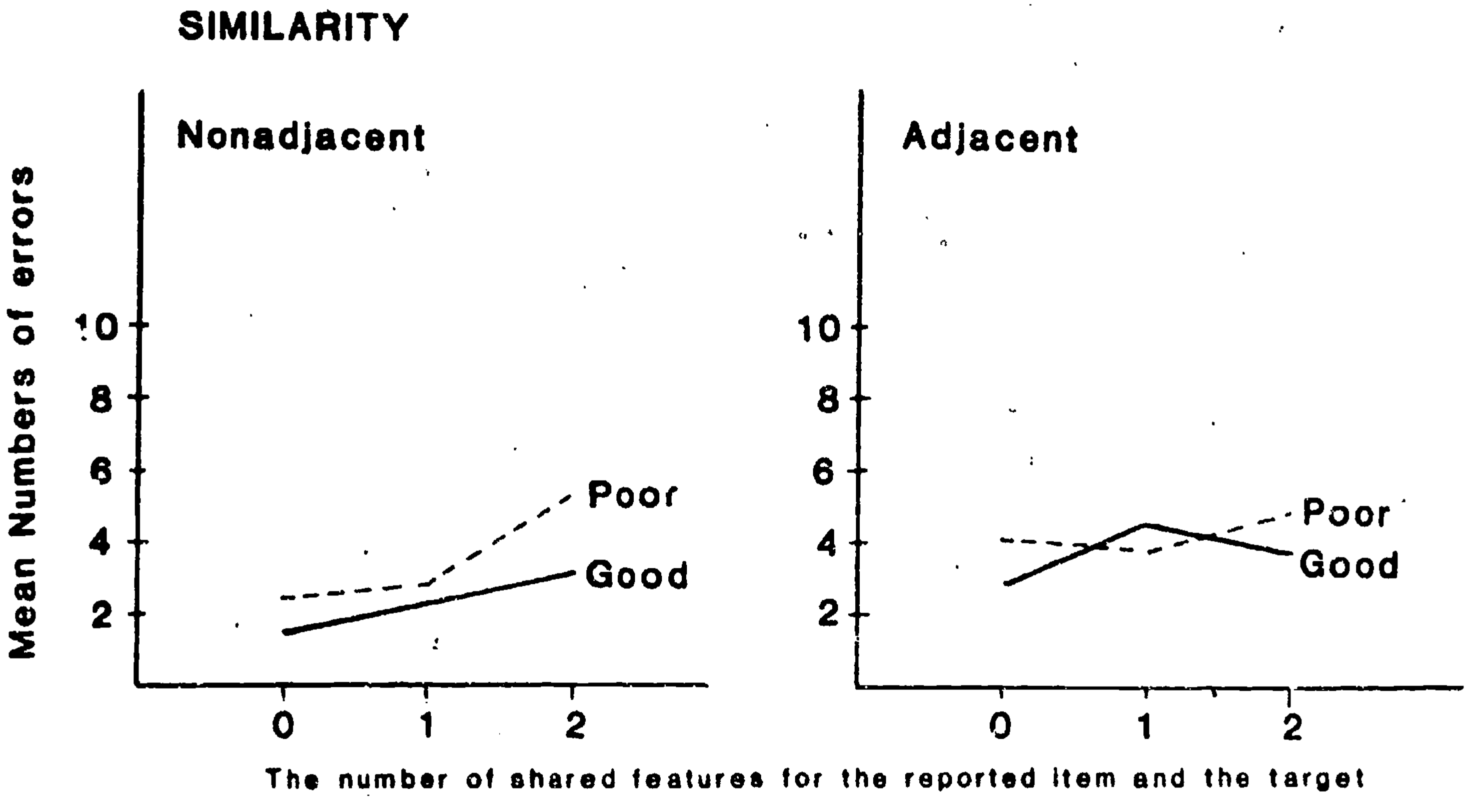


Figure 3. Experiment 1: Analysis of feature similarity effects for adjacent and nonadjacent types of errors. The number of shared features for the reported item and the target.

more likely to be strictly order confusions, while more distant mistakes, which are subject to effects of phonetic similarity, are more likely to be "item" confusions reflecting poor retention of information. This pattern was obtained equally for good and poor readers as indicated by the lack of an interaction between reading group, similarity, and adjacency ($p > .5$).

In sum, in this experiment poor readers were found to recall significantly less information than did the good readers. However, like good readers, their errors consisted largely of information that had been in the string, reported in the wrong order, rather than errors of omissions or substitutions. Further analyses showed that these order errors revealed significant effects of adjacency and of phonetic similarity. Thus the errors of poor readers show the same systematic effects of processing as do the errors of good readers, but occur at a higher rate. The implication of this would seem to be that poor readers employ the same coding strategy as do good readers but less effectively.

Experiment 2

To determine the generality of the results to other consonants and other classes of phonemes, good and poor readers were tested on a second set of items consisting of syllables that started with the consonants /Mə/, /Nə/, /Bə/, and /Kə/. This task was designed to allow a further investigation of the phonetic factors in STM. In addition to a same-vowel condition, a mixed-vowel set was employed to further explore the nature of order errors. In this condition, it is possible to analyze whether order errors consist of transpositions of phonetic segments (e.g., /Mɪ/, /Næ/ for /Nɪ/, /Mæ/) or of syllable misorderings (e.g., /Mæ/, /Nɪ/ for /Nɪ/, /Mæ/. In this way we hoped to make a more fine-grained analysis of the processing strategies of good and poor readers.

Subjects

The same subjects were recruited for participation in Experiment 2, conducted in the spring of the following school year (3rd grade). Four children were no longer available, two good readers and two poor readers. The remaining children were reevaluated for inclusion in the study, in accord with the criteria outlined for Experiment 1.[2] For all subjects, placement in a reading group was the same as it had been the previous year. Additional children were screened to increase the number of subjects in each group. Two children qualified as good readers and four as poor readers bringing the group sizes to 14 good readers and 16 poor readers.

As before, the reading groups were non-overlapping in reading level. The 14 good readers had a mean reading grade level of 8.7. The 16 children who were labeled poor readers had a mean reading grade level of 2.9. The PPVT IQ scores did not differ significantly for the reading groups: good readers, $\bar{x} = 113$; poor readers, $\bar{x} = 111$. Nor did the ages significantly differ: good readers, $\bar{x} = 107.2$ mos.; poor readers, $\bar{x} = 108.3$.

Materials and Procedure

Experiment 2 was designed to be exactly parallel to Experiment 1 with different stimulus sets. A trial again consisted of four nonsense syllables presented auditorily, and the subjects were asked to repeat the list in the order of presentation. There were now two conditions, a same-vowel set and a mixed-vowel set. The construction of the test sequences was identical to Experiment 1. The initial consonants of the test items in each set again had 0, 1, or 2 phonological features in common with the other stimuli selected (see Table 2). In the same-vowel condition the stimuli were /Mə/, /Nə/, /Bə/, and /Kə/. In the non-rhyme set, the consonants were randomly paired with the vowels /ɪ/, /ɛ/, /ɔ/, /ə/, /ɑ/, /əɪ/, /ɔɪ/, /eɪ/, /æ/, and /u/ (with the stipulation that CV combinations sounding like real words were excluded).

Table 2

Experiment 2: The consonant pairs described in terms of shared distinctive features

Consonant Pairs	Voicing	Place	Consonant	Number of Shared Features
	Distinctive Feature			
MN	+	−	+	2
MB	+	+	−	2
NB	+	−	−	1
BK	−	−	+	1
MK	−	−	−	0
NK	−	−	−	0

The preparation of stimulus tapes and the method of testing also mirrored the procedures adopted in Experiment 1, except that each child was tested for three sessions (one screening session and two memory-task sessions). In each memory-task session a subject would have three practice trials (repeated if necessary) and six test trials for one condition (e.g., same vowel), followed by three practice trials (or more) and six test trials for the other condition (e.g., mixed vowel). The order of test conditions was reversed for the second session. Within each reading group, half of the subjects began with the same vowel condition, half with the mixed vowel set.

Results and Discussion

The scoring methods used in Experiment 1 were employed. The results will be presented jointly for the two vowel conditions. First, the correct response data will be discussed, followed by the error analysis results.

Analysis of correct responses. In Figure 4 the mean number correct at each serial position is plotted for the two vowel sets, same vowel and mixed vowel. In these plots, the data have been analyzed only for consonant information. Thus, in this analysis the vowel responses were not scored. Except for recall being a little better for all subjects, as expected for older children, the results replicate those in Experiment 1. Good readers were superior to poor readers in recall for both the same vowel set, *F*(1,28) = 14.25, *p* < .001, and the mixed vowel set, *F*(1,28) = 5.64, *p* = .05. In addition, the serial position effect was significant for both vowel conditions: same vowel, *F*(3,26) = 19.52, *p* < .001; mixed vowel, *F*(3,26) = 18.63, *p* < .001.

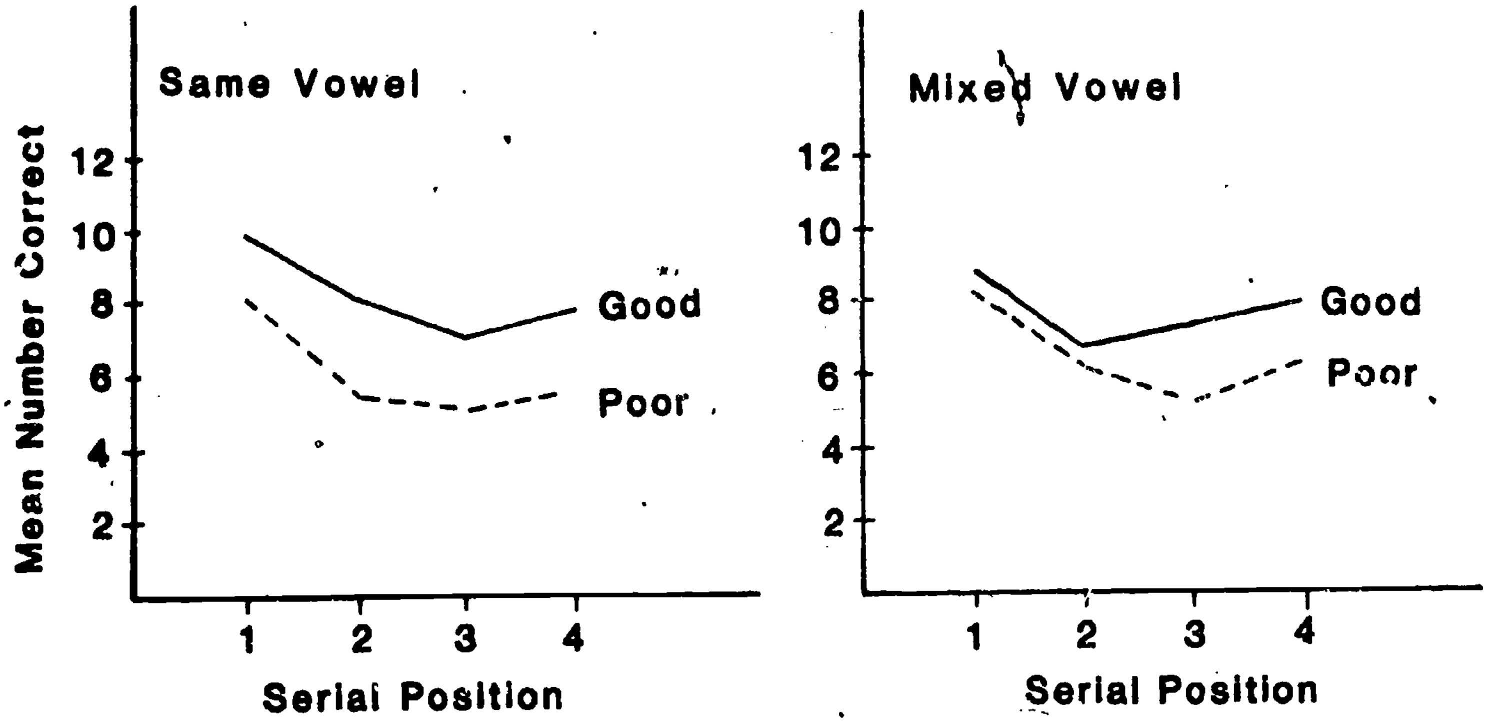


Figure 4. Experiment 2: The mean number of consonants correctly reported plotted by serial position.

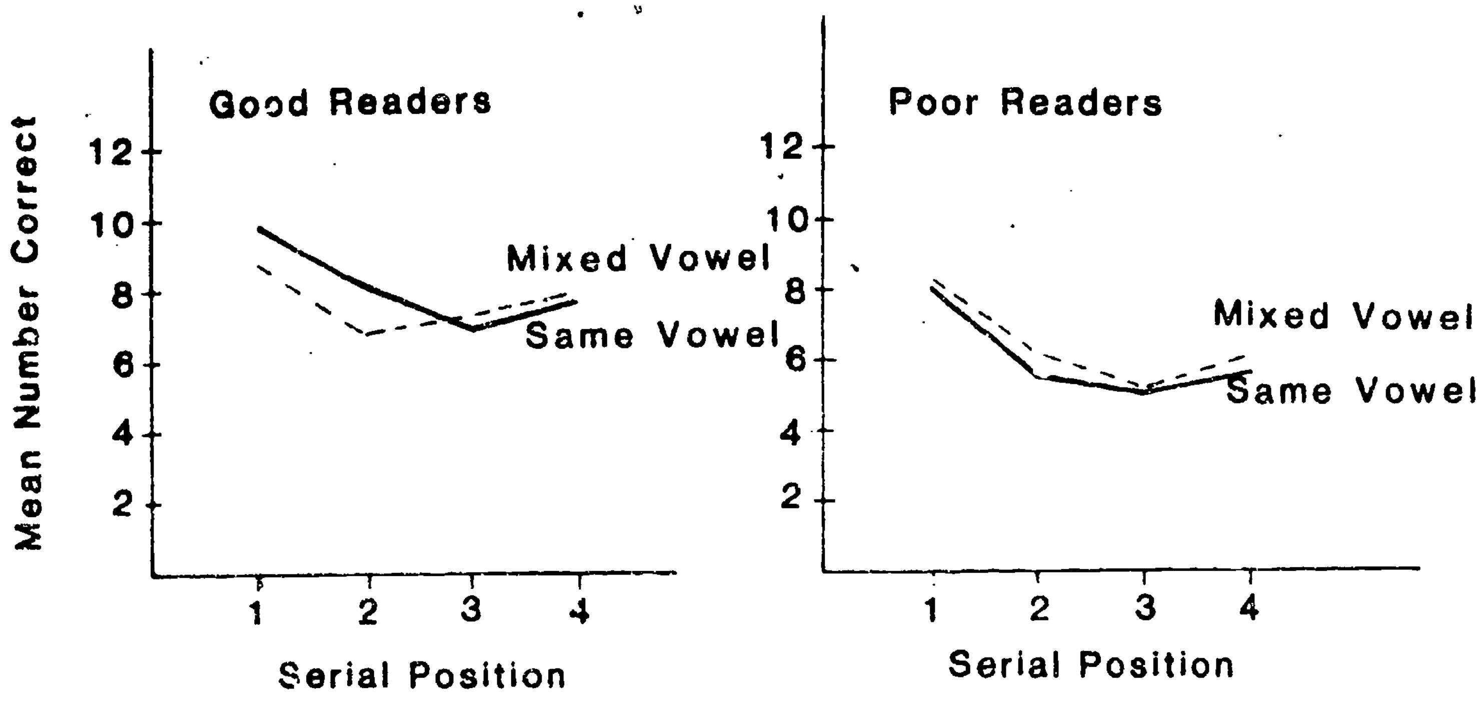


Figure 5. Experiment 2: The mean number of consonants correctly reported replotted for each reading group.

Interestingly, recall of consonants appears to be independent of the vowel environment. There was a striking lack of difference in error rate for consonants for the two vowel sets, $F(1,38) = .00$, $p > .99$, as shown in Figure 5. That this held for both reading groups is supported by the lack of a group x test interaction, $F(1,28) = 2.43$, $p = .13$.

In most memory studies, the consonants are not scored in isolation, but rather the entire response is scored as correct or not. We will now present the data in this fashion, counting a response as correct only if both the consonant and vowel were accurately reported. Scoring the entire syllable, a substantial difference in accuracy emerges for the memory sets, $F(1,28) = 146.49$, $p < .001$. This can be seen in Figure 6. Of course, the error rate for the same-vowel set is the same as plotted for the consonant scoring, since subjects do not make mistakes on the vowels in this task. This is not so for the mixed vowel set where the amount of information to be recalled is much greater. Subjects must retain both the consonant and the vowel, and with nonsense syllables this cannot be facilitated by semantic information. The increased memory load is reflected in the lower accuracy for the mixed vowel condition.

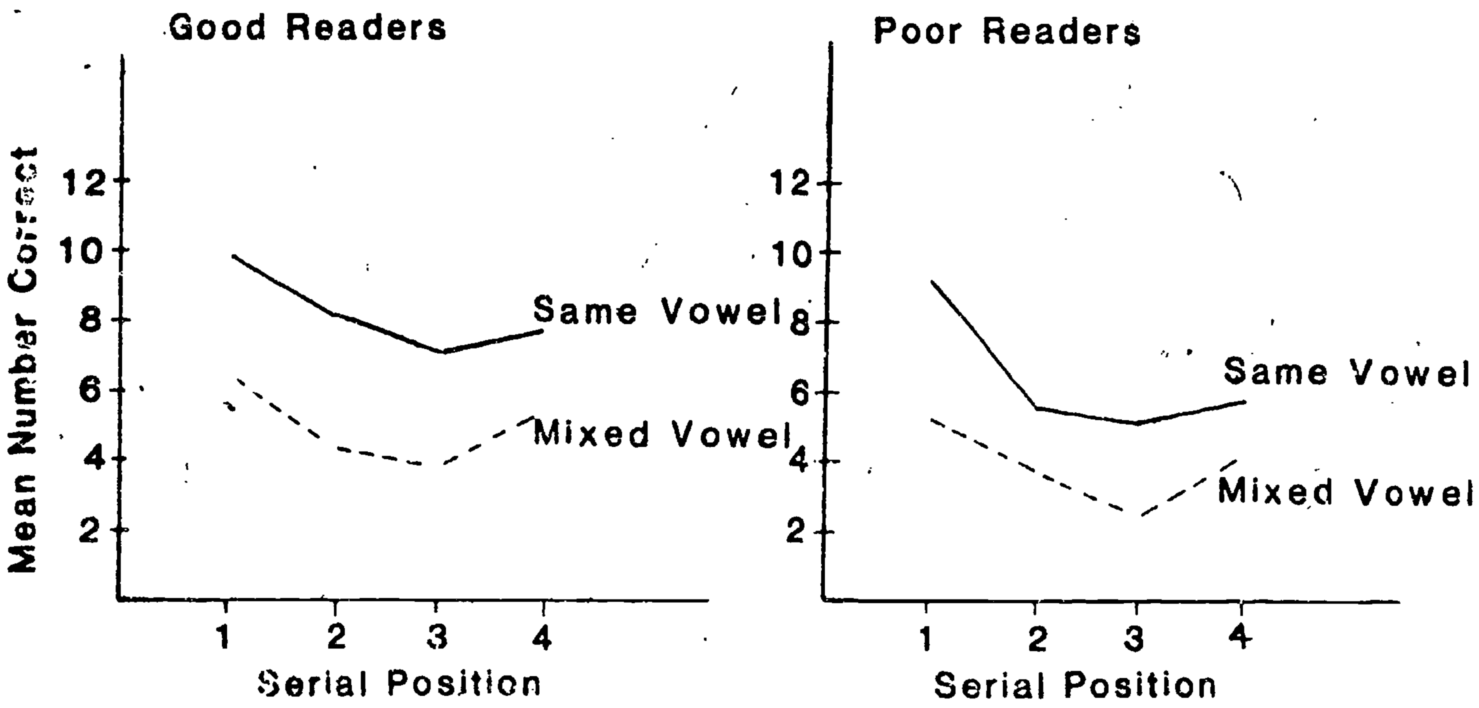


Figure 6. Experiment 2: The mean number of correct responses (total syllable) reported plotted for each reading group.

Good readers show a greater improvement on the easier memory task (same vowel set) than do the poor readers, $F(1,28) = 8.59$, $p = .007$. The particular pattern of errors by good and poor readers for vowels and consonants will be discussed below in the error analysis section. Further, the results of the same vowel and mixed vowel sets in relation to other studies investigating the effects of rhyme on recall will be addressed in the conclusion.

In sum, the results of the correct response analyses show that good readers have performed significantly better on both conditions (same and mixed-vowel) and for both scoring techniques (consonant alone and whole syllable).

Error analysis. Turning to analyses of the types of errors, we found, as in Experiment 1, that the majority of errors consisted of misorderings of the items in the string, rather than substitutions of new items or omissions. This effect was significant both for the same vowel condition, $F(1,28) = 135.11$, $p < .001$, and the mixed vowel condition, $F(1,28) = 196.08$, $p < .001$.

However, in the mixed vowel set, in which we can examine vowel errors, the errors for consonants and vowels differed somewhat. As noted, the reading groups differed significantly on the consonant errors, $F(1,28) = 5.64$, $p = .025$, but the difference was not obtained for vowel errors, $F(1,28) = 2.51$, $p = .124$. Although both groups produced many errors on the vowels, error rate did not distinguish the groups.

Table 3 displays the ways in which the consonants and vowels vary as to error type. First, as stated earlier, it can be seen that for the consonants, very few errors consisted of substitutions or omissions. With such a limited data set, this is as would be expected. For the vowels a larger set of stimuli was possible in any particular string, and a fair number of substitution errors occurred. Second, few of the errors for either reading group consisted of entire syllable misorderings, and no group difference was observed for this error type, $F(1,28) = 1.90$, $p = .179$. Third, the majority of errors consist of transpositions of the available consonants and vowels, which create new syllables. The significant difference between good and poor readers arises from the greater frequency of transposition errors for the poor readers.

Table 3

Nonrhyme condition, Analysis of Errors

Consonant Errors

Reading Group	$\bar{x}$ Total Errors for Consonants	$\bar{x}$ Errors: Whole Syllable in Wrong Position	$\bar{x}$ Errors: Transposition (Consonant Wrong Position With New Vowel)	$\bar{x}$ Errors: Omission and Substitution
Good	16.93	2.07	13.07	1.79
Poor	22.00	3.44	17.06	1.50

Vowel Errors

Reading Group	$\bar{x}$ Total Errors for Vowels	$\bar{x}$ Errors: Whole Syllable in Wrong Position	$\bar{x}$ Errors: Transposition (Vowel Wrong Position With New Consonant)	$\bar{x}$ Errors: Omission and Substitution
Good	15.36	2.07	6.00	7.29
Poor	19.13	3.44	8.50	7.19

Let us next look, as in Experiment 1, at the effects of adjacency and phonological similarity on the occurrence of consonant errors. It is clear from Figure 7 that there is a pronounced effect of adjacency on errors in both vowel conditions. For both good and poor readers, transpositions more often involved consonants from adjacent syllables than from nonadjacent stimuli (same vowel set, $F(1,28) = 104.72$, $p < .001$; mixed vowel set, $F(1,28) = 41.71$, $p < .001$). Further, there were also significant effects of phonetic feature similarity on the incidence of transposition errors (same vowel set, $F(2,27) = 6.61$, $p = .005$; mixed vowel set, $F(2,27) = 5.33$, $p = .011$). Order errors were thus more likely to occur between stimuli that shared phonetic information. In this experiment, in contrast to the first experiment, the effects of phonetic similarity were evident both for the adjacent errors and for the nonadjacent errors. Therefore, in Figure 8, the data for similarity effects are combined for these two error types. The reader will recall that the effects of similarity had been more pronounced in the first experiment for the nonadjacent errors. It is not clear whether this difference between Experiment 1 and Experiment 2 may indicate a developmental increase in sensitivity to phonological factors relative to adjacency effects, or whether it may be related to the particular stimuli used in these memory tasks.

In any case, to summarize, although the poor readers made more errors than good readers, the majority of errors for both consisted of reorderings of the phonetic information in the string. These recombinations showed strong influences of adjacency and phonetic similarity reflecting, we presume, the underlying processing strategies. The kinds of errors suggest that the inferior performance of poor readers on these short-term memory tasks is not the consequence of a different coding strategy, but rather of a lesser degree of skill with a phonetic strategy.

Conclusion

Our goal was to conduct studies that would allow us to determine the coding processes of good readers and poor readers on short-term memory tasks. Previous work (Brady et al., 1983) had provided preliminary evidence that poor readers, like good readers, use a phonetic code in STM. In the present experiments this question was directly evaluated using nonsense strings in which the phonetic similarity of the items was controlled.

Good and poor readers from the second grade (Experiment 1) and from the third grade (Experiment 2) were tested in experiments with differing sets of consonant/vowel syllables. The results of the two experiments were congruent: The majority of consonant errors consisted of transpositions of one item from the sequence for another, with significant effects of phonetic similarity. This pattern held for both good readers and poor readers. The two groups differed, however, in the occurrence of consonant transposition errors, which were significantly more frequent for the children with reading problems. Both reading groups also showed significant effects of adjacency: order confusions for all stimuli were more likely to occur between adjacent items than between nonadjacent items. This pattern was produced by good readers and poor readers and probably reflects the demands of a serial order task.

In the second experiment a mixed-vowel condition was added that allowed an examination of vowel errors. While both good and poor readers produced a fair number of errors for vowels, the error rate did not distinguish the reading groups.

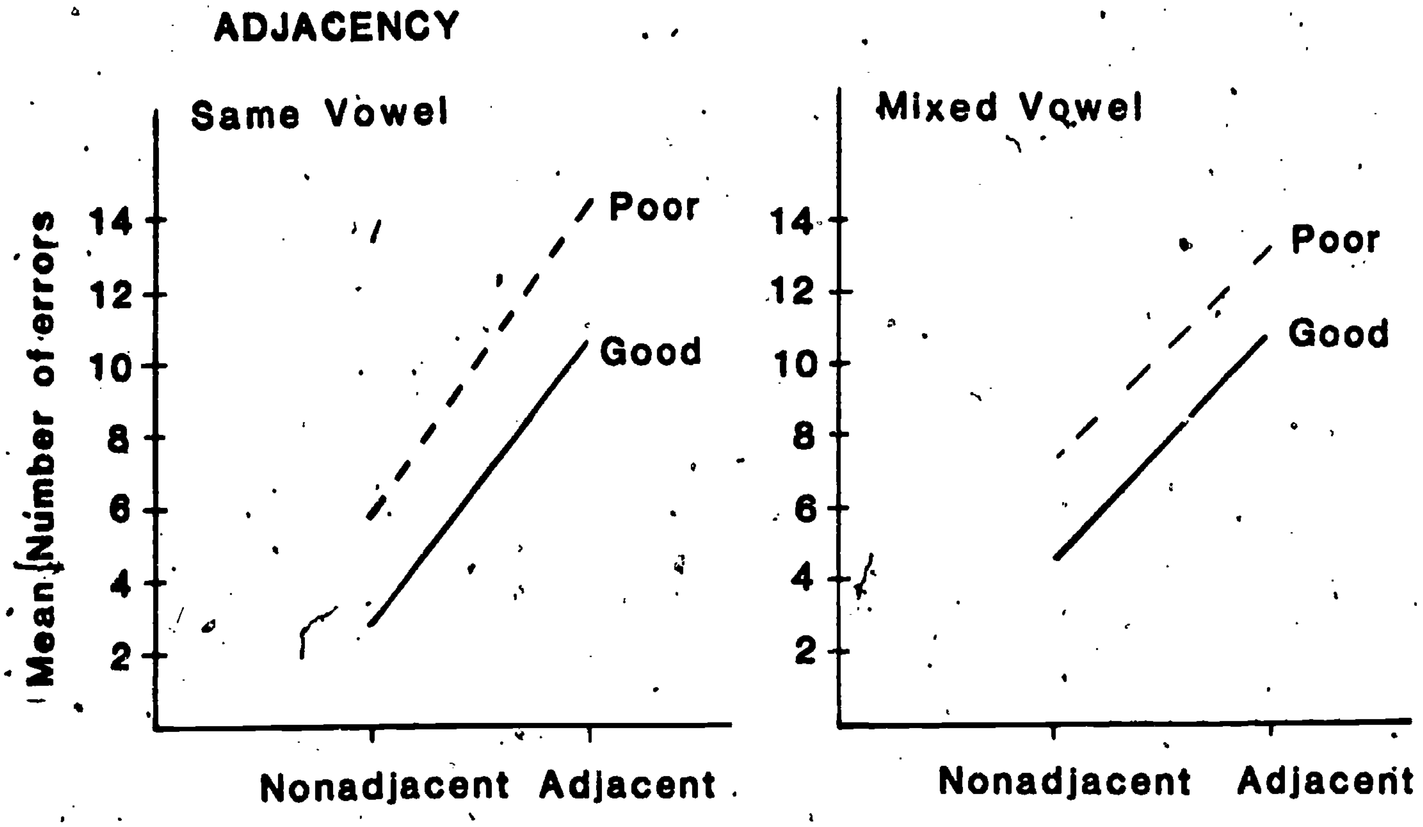


Figure 7. Experiment 2: The mean number of times the error consisted of an item from a nonadjacent or an adjacent position.

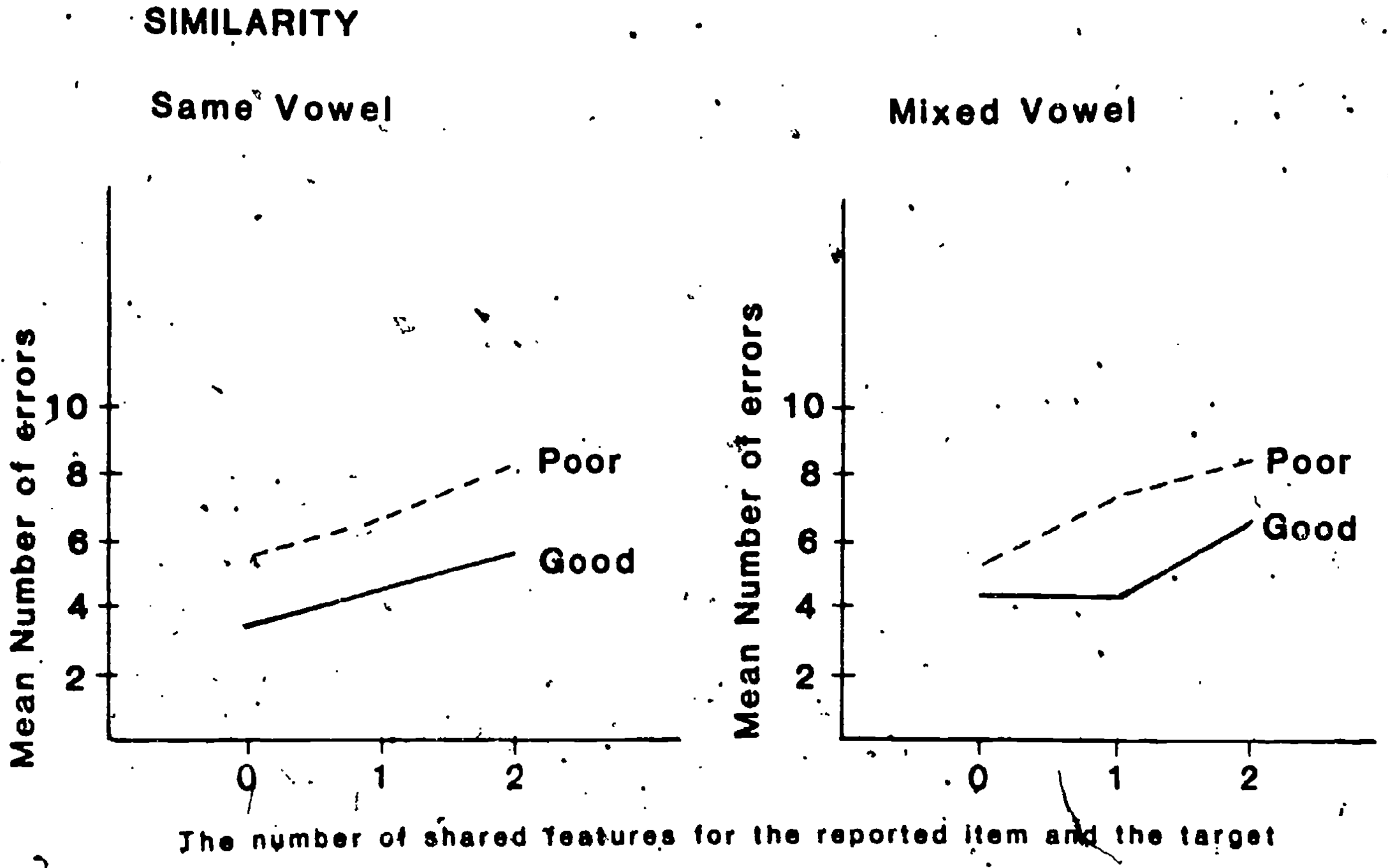


Figure 8. Experiment 2: Analysis of feature similarity effects.

A comment is in order about the relative difficulty of the same vowel (rhyming) and mixed vowel (nonrhyming) sets. Our results are seemingly at odds with what is generally found. As mentioned in the introduction, adult subjects and good readers usually have better recall for non-rhyming items than for lists of rhyming stimuli (e.g, Baddeley, 1966; Shankweiler et al., 1979). The greater confusion in recall when the items are phonetically similar has been taken as reflecting the coding processes in STM. However, these findings have been obtained for longer sequences than were used in the present experiment. The effects of rhyme may well depend on how taxed the system is (a similar idea has been expressed by Hall, Wilson, Humphreys, Tinzmann, & Bowyer, 1983). The everyday experience of rhyme facilitating recall may be related to this. In the present task, sequences were deliberately kept short so as to optimize the subject's ability to recall the correct number of stimuli. We thought this would facilitate the examination of order errors and transposition errors in the reported items. For longer strings subjects increasingly fail to recall the entire sequence. In that case, partial report by the subject (e.g., giving 4 out of 7 stimuli) permits a less structured analysis of errors.

That rhyme effects are tied to task factors has also been suggested by results obtained for adults (see Watkins, Watkins, & Crowder, 1974). In a paradigm similar to the present one using short strings of nonsense syllables, Ellis (1980) did not find significant differences in the error rates for an all-same vowel condition and for an all-different vowel condition (though with other conditions there was a significant effect of vowel environment on error rate). Thus, the particular type of STM task used in the present experiment has not been found to produce the standard rhyme effect with normal adults.

Analyses of previous STM studies with children varying in reading ability shows that the levels of recall on STM tasks have consistently distinguished reading groups. However, the effects of rhyme have proved to be somewhat labile. As for adults, task factors appear to influence the relative difficulty of rhyming strings. Hanson, Liberman, and Shankweiler (1984) also found repeated rhyming strings to be easier for subjects. In their task they also used short sequences (4 items) with the same stimuli repeated on each trial in varying order. In addition, the "rhyme effect" appears to be sensitive to subject characteristics (Hall et al., 1983) and to age effects (Olson, Davidson, Kliegl, & Davies, 1984). It is evident that additional work is necessary to understand the basis of the traditional rhyme effect in STM.

In the present study, we are asking a different question: can we find out whether good and poor readers employ different strategies in STM? What is critical is not so much the direction of the effect of rhyme but whether poor readers are susceptible to effects of phonetic similarity. To summarize our findings on that question, error analyses revealed that both good and poor readers show effects of phonetic similarity from which we infer that both groups use a phonetic coding strategy. The inferior performance of the poor readers arose from a higher incidence of errors involving transpositions of consonants sharing phonetic features. The shared features can be presumed to place demands on phonetic coding skills, thus these results suggest that the STM deficit associated with poor readers is related in some degree to greater difficulty in establishing or maintaining a phonetic code.

The factors of phonetic similarity and adjacency have been noted as important aspects of STM processing for adults as well. Several authors have noted that errors in recall by adults show effects of feature similarity (Cole, Haber, & Sales, 1968; Cole, Sales, & Haber, 1969; Hintzman, 1967; Wickelgren, 1966). Further, Ellis (1980) specifically documented effects of phonetic parameters on the occurrence of transposition errors by adults on tasks similar to those in the present experiments. Hitch (1974) and Ryan (1969) also reported a strong effect of adjacency in the recall errors of adults.

Thus the same processing strategies appear to be at work for adults, children who are good readers, and children who are poor readers. Yet a difference exists between groups in recall level, with adults performing better than children, and good readers recalling more than poor readers. We are suggesting that the STM deficit of poor readers is related to less efficient processing of phonetic information by those children, perhaps reflecting a maturational lag (Mann & Liberman, 1984; Satz & Sparrow, 1970). In a developmental study, Olson et al. (1984) reported the same improvements in recall and sensitivity to phonetic factors in poor readers as in good readers, but at an older age.

Case, Kurland, and Goldberg (1982) have offered an account of why younger children in general perform less well that may also apply to poor readers. These authors argue that as a child gets older, basic encoding and retrieval operations in STM become more efficient, resulting in more functional storage space (and in a concomitant increase in short-term memory capacity). In support of this interpretation, they report a linear relationship between increases in memory span and increases in speed of word repetition for normal children 3 to 6 years old. Their position is buttressed by an additional experiment in which adults were forced to count in an unfamiliar language. The speed of counting for the adults was now equal to the rate of six-year-olds, and their memory span correspondingly dropped.

It may be worth noting that individual differences in memory span are found throughout the lifespan. Furthermore, there are some indications that phonological skills also vary for adults and that these two findings may be related. For example, Baddeley, Thomson, and Buchanan (1975) found that adults' memory span could be predicted on the basis of the number of words that the subject could read in approximately 2 seconds.

To summarize, the present work indicated that children who are poor readers are using the same phonetic processes in STM as are good readers or adults, but less efficiently. We would like to explore further the relationship of phonetic processing skills to short-term memory to determine whether that relationship can account for the developmental changes in short-term memory that have been observed and for the memory differences for children differing in reading ability.

References

Alegria, J., & Pignot, E. (1979). Genetic aspects of verbal mediation in memory. Child Development, 50, 235-238.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. Quarterly Journal of Experimental Psychology, 18, 362-365.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575-589.

Bauer, R. (1977). Memory processes in children with learning disabilities. *Journal of Experimental Child Psychology*, *24*, 415-430.

Brady, S., Shankweiler, D., & Mann, V. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology*, *35*, 345-367.

Byrne, B., & Shea, P. (1979). Semantic and phonetic memory in beginning readers. *Memory & Cognition*, *7*, 333-338.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*, 386-404.

Cole, R. A., Haber, R. N., & Sales, B. D. (1968). Mechanisms of aural encoding: I. Distinctive features for consonants. *Perception & Psychophysics*, *3*, 281-284.

Cole, R. A., Sales, B. D., & Haber, R. N. (1969). Mechanisms of aural encoding: II. The role of distinctive features in articulation and rehearsal. *Perception & Psychophysics*, *6*, 343-348.

Conrad, R. (1971). The chronology of the development of covert speech in children. *Developmental Psychology*, *5*, 398-405.

Conrad, R. (1972). Speech and reading. In J. Kavanagh & I. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin*, *89*, 63-100.

Drewnowski, A. (1980). Memory functions for vowels and consonants: A reinterpretation of acoustic similarity effects. *Journal of Verbal Learning and Verbal Behavior*, *19*, 176-193.

Drewnowski, A., & Murdock, B. B. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 319-332.

Dunn, L. M. (1981). *Peabody Picture Vocabulary Test--Revised*. Circle Pines, MN: American Guidance Service, Inc.

Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior*, *19*, 624-634.

Hall, J. W., Wilson, K. P., Humphreys, M. S., Tinzmann, M. B., & Bowyer, P. M. (1983). Phonemic-similarity effects in good vs. poor readers. *Memory & Cognition*, *51*, 520-527.

Hanson, V. L., Liberman, I. Y., & Shankweiler, D. (1984). Linguistic coding by deaf children in relation to beginning reading success. *Journal of Experimental Child Psychology*, *37*, 378-393.

Healy, A. F. (1975). Short-term retention of temporal and spatial order: *Bulletin of the Psychonomic Society*, *5*, 57-58.

Hintzman, D. L. (1967). Articulatory coding in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *6*, 312-316.

Hitch, G. H. (1974). Short-term memory for spatial and temporal information. *Quarterly Journal of Experimental Psychology*, *26*, 503-513.

Hogaboam, T., & Perfetti, C. A. (1978). Reading skill and the role of verbal experience in decoding. *Journal of Educational Psychology*, *20*, 717-729.

Holmes, D. R., & McKeever, W. F. (1979). Material specific serial memory deficit in adolescent dyslexics. *Cortex*, *15*, 51-62.

Jorm, A. R. (1983). The cognitive and neurological basis of developmental dyslexia: A theoretical framework and review. *Cognition*, *7*, 19-33.

Katz, R. B. (1982). *Phonological deficiencies in children with reading disability: Evidence from an object-naming task.* Doctoral dissertation, University of Connecticut.

Katz, R. B., Healy, A. F., & Shankweiler, D. (1983). Phonetic coding and order memory in relation to reading proficiency: A comparison of short-term memory for temporal and spatial order information. *Applied Psycholinguistics*, *4*, 229-250.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology*, *32*, 474-484.

Liberman, I. Y., Mann, V. A., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and non-linguistic material in relation to reading ability. *Cortex*, *18*, 367-375.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading: The proceedings of the CUNY Conferences* (pp. 207-225). Hillsdale, NJ: Erlbaum.

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory: Can they presage early reading problems? *Journal of Learning Disabilities*, *17*, 592-599.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. *Memory & Cognition*, *8*, 329-335.

Mark, L. S., Shankweiler, D., Liberman, I. Y., & Fowler, C. A. (1977). Phonetic recoding and reading difficulty in beginning readers. *Memory & Cognition*, *5*, 623-629.

Olson, R. K., Davidson, B. J., Kliegl, R., & Davies, S. E. (1984). Development of phonetic memory in disabled and normal readers. *Journal of Experimental Psychology*, *37*, 187-206.

Ryan, J. (1969). Grouping and short-term memory: Different means and patterns of grouping. *Quarterly Journal of Experimental Psychology*, *21*, 137-147.

Satz, P., & Sparrow, S. (1970). Specific developmental dyslexia: A theoretical framework. In D. J. Bakker & P. Satz (Eds.), *Specific reading disability: Advances in theory and method.* Rotterdam: Rotterdam University Press.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 531-545.

Torgesen, J. K. (1982). The study of short-term memory in learning-disabled children: Goals, methods and conclusions. In K. Gadow & I. Bialov (Eds.), *Advances in learning and behavioral disabilities* (Vol. I). Greenwich, CT: JAI Press.

Vellutino, F., Pruzek, R., Steger, J., & Meshoulam, U. (1973). Immediate visual recall in poor and normal readers as a function of orthographic-linguistic familiarity. *Cortex*, *8*, 106-118.

Watkins, M. J., Watkins, O. C., & Crowder, R. G. (1974). The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning and Verbal Behavior*, *13*, 430-447.

Wickelgren, W. A. (1966). Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America*, *39*, 388-398.

Winbury, N. (1984). *A developmental investigation of the use of phonetic and semantic memory codes in good and poor readers.* Unpublished doctoral dissertation, University of Connecticut.

Woodcock, R. W. (1973). *Woodcock Reading Mastery Tests*. Circle Pines, MN: American Guidance Services, Inc.

Footnotes

[1] In the remainder of the analyses for experiments 1 and 2, the data were likewise reanalyzed controlling for age and IQ. In no case was the significance of the differences between reading groups reduced when age and IQ were controlled.

[2] The criteria for inclusion were adjusted for children a year older. Subjects were selected whose ages fell within the range of 100 to 112 months.

LONGITUDINAL PREDICTION AND PREVENTION OF EARLY READING DIFFICULTY*

Virginia A. Mann†

Abstract. The results of many studies suggest that early reading problems are associated with deficiencies in certain spoken language skills. Children who encounter reading difficulty tend to be less able than matched good readers to perceive spoken words under noisy conditions, less able to retain linguistic material in temporary memory, less able to comprehend certain spoken sentences accurately, and less fully aware of the phonological structure of spoken words. This paper summarizes these findings, and places them in the context of the requirements of skilled reading. The results of two longitudinal studies are reviewed, which show that inferior performance in kindergarten tests of language skills may presage future reading problems in the first grade. Based on these studies, procedures are suggested for kindergarten screening and for some ways of aiding children who, by virtue of inferior performance on the screening tests, might be considered at risk for early reading difficulties.

The focus of this paper is the prediction and prevention of a specific, and quite prevalent, form of learning disability: early reading difficulty. The contention, which will be evidence from a variety of experiments, is that deficiencies in certain spoken language skills often limits the attainment of beginning reading skills. From the assumption that skilled reading involves decoding a written representation of one's spoken language, it follows that linguistic skills should be among the critical prerequisites for successfully learning to read. The view that reading skill is derived from primary language skill is evident from a consideration of what skilled reading is all about. It is also supported by the findings of experimental investigations of factors related to reading ability in adults. It provides the theoretical perspective that has guided investigations by the reading research group at Haskins Laboratories. Finally, it has led to the findings presented below that certain language deficiencies in kindergarteners are prognostic of early reading problems.

---

*Annals of Dyslexia, 1984, 34, 117-136.

†Also Bryn Mawr College.

Acknowledgment. This paper was presented at the 34th Annual Conference of the Orton Dyslexia Society, November, 1983, San Diego, California. It was prepared while the author was a Fulbright Fellow at the Research Institute of Logopedics and Phoniatrics at the University of Tokyo, Tokyo, Japan. Much of the research herein described was supported by NICHD Grant HD-01994 and BRS Grant 05596 to Haskins Laboratories. I would like to acknowledge the helpful cooperation of the children and teachers at the Bolles School in Jacksonville, FL, and to express my gratitude to Donald Shankweiler for his comments on an earlier draft, as well as to Isabelle Liberman for insightful conversations.

In particular, two linguistic factors are consistently associated with reading ability in beginning readers (see Mann & Liberman, 1984, or Mann, in press). These include children's degree of sophistication about the phonological structure of language, and their ability to process spoken language fully. I will now turn to the task of elucidating and discussing some of the research that concerns each of these factors, to illustrate how it informs our understanding of early reading difficulty, its prediction and prevention.

Phonological Sophistication and Phonetic Processing: Factors in Skilled Reading

Like all orthographies, the English alphabet functions as a symbol system that transcribes certain units of the spoken language, and like all orthographies it appeals to the reader's intuitive appreciation of some aspect of linguistic structure (Hung & Tzeng, 1981; Liberman, Liberman, Mattingly, & Shankweiler, 1980). In actuality, English orthography is a morphophonological transcription that represents the word as a sequence of systematic phonemes, while, at the same time, capturing its constituent morphemes and underlying phonology. It therefore maps onto a deep, abstract level of language that corresponds rather closely to the way generative phonologists assume that words are represented in the ideal speaker-hearer's mental dictionary, or lexicon (Chomsky, 1964). If this characterization is accurate, the most effective strategy for the reader of English would be to recover the abstract lexical representation that a given string of letters "stands for," and with it, the word's semantic and syntactic extensions. Readers may also recover the phonetic representations of words by applying the phonological rules of English to morphophonological representations--rules that otherwise relate phonetic representations to morphophonological ones.

There is an advantage and a disadvantage to the way the alphabet represents English, and both of these follow from the nature of the relationship between letter sequences and spoken words. The advantage is that knowledge of this relationship between printed and spoken language allows the reader to decode not only highly familiar words, but also less familiar ones, and even words that have never been seen nor heard before. Whereas a skilled reader of a logography must have memorized at least two thousand distinct characters in order to read a newspaper, a skilled reader of English need only know a limited set of phoneme to grapheme correspondences and the phonological and morphological rules of English. But the disadvantage of the alphabet is that it requires phonological sophistication--a relatively fine-grained level of intuitive appreciation about the phonological structure of spoken language. To take full advantage of the alphabet, would-be-readers must somehow access their tacit knowledge of phonemes, morphemes, and phonological rules and apply that knowledge in an explicit, artificial fashion not required for spoken language (Mattingly, 1972). Such extensive phonological sophistication need not be achieved by the readers of a logography, for example, who need only know that their spoken language consists of words. Readers of the alphabet, however, must not only know about words, but also about the internal structure of words; that is, they must know about syllables and phonemes, and about the complex phonological rules that relate the phonetic units we produce and perceive to the abstract morphophonological representations that the letters of words "stand for." Otherwise, they cannot realize the virtues of the alphabet.

Theoretical considerations, then, reveal the relevance of phonological sophistication to effective use of the English alphabet. The further importance of another set of linguistic factors, spoken language processing skills, is illustrated by some experimental evidence about the skilled reading of words, sentences, and paragraphs. The question of whether speech recoding mediates lexical access from print has occupied much of the research on skilled reading (see, Crowder, 1982, for a recent review). Current evidence favors "dual access" of the lexicon by phonetic and visual processes operating in parallel. Yet regardless of how the lexicon is accessed, it is, at base, the morphophonological representation of a word that is being accessed, and with it, the word's semantic extensions and syntactic properties and its phonetic representation.

From the point of lexical access onward, the involvement of speech processes in reading is quite clear (Perfetti & McCutchen, 1982; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). First of all, there is much evidence that temporary memory for orthographic material (including isolated letters, printed nonsense syllables, and printed words) involves recoding the material into some kind of phonetic representation. Evidence that phonetic representation is employed in the service of temporary memory for such material can be found in the nature of the errors subjects make, and in the way that a phonetic manipulation, such as creating an inordinate density of rhyming items, can penalize performance (cf. for example, Baddeley, 1978; Conrad, 1964, 1972; Drewnowski, 1980; Levy, 1977). Adult subjects also appear to rely on phonetic representation when they are required to comprehend written sentences (Kleiman, 1975; Levy, 1977; Slowiaczek & Clifton, 1980; Tzeng, Hung, & Wang, 1977). Moreover, when reading sentences and paragraphs, they appear to employ not only the temporary memory system, but also the parsing system that supports recovery of the syntactic structure of spoken sentences and discourse. This is evidenced by the significant positive correlations between reading and listening comprehension (cf. Curtis, 1980; Daneman & Carpenter, 1980; Jackson & McClelland, 1979).

To summarize, theoretical considerations and experimental evidence reveal that the critical determinants of skilled reading of English include sophistication about phonological structure, and the adequacy of certain processes integral to spoken language comprehension. This recognition can now provide a meaningful framework within which to consider the process of beginning reading, and the problem of early reading difficulty.

Language Skills and Beginning Reading

What is required for success in learning to read? Obviously beginning readers of any orthography must be able to differentiate and remember the various orthographic shapes. Yet they must also differentiate and remember spoken words, phrases and sentences, because without these, there would be nothing for the orthography to transcribe. The well-known difficulties of congenitally deaf readers are one form of proof of the importance of spoken language skills for beginning readers. Further proof can be found in the relation between spoken language processing skills and success in learning to read.

Another requirement for successful beginning reading of the alphabet, in particular, is phonological sophistication (Liberman, Liberman, Mattingly, & Shankweiler, 1980). Alphabetic transcription necessitates that the child not only process spoken language effectively, but also be sophisticated about the phonological units of language. For example, successful beginning readers need not only distinguish words like "cat" and "hat," but be capable of holding them in memory, so that they can comprehend the differences between "A cat is on the hat," and "A hat is on the cat." They must further possess the linguistic sophistication that allows them to perceive the phonological relationship between "cat" and "hat"--that, among other things, these words differ in one phoneme, the first, and share a phoneme, the final one, which is the same as the initial phoneme in "top." Without this and other aspects of phonological sophistication, the alphabet will remain a mystery to them, and its virtues unrealized. Research reveals, however, that this and other aspects of phonological sophistication pose a difficulty for many young children, particularly those who incur early reading problems.

Having made these preliminary points about the requirements of beginning reading, let me turn to the problem of early reading difficulty, discovering its associated language deficiencies and predicting its occurrence. The past decade has witnessed considerable interest in these matters, and many studies of the psychology of early reading problems have uncovered an association between difficulty in learning to read, and difficulty within the two domains of spoken language processing skills and phonological sophistication.

The Relation between Spoken Language Skills and Reading Difficulty

Spoken language processing skills are important to beginning readers of all orthographies, and in accordance with this fact, a link between early reading ability and spoken language processing ability has been established for more than one alphabetic orthography (cf. Mann, 1982; Stanovich, 1982a, 1982b), and for syllabaries and logographies as well (cf. Stevenson, Stigler, Lucker, Hsu, & Kitamura, 1982). For the sake of brevity, this presentation will focus exclusively on the language processing problems found among poor readers of English.

As for such problems, it is by now quite clear that poor readers in the early elementary grades (i.e., children reading a half-year or more below grade expectation) do not suffer from a general impairment in perception, or in learning and memory, so much as from a language impairment that specifically penalizes certain phonological processing skills. For example, poor readers tend to be equivalent to good readers (i.e., children reading a half-year or more above grade expectation) in audiometry scores and nonverbal auditory perception, yet are inferior in ability to identify spoken words that are partially masked by noise (Brady, Shankweiler, & Mann, 1983). Poor readers also appear to have some other difficulties in recovering the phonetic representation of words, as evidenced by their difficulty with object and letter naming tasks (cf. Denckla & Rudel, 1976; Katz, 1982). Furthermore, they have a specific difficulty with short-term memory for verbal material. For example, poor readers do less well than good readers when temporarily remembering printed nonsense syllables, but not photographs of faces, or other purely visual materials (Liberman, Mann, Shankweiler, & Werfelman, 1982). They also tend to have difficulty recalling strings of spoken digits, spoken words, and even the words of spoken sentences (Mann, Liberman, & Shankweiler, 1980), yet have no such difficulty recalling nonverbal stimuli in a block-tapping task (Mann & Liberman, 1984).

One deficiency that is basic to reading and other language skills is a deficiency in use of phonetic representation in short-term memory (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Mann, 1984; Mann, Shankweiler, & Smith, 1984). It is this deficiency that limits the poor readers' ability to recall immediately such verbal material as syllables and words, phrases and sentences, regardless of whether the material is heard or read. That phonetic representation is problematic for poor readers can be seen in the results of studies that have manipulated a phonetic characteristic of the material being recalled, namely, the density of rhyming items. Normally, when the to-be-recalled items do not rhyme, good readers excel with respect to poor readers. However, when all of the items rhyme, the advantage of the good readers is greatly reduced or even eliminated, because for them, as for adults, the presence of phonetic confusability penalizes their ability to remember the words in order. Poor readers, in contrast, are less penalized by the presence of rhyme, that is, they are not as susceptible to the stress on phonetic representation. This result was originally demonstrated for recall of letter strings by eye and by ear (Shankweiler et al., 1979), and has been extended to recall of spoken word strings and spoken sentences (Mann et al., 1980). Taken together, these two findings, that poor readers' inferior short-term memory performance is confined to verbal material, and that there are consistent discrepancies between good and poor readers' susceptibility to the effects of rhyme, support a conclusion that poor readers are somehow lacking in ability to retain the full phonetic representation of words in short-term memory.

One consequence of a lack of effective use of phonetic representation on the part of poor readers would seem to be a difficulty with the comprehension of certain spoken sentences as well as with the repetition of sentences (Mann et al., 1984). For example, when required to act out the meaning of sentences that contain relative clauses, poor readers tend to make more mistakes than good readers because they made relatively more of the kinds of errors that slightly younger children make. On the basis of this finding (and other work we have done that suggests that poor readers do not always comprehend sentences less well than good readers), we have suggested that difficulties with phonetic representation may retard certain aspects of syntactic development among poor readers (Mann et al., 1984). Thus, out of a primary difficulty with phonetic representation may come second-order difficulties with other aspects of language development, including syntactic development and, ultimately, reading acquisition.

Having shown a connection between early reading difficulty and deficits in phonological processing skills, I will now turn to the results of two longitudinal studies showing that difficulty with certain phonological processes can often be found as antecedents of reading failure.

Phonological Processing Skills Can Presage Future Reading Ability

The first study to be described (Mann & Liberman, 1984) reveals that those kindergarten children who make less effective use of phonetic representation in a word-string recall task are likely to become the poorer readers of their first-grade class. The subjects were a population of 62 kindergarten children whom we followed longitudinally for one year. During May of the kindergarten year, we assessed IQ, phonological abilities (to be described later), use of phonetic representation, as indexed by the ability to repeat strings of rhyming words and strings of nonrhyming ones, and also nonverbal short-term memory, as indexed by their ability to repeat a block-tapping sequence on the Corsi Blocks. In May of the following year, when the children were at the end of the first grade, we again tested their memory, and also tested their reading ability. At that time, the teachers rated the children as good, average, or poor in reading ability.

Our finding was that children in the three reading groups did not differ in age, nor had they differed in kindergarten measures of IQ. Likewise, they did not differ in nonverbal memory performance as measured by the Corsi block test, either when they were kindergarteners or when they were in the first grade.. What we did find was that children who differed in reading ability significantly differed in their ability to repeat a string of spoken words. In addition, as we had discovered in the past, the extent of difference among children in the three reading groups was greatest in the case of phonetically nonconfusable words. Most importantly, these differences had been present before the children entered the first grade. As kindergarteners, the future poor readers made significantly more errors than the future good readers on the word strings, and their performance was not penalized by the presence of rhyme in the way that the performance of good readers was. Hence we concluded that the future poor readers did not make as effective use of phonetic representation as the future good readers did, and that this deficiency presages reading difficulties in the first grade. (For a more thorough report of this study and its materials, see Mann & Liberman, 1984).

The results of a second, newly completed longitudinal study make much the same point, extending the demonstration that tests of phonological processing can presage reading success. The results also reveal that screening can be conducted at an earlier time in the school year, and still effectively predict future reading ability. The subjects were 44 children tested during January of the kindergarten year and again the following January when they were in the first grade. As kindergarteners, they received an IQ test (the Peabody Picture Vocabulary Test), a verbal memory test (involving immediate, verbatim recall of seven strings of four unrelated words), a naming test (rapid naming of a randomized sequence of the capital letters of the alphabet as a measure of access to phonetic representations), and a syntactic test (manipulating toy dolls to enact the meaning of eight active sentences and eight passive sentences). They also received two tests of phonological sophistication, which will be described in the following section. As first graders, the reading ability of each child was established by administering the word recognition and word attack subtests of the Woodcock, and by asking the teachers to rate the child as good, average, or poor in reading ability. (Statistical evaluation of the results of this study can be found in the Appendix.)

The general profile of this population is summarized in Table 1, where the children are grouped according to teachers' ratings of their reading ability. Children in the three reading groups did not differ in age, or IQ, but did differ in the sum of raw scores on the two Woodcock tests. A summary of children's performance on the kindergarten tests of language processing appears in Table 2. The future poor readers were slower at naming the letters and made more errors than the future good readers did. As in our previous longitudinal study, the future poor readers also recalled fewer words in the verbal memory test than either good or average readers. Thus, these two tests of phonological processing skill, letter naming and temporary verbal memory, proved capable of distinguishing the future poor readers from the other children in their classrooms. In contrast, the third test that appears in Table

Table 1

Children Participating in a Longitudinal Study: Basic Profile

Reading Ability (Rated by first-grade teachers)	Kindergarten Testing: Mean Age (in months)	Kindergarten Testing: Mean IQ (Peabody)	First-Grade Testing: Mean Woodcock Raw Score (Word ID + Word Attack)
Good Readers N=10	69.2	118.5	94.8
Average Readers N=22	72.7	118.1	45.2
Poor Readers N=12	72.2	116.7	16.2

Table 2

Performance on Kindergarten Tests of Linguistic Processing in Relation to First-Grade Teachers' Ratings of Reading Ability

Reading Ability (first-grade rating)	Letter Naming (mean speed in s)	Letter Naming (mean errors)	Verbal Memory (mean words correct; max.=28)	Passive Sentence Comprehension (mean items correct; max.=8.0)
Good	21.3	0.0	22.8	8.0
Average	30.7	0.8	16.9	6.5
Poor	46.4	3.3	13.0	6.5

2, the test of syntactic ability, was only moderately successful. While it appeared to distinguish the future poor readers from the good readers, it did not distinguish them from the average ones.

Taken together, the two longitudinal studies support the conclusion that tests of language processing skill are better predictors of future reading ability than age, comparable tests of nonverbal memory, or tests of IQ. Tests of phonological processing skills, in particular, appear to be the best predictors of future reading ability although more work is needed before a final conclusion can be reached. Let me now turn to those additional findings from each study that reveal that tests of phonological sophistication can also predict future reading ability.

Phonological Sophistication Can Presage Future Reading Ability

The first point to be made in this section is that the importance of phonological sophistication to early reading ability is evident in the oral reading errors that good and poor beginning readers make. Linguistic analyses of such errors (see, for example, Fischer, Liberman, & Shankweiler, 1977; Shankweiler & Liberman, 1972) have shown that the reading difficulties of most children, including those diagnosed as dyslexic, tend not to involve deficiencies in visual perception or memory, so much as in phonological sophistication. Errors involving letter and sequence reversals are relatively infrequent, as compared to errors that reflect a problem in relating the structure of the printed word to the phonological structure of the spoken word.

Additional evidence that phonological sophistication is a special problem for poor readers can be found in the results of a study I recently conducted in collaboration with Isabelle Liberman and Hyla Rubin. The subjects were 62 third-graders, who were divided into three reading-ability groups according to their teachers' ratings. The study involved having children read the words of Galistel's GE Test of Coding Skills, in which words and phonologically plausible nonwords are arranged into ten categories according to the complexity of the phonological relation between the printed and spoken word. Our interest was in the specific types of words that caused children difficulty. All children made some errors on the Galistel test, and, as would be expected, the poorer readers made more errors than the better ones. Apropos of the point that poor readers are lacking in phonological sophistication, it was found that those categories that placed the greatest demands on phonological skill were inordinately difficult for these children.

Another finding that makes a similar point is summarized in Figure 1. The figure displays an analysis of children's responses in oral reading according to the familiarity of individual words. This analysis was done by noting which of the test words were included in the Cheek basal list for the first and second grades. It can be seen that all children were quite successful in reading the highly familiar Cheek words. More errors occurred on the other categories: words not included in the Cheek list and phonological plausible nonwords. If visual memory were a problem for the poor readers, we would have expected them to do inordinately poorly on the presumably less familiar words not on the Cheek list as compared to the better readers, but not necessarily on the nonwords, which none of the children had seen before. However, what we found is that the poorer readers had inordinate difficulty with both the non-Cheek words and the nonwords. That is, they were distinguished by their poor ability to read any items that place demands on phonologically-based analytic abilities and not by a lack of visual memory, per se. It may also be noted in Figure 1 that the nature of reading errors is the same for boys and girls. This is in agreement with our earlier findings in regard to absence of gender-specific patterns of deficit: the nature of reading difficulty does not depend on the sex of the child (Liberman & Mann, 1981). (Statistical analysis of the data appears in the Appendix.) The errors of poor readers, then, reflect some lack of phonological sophistication.

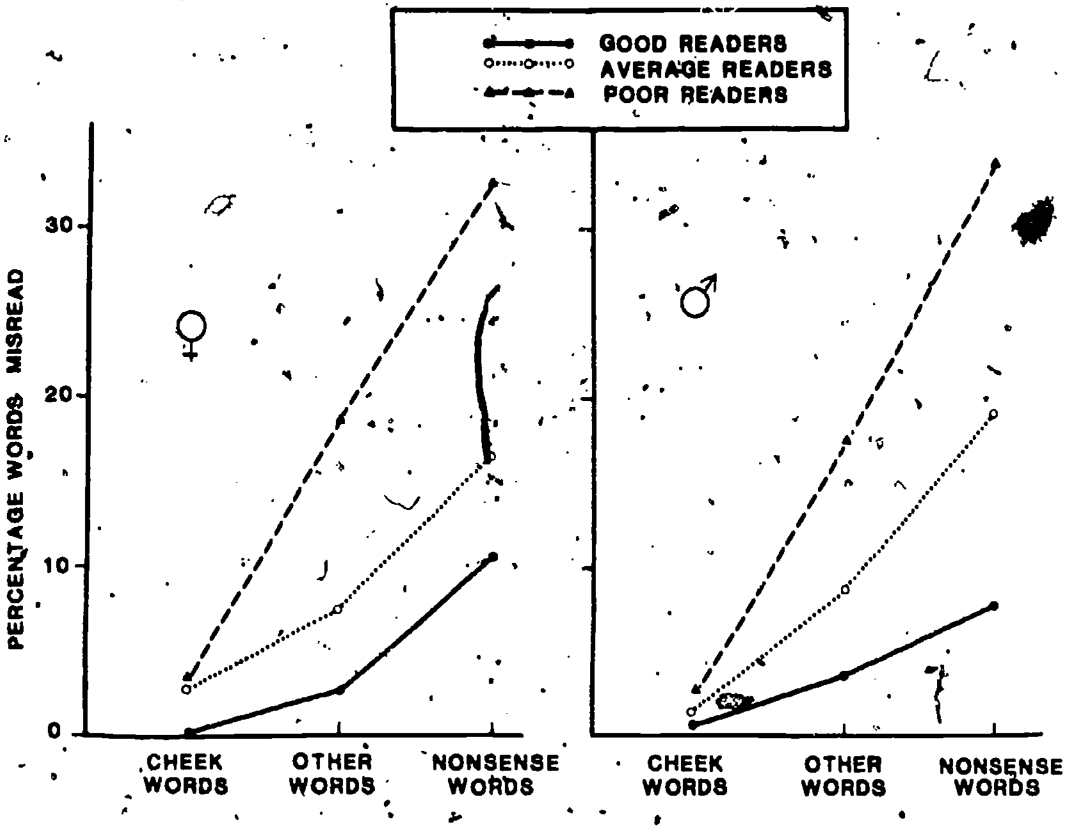


Figure 1. The proportion of words misread by girls and boys, as a function of reading ability and stress on phonological sophistication.

As will become apparent, longitudinal work further reveals that phonological deficiencies can antedate reading difficulty in the first grade. Before presenting that work in detail, however, it is appropriate for me to make a few observations about the development of phonological sophistication and its relation to reading instruction. A variety of evidence suggests that there is a reciprocal relationship between learning to read an alphabetic orthography and awareness about phonemes. (But, as we shall see, matters are somewhat different in parts of the world where a nonalphabetic form of writing is in use.) First of all, research in the U.S. and England indicates that four- and five-year-old children are generally lacking in awareness about phonemes, and a spurt in their phonological sophistication occurs at age six, when most of them begin to receive reading instruction (Bradley & Bryant, 1983; Liberman, Shankweiler, Fischer, & Carter, 1974). Second, in Portugal, where the writing system is, of course, also alphabetic, it has been found that most illiterate adults cannot add or delete initial phonemes in spoken utterances as well as literate ones can (cf. Morais, Cary, Alegria, & Bertelson, 1979). Third, in Japan, where nonalphabetic orthographies are employed, I have found that first-grade children cannot count, delete or reverse phonemes as easily as

American children of the same age. Nonetheless, some other work provides evidence that reading instruction is not the only determinant of phonological sophistication. For example, I have found that some Japanese children are aware of phonemes, regardless of their lack of exposure to the alphabetic principle. It has also been noted that some English-speaking children fail to acquire phonological sophistication despite considerable instruction in the use of the alphabet (Bradley & Bryant, 1978). (It is for all of these reasons, perhaps, that studies employing widely diverse subject populations, school systems, and measurement devices indicate a strong correlation between lack of phonological sophistication in kindergarten and later success in learning to read).

The first longitudinal study described in the previous section is a case in point. In that study (Mann & Liberman, 1984), we assessed the phonological sophistication of kindergarten children by requiring them to induce the rules of a game that involved counting the number of syllables in spoken words. Syllable counting was measured instead of phoneme counting, because awareness of syllable-sized units can be expected to precede awareness of phonemes and is probably a natural cognitive achievement of sorts, since it can be present in preschool children (Liberman et al., 1974). Moreover, unlike phoneme awareness, syllable awareness is not strongly facilitated by a phonics-based program of reading instruction (Alegria, Pignot, & Morais, 1982). We found that the future poor readers, as kindergarteners, scored lower on the syllable counting task, often performing at chance level, and rarely achieved the six correct responses in a row needed to pass criterion. The future good readers tended to receive the highest scores, to do considerably better than chance, and most of them had passed criterion. The performance of the average readers fell in between these two extremes.

Turning now to the second longitudinal study, we measured both syllable- and phoneme-awareness under the guise of a "talking backwards" game. Syllable awareness was measured by requiring children to reverse the order of the syllables in a two- or three-syllable nonsense word; phoneme awareness was measured by requiring children to reverse the order of the phonemes in a two-phoneme nonsense syllable. The results are shown in Table 3, where it can be seen that, on each test, the future poor readers did worse than the future average readers and the good readers did best of all. The strongest differences, however, and the ones that are the most predictive, involve performance on the phoneme reversal test.

Table 3

Performance on Kindergarten Tests of Linguistic Awareness in Relation to First-Grade Teachers' Ratings of Reading Ability

Reading Ability (first-grade rating)	Syllable Reversal (mean items correct; max.=16)	Phoneme Reversal (mean items correct; max.=16)
Good	14.6	5.2
Average	14.4	0.8
Poor	13.5	0.0

To summarize, then, in addition to tests of phonetic processing ability, tests that require manipulation (i.e., counting, reversal) of the syllables or phonemes in spoken words can also effectively presage future reading ability. Perhaps the best predictor of this type would be a test involving some manipulation of phoneme-sized units, although this has the potential disadvantage of confounding differences in native ability with differences in extent of exposure to reading instruction.

Some Remarks on the Prediction and Prevention of Early Reading Difficulty)]

As has been noted elsewhere (Mann & Liberman, 1984) the primary contribution of our longitudinal research is to suggest that, among kindergarteners, the status of certain phonological skills--verbal short-term memory, letter naming ability, awareness about syllables, and awareness about phonemes--may presage first-grade reading ability. Tests of these skills might therefore be used as part of a kindergarten screening battery. In this light, I will consider some of the practical implications of each of the longitudinal studies I have described.

The first indicates that measures of two skills, performance in recalling a string of nonrhyming words and performance in counting the number of syllables in spoken words, can together account for about a quarter of the total variance in children's first-grade reading ability. The success of these two measures lies not in their ability to predict fine differences in ability, but in their ability to predict the extremes of reading ability. A child who does well on both tasks is not at risk for future reading problems, whereas children who fall within the lower quartile of a kindergarten population in their performance on both tasks have a significant likelihood of encountering reading difficulty.

A somewhat finer grade of predictive success can be achieved using the results of the second, more recent, longitudinal study. When kindergarten performance on three measures, letter naming speed, accuracy of word string recall, and accuracy in reversing two-phoneme utterances, are entered into a regression equation, they account for 74% of the variance in raw scores on the Woodcock tests. Hence children who rank in the lower quartile of the class in letter naming ability, verbal memory, and phoneme awareness should surely be considered at risk. As for those who are deficient in only one or two of these skills, with future research it should be possible to determine the relative importance of these factors in terms of their contribution to the likelihood of a child's encountering difficulty in learning to read.

The development of tests that successfully identify children at risk for reading problems is surely an accomplishment. However, what is one to do with the child considered at risk? Teachers and others interested in the question of how to prevent reading problems would do well to read what Isabelle Liberman has written on this subject (Liberman, 1982; Liberman, Shankweiler, Blachman, Camp, & Werfelman, 1980). Here, I would like to focus briefly on several points, most of which have been made elsewhere (see, for example, Liberman, 1982, or Mann & Liberman, 1984).

Considering first the child who is lacking in the phonological processing skills necessary for effective verbal short-term memory and letter-naming ability, it must be said that the prospects for remediation of these deficiencies have not really been explored. There is considerable reason to consider the possibility that deficient phonological processing skill is a consequence of a specific maturational lag in language development. However, the concept of maturational lag typically implies that the poor readers will "catch up" to the good readers, given enough time, and this implication is inconsistent with findings that specific verbal deficiencies sometimes persist, as in adolescents suffering from developmental dyslexia (cf. Mann, in press, and Mann & Liberman, 1984, for appropriate references). Thus it is possible that the language processing deficiencies we may detect in the kindergarten child will be of a permanent nature. This is not to say that remediation is a hopeless cause, for we may still expect that, through appropriate intervention, the extent of the deficiency can be lessened. Unfortunately, research has not yet specified the exact form of remediation that is most desirable. It is logical to think that children might be helped by practice in naming letters and objects as well as by practice in learning nursery rhymes, and stories by heart. Such practice might help children to exercise those language processing that they do possess. Yet it must be kept in mind that remedying a specific symptom need not remedy the underlying cause of that symptom. Clearly much research is needed in this area.

The prospect may be brighter, however, with regard to remediating deficiencies in phonological sophistication. While it is true that some aspects of phonological sophistication, such as initial awareness about syllables, tend to be natural cognitive achievements, much of the development of phonological awareness may be facilitated (if not precipitated) by experience that encourages the child to manipulate phonological structure. For some children this experience may involve no more than learning the correspondences between certain written and spoken words. Even minimal exposure may be enough to enable some children to discover the alphabetic principle for themselves. This is probably true of unexpectedly precocious readers, and of most children who survive the basal method of beginning reading instruction. Yet other children don't discover that principle for themselves, and may need some systematic training in order to achieve the level of sophistication about phonemes and phonological rules that is required for skilled reading of English. With all due respect to Socrates, it isn't really good educational policy to make all children reinvent the alphabet for themselves--we should let them in on the secrets of alphabetic transcription as early as possible.

How is this to be done? To begin with, children should be read to, and their attention should be directed to the printed words that correspond to the spoken words of their favorite story. Teachers and parents can use many indirect methods to draw children's attention to phonological structure--teaching nursery rhymes and poetry, for example, or encouraging secret languages like "pig latin" or "talking backwards." They can, for example, give children special nicknames that involve some systematic manipulation of phonological structure (such as reversing the order of syllables, or dropping the first phoneme), and then ask the children to invent similar nicknames for their siblings and friends. Once attention is directed to phonological units, direct awareness training can be instigated through counting games or elision games, starting at the less abstract level of the word and working down to the level of the phoneme. Finally, phoneme awareness and reading could be introduced with the procedure of Elkonin.

The Elkonin procedure has been described elsewhere (Liberman, 1982; Liberman et al., 1980; Mann & Liberman, 1984) and for the sake of brevity I will only review its merits. It provides a linear visuospatial structure to which the temporal sequence of phonemes in a spoken word can be related. It gives the child the actual number of phonetic segments in a word so that uninformed guessing is not necessary. Explicit naming of pictures is required and can exercise the child's ability to access the phonetic representation of a word rapidly. Since the picture is always present, and only one is considered at a time, demands on verbal short-term memory are minimal. For all of these reasons, the Elkonin procedure is especially advantageous for use with children who, by virtue of inferior phonological sophistication, naming ability, and verbal short-term memory, have been identified at risk for future reading problems. If adopted for general use, it could help to ameliorate reading difficulty, and might be expected to speed the progress of any beginning reader.

References

Alegria, J., Pignot, E., & Morais, J. (1982). Phonetic analysis of speech and memory codes in beginning readers. Memory & Cognition, 10, 451-456.

Baddeley, A. D. (1978). The trouble with levels: A reexamination of Craik and Lockhardt's framework for memory research. Psychological Review, 85, 139-152.

Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organization as a possible cause of reading backwards. Nature, 271, 746-747.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read--a causal connection. Nature, 301, 419-421.

Brady, S., Shankweiler, D., & Mann, V. (1983). Speech perception and memory coding in relation to reading ability. Journal of Experimental Child Psychology, 35, 345-367.

Chomsky, N. (1964). Comments for project literacy meeting. Project Literacy Report No. 2, pp.1-8. Reprinted in M. Lester (Ed.), Reading in applied transformational grammar. New York: Holt Rinehardt and Winston.

Conrad, R. (1964). Acoustic confusions in immediate memory. British Journal of Psychology, 55, 75-84.

Conrad, R. (1972). Speech and reading. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye: The relationships between speech and reading. Cambridge, MA: MIT Press.

Crowder, R. G. (1982). The psychology of reading. New York: Academic Press.

Curtis, M. E. (1980). Development of components of reading skill. Journal of Educational Psychology, 72, 656-669.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, 450-466.

Denckla, M. B., & Rudel, R. G. (1976). Naming of object drawings by dyslexic and other learning-disabled children. Brain and Language, 3, 1-15.

Drewnowski, A. (1980). Memory functions for vowels and consonants: An interpretation of acoustic similarity effects. Journal of Verbal Learning and Verbal Behavior, 19, 176-193.

Fischer, F. W., Liberman, I. Y., & Shankweiler, D. (1977). Reading reversals and developmental dyslexia: A further study. Cortex, 14, 496-5[illegible].

Hung, D. L., & Tzeng, O. J. L. (1981). Orthographic variations and visual information processing. Psychological Bulletin, 90, 377-414.

Jackson, M., & McClelland, J. L. (1979). Processing determinants of reading speed. Journal of Experimental Psychology: General, 108, 151-181.

Katz, R. B. (1982). Phonological deficiencies in children with reading disability: Evidence from an object naming task. Unpublished doctoral dissertation, University of Connecticut.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. Journal of Experimental Child Psychology, 32, 474-484.

Kleiman, G. (1975). Speech recoding in reading. Journal of Verbal Learning and Verbal Behavior, 14, 323-339.

Levy, B. A. (1977). Reading: Speech and meaning processes. Journal of Verbal Learning and Verbal Behavior, 16, 623-638.

Liberman, I. Y. (1982). A language-oriented view of reading and its disabilities. In H. Myklebust (Ed.), Progress in learning disabilities (Vol. 5). New York: Grune and Stratton.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. Kavanagh & R. Venezky (Eds.), Orthography, reading and dyslexia. Baltimore: University Park Press.

Liberman, I. Y., & Mann, V. A. (1981). Should reading remediation vary with the sex of the child? In A. Ansara, N. Geschwind, A. Galaburda, N. Albert, & N. Gartrell (Eds.), Sex differences in dyslexia. Towson, MD: The Orton Society.

Liberman, I. Y., Mann, V. A., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and non-linguistic material in relation to reading ability. Cortex, 18, 367-375.

Liberman, I. Y., Shankweiler, D., Blachman, B., Camp, L., & Werfelman, M. (1980). Steps towards literacy. Report prepared for Working Group on Learning Failure and Unused Learning Potential. President's Commission on Mental Health, Nov. 1, 1977. In P. Levinson & C. H. Sloan (Eds.), Auditory processing and language: Clinical and research perspectives. New York: Grune and Stratton.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. Journal of Experimental Child Psychology, 18, 201-212.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), Toward a psychology of reading: The proceedings of the CUNY Conferences (pp. 207-225). Hillsdale, NJ: Erlbaum.

Mann, V. A. (1982). Reading skill and language skill. Haskins Laboratories Status Report on Speech Research, SR-69, 151-170.

Mann, V. A. (1984). Reading skill and language skill. Developmental Review, 4, 1-15.

Mann, V. A. (in press). Why some children encounter reading problems: The contribution of difficulties with language processing and linguistic sophistication to early reading difficulty. To appear in B. Wong & J. Torgessen (Eds.), Learning disability: Some new perspectives. New York: Academic Press.

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory: Can they presage early reading success? Journal of Learning Disabilities, 17, 592-599.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. Memory & Cognition, 8, 329-335.

Mann, V. A., Shankweiler, D., & Smith, S. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. Journal of Child Language, 11, 627-643

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

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

Perfetti, C. A., & McCutchen, D. (1982). Speech processes in reading. *Speech and Language: Advances in Basic Research and Practice, 7,* 237-269.

Shankweiler, D., & Liberman, I. Y. (1972). Misreading: A search for causes. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading.* Cambridge, MA: MIT Press.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 531-545.

Slowiaczek, M. L., & Clifton, C. (1980). Subvocalization and reading for meaning. *Journal of Verbal Learning and Verbal Behavior, 19,* 573-582.

Stanovich, K. (1982a). Individual differences in the cognitive processes of reading: I. Word decoding. *Journal of Learning Disabilities, 15,* 485-493.

Stanovich, K. (1982b). Individual differences in the cognitive processes of reading: II. Text-level processes. *Journal of Learning Disabilities, 15,* 549-554.

Stevenson, H. W., Stigler, J. W., Lucker, G. W., Hsu, C-C., & Kitamura, S. (1982). Reading disabilities: The case of Chinese, Japanese and English. *Child Development, 53,* 1164-1181.

Tzeng, O. J. L., Hung, D. L., & Wang, W. S-Y. (1977). Speech recoding in reading Chinese characters. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 621-630.

Appendix

Statistical Evaluation of Experimental Results

Longitudinal Study of Language Processing Skills, Linguistic Awareness, and Reading Ability. Statistical analyses of the data summarized in Tables 1-3 include t-tests of differences between the scores of good and poor readers, and Pearson correlations between various scores and a measure of reading ability. Turning first to the data in Table 1, the good and poor readers differed in the sum of raw scores on the Woodcock Word Attack and Word Identification, $t(20)=5.3$; $p<.002$, although they did not differ in age, or in IQ at the .05 level of confidence. As for the data in Table 2, as kindergarteners, the future good and poor readers differed in all four measures of language processing: 1) speed of letter naming, $t(20)=3.32$, $p<.01$, 2) errors in naming the letters, $t(20)=5.91$; $p<.0002$, 3) verbal memory, $t(20)=2.2$, $p<.05$ and 4) comprehension of passive sentences, $t(20)=3.6$; $p<.01$. Pearson product moment correlations revealed significant associations between the first-grade sum of raw scores on the Woodcock Word Identification and Word Attack Subtests and the kindergarten measures of: letter naming speed, $r(44)=-.42$, letter naming errors, $r(44)=-.52$, and verbal memory, $r(44)=.56$, all of which are significant beyond the .05 level. The correlation between Woodcock scores and comprehension of passive sentences failed to reach significance at the .05 level of confidence. Finally, as for the data in Table 3, which concerns performance on the two tests of language awareness, good and poor readers significantly differed in performance on the phoneme reversal test, $t(20)=9.2$, $p<.0002$, but not on the syllable reversal test.

Likewise, a significant correlation existed between Woodcock scores and performance on the phoneme reversal test, $r(44)=.75$, although the correlation between Woodcock scores and performance on the syllable reversal test failed to reach significance at the .05 level of confidence.

Reading Errors Among Good, Average and Poor Readers in the Third Grade. The teacher-ratings of reading ability are confirmed by the finding that, when raw scores on the Woodcock Word Identification and Word Attack Tests were summed, the poor readers had correctly read an average of 133.2 words, whereas average readers had read 156.2 and good readers, 175.6. Statistical analysis of the reading errors made on the GE test (summarized in Figure 1) consisted of an analysis of variance involving reading ability, sex, GE category and Cheek category. That analysis revealed that children in the three reading groups differed in their overall performance, with the better readers making fewer errors than the poorer ones, and the average readers falling between, $F(2,56)=31.55$, $p<.0001$. Certain parts of the GE test were harder than others, $F(9,504)=75.38$, $p<.0001$, but the poorer readers encountered inordinate difficulty with these parts of the test relative to the better readers, $F(18,504)=6.28$, $p<.0001$. In general, the Cheek words were easier to read than non-Cheek words, which, in turn, were easier than the phonologically plausible nonwords, $F(2,112)=205.5$, $p<.0001$. Most importantly, as compared to the better readers, the poorer readers encountered much less difficulty with the Cheek words than with the words that were not on the Cheek list, and with the phonologically plausible nonwords, $F(4,112)=22.9$. This basic pattern of results was not a function of the sex of the child, as the main effect of sex, and all interactions involving sex, fail to reach significance at the .05 level of confidence.

TEMPORARY MEMORY FOR LINGUISTIC AND NONLINGUISTIC MATERIAL IN RELATION TO THE ACQUISITION OF JAPANESE KANA AND KANJI

Virginia A. Mann†

Abstract. It has been found that good and poor beginning readers of the alphabet perform equivalently in temporary memory for such nonlinguistic visual material as abstract designs and faces. Good readers excel, however, in temporary memory for such linguistic material as printed or spoken words, because they make more effective use of phonetic representation. To determine whether linguistic and nonlinguistic memory skills have the same relationship to reading skills among good and poor beginning readers of nonalphabetic orthographies, the present study focused on beginning readers of Japanese Kana and Kanji. Two experiments employed the recurring recognition paradigm of Kimura to assess the relationship between temporary memory for various types of material and Japanese children's reading ability in the second grade. The first experiment examined temporary memory for spoken nonsense words, and revealed that good and poor readers of Japanese differ in the use of nonalphabetic orthographies. The second explored the relationship between memory for nonlinguistic visual material and orthographic material including: 1) abstract designs, 2) faces, 3) Hirigana, and 4) Kanji. It revealed that children who differed in reading ability differed in memory for Kana and Kanji, but, unlike good and poor readers of the alphabet, also differed in memory for the abstract designs. In particular, memory for Kana was significantly related to that for Kanji and spoken syllables, but not abstract designs or faces, whereas memory for Kanji was significantly related to that for nonsense designs and spoken syllables, but not faces. The implication is that for nonsense designs and spoken syllables, but not faces. The implication is that effective use of phonetic representation contributes to successful acquisition of all orthographies, whereas the importance of nonlinguistic memory skills can depend on the nature of the orthography at hand.

---

†Also Bryn Mawr College.

Acknowledgment. This study was completed under partial support from a Fulbright Fellowship, at the Research Institute of Logopedics and Phoniatrics, University of Tokyo. I would like to express my gratitude to Dr. Masayuki Sawashima, for his help throughout this project and to thank my research assistant, Dr. Seishi Hibi, for his valuable insights. The help of Dr. Tazuko Uyeno is also gratefully acknowledged. This study could not have been completed without the very gracious assistance of Ms. Shizuko Fukuda, acting principal of the elementary school attached to Ochanomizu University and the children and teachers of the second grade.

Temporary memory is important to the would-be reader of any orthography. Whether material is written in an alphabet, a syllabary, or a logography, the reader who intends to comprehend its full meaning must be able to retain the information represented by individual characters until such larger units as words, sentences, or paragraphs can be apprehended. How beginning readers of syllabaries and logographies meet these temporary storage requirements is the issue at stake in this study, which examines the relation between temporary memory skills and success in learning to read Japanese. Several different types of material are to be considered, as an accumulating body of evidence from the psychological (Ellis, 1975; Woodhead & Baddeley, 1981) and neuropsychological literature (see, for example, Kimura, 1963; Milner & Taylor, 1972; Warrington & Shallice, 1969) reveals that temporary memory can involve separable components. One such component employs phonetic representation as a means of retaining linguistic material such as names of objects, spoken or printed words, etc., and is localized within the left, language dominant hemisphere. This stands in contrast to another component, which employs nonlinguistic representations as a means of retaining such nonlinguistic materials as abstract designs and faces, and is localized within the right hemisphere. The question to be asked over the course of two experiments is whether the linguistic and nonlinguistic components of temporary memory make equivalent contributions to success in learning to read Kana and Kanji.

In recent years, studies in America and Europe have asked whether success in learning to read alphabetic orthographies is related to the ability to remember certain types of information. These studies reveal that not all temporary memory abilities tend to distinguish good and poor beginning readers of the alphabet. For example, good and poor readers in the second grade do not significantly differ in the ability to remember such nonlinguistic visual material as photographs of people's faces, or abstract visual designs (see, for example, Liberman, Mann, Shankweiler, & Werfelman, 1982; Vellutino, Steger, DeSetto, & Phillips, 1975). Yet it is quite evident that good readers surpass poor readers in temporary memory for syllables, words, and sentences--whether these are heard or read (see, for example, Byrne & Shea, 1979; Mann, 1984; Mann, Liberman, & Shankweiler, 1980 Mark, Shankweiler, Liberman, & Fowler, 1977; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). This has been explained by appeal to evidence that superior readers make effective use of phonetic representation in temporary memory (Shankweiler et al., 1977).

Attempts to clarify and explain the association between effective use of phonetic representation and early reading ability have tended to focus on beginning readers of English (see, for example, Brady, Shankweiler, & Mann, 1983; Katz, Shankweiler, & Liberman, 1981; Mann, in press; Mann & Liberman, 1984; Mann et al., 1980; Mann, Shankweiler, & Smith, 1984; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979), or of other alphabetically-transcribed languages such as French (Alegria, Pignot, & Morais, 1982), Swedish (Lundberg, Oloffson, & Wall, 1980) and Dutch (Mann, 1982). One possible explanation is that learning to decode a phonetic transcription of spoken language places certain demands on memory for phonetic material (see, for example, Shankweiler & Liberman, 1976). If so, the association between use of phonetic representation and reading skill might be restricted to readers of the alphabet (conceivably extending to readers of a syllabary, since a syllabary is a type of phonological transcription) but not to readers of a logography, since logographies do not transcribe the phonological structure of spoken words directly. Another possibility is that phonetic representation is critical to all language processing, spoken and written alike, because it meets the temporary storage requirements involved in recovering phrases, sentences and paragraphs from sequences of individual words (see Liberman, Liberman, Mattingly, & Shankweiler, 1980; or Mann, 1984). In this case, the relationship between phonetic representation and reading ability should extend to readers of any orthography--alphabet, syllabary or logography--because all of these require that readers recover phrases, sentences, etc. one way or another.

One straightforward means of determining why use of phonetic representation is associated with reading skill is to examine the use of phonetic representation by good and poor readers of syllabaries and logographies. Japanese has the virtue of using both; however, relatively little attention has been devoted to the temporary memory skills associated with acquiring Japanese. As some type of temporary memory should be essential to readers of syllabaries and logographies, we would expect some relationship between temporary memory skills and success in learning to read Japanese. It is even possible that the relationship between reading and memory skill will involve linguistic memory, and phonetic memory in particular, given some evidence that for both Japanese children and American children, memory for the meaning of spoken text, as well as serial memory for words and digits, is associated with reading ability in the fifth grade (Stevenson, Stigler, Lucker, Hsu, & Kitamura, 1982). Yet there is no direct evidence about the use of phonetic representation by children who are in the early grades in Japan.

Certainly it is possible that effective use of phonetic representation characterizes good readers of nonalphabetic orthographies just as it characterizes good readers of the alphabet. This follows from a consideration of the fundamental nature of all orthographies, and from an observation about a coding strategy that is common to skilled readers of Chinese and English. All orthographies function to transcribe spoken language, hence it would be parsimonious if reading drew upon some of the processes that otherwise support spoken language use. Skilled readers who are attempting to remember written words in order to comprehend written sentences and paragraphs might rely on phonetic representation, as phonetic representation fulfills the temporary memory requirements of comprehending spoken sentences and paragraphs. Confirmation that this is indeed the case has been provided by experimental studies showing that both the temporary memory for orthographic material (including isolated letters, printed nonsense words, and real words and sentences) and the comprehension of written text involve recoding print into a phonetic representation. Most importantly, phonetic representation is employed in the service of temporary memory and comprehension whether subjects are reading the English alphabet (see, for example, Daneman & Carpenter, 1980; Kleiman, 1975; Levy, 1977; Meyer, Schvaneveldt, & Ruddy, 1974; Slowiaczek & Clifton, 1980) or the Chinese logography (Hung & Tzeng, 1981; Tzeng, Hung, & Wang, 1977).

However, it is nonetheless possible that the acquisition of nonalphabetic writing systems places a less severe demand on phonetic memory than the alphabet does. Before beginning readers can begin to comprehend phrases and sentences, they must learn to decode individual words. While all orthographies serve to transcribe the words of spoken language in one way or another, they differ in the nature of the units they transcribe: alphabets transcribe phonemes, syllabaries such as the Japanese Kana transcribe syllables, and logographies such as Chinese and Japanese Kanji transcribe words. It could be argued that these differences have consequences on the importance of phonetic representation to children's initial acquisition of word decoding skills. The beginning reader of the alphabet, for example, who is attempting to recognize a written word like "kitten," must be able to integrate the phonemes that the letters represent, and this may require effective use of phonetic representation. For the reader of a syllabary, integrating a sequence of syllables like "neko" into a word may also involve phonetic representation, but the demand could be milder than that of the alphabet, insofar as syllables are less abstract phonological units than phonemes, and there are typically fewer of them in a word. Finally, for the reader of a logography, word decoding may place almost no demand on phonetic representation. Since the characters of logographies transcribe words on a one-to-one basis, recognizing a word could be an all-or-none process that does not require that phonemes or syllables be retained in temporary memory.

Thus it is an open question whether use of phonetic representation will distinguish good and poor beginning readers of Japanese. Likewise, it is unknown whether, like good and poor readers of the alphabet, good and poor readers of Kana and Kanji tend to possess equivalent nonlinguistic memory skills. Here it is conceivable that the acquisition of the Kana syllabaries and the Kanji logography could place a certain demand on visual memory systems. Syllabaries, to some extent, and logographies, in particular, involve considerably more orthographic units than the alphabet. Thus would-be readers of Japanese must encode and remember more visual shapes. For mature readers of Japanese, the ability to read Kana and Kanji, like that to read the alphabet, tends to be associated with the integrity of the linguistic faculties of the left hemisphere (Sasanuma, 1975; Sasanuma & Fujimura, 1971). Hence, it appears unlikely that the visual memory demands of Kana and Kanji cause skilled readers of Japanese to place inordinate reliance on nonlinguistic memory skills. Nonetheless, for the beginning reader who is acquiring an initial knowledge about orthographic characters, it is possible that learning to remember more-or-less abstract patterns of lines and curves could demand effective use of nonlinguistic memory skills. If so, good beginning readers of Japanese may surpass poor readers in their ability to hold certain types of nonlinguistic material in temporary memory.

A rationale has been offered that the type of temporary memory abilities that are crucial to success at learning to read may depend on the type of orthography at hand. Effective use of phonetic representation could contribute to the attainment of early reading skill among American children because it fulfills the temporary storage requirements of all language processing, or because it fulfills certain specific requirements of learning to decode a phonographic transcription. Likewise, effective visual memory skills, which bear little association to American children's skills in learning to read, could be of limited utility only to readers of alphabetic orthographies, or could be of limited utility to all beginning readers. In the two experimental studies that follow, an attempt was made to discern whether phonetic representation and various types of nonlinguistic memory abilities distinguish good and poor beginning readers of Kana and Kanji. The design is prompted by a previous study of American children that used the recurring recognition paradigm of Kimura (1963) to assess good and poor readers' ability to remember alphabetically written material, visual nonsense designs, and photographs of unfamiliar faces (Liberman et al., 1982). The first experiment extends this methodology to the use of spoken nonsense materials, in order to determine whether effective use of phonetic representation distinguishes good and poor readers of Japanese in the second grade. The second experiment compares memory for the two types of orthographic material, Kana and Kanji, with that for two types of nonlinguistic visual material, abstract designs and faces.

Experiment 1

Methods

Subjects. The subjects were second-grade children attending the primary school attached to Ochanomizu University in Tokyo, Japan. All available children participated, including 50 girls and 50 boys, of mean age 86.6 months (sd.=3.5 months). At the completion of the study, each child was rated by his or her classroom teacher as either good, average, or poor in reading ability.

Materials. The materials comprised 52 two-mora (i.e., disyllabic) pseudoword items that were phonologically plausible according to the intuitions of five native speakers of Japanese (three members of the staff of the Institute for Logopedics and Phoniatrics, a Japanese linguist, and a teacher of Japanese). The pseudowords were so constructed that all Japanese consonants and vowels were represented in a variety of combinations (with the exception of consonant-[y] clusters as in [kyo]). Both V and CV mora occurred with the restriction that no mora occur more than once in either initial or final position.

Memory for these materials was assessed according to the recurring recognition paradigm of Kimura (1963). This required that 80 test items be constructed, each item selected from the pool of 52 items. Four items were repeated eight times each (the recurring items) and the remaining 48 were used once each (the nonrecurring items). In compiling the test, the items were divided into eight sets of ten items each, with each set containing the four recurring items randomly interspersed with six of the nonrecurring items. The first set of ten items constituted the inspection set, the remaining seven constituted the recognition set of 70 test items. The test was administered by a male native speaker of Japanese who read each item aloud with a flat intonation at a rate of one every five seconds.

Procedure

All children were tested in their classrooms while seated with their classmates in their normal seating arrangement. Testing was completed during a single session conducted in the early afternoon after school hours. The instructor told the children that he would read to them some words they had never heard before (i.e., nonsense words) and that their task was to listen to the initial set of ten words and to try to remember each of them; afterwards they would hear the test set of 70 words, and were to mark their responses on a sheet numbered from 1 to 70 with a single box next to each number. They were to put an "O" in the corresponding box if the word had occurred in the initial set; otherwise they should put an "X" in the box. At this point, presentation of the ten initial items began, immediately followed by presentation of the 70 test items. To aid children in the correct use of the response sheet, the instructor said the number of each test item prior to reading the item aloud.

Results and Discussion

Following the methodology employed in the previous study of American children (Liberman et al., 1982), the data were analyzed in terms of the total number of correct responses, summing over the seven sets of items, and including both the number of correct recognitions of recurring items and the number of correct rejections of nonrecurring items. To determine whether this data reduction procedure masked any critical findings, the data obtained from one of the three classrooms of children (n=33) were subjected to a more detailed analysis. That analysis indicates that a consideration of early vs. late sets, or of recognition vs. rejection would not alter the basic pattern of findings and the interpretation of them.

The main purpose of this experiment was to determine whether, like beginning readers of alphabetic orthographies, children who differ in the ability to read Japanese differ in memory for nonsense words. The standpoint from which I will attempt to answer this question is the teachers' ratings of the children's reading ability, a measure that does not separate skill in reading the Kana syllabaries from that in reading the Kanji logography. Attention to the separate contribution of each skill will be a concern of Experiment 2; here, the analysis focuses on the question of whether reading ability, in general, is related to memory for spoken nonsense words.

The answer is affirmative: good readers surpassed poor readers in memory for the nonsense words, as the following analyses will show. On the average, performance was significantly better than the chance level of 50% correct, $\underline{t}(99)=40.0$, $\underline{p}<.001$, although children differed considerably in the accuracy of their responses. The highest score was 65 items correct (out of 70) the lowest was 36, and the mean was 57.0, or 81% correct. The 15 children whom their teachers deemed to be good readers achieved a mean score of 62.4 correct, which is significantly better than the mean score of 45 correct achieved by the 10 children who were deemed to be poor readers, $\underline{t}(23)=6.18$, $\underline{p}<.001$. When the scores of all children (good, average, and poor readers) are considered, the relation between memory performance and reading ability is positive, and a modest, but significant correlation is found, $r(100)=.25$, $\underline{p}<.006$.

The results of this experiment therefore suggest that children who are good beginning readers of Japanese, like those who are good beginning readers of English (Byrne & Shea, 1979), tend to excel poor readers in temporary memory for spoken nonsense words. The implication is that children who are particularly successful readers of nonalphabetic orthographies tend to make more effective use of phonetic representation than children who read less well. Whether use of phonetic representation is of equal importance to successful acquisition of the Kana syllabary and the Kanji logography remains to be explored in the second experiment, along with the question of whether memory for nonlinguistic material is related to successful acquisition of either Kana or Kanji.

Experiment 2

Methods

Subjects. The subjects of Experiment 2 were the same 100 children who participated in Experiment 1.

Materials. The materials included four different types of visual stimuli: abstract designs, photographs of faces, Hiragana, and Kanji. Memory for each type of stimulus was assessed separately, using tests modeled on Kimura's (1963) recurring recognition paradigm. A description of each of the materials follows: 1) The Kimura abstract designs were used according to a modified test procedure by Liberman et al. (1982). These consisted of irregular, nonrepresentational line drawings. 2) The face materials were black and white photographs of male faces in half profile taken from a high school yearbook. Half were looking to the left, and half to the right. Preparation of the materials was as in Liberman et al. (1982). In order to minimize distinguishing details that might lend themselves to verbal labeling, none of the photographs showed teeth, facial hair, eyeglasses, or distinctive marks such as scars, etc. In addition, a uniform mask was applied to each picture to cover hair and background detail. 3) The Hiragana materials were meaningless, phonetically plausible digraphs that combined two characters, one above the other, which had been photographed from a set of flash cards. Each digraph was a transcription of one of the 52 basic items employed in Experiment 1 and in the test sequence, different stimuli recurred and a a different order of presentation was employed. 4) The Kanji materials were chosen from the approved set of Kanji that children master in the first grade. The characters were photographed from large-sized prototypes contained in a standard dictionary.

The preparation of the memory test for each type of material was as described in Experiment 1. From each set of 52 stimulus items, a test set of 80 items was constructed, with four of the items recurring eight times each, and 48 of the stimuli occurring once. The items were divided into eight sets of ten; within each set, the four recurring items were interspersed with six nonrecurring ones. The first set of ten constituted the inspection set, and the remaining seven sets contained the 70 test items.

Procedure

Experiment 2 was run togther with Experiment 1, in two sessions one week apart. The Kana digraphs and the faces were presented in the first session; the Kanji, and the nonsense designs were presented in the second. All stimuli were projected on a large screen at the front of the room by means of a Kodak carousel projector. The subtended angle was sufficiently great for easy visibility from all parts of the room.

The procedure was analagous to that in Experiment 1 and was the same for all four types of materials. The instructor began each test by telling the child that some Kana (or Kanji, etc.) would be shown on the screen at the front of the class. The task was to look carefully and to try to remember each item as it appeared. As in Experiment 1, the children were given a response sheet that contained boxes numbered from 1-70, and were told that they would use the sheet to mark each new item with an "X" and each previously seen item with an "O." The ten inspection items were then presented at a rate of approximately one every five seconds, followed by the 70 test items at the same rate of presentation. The instructor said the number of each test item as it appeared on the screen, to insure that the children would not lose their place on the response sheet.

Results and Discussion

As in Experiment 1, the data were scored in terms of the total number of correct responses given to each type of material, and scores were considered in relation to children's designation as good, average, or below average in reading ability. Although performance on the various types of material was always significantly better than chance ($t(99)=29.0$ for nonsense designs; $t(99)=20.4$ for faces; $t(99)=50.5$, for Kanji; $t(99)=40.3$ for Kana; all of which are significant at the $p<.001$ level), the level of performance varied among children and across the various types of items. Average scores were highest for Kanji (67.5 items, 96.5% correct) followed by Kana (62.7 items, 89.6%), abstract figures (59.0 items, 84.3% correct) and faces (57.0 items, 81%).

The main purpose of Experiment 2 was to determine the relation between reading ability, memory for nonlinguistic visual information, and memory for Kana and Kanji. To that end, the first analysis concentrated on comparing the mean performance of the 15 good and the 10 poor readers. This has the virtue of permitting a direct comparison of the present results with those obtained in Liberman et al.'s (1982) study of American children. For convenience, the American data appear in Figure 1 so that they may be compared with the present results, which appear in Figure 2.

Turning first to the orthographic materials, it can be seen on the right side of Figure 2 that the good readers of Japanese achieved superior scores on both the Kana, $t(23)=18.7$, $p<.001$, and Kanji materials, $t(23)=6.12$, $p<.001$. Note also that, whereas the good readers had achieved equivalent scores on the two types of orthographic material ($p>.01$), the poor readers were markedly worse on the Kana digraphs than on the Kanji, $t(9)=7.18$, $p<.01$. Consequently, the extent of difference between children in the two reading groups is greater in the case of the Kana materials.

As for the two types of nonlinguistic material, the left side of Figure 2 reveals that, unlike the beginning readers of English, whose data appear in Figure 1, the good readers of Japanese surpassed the poor readers in their performance on the abstract designs, $t(23)=4.76$, $p<.01$. Like beginning readers of English, however, the good readers of Japanese did not significantly differ from the poor readers in memory for faces. Although the good readers tended to achieve slightly higher scores, the difference failed to reach significance, $p>.1$.

Further analysis involved a series of Pearson product-moment correlations computed to assess the interrelations between children's performance on the various types of materials employed in Experiments 1 and 2, and the teachers' ratings of their reading ability. That analysis revealed that performance on the Kana materials was significantly correlated with performance on the Kanji materials, $r(100)=.19$, $p<.03$, with the teachers' ratings, $r(100)=.32$, $p<.001$, and with performance on the spoken utterances employed in Experiment 1, $r(100)=.27$, $p<.01$. Performance on the Kanji materials was likewise correlated with the teachers' ratings, $r(100)=.36$, $p<.01$, and performance on the spoken utterances of Experiment 1, $r(100)=.30$, $p<.01$. Neither the correlation between memory for Kanji and teacher ratings, nor that with memory for spoken utterances was significantly different from the correlations obtained in the case of the Kana materials.

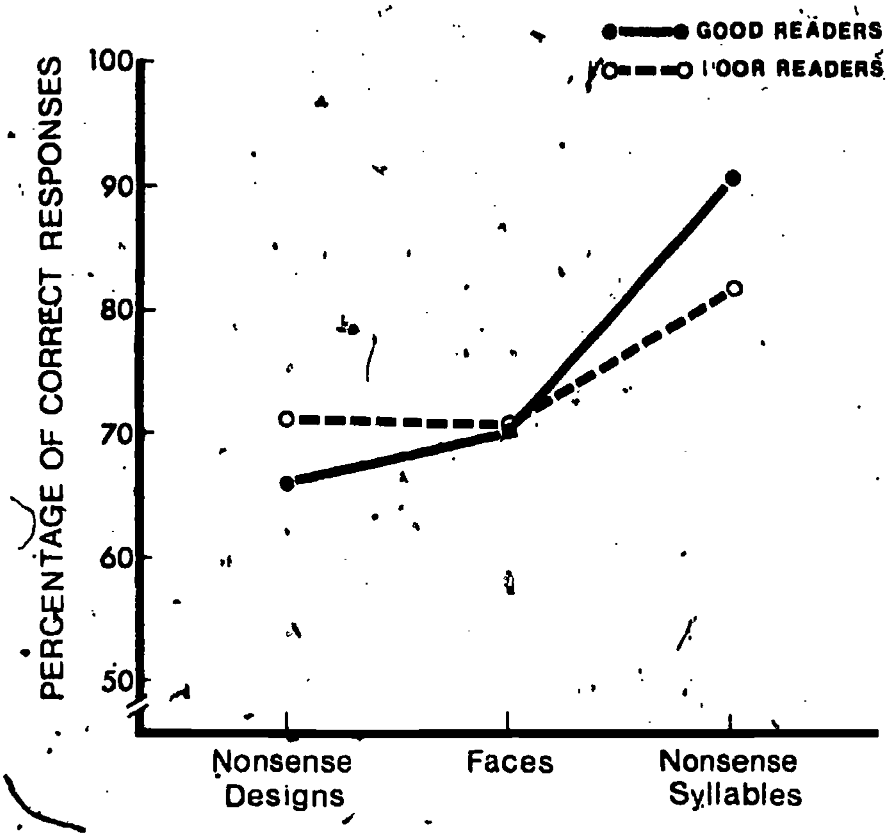


Figure 1. Mean percentage of correct responses made by good and poor readers of English on nonsense designs, American faces and printed nonsense syllables.

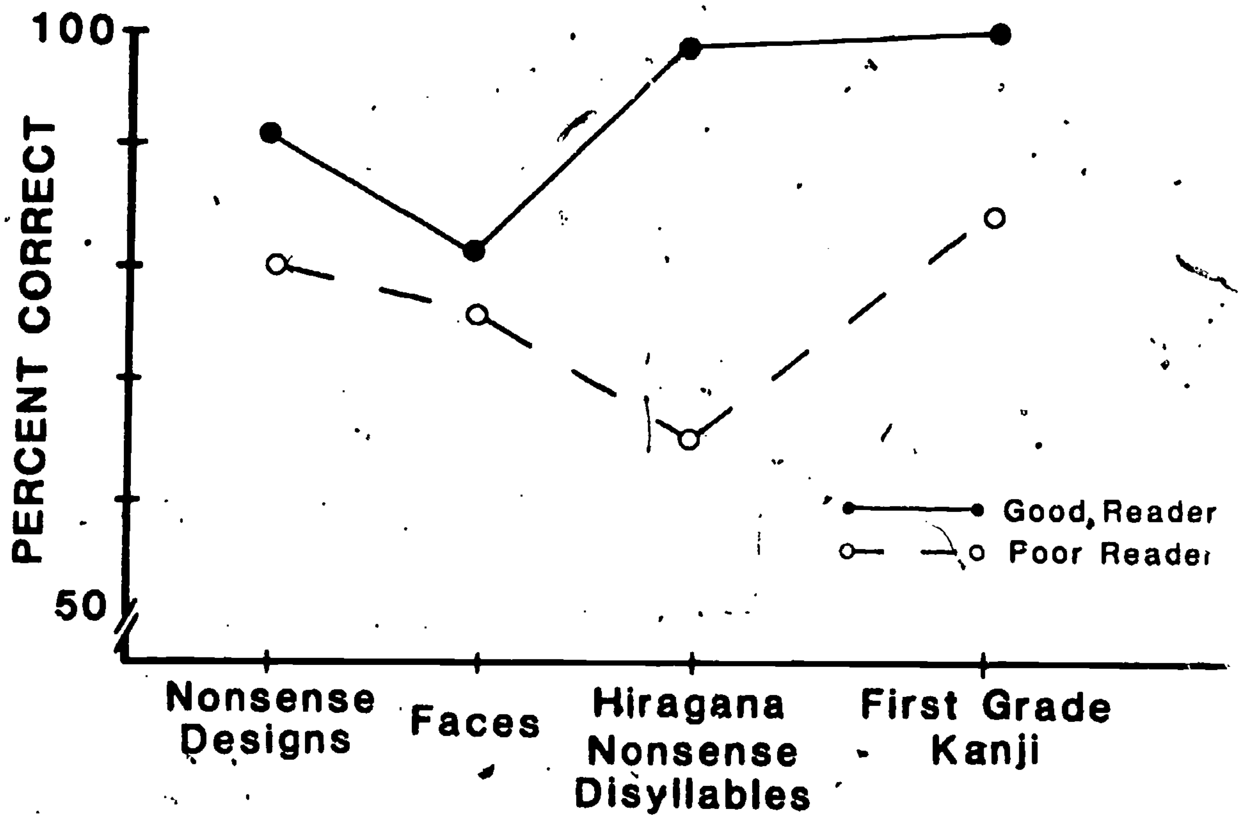


Figure 2. Mean percentage of correct responses made by good and poor readers of Japanese on nonsense designs, Japanese faces, Kana digraphs, and Kanji.

There was one difference, however, between memory for Kanji and that for Kana. Memory for Kanji significantly correlated with memory for the nonsense designs, r(100)=.18, p<.03, whereas memory for Kana did not (p>.1). Thus, the differences between good and poor readers' performance on the abstract designs would appear to be more closely associated with their differences in memory for Kanji than with that for Kana.

Further analyses of the data revealed that memory for the abstract designs was correlated with the teachers' ratings, r(100)=.36, p<.01. It was also correlated with memory for the faces, r(100)=.26, p<.01. Neither memory for Kana nor memory for Kanji correlated with memory for faces (p>.1), and all other correlations failed to reach significance at the .05 level of confidence.

Discussion

Different components of memory may accomplish temporary memory for linguistic and nonlinguistic materials. As is evident from the present results, and those obtained in previous studies of American children (Liberman et al., 1982; Mann & Liberman, 1984), children who possess superior skills in one domain need not possess superior skills in another. Skill in remembering faces, for example, may have little to do with skill in remembering printed or spoken words, which agrees with what is known about the memory skills of adults (Woodhead & Baddeley, 1981).

The present study concerned the types of temporary memory skills that are most pertinent to children's ability to learn to read Japanese, a language whose orthography comprises a syllabary (Hiragana) and a logography (Kanji) instead of an alphabet. The possibility that different memory skills might make different contributions to early reading success follows from findings about American children learning to read an alphabetic orthography (Liberman et al., 1982; Mann, 1984; Mann et al., 1980). Children who are good beginning readers of English tend to surpass poor beginning readers in use of phonetic representation (see Mann, 1984, for a review). However, as documented in Figure 1, good beginning readers of English do not surpass poor beginning readers in memory for abstract visual designs or in memory for faces. The benefits of superior use of phonetic representation, and the relative neutrality of nonlinguistic memory skills, could reflect the narrow demands of learning to read a phonetic transcription, or the broader demands of learning to read any written representation of spoken language. It was to decide between these alternatives that the present study was conducted. It sought to gain the broader perspective on reading acquisition that is available through study of children learning to read a syllabary and a logography instead of the alphabet.

It has been claimed that all children learn to read Japanese (except those markedly deficient in intelligence), and that reading difficulty is a problem peculiar to Western children learning to read the alphabet (Makita, 1968, 1974). A recent study, however, offers evidence that reading disabilities may occur as often in Japan as in America (Stevenson et al., 1982). Thus it would appear that early reading difficulty can occur in syllabaries and logographies as well as in the alphabet. With this finding in mind, the present study focused on a population of second-grade children whom their teachers rated as good, average, or poor in reading ability.

Confirmation that the children rated as good readers were truly better readers than the children rated as poor readers can be had from the data that appears on the right side of Figure 2. The children who were good readers remembered both the Kana and Kanji materials significantly better than the poor readers, which would be consistent with the fact that they possess a superior ability to read each type of character. A further finding about the reading ability of children in each group is evident in Figure 2. The poor readers encountered more difficulty in remembering the Kana digraphs, whereas the good readers were equally accurate on the Kana and Kanji materials. The fact that the Kana digraphs were nonsense materials could be the source of the poor readers' problems with these materials, as could be the fact that children had not had any practice in reading and remembering such materials. Alternatively, Kana may have been problematic because learning to decode an orthography that transcribes the abstract phonological subcomponents of words could be more demanding than acquiring one that transcribes language at the level of the word. Future research is necessary to clarify the basis of poor readers' difficulties with the Kana digraphs.

Let me now turn to the question of whether linguistic memory skills are related to success at learning to read Kana and Kanji. The answer is affirmative: children who differ in reading ability tend to differ in the ability to remember spoken nonsense words as well as in the ability to remember Kana and Kanji. Moreover, their memory for the nonsense syllables was equally related to their memory for Kana and Kanji, which implies that the importance of phonetic representation is not limited to orthographies that involve some type of phonological transcription. The implication, then, is that effective use of phonetic representation characterizes superior beginning readers, whether they are learning to read an alphabet, a syllabary, or a logography. Reading success in all orthographies may be influenced by the ability to recode orthographic material into a phonetic representation, which is consistent with findings that mature readers do so whether the language they read employs an alphabet, or not (see, for example, Conrad, 1964; Levy, 1977; Slowiaczek & Clifton, 1980; Tzeng et al., 1977).

With respect to the relation between memory for orthographic and nonlinguistic visual materials, there are commonalities between beginning readers of English and Japanese, but there is also an interesting difference. The commonality is that, like American children, Japanese children who differ in reading ability tend not to differ in memory for faces. Thus it cannot be concluded that good readers possess a superior memory, in general. The further implication is that at least one aspect of nonlinguistic memory skill may not be particularly relevant to success in learning to read any orthography. The difference between beginning readers of Japanese and English is that the good readers of Japanese surpassed the poor ones in memory for the abstract designs.

Two observations suggest a plausible explanation of the orthography-specific relationship between reading ability and memory for abstract designs. The first is that, as can be seen from a comparison of Figures 1 and 2, Japanese children tended to surpass American children in memory for the nonsense designs (a mean score 58 items correct, as compared to 49 items correct, respectively). It may be the case that the Japanese children employed a more effective strategy for remembering these materials. The second observation concerns the nature of this hypothetical strategy. During testing, I noted that, unlike American children, many Japanese children attempted to trace the designs with their fingers or even with a motion of their head. My hypothesis is that they were encoding the design into a graphomotor representation, in much the same way that they might encode an unfamiliar Kanji character when the teacher first presents it in class. A graphomotor coding strategy could be encouraged by the teacher's instructions in Kana and Kanji, as these always involve presenting characters as a sequence of strokes that the children must copy and memorize. Applied to the abstract designs, a graphomotor coding strategy can explain why memory for the Kanji correlated with that for the designs. (However, one would have to explain the absence of a correlation between memory for nonsense designs and that for Kana digraphs.) It could further account for the correlation between the teachers' ratings of reading ability and performance on the nonsense designs. Finally, it can account for the superior performance of the Japanese children, in general: the American children, by virtue of their education, would have been less likely to make systematic use of a graphomotor coding strategy as a means of remembering the nonsense designs, and therefore would have been less successful.

If these arguments are accepted, the implication is that good beginning readers of Japanese, in addition to making more effective use of phonetic representation, may also make more effective use of graphomotor representation. Further research is needed to confirm the use of a graphomotor strategy by the Japanese children, and whether graphomotor coding continues to characterize skilled readers of Japanese beyond the early elementary grades. Findings that the mature reading of Kanji and Kana is more disrupted by damage to the left, language dominant, hemisphere than by damage to the right (Sasanuma, 1975; Sasanuma & Fujimura, 1971) would seem inconsistent with a view that some nonlinguistic coding strategy is fundamental to skilled reading of Japanese. Perhaps coding strategies change with age and reading experience, or perhaps the Kimura figures are processed differently by all Japanese subjects.

To return to the major findings of this study, two experiments have revealed that, for second graders who are learning to read Japanese, use of phonetic representation in temporary memory is pertinent to the ability to read well; and to the ability to remember Kana and Kanji materials for a brief period of time. Memory for Kana is related to memory for spoken nonsense words, but not to memory for nonlinguistic materials such as faces and abstract designs. In contrast, memory for Kanji is related not only to memory for spoken nonsense words, but also to memory for nonsense designs, which would seem to imply partial reliance on a graphomotor coding strategy. It can be concluded that acquisition of both Kana and Kanji, like that of the alphabet, makes demands on the linguistic component of temporary memory. Acquisition of Kanji, however, contrasts with that of other orthographies, insofar as it makes an additional demand on a nonlinguistic component. The outcome of having to master both Kana and Kanji is that, for the beginning reader in Japan, both phonetic and graphomotor memory abilities are associated with early reading success.

References

Alegria, J., Pignot, E., & Morais, J. (1982). Phonetic analysis of speech and memory codes in beginning readers. Memory & Cognition, 10, 451-456.

Brady, S., Shankweiler, D., & Mann, V. (1983). Speech perception and memory coding in relation to reading ability. Journal of Experimental Child Psychology, 35, 345-367.

Byrne, B., & Shea, P. (1979). Semantic and phonetic memory in beginning readers. Memory & Cognition, 7, 333-338.

Conrad, R. (1964). Acoustic confusions in immediate memory. British Journal of Psychology, 55, 75-84.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory. Journal of Verbal Learning & Verbal Behavior, 19, 450-466.

Ellis, H. D. (1975). Recognizing faces. British Journal of Psychology, 66, 409-426.

Hung, D. L., & Tzeng, O. J. L. (1981). Orthographic variations and visual information processing. Psychological Bulletin, 90, 377-414.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. Journal of Experimental Child Psychology, 32, 474-484.

Kimura, D. (1963). Right temporal lobe damage. Archives of Neurology, 8, 264-271.

Kleiman, G. M. (1975). Speech recoding in reading. Journal of Verbal Learning and Verbal Behavior, 14, 323-339.

Levy, B. A. (1977). Reading: Speech and meaning processes. Journal of Verbal Learning and Verbal Behavior, 16, 623-638.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading, and dyslexia. Baltimore, MD: University Park Press.

Liberman, I. Y., Mann, V. A., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and non-linguistic material in relation to reading ability. Cortex, 18, 367-375.

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.

Makita, K. (1968). Rarity of reading disability in Japanese children. American Journal of Orthopsychiatry, 38, 599-614.

Makita, K. (1974). Reading disability and the writing system. In J. E. Merritt (Ed.), New horizons on reading. Newark, DE: IRA Press.

Mann, V. A. (1982). Reading skill and language skill. Haskins Laboratories Status Report on Speech Research, SR-69, 151-170.

Mann, V. A. (1984). Reading skill and language skill. Developmental Review, 4, 1-15.

Mann, V. A. (in press). Why some children encounter reading problems: The contribution of difficulties with language processing and linguistic sophistication to early reading difficulty. To appear in B. Wong & J. Torgessen (Eds.), Learning disability: Some new perspectives. New York: Academic Press.

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory: Can they presage early reading problems? Journal of Learning Disabilities, 17, 592-599.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. Memory & Cognition, 8, 329-335.

Mann, V. A., Shankweiler, D., & Smith, S. T. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. Journal of Child Language, 11, 627-643.

Mark, L. S., Shankweiler, D., Liberman, I. Y., & Fowler, C. A. (1977). Phonetic recoding and reading difficulty in beginning readers. Memory & Cognition, 5, 623-629.

Meyer, D. E., Schvaneveldt, R., & Ruddy, M. (1975). Loci of contextual effects on visual word-recognition. In P. M. A. Rabbitt (Ed.), Attention and performance V (pp. 98-118). London: Academic Press.

Milner, B., & Taylor, L. (1972). Right hemisphere superiority in tactile pattern recognition after cerebral commissurotomy: Evidence for nonverbal memory. Neuropsychologia, 10, 1-15.

Sasanuma, S. (1975). Kana and Kanji processing in Japanese aphasics. Brain and Language, 2, 369-383.

Sasanuma, S., & Fujimura, O. (1971). Selective impairment of phonetic and non-phonetic transcription of words in Japanese aphasic patients: Kana vs. Kanji in visual recognition and writing. Cortex, 7, 1-18.

Shankweiler, D., & Liberman, I. Y. (1976). Exploring the relations between reading and speech. In R. M. Knights & D. K. Bakker (Eds.), Neuropsychology of learning disorders: Theoretical approaches. Baltimore: University Park Press.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. Journal of Experimental Psychology: Human Learning and Memory, 5, 531-545.

Slowiaczek, M. L., & Clifton, C. (1980). Subvocalization and reading for meaning. Journal of Verbal Learning and Verbal Behavior, 19, 573-582.

Stevenson, H. W., Stigler, J. W., Lucker, G. W., Lee, S., Hsu, C., & Kitamura, S. (1982). Reading disabilities: The case of Chinese, Japanese, and English. Child Development, 53, 1164-1181.

Tzeng, O. J. L., Hung, D. L., & Wang, W. S-Y. (1977). Speech recoding in reading Chinese characters. Journal of Experimental Psychology: Human Learning and Memory, 3, 621-630.

Vellutino, F. R., Steger, J. A., DeSetto, L., & Phillips, F. (1975). Immediate and delayed recognition of visual stimuli in poor and normal readers. Journal of Experimental Child Psychology, 19, 223.

Warrington, E. K., & Shallice, T. (1969). The selective impairment of auditory verbal short-term memory. Brain, 92, 885-896.

Woodhead, M. M., & Baddeley, A. D. (1981). Individual differences and memory for faces, pictures, and words. Memory & Cognition, 9, 368-370.

131

COGNITIVE PROCESSES IN READING: WHERE DEAF READERS SUCCEED AND WHERE THEY HAVE DIFFICULTY*

Vicki L. Hanson

The act of reading involves the recognition of individual words and the integration of the meanings of those words for text comprehension. The present paper reports on studies with deaf college students, focusing first on their word recognition processes and then on their short-term memory representation of those words.

The approach of this paper will be to focus on analytic reading, in which the reader takes advantage of the linguistic information reflected in the orthography and performs a grammatical analysis on the words of a sentence, thus leading to comprehension (Mattingly, 1980). While some have taken the position that reading need not involve such linguistic mediation, there is a great deal of evidence in the literature indicating that such analytic processing promotes acquisition of reading among beginning readers and facilitates reading (especially of difficult material) for more advanced readers (Gleitman & Rozin, 1973; Liberman, 1983). Evidence of this sort provides the motivation for focusing on analytic reading in the present paper.

The orthography of English is an alphabetic writing system that reflects the morphophonemic structure of the language (Chomsky & Halle, 1968; Klima, 1972; Venezky, 1970). Hearing college students exploit this structure in the reading of words, even in the reading of those words that are familiar (Brooks, 1977; Massaro, Taylor, Venezky, Jastrzembski, & Lucas, 1980). Similarly, deaf college students are sensitive to orthographic structure (Hanson, 1983; Hanson, Shankweiler, & Fischer, 1983), and they take advantage of this structure to facilitate word recognition (Hanson, 1982b, 1983). For example, in a study in which deaf students were presented printed letter strings that were orthographically regular (e.g., REMOND, SIFLET) or orthographically irregular (e.g., RDEMNO, EFLSTI), these deaf students were found to recall letters of the orthographically regular strings more accurately than those of the irregular strings (Hanson, 1983). Similar results were obtained in an experiment investigating the recognition of fingerspelled words in which deaf adults were asked to report the letters of fingerspelled strings that were orthographically regular (e.g., S-N-E-R-G-L-I-N) or orthographically irregular (e.g., F-T-E-R-N-A-P-S). They more accurately reported the letters of the regular strings (Hanson, 1982b).

This superior performance on the regular strings suggests that skilled deaf readers, like skilled hearing readers, are sensitive to orthographic

---

*To appear in D. S. Martin (Ed.), Cognition, education, and deafness: Trends in research and instruction. Washington, DC: Gallaudet College Press.

Acknowledgment. Preparation of this article was supported by National Institute of Neurological and Communicative Disorders and Stroke Grant NS-18010.

structure. As further support for this suggestion, nearly all of the incorrect letter reports in the experiment on fingerspelling were found to be orthographically permissible. Particularly striking was the finding that for the orthographically irregular strings, many of the errors in reporting letters tended to result from the subjects' attempts to regularize the spelling of these strings. For example, in recalling the string F-T-E-R-N-A-P-S, some subjects omitted the letter that made the sequence irregular and wrote fernaps. Others added a letter to make the sequence regular and wrote afternaps, while others rearranged the letters of the sequence and wrote ferntaps.

So far, this discussion has concentrated on reading at the level of the single word. But, in addition to the ability to deal with the structure of individual words, reading requires holding words and their order of arrival in memory long enough to permit sentence comprehension. Short-term memory studies have been used to examine the nature of the internal representation (or code) used by deaf readers to mediate this comprehension process.

In studies of short-term memory with deaf college students, two primary findings have emerged. The first has been that these students, particularly the better readers, tend to use a speech-based code in the short-term retention of printed English words (Hanson, 1982a; Lichtenstein, in press). These results are consistent with Conrad's (1979) finding that the better deaf readers among high school age students tend to use a speech-based code. These results extend Conrad's work, however, in an important way: While the students tested by Conrad attended schools that were strictly oral in their educational approach, the college students tested in these more recent studies have had manual language experience, some even being native signers of American Sign Language. The second finding to emerge from the short-term memory studies with college students has been that deaf readers have difficulty in using a speech code. Even deaf readers who do use it, use it less efficiently than hearing readers (Hanson, 1982a; Lichtenstein, in press).

Given this difficulty in using a speech-based code, why might the better adult readers tend to prefer it over a manual code? A partial answer to this question is suggested by research on the retention of serial order information. Since English is a language in which word order carries critical syntactic information, the retention of word order during sentence comprehension is essential. In an experiment comparing the memory of deaf and hearing college students for sequences of printed English words, deaf college students had poorer recall only when they were required to recall the words in their order of occurrence; the deaf students were comparable to the hearing students when recall of order was not required (Hanson, 1982a). Thus, the deaf students had specific difficulty in retaining information about the order in which words were presented. The extent of this difficulty appears to be related to deaf individuals' ability to use a speech code; in tests of short-term memory, those students with the larger memory spans have shown the greatest use of a speech code (Conrad, 1979; Hanson, 1982a; Lichtenstein, in press). These results suggest that the retention of word order information depends on the ability to use a speech code.

In summary, this research suggests that deaf college students are quite proficient at using the orthographic structure of words in word recognition, but that even these readers experience persistent difficulties in the short-term memory processes that mediate comprehension. The difficulties appear related, at least in part, to inefficient use of a speech-based code.

The question remains as to the nature of the speech-based representation used by deaf readers and how this representation is developed. Deaf readers could acquire information about a speech-based code from the orthography or through speaking and lipreading. It may be the case that deaf readers' ability to use some form of speech-based code is not well reflected in the intelligibility ratings of their speech. These intelligibility ratings are based on listeners' ability to understand the deaf speakers' utterances, not on the deaf individuals' ability to utilize speech in reading. Further research needs to be directed at determining how an effective speech-based code might be acquired by deaf individuals for the purpose of reading.

References

Brooks, L. (1977). Visual patterns in fluent word identification. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading* (pp. 143-181). Hillsdale, NJ: Erlbaum.

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

Conrad, R. (1979). *The deaf school child*. London: Harper & Row.

Gleitman, L. R., & Rozin, P. (1973). Phoenician go home? (A reply to Goodman). *Reading Research Quarterly*, *8*, 494-501.

Hanson, V. L. (1982a). Short-term recall by deaf signers of American Sign Language: Implications for order recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8* 572-583.

Hanson, V. L. (1982b). Use of orthographic structure by deaf adults: Recognition of fingerspelled words. *Applied Psycholinguistics*, *3*, 343-356.

Hanson, V. L. (August, 1983). *Access to orthographic regularities by deaf readers*. Paper presented at the meeting of the American Psychological Association, Anaheim, California.

Hanson, V. L., Shankweiler, D., & Fischer, F. W. (1983). Determinants of spelling ability in deaf and hearing adults: Access to linguistic structure. *Cognition*, *14*, 323-344.

Klima, E. S. (1972). How alphabets might reflect language. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye* (pp. 57-80). Cambridge, MA: MIT Press.

Liberman, I. Y. (1983). A language-oriented view of reading and its disabilities. In H. Myklebust (Ed.), *Progress in learning disabilities* (Vol. 5, pp. 81-101). New York: Grune & Stratton.

Lichtenstein, E. H. (in press). The relationships between reading processes and English skills of deaf college students: Parts I & II. *Applied Psycholinguistics*.

Massaro, D. W., Taylor, G. A., Venezky, R. L., Jastrzembski, J. E., & Lucas, P. A. (1980). *Letter and word perception: Orthographic structure and visual processing in reading*. Amsterdam: North-Holland.

Mattingly, I. G. (1980). Reading, linguistic awareness and language acquisition. Paper presented at the International Reading Association Seminar, Linguistic Awareness and Learning to Read, Victoria, B.C., June, 1979. *Haskins Laboratories Status Report on Speech Research, SR-61*, 135-150.

Venezky, R. L. (1970). *The structure of English orthography*. The Hague: Mouton.

DISCOVERING MESSAGES IN THE MEDIUM: SPEECH PERCEPTION AND THE PRELINGUISTIC INFANT*

Catherine T. Best†

To parents and those who work with infants, one of the most remarkable developments during the first year of life is the rapid growth of vocal communication skills prior to language. Initially, the infant communicates by cries, but during the second half-year, the speechlike sounds of babbling emerge. By twelve months, babbling not only conveys feelings and needs to others but may also express infants' observations of regularities in the events and objects of their world.

The sounds that infants make are but one facet of their progress toward verbal communication. More hidden from our view, yet also important, is their perceptual grasp of the speech around them. In this chapter on infant speech perception, speech will be considered as the medium through which language is expressed vocally, much like the sounding of musical instruments is the medium through which a symphony is expressed. Both language and symphony are structurally complex systems, with many levels of concurrent organization, which are reflected in the organization of the medium. Speech carries the multiple messages that can be conveyed verbally, and hence carries information about the complex structural organization of vocal communication, which includes not only the structure of words and sentences but also broader aspects, such as stress, conversational rhythm, and voice characteristics (e.g., speaker gender, age, identity, and emotional state). It also carries information about finer grained structures such as consonants, vowels, and syllables, and the vocal gestures that produce them.

A listener's knowledge about the organization of vocal communication sets limits on which messages can be recognized in speech. That is to say, one

---

*In H. E. Fitzgerald, B. M. Lester, & M. W. Yogman (Eds.), Theory and research in behavioral pediatrics, Vol. 2 (pp. 97-145). New York: Plenum Press, 1984.

†Also Wesleyan University

Acknowledgment. I would like to thank Steven S. Braddon, Kristine S. MacKain, Roy Pea, Carol Fowler, Michael Studdert-Kennedy, and the editors of this book for their thoughtful comments on an earlier version of this chapter. Special thanks are due to Kristine S. MacKain for her helpful discussion about intermodal speech perception by infants, and to Steven Braddon for discussion and encouragement throughout the project. Work on the chapter was supported in part by a Spencer Young Faculty grant and Biomedical Research Grant 42-USC 241 awarded through Teachers College; the work was initiated while the author was supported by an NIH Fellowship (NS-05877) completed at Haskins Laboratories.

must know or learn which structures to listen for within the wealth of information that the medium reflects about the events forming a vocalization (stated in the spirit of E. J. Gibson, 1977, and J. J. Gibson, 1966). Although infants may recognize some aspects of nonlinguistic structures in vocal communication, they are limited in their recognition of language structure as such. To use any language, then, they must still discover the existence and meaning of many of the messages in speech, particularly those of words and phrases. But where do they start and how do they proceed in their discoveries during the first year? For infants who have not yet discovered the word, what messages are perceptually available to them in speech that might expedite that discovery?

The central concern of this chapter is with the nature of messages that prelinguistic infants may hear in human speech, at the level of the finer grained structures that we know as consonants and vowels. This issue has been addressed in the last twelve years of research on infant speech perception. Two basic themes about the information that infants perceive in speech have emerged. Both themes presume that some innate mechanism(s) of the auditory perceptual system fully account for infant speech perception, thus implying a mechanistic view that the young perceiver's role in seeking information in speech is rather passive. According to the first theme, infants possess species-specialized perceptual mechanisms that are tuned to linguistic contrasts among phonemes, those individual consonants and vowels we adults often associate with letters or letter combinations in words. The other theme proposes that infant speech perception is shaped by the auditory system's response to acoustic components of the stimulus; that is, the perceptual process is stimulus-bound and intrinsically neutral with respect to speech versus other sounds. These neutral acoustic attributes include the bits of noise and frequency changes, interspersed with silent gaps and humming or buzzing, which comprise a physical description of the speech signal.

It will be argued here that neither theme adequately explains how infants perceive the speech they normally hear during development. In their stead, the features of a third perspective based on ecological considerations (see also Fowler, Rubin, Remez, & Turvey, 1980; Summerfield, 1978) will be outlined, which posits a more active, information-seeking role for the perceiver. This alternative view is that infants actively attend for information in the speech medium about the natural forces that structured it, particularly how it was shaped by the human vocal tract. For the sake of simplicity, this perceptual focus on how speech is structured by its vocal source will be referred to as speech source perception. This term is offered rather than articulatory perception (see Studdert-Kennedy, 1981a; Summerfield, 1978), in order to encompass not only the articulatory gestures of the mouth and tongue but also the anatomical structure of the human vocal tract and its variations according to speaker characteristics (e.g., sex, age, and emotional state). As will be argued, a theory focused on the vocal tract sources of the speech medium's acoustic structure has greater potential than the other two themes for explaining how and why infants might begin to develop language based on the speech they hear. In short, it would provide the infant a more direct avenue by which to discover and produce words.

But how is this vocal source information conveyed in speech? Simply stated, for now, the acoustic properties of sounds are determined by the structure and movements of the sound-making object, including the human vocal tract (Fant, 1960; Flanagan, 1973). Thus, speech carries information about vocal configurations and gestures (Cooper, 1981; Dudley, 1940; Paget, 1930); the speech source view proposes that this vocal tract information is available to perception. This view will be explained in more detail, with support from recent research with adults and infants. It will also be suggested that the specialization of the human left cerebral hemisphere for language reflects an attunement to detect information in speech about the articulatory gestures of the speaker's vocal tract. The chapter concludes by noting that speech source perception is only one contribution to the infant's development of language. In order to discover words and develop language, infants must also learn about the broader aspects of language from the natural context in which speech occurs.

1. Setting the Context

1.1 Language and the Prelinguistic Infant

Prelinguistic infants, by definition, do not yet produce true words; that is, their vocalizations apparently do not refer to objects and events in the way that the words of the adult language community do. It should be noted that it is not possible to draw a sharp chronological division between prelinguistic and linguistic periods in the development of either speech production or speech perception. Generally, however, the first year of life is considered to be prelinguistic.

Two complementary and interdependent questions provide a guide for understanding the prelinguistic antecedents of language development: What is vocal language that a prelinguistic infant may come to know it? and What is the prelinguistic infant that she or he may come to know vocal language? (adapted from McCulloch, 1965). The next few sections will focus on the former question, to frame the subsequent discussion of infant speech perception research. They will describe the basic characteristics of speech that are important for understanding the task facing a prelinguistic perceiver. Once the stage has been set, the three theoretical views about the way infants process speech will be described in greater depth.

1.2 What Is Vocal Language

What type of information or messages does speech carry that prelinguistic infants might perceive? Prelinguistic infants do not yet produce words, nor do most infants under 9-10 months yet comprehend spoken words (e.g., Lenneberg, 1967). Thus, we should not expect younger infants to perceive any information that is defined by word meanings. Nevertheless, some coherent information in human speech must be available to prelinguistic infants, for they do eventually discover words.

To discover words in the speech directed toward them, presumably infants would, in part, have to (a) disembed from continuous speech the recurring subpatterns that become familiar words; (b) recognize the invariance in the pattern of a word, across the variations in acoustic detail that occur when it is produced by different speakers or in different contexts; and (c) recognize the relevant differences that do specify meaningfully different patterns. And in order to produce words, they would also have to recognize how to imitate or approximate subpatterns from a language-user's speech, even though the acoustic output of their own smaller and differently proportioned vocal tracts differs substantially from that provided by their older models (Goldstein, 1979; Liberman, Harris, Wolff, & Russell, 1971).

In language research with adults, phonemes (consonants and vowels) have often been considered to be the building blocks of words. According to that perspective, the achievement of the perceptual tasks previously listed would seemingly be founded on perception at the phonemic level of speech. To date, most infant speech perception research has focused on phonemes or their combinations in syllables, on the apparent assumption that perception of these subword units must be precursory to the perception of words.

Language users easily recognize words and phonemes when listening to conversational speech. However, these recognitions are no small feat for prelinguistic listeners, who lack a language system that could help them solve some apparent puzzles in adult speech perception. The source of these puzzles, which lies in the acoustic characteristics of speech, is discussed next. (For more extensive discussions of adult speech perception, see Fowler et al., 1980; Liberman, 1982; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Pisoni, 1978).

1.3 The Puzzles in the Acoustic Shape of Speech

To imagine the infant's difficulty, recall listening to a stretch of conversation in an unfamiliar language. Foreign speech typically seems like a relatively continuous flow of sounds, in which the quality of some phonemes may be unfamiliar (e.g., the /r/ of French or Spanish), and the boundaries of individual words often may be undecipherable. This can make it difficult for a listener to recognize a foreign word uttered in different sentences and by different people. The infant's problem is compounded because, in contrast to a language user, infants presumably do not know what words are, so this concept cannot guide their discovery of word boundaries in the flow of conversational speech. Similarly, a language user may have difficulty recognizing the precise qualities of an unfamiliar foreign phoneme when it is uttered in different words and by different people. Yet, relative to mature language users, to infants all of the phonemes occurring even in their native language environment would be comparatively unfamiliar. Infants also presumably lack certain concepts about the linguistic role of phonemes that may guide the language user's recognition of individual phonemes in conversational speech.

1.3.1 Acoustic continuities and discontinuites in running speech. One reason a sentence spoken in an unfamiliar language sounds indivisible is that utterances in natural conversation are a fairly continuous stream of sound. This is partly attributable to the cohesive intonation, or pitch contour of the voice. But it results also from the vocal-tract movement trajectories that interconnect the adjacent words in sentences or phrases. In conversation, speakers rarely pause between words, instead usually moving in connected fashion from one to the next, just as a runner usually adjusts to changes in terrain or direction without pausing between step cycles. Sometimes neighboring words in informal speech even become contracted (e.g., "what are you ..." becomes "wadaya ..." or "whatcha ..."). Thus, the raw acoustic properties of conversational speech do not always reveal clear boundaries between words.

Conversely, the vocal tract can make other relatively rapid adjustments that do cause obvious acoustic discontinuities. These breaks can occur within as well as between words, however. For instance, in the word "so" there is a rather sudden change from the lack of vocal-cord vibration during the voiceless /s/ to the onset of vibration for the voiced sound /o/. This causes an acoustic break between the noiselike, aperiodic hiss of the /s/ and the voiced acoustic periodicity of the vowel. The paradox is that a knowledgeable listener perceives these discontinuites as an integral part of a word, where appropriate, rather than as breaks between words.

The speech properties just discussed can be seen in the spectrogram in Figure 1. A spectrogram is one way of visualizing the acoustic components of speech. As indicated earlier, these acoustic characteristics are determined by the structure and movements of the vocal tract (Fant, 1960; Flanagan, 1973).

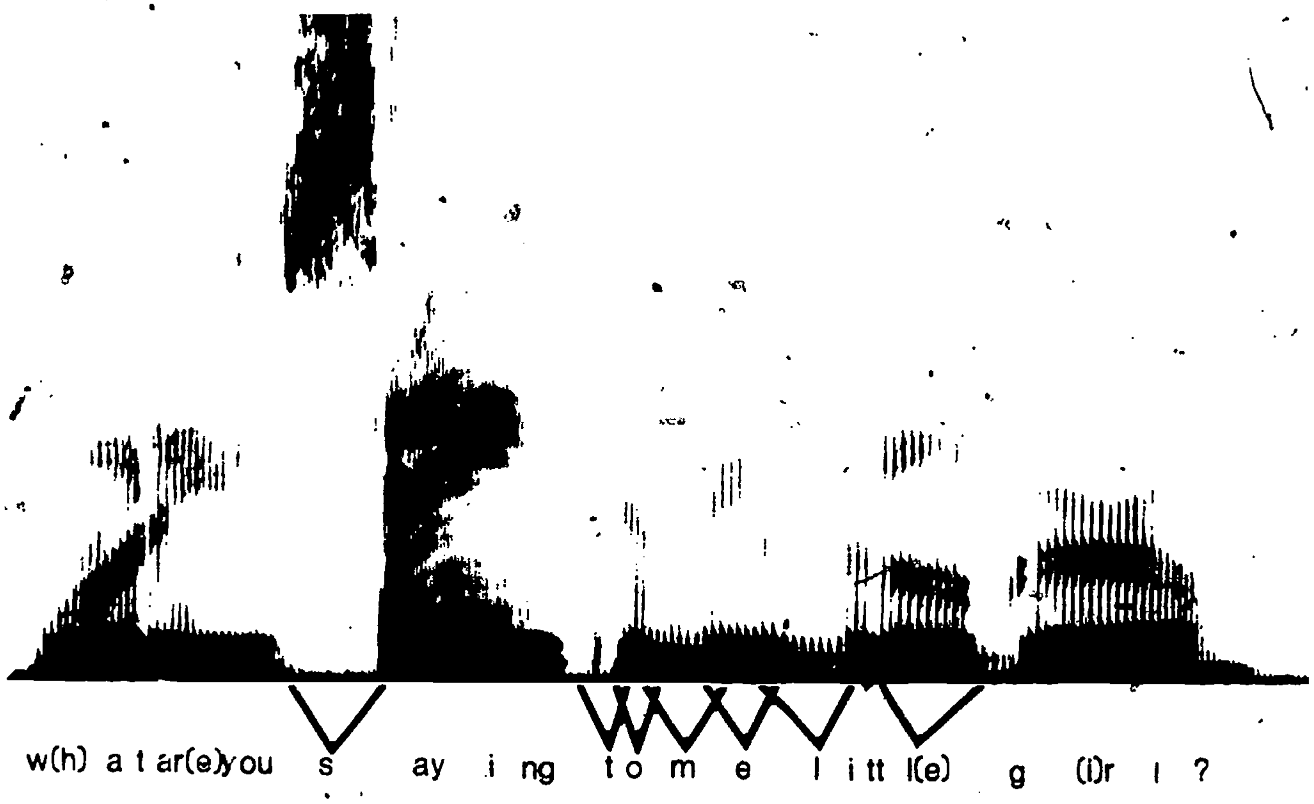


Figure 1. Spectrogram of the sentence "What are you saying to me, little girl?" spoken by a man to a young baby.

The spectrographic analysis shows the relative acoustic intensity (darkness level) of the frequency components in the speech signal (ordinate) as they change over time (abscissa). The wide, horizontally varying bars of increased density within the dark vertical striations are called <u>formants</u>, the lowest being referred to as the first formant (F1), and they correspond to the time-varying resonant frequencies of various relatively hollow spaces or chambers in the vocal tract. In the vowel "ee," for example, a small resonating chamber is formed at the front of the mouth between the edges of the tongue blade pressed against the upper teeth and the close approach of the soft palate to the base of the tongue, while a relatively large resonating chamber forms at the back of the mouth behind the base of the tongue. This results in a low-frequency F1 and a high-frequency second formant (F2).

The vertical striations in which the formats appear represent the individual energy pulses emitted by each vocal-fold vibration. The more closely packed the striations are, the briefer the periods between pulses and hence the higher the pitch of the voice. In Figure 1, the man raised his voice pitch substantially for the word "saying" and then dropped it for "(to) me little," raising it again toward the end of "girl." This degree of pitch modulation is more exaggerated than normal, and often occurs when parents talk playfully to their babies (Kaye, 1980). The dappled, nonstriated patches represent aperiodic acoustic noise produced by air turbulence at some point in the vocal tract as with the tongue-tip constriction near the upper front teeth for the /s/ of "saying."

1.3.2 Continuities and discontinuities among and within phonemes. Since infant speech research has focused primarily on phonemes, the sentence in Figure 1 is printed beneath the spectrogram for reference. The letters for the vowels and consonants are roughly lined up under the midpoint of their portion of the acoustic signal. The match between phoneme and acoustic information is only approximate, however, because of inherent difficulties in determining the acoustic span of a phoneme. At times, discontinuities appear to fall within rather than between phonemes, as was true at the level of words in a sentence. For instance, the /t/ in "to" encompasses an aperiodic noise burst, the following brief period of breathy aspiration, and the subsequent rapid transitions of the formant frequencies (i.e., the upglide at onset of the lowest formant and downglides in the pale higher formants). Again, the knowledgeable listener may paradoxically perceive acoustic discontinuities as integral to a unitary phoneme.

Another difficulty with matching phonemes to acoustic segments derives from the temporal overlap of adjacent phonemes. Speakers do not simply finish one phoneme and, at precisely that time, begin the next. Interconnected trajectories characterize the neighborly relations not only of words, but also of vowels and consonants, for example, the trajectory in "me" from the lips being closed for /m/ to the lips being open and tongue blade high for "ee." The structure of the vocal tract does not permit an instantaneous change from one configuration to another, just as a runner's leg cannot instantaneously change from flexed to extended position. In the sample sentence, not only is a discrete boundary missing between the words "to" and "me," but there is none within "to" to define the end of /t/ and the beginning of the vowel. Yet a perceiver familiar with the language can identify individual phonemes as well as individual words.

The trajectories between target vocal-tract configurations, however, account only partially for the acoustic interconnection between phonemes. Often, vocal-tract adjustments for a phoneme begin one to several segments ahead of it, or persist one to several segments beyond. In other words, there is some coarticulation among nearby phonemes (e.g., Bell-Berti & Harris, 1979; Fowler, 1980). While pronouncing the /t/ in "too," a speaker usually is already rounding his lips appropriate for the "oo." This lip-rounding is not a standard property of /t/--for the /t/ in "tee," the corners of the lips are instead pulled back slightly for the following "ee."

1.3.3 Phonetic context effects and acoustic variability. Coarticulation among phonemes causes the acoustic characteristics of any item to be assimilated to its neighbors. The articulatory difference between the two /t/'s results in "too" beginning with a somewhat lower frequency noise burst than "tee." The paradox or puzzle is that, although the perceiver recognizes an invariant identity for a vowel or consonant across various phonemic contexts, there is no clearcut invariance in its raw acoustic properties.

Movement trajectories also contribute to this acoustic variability problem. Their shapes are determined by the vocal configurations they interconnect and there are rarely definable boundaries in conversational speech between a static configuration and a trajectory into or out of it. Figure 1 indicates that, because of the difference in the surrounding phonemes, the first and the last /l/ in "little" are acoustically different, both in the flanking formant trajectories and in the exact frequencies of the flatter formants midway through the "segment."

1.3.4 <u>Vocal tract variations and perceptual normalization</u>. Not illustrated in the figure is a broader problem of acoustic variability: the acoustic contextual variation caused by different speakers. Of importance are the differences found between males and females, or between children and adults. On the average, female vocal tracts are smaller than those of males, which biases the acoustics of female speech toward higher frequencies in voice pitch and in formant frequencies. More important, though, are the age and gender differences in proportional relations among vocal-tract areas. The ratio between the distance from the vocal cords to the base of the tongue, versus the distance from the lips to the base of the tongue, is greatest for adult males and smallest for young infants (Goldstein, 1979; Lieberman, Harris, Wolff, & Russell, 1971).

Because formant frequencies are determined by the sizes of the vocal resonating chambers, these vocal-tract ratio differences cause age and gender differences in the proportional relations among formant frequencies for a given vowel. It has been impossible thus far to derive a simple mathematical formula for the formant frequency relations of a vowel produced by proportionally differing vocal tracts. In other words, there is no invariant acoustic description of formant frequency relations across men's, women's, and children's utterances of the vowel (Bernstein, 1981; Broad, 1981; Kent & Forner, 1979). The puzzle is that listeners, at least those familiar with the language, immediately hear the vowel's identity across a variety of vocal tracts differing in size and proportion. This perception of constancy in the face of speaker-specific acoustic variations has been referred to as the vocal-tract <u>normalization</u> problem.

1.3.5 <u>Summary</u>. These acoustic properties thus pose a number of difficulties for the perceptual capture of words or phonemes from conversational speech, even for adults listening to their native language. These difficulties can cause a sentence spoken in an unfamiliar language to sound like a rather undivided flow, when the listener is not prepared to handle them. However, when adults listen to their native language, they can identify discrete phonemes and words. Most likely, this is because they already know the phonemes and many of the words in their language, as well as the permissible ways by which items of either type can combine. Infants, on the other hand, do not have this knowledge of language. Yet they must be able to "solve these puzzles" in perceiving speech in order to ultimately discover words, since the words directed to infants are usually embedded in phrases or sentences (Kaye, 1980) and are presented to them by a variety of people in different speech contexts. What might infants perceive in speech, at the level of the phoneme, that could help them to recognize discrete words within the flow of sound? For consideration of this question, the discussion now turns to the infant speech perception literature.

2. The Foundations of Infant Speech Perception Research

Research on infant speech perception has largely been guided by two theoretical approaches to one overriding issue and its underlying assumptions, as indicated earlier. Although the questions and discussion presented thus far have been oriented around the eventual discovery of words by infants who initially have very limited knowledge about vocal communication, this has not been the major issue in research and theory on infant speech perception. Instead, the primary theoretical focus has been on how infants solve the acoustic puzzles of phoneme perception (e.g., Jusczyk, 1981a, 1981b). Its main underlying assumptions have been that (a) the basic perceptual unit in speech is phoneme-sized, (b) the speech percept derives from intraperceiver transformation(s) of the acoustic properties of the stimulus, and (c) the source of the transformation is an innate mechanism of the auditory system.

Much of the research has been generated by a controversy over the nature of the transformation(s) and supporting mechanism, which can be traced to a similar controversy in experimental work on adult speech perception. On one side is the phonetic interpretation of infant speech perception, which posits that the perceptual mechanism is uniquely human by nature and differs qualitatively from the means for perceiving other sounds. Proposals about the exact properties of the specialized phonetic mechanism have ranged from a comparator that matches incoming speech sounds to the neuromotor commands for producing them (the motor theory of speech perception: Liberman et al., 1967) to innate categories of linguistic features of phonemes (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971) that may be mediated by innately tuned neural feature-detectors (e.g., Cutting & Eimas, 1975).

On the other side of the controversy is the psychoacoustic approach, which presumes the machinery for the speech-to-percept transformation to be neither uniquely human nor limited to the perception of speech. In the psychoacoustic view, the general organization of the mammalian (or primate) auditory system yields an invariant stimulus-bound response whenever a given acoustic property occurs, regardless of the class of sound (e.g., speech vs. nonspeech) to which the individual is listening.

In the following summary and interpretation of the literature, it will be argued that the psychoacoustic-phonetic controversy in infant speech perception research is misguided. Both views are inadequate because they fail to consider the relation between infant and language. Following that review and discussion, a promising alternative theoretical perspective on infant speech perception will be described: the ecologically motivated (e.g., J. J. Gibson, 1966; Summerfield, 1978) speech source perception view outlined earlier. But at this point, the issue that has guided existing infant speech perception research, the psychoacoustic-phonetic controversy, must be placed in proper historical perspective.

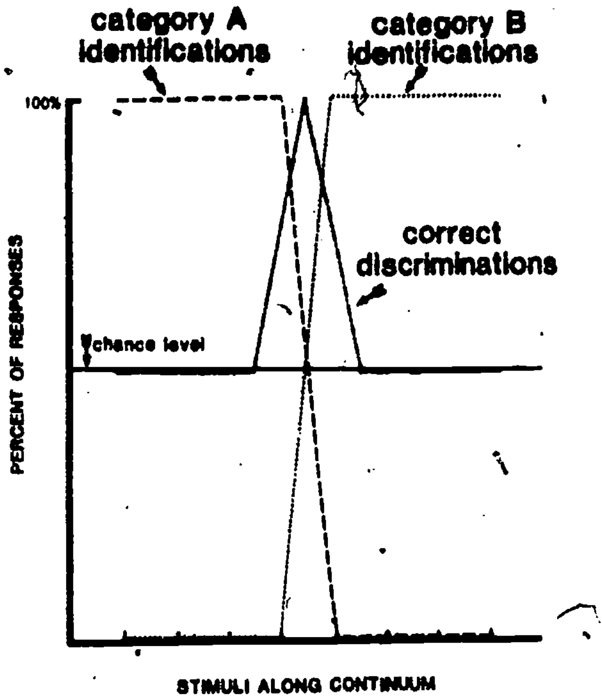
2.1 The Empirical Beginnings

Research on infant perception of phonemes began with two reports in 1971, both of which gave a phonetic interpretation to the underlying processes. Each study employed a variant of the habituation paradigm to map the limits of infants' phoneme categories via their discriminations of syllable pairs differing in initial consonant. One of the studies found that 5- to 6-month-olds can discriminate natural utterances of /ba/ and /ga/ (Moffitt, 1971). The consonants in the tested syllables are both voiced stop consonants (along with /d/; the voiceless stops are /p/, /t/, and /k/), but they differ in place of articulation. Since the infants discriminated between consonants that differed solely in place of articulation, the author concluded that "linguistic-perceptual capacities are present during early life" (p. 717).

The other study (Eimas et al., 1971) has received the preponderant attention in subsequent research. In this study, 1- and 4-month-olds were presented with computer-synthesized versions of /ba/ and /pa/, which differ in the articulatory property called <u>voice onset time</u> (VOT). That is, in the voiceless /p/ of /pa/, the vocal cords begin vibrating later with respect to the lip-opening gesture than is the case with the voiced /b/ of /ba/. Therefore, /p/ has a longer VOT than /b/. The acoustic consequences of articulatory differences in VOT are many (Lisker, 1978), but research attention has primarily focused on the time difference between the consonantal noise burst and the onset of periodic voicing, referred to here as <u>acoustic</u> VOT. It is usually confounded with other acoustic differences between a voiced-voiceless consonant pair.

Computer synthesis was used in the Eimas et al. study to produce a systematic series or continuum of syllables, which varied in equal-sized steps along the acoustic VOT dimension. Such acoustically controlled continua usually cannot be produced by a human speaker because of mechanical constraints on possible vocal tract movements (although in the case of stop voicing, human speakers can produce a range of different acoustic VOT values). Adult listeners typically fail to hear the gradual steps of acoustic change along synthetic continua between two contrasting consonants. Instead, they identify all stimuli as exemplars of one or the other phoneme category, and a sharp boundary on the continuum separates the two perceptual categories. Adult listeners also discriminate between acoustically different pairs of synthetic stimuli much better when the members are from different phoneme categories than from the same category. This pattern of identification and discrimination results has been termed <u>categorical perception</u> (see Figure 2). It was originally taken as evidence for a specialized phonetic mode of perception, since the nonspeech continua that had been similarly studied were perceived continuously (i.e., no clear labeling boundary or no clear performance peak in discrimination ability) rather than categorically (e.g., Liberman et al., 1967; Repp, 1982).

To determine whether infants also perceive speech categorically, and by presumption phonetically, Eimas et al. (1971) tested their discrimination of synthetic syllables that differed in acoustic VOT, but either did or did not differ according to American English /pa/ versus /ba/. The within-category pairs differed by the same magnitude of acoustic VOT as did the between-category pairs. The infants discriminated the between-category difference much better than the within-category difference, in agreement with the adult phoneme boundary. This finding led the authors to conclude that infants have an innate capacity to perceive speech linguistically; that is, in terms of adult phoneme categories.



CATEGORICAL PERCEPTION

Figure 2. A schematic diagram of ideal results of categorical perception tests.

These two early studies were scientifically intriguing, and encouraged a profusion of infant speech research that still continues. Some researchers essentially accepted the premise that the means by which acoustic properties are translated to a percept is speech-specialized, that is, specifically phonetic in nature. They went on to explore the range and nature of the phoneme contrasts that young infants discriminate. As the psychoacoustic-phonetic controversy indicates, however, other researchers took issue with the phonetic perspective. The studies of this latter group have attempted to show that categorical perception is due to general psychoacoustic mechanisms, which respond to particular acoustic attributes whether they appear in speech or nonspeech sounds, but which occur most frequently in speech (e.g., Blumstein, 1980; Stevens & Blumstein, 1978).

2.2 Infant Perception of Phoneme Contrasts

Many of the early studies employed synthetic syllable continua, since explaining categorical perception has been a key interest in the psychoacoustic-phonetic controversy. According to the definition of categorical perception, percentage-correct discrimination of stimuli from a continuum must show a performance peak at the position of the category boundary for labeling of those stimuli (see Figure 2) (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). Therefore, both identification and discrimination data are needed to assess categorical perception. However, because there is no currently acceptable test for infants' identification of sounds, the conclusions of the infant research are based on discrimination data only (see Jusczyk, 1981a). Since the infant discrimination data cannot be compared to infant identification responses, it may be better to refer to the peaks and troughs in their discrimination of synthetic speech continua by some other term, such as perceptual boundary effect (Kuhl, 1981b; Wood, 1976).

A number of other infant studies from the phonetic perspective have employed natural rather than synthetic speech stimuli, thus focusing on discrimination of contrastive spoken exemplars rather than on categorical discrimination. In both types of study, the discrimination data have been derived from either habituation tests or tests of generalization of conditioned operant responses, and the subjects have usually been between 1 and 7 months of age (prior to the reported onset of word comprehension). The consonant contrasts that were tested have usually occurred in syllable-initial position, which makes phoneme discrimination in natural speech easier for young children than do medial or final positions (Schvachkin, 1973).

The categorical discrimination studies have found that infants show a boundary effect for a variety of synthetic consonant contrasts, with their discrimination peak usually occurring at or near the adult American English identification boundary. These contrasts include stop consonant voicing (specially, /p/ vs. /b/, and /d/ vs. /t/) as cued by acoustic VOT (Eimas, 1975a; Eimas et al., 1971; Streeter, 1976) or, by another naturally occurring cue, the extent of frequency change in the first formant (F1) transition (J. L. Miller & Eimas, 1981).

Infants also show a boundary effect for place of articulation differences between stop consonants (/b/ vs. /d/ vs. /g/) (C. L. Miller, & Morse, 1976; C. L. Miller, Morse, & Dorman, 1977; Morse, 1972; Williams & Bush, 1978), even when the consonants occur in the middle of vowel-consonant-vowel (VCV) syllables (Jusczyk & Thompson, 1978) or in the final position of VC or CVC syllables (Jusczyk, 1977). Infant boundary effects have also been found for place of articulation distinctions between liquid consonants (/r/ vs. /l/, Eimas, 1975b). There is an infant boundary effect for place of articulation differences among fricatives (voiceless: /f/ vs. "th" as in "thanks"; voiced: /v/ vs. "th" as in "that"), although the infants' discriminations may differ in some respects from adults' (Jusczyk, Murray, & Bayly, 1979). Infants also show a boundary effect for manner of articulation differences between /b/-/w/ (same place of articulation, but stop vs. semivowel manner: Eimas & Miller, 1980a; Hillenbrand, Minifie, & Edwards, 1979) and between /b/-/m/, although they do not discriminate the latter contrast as categorically as adults, that is, they show moderate discrimination of the within-category pairs (Eimas & Miller, 1980b).

Adults perceive synthetic continua between vowel contrasts in a more continuous or less categorical fashion than consonants, unless the vowels are severely shortened in duration (e.g., Crowder, 1973; Liberman et al., 1967; Pisoni, 1973, 1975). Infants show similar effects in perception of the "ee-ih" vowel distinction (Swoboda, Morse, & Leavitt, 1976; Swoboda et al., 1978).

Consistent with the boundary effect findings, infants discriminate a fairly wide array of the natural consonant and vowel contrasts in spoken English. They discriminate natural stop voicing distinctions (Trehub & Rabinovitch, 1972), as well as some contrasts that children often do not produce correctly until late in phonological development (3-5 years), and which often cause persistent articulatory difficulties: certain fricative place (/s/-"sh") and voicing (/s/-/v/) contrasts (Eilers & Minifie, 1975; Eilers, Wilson, & Moore, 1977), the place contrast between the liquid consonants /w/-/r/ (Eilers, Oller, & Gavin, 1978), and the consonant clusters /sl/-/spl/ (Morse, Eilers, & Gavin, 1982). Infants also discriminate the naturally produced vowels "ee" versus "ih" (Eilers et al., 1978), as well as the "ee-ah-oo" triad, whether they occur in isolation or in CV syllables (Trehub, 1973).

To summarize, young prelinguistic infants discriminate many consonant contrasts and often show a boundary effect akin to the adult category boundary. They even discriminate some consonant contrasts that children produce late and often misarticulate. Infants also discriminate vowel contrasts, and do so less categorically than consonants, again like adults. (For more extensive discussion of the theoretical particulars, see Aslin & Pisoni, 1970; Cutting & Eimas, 1975; Eilers, 1980; Eilers & Gavin, 1981; Eimas, 1974a, 1975a; Jusczyk, 1981a, 1981b; Kuhl, 1978, 1980, 1981a; Mehler & Bertoncini, 1978; Morse, 1978; Trehub, Bull, & Schneider, 1981; and Walley, Pisoni, & Aslin, 1981).

The performance pattern does not, however, indicate whether infants discriminate by psychoacoustic or phonetic means. Initially, researchers who took a phonetic view assumed that the boundary effect was evidence for a speech-specialized perceptual process. But that assumption could be questioned, and was submitted to test by researchers on both sides of the theoretical dichotomy. The alternative posed by the psychoacoustic perspective was, of course, that the perceptual boundary might be an attribute of the auditory system's response to the acoustic properties, rather than the phonemic identities, of the speech sounds.

2.3 Is the Boundary Effect "Phonetic" or "Psychoacoustic"?

A direct test of this question was to see whether infants show a discrimination boundary effect for some phoneme-differentiating acoustic property even when it occurs outside a speech context. Both Morse (1972) and Eimas (1974b, 1975b) isolated the major acoustic cue that had been manipulated to produce a place of articulation continuum, by stripping away the other acoustic properties that were shared by the contrasting phonemes, and presented young infants with the isolated cue continuum to discriminate. Morse (1972) tested infants with the isolated F2 for the /ba/-/ga/ contrast, whereas Eimas tested the isolated F2 cue for /da/-/ga/ (1974b), and the isolated F3 cue for /ra/-/la/ (1975b). Although in each case the isolated formants were the sole acoustic property that distinguished the phoneme contrast, and hence were crucial to adult categorical perception of that contrast, outside of their natural context they sounded like nonspeech "bleats."

The argument was that if the infant boundary effect reflects a uniquely speech-related perceptual specialization, it should occur in the perception of a particular acoustic cue only when that cue actually specifies a phoneme distinction. The infants in the Morse and Eimas control studies did show a boundary effect for the full syllables but failed to show one for the isolated formants, and in fact discriminated the latter very poorly. The authors interpreted their findings as support for the phonetically specialized nature of infant boundary effects.

The psychoacoustic perspective offered an alternate interpretation, however. The crucial stimulus attribute for psychoacoustically based boundaries could be the interrelation between the distinctive acoustic cue and the nondistinctive information provided in the other formants (e.g., Jusczyk, 1981a). Such a relational attribute would be destroyed by presenting the distinctive cue in isolation. Therefore, the Morse and Eimas studies could not definitively answer the controversy.

A more appropriate nonspeech control should maintain the interrelations among the acoustic features involved in a phoneme distinction. One such nonspeech distinction is the difference in risetime, or time from onset of a sound until it reaches its maximum intensity, between plucked versus violin-like sounds, which is analogous to the "sha-cha" distinction in speech. Adults had been reported to perceive the pluck-bow distinction categorically, even though it is nonspeech (Cutting & Rosner, 1974). Jusczyk, Rosner, Cutting, Foard, and Smith (1977) extended that finding to infants and presented it as evidence that infant perceptual boundary effects have a psychoacoustic basis. However, subsequent replications of the adult study uncovered a stimulus problem. The acoustic differences among the original pluck-bow stimuli were not of equal magnitude throughout the continuum; when this source of acoustic discontinuity was removed, adults no longer perceived pluck-bow categorically (Rosen & Howell, 1981). Since Jusczyk et al. employed the original pluck-bow continuum, the infant findings must be questioned (Jusczyk, 1981a, 1981b).

A second infant nonspeech study was subsequently run, using tone onset time (TOT) differences between the individual tones of a two-tone chord, which is an analogue for the acoustic VOT distinction in speech (Jusczyk, Pisoni, Walley, & Murray, 1980). Adults show a sharp TOT boundary in line with their boundary for acoustic VOT (Pisoni, 1977). Thus, the phonetic uniqueness of adult categorical perception has been called into question by the TOT results, along with similar reports on other nonspeech contrasts (e.g., J. D. Miller, Wier, Pastore, Kelly, & Dooling, 1976). In the Jusczyk et al. (1980) study, the infants discriminated TOT differences nearly categorically, leading the authors to conclude that earlier reported VOT boundary effects may not be unique to speech perception by infants either.

Nonetheless, the TOT findings also fail to offer a definitive choice between the phonetic and the psychoacoustic interpretations of infant perceptual boundaries. In contrast to adults, the infants failed to discriminate TOT as categorically as acoustic VOT, and the position of their TOT boundary differed significantly from their acoustic VOT boundary. Jusczyk et al. (1980) claim that this does not damage the general psychoacoustic stand, since TOT only partly captures the articulatory VOT distinction (i.e., the psychoacoustic key could be some other, untested acoustic attribute of articulatory VOT). However, the developmental data are at odds with this logic. The TOT and acoustic VOT boundaries do match for adults, indicating that a perceptual change must occur between infancy and adulthood. Yet the infant and adult VOT boundaries match. Therefore, it is the TOT boundary that changes developmentally, and not the VOT boundary, in contradiction to the claim that infants come to perceive distinctions via psychoacoustic means (Jusczyk, 1981b).

In any event, the evidence from the infant nonspeech perception research is weak. The currently available nonspeech control studies present a theoretical stalemate between the phonetic and psychoacoustic claims about the nature of infant speech perception. However, the proponents of the controversy have argued that the answer may lie elsewhere in the premises of the psychoacoustic-phonetic distinction. It might be settled by exploring whether phonetic perception is uniquely human, a claim made by the phonetic perspective that is rejected by the psychoacoustic perspective.

2.4 Is the Boundary Effect Uniquely Human?

Empirical attacks on the claim that humans are the sole possessors of categorical phoneme perception have involved assessing whether other mammals or primates show abrupt shifts in perceptual sensitivity around the phoneme boundaries. It was reasoned that, if animals showed a boundary effect, general psychoacoustic factors must then account for categoricity in speech perception, since by definition infrahumans cannot perceive in a humanly specialized manner. The relevance of this research to infant speech perception is that infants and animals are nonusers of language, but only infants are human and have the capacity to develop human language.

Researchers of animal speech perception have reported boundary effects, similar to those found with infants, for discrimination of the stop consonant voicing distinction by chinchillas (South American rodent) (Kuhl, 1981b; Kuhl & Miller, 1975; J. D. Miller, Henderson, Sullivan, & Rigden, 1978), and by rhesus monkeys (Waters & Wilson, 1976) and for the stop consonant place of articulation distinction by monkeys (Morse & Snowden, 1975; Sinnott, Beecher, Moody, & Stebbins, 1976). Chinchillas also discriminate the vowels "ah" and "ee" (Burdick & Miller, 1975); as do dogs (Baru, 1975), and exhibit boundary effects in go- no-go categorizations of voicing among stop consonants, with their boundaries falling at the position of human adult boundaries (Kuhl & Miller, 1978). Thus, the psychoacoustic interpretation is that the perceiver's knowledge of language is not a necessary precondition for the boundary effect, and apparently neither is membership in the human species.

Researchers who support the psychoacoustic view of speech perception have used the animal findings to propose the following picture of the evolution of speech perception and production: the mammalian auditory system has specialized notches in sensitivity for certain regions of certain acoustic dimensions. These psychoacoustic specializations placed selective pressures on the choice of phoneme contrasts by human languages. The production of the phonemes chosen must have capitalized on just the acoustic domains that are most neatly suited to mammalian psychoacoustic specializations (Kuhl, 1981b; Kuhl & Miller, 1975, 1978; J. D. Miller, 1977; Stevens, 1972). By extension, human infants possess those same psychoacoustic sensitivities (Aslin & Pisoni, 1980; Jusczyk, 1981a, 1981b; Kuhl, 1978; Walley, Pisoni, & Aslin, 1981), and their attention is thus captured by the acoustic attributes of the language in their environment.

However, the claims of this psychoacoustic proposal belie the clarity of the infrahuman data. The animal data fail in several ways to match those of human adults. Recall the earlier argument that a determination of categorical perception depends on data from labeling and from discrimination tests; the animal research has necessarily relied only on discrimination data. Indeed, animal discrimination boundaries are less sharply defined than human adult phoneme boundaries (Kuhl & Miller, 1978). In other words, animals are noticeably better than human adults at discriminating within-category acoustic differences in place of articulation continuum, and worse at discriminating between-category differences, indicating lowered categoricity (Morse & Snowden, 1975). The between-species categoricity difference is consistent with the greater sensitivity of human adults to formant-onset frequency changes, by a factor of about two (Sinnott et al., 1976). In addition, only human adults show reaction time increases, large ones in fact, when making within-category discriminations (Sinnott et al., 1976). Finally, the absolute position of the human adult category boundary is more stable than the monkey boundary, in the face of variations in the acoustic range covered by a synthetic phoneme continuum (Waters & Wilson, 1976). Humans obviously do show speech-relevant perceptual specializations beyond the limits of the other mammals tested.

The animal and nonspeech research may suggest that the boundary effect is not absolutely speech-specific and species-specific. However, there are clear and unexplained differences remaining between the human adults' perception of speech and the control studies on animal speech perception and infant non-speech perception. Thus, the basic theoretical choice between the phonetic and psychoacoustic explanations of speech perception, especially in infants, is still open. If studies of the boundary effect have failed to solve the psychoacoustic-phonetic quandary, then possibly a more abstract characteristic of speech perception would (such as the perceptual constancy of phoneme identity across phonemically irrelevant acoustic variations).

2.5 Phonemic Perceptual Constancy in Infants

Recall the earlier discussion of some puzzles in the fit between speech acoustics and perception, particularly the lack of satisfactory acoustic descriptions for the invariant identity of a phoneme across different contexts of surrounding phonemes (the acoustic variability problem) or as uttered by differently proportioned vocal tracts (the normalization problem). In spite of these puzzles, adults perceive the identity of a phoneme spoken in widely different words and by different vocal tracts with seeming immediacy and effortlessness.

To see whether infants show a similar perceptual constancy, Fodor and colleagues (Fodor, Garrett, & Shapero, 1970; Fodor, Garrett, & Brill, 1975) trained them to respond operantly to a pair of vowel-differing syllables that either began with the same constant (e.g., "pee"-"poo") or began with different constants (e.g., "pee"-"kah"). In both conditions, the consonants differed acoustically because of the change in vowel context. The authors wanted to assess whether, despite the acoustic variability, the infants learned the operant response more easily for the consonantal match. The infants were later tested on a new syllable (e.g., "pah"), to determine whether they generalized the learned operant response more consistently from the consonant-matched pair to a new syllable beginning with the same consonant. The infants did learn and generalize more consistently for consonantal matches than consonantal mismatches. If they had learned to associate syllable pairs simply by remembering the pairing of their dissimilar acoustic properties, they should have responded to the mismatched-consonant pairs as consistently as to the consonant-matched pairs. The authors concluded that these prelinguistic infants had maintained perceptual constancy for consonantal identity in the face of concurrent acoustic variations, and that this ability must depend "on innately determined phonological identities" (p. 180).

There are two problems with this claim. Their task was difficult for infants to learn, and several attempted replications or extensions have failed. In addition, the psychoacoustic perspective offered an alternative interpretation: The perceptual constancy might reflect a response to some (yet uncovered) higher order acoustic invariant shared by the varying instances of a given consonant (Kuhl, 1980, 1981b). Using a different technique than Fodor et al., Kuhl and her colleagues found perceptual constancy in young infants for the vowels "ah" versus "ee" (Kuhl, 1979) and for the fricatives /f/ versus /s/ (Holmberg, Morgan, & Kuhl, 1977) across different neighboring phoneme contexts and different speakers. Thus, consistent with the Fodor et al. report, the infants solved the acoustic invariance problem. Since perceptual constancy was maintained across different speakers the infants also solved the normalization problem. But whereas Fodor et al. favored the phonetic viewpoint, Kuhl and others favor the psychoacoustic viewpoint (e.g., Jusczyk, 1981b; Walley et al., 1981), in part because chinchillas and dogs show perceptual constancy for "ah" versus "ee" across speakers and pitch contours (Baru, 1975; Burdick & Miller, 1975). Chinchillas also show such constancy for /t/-/d/ even in different vowel contexts (Kuhl & Miller, 1975).

The perceptual constancy findings thus indicate that infants can somehow solve two seemingly knotty acoustic puzzles to reach an important perceptual aspect of phoneme identity. Once again, the findings apparently do not allow a theoretical choice between the phonetic and psychoacoustic explanations of infant speech perception. Also, as will be argued in the next section, neither is the theoretical choice decided by considerations about the innateness of infant phoneme perception.

2.6 Innateness of Infant Phoneme Perception Effects

A pervasive notion on both sides of the psychoacoustic-phonetic controversy has been that a boundary effect or perceptual constancy effect is innate if infants show it "at the earliest age tested" (e.g., Aslin & Pisoni, 1980; Eimas, 1975a; Eimas et al., 1971; Jusczyk, 1981a, 1981b; Kuhl, 1978). Curiously, the empirical foundation for this belief includes almost no data before 1 month, a handful of studies on 2-month-olds, and many studies that have collapsed data across 1-4 months, 4-6 months, 6-8 months, or 10-12 months. Few have compared different ages groups (cf. Best, Hoffman, & Glanville, 1982; Eilers, Wilson, & Moore, 1977; Werker, 1983; Werker & Tees, 1982).

This view apparently assumes that "an infant is an infant is an infant," across at least the first 6 months of life. In studies that averaged over several months of age, the "earliest age" cannot be trusted since it refers to the youngest infant they tested, even though group data were reported and appropriate age analyses were almost never run (but see 1- versus 4-month age differences in Eimas et al., 1971). Thus, it is nearly impossible to assess which, if any, perceptual boundaries are innate, or presumably biologically determined and inborn. All we can note are the ages below which a given perceptual effect has <u>not</u> been shown; for this review, the conservative assumption will be that the average age tested is the correct "earliest age" to show the reported effect.

The literature offers the following observations: First, infant boundary effects matching the adult findings have been shown only by a mean of 2½ months (e.g., Eimas, 1974b, 1975b; Morse, 1972). Conversely, newborns and 1-month-olds do not discriminate VOT absolutely categorically (Jusczyk et al., 1979) nor do infants under 3-4 months discriminate phoneme contrasts under demands on short-term memory (Best et al., 1982; Morse, 1978). These data hint at a perceptual change sometime between 1 and 2½ months, which is consistent with widespread biobehavioral and social changes around 6-10 weeks (e.g., Clifton, Morrongiello, Kulig, & Dowd, 1981; Emde & Robinson, 1979; Haith, 1979). These biobehavioral changes at 6-10 weeks include vocal behavior (Oller, 1980; Stark, 1980), suggesting that early changes in speech perception should be further explored (see Werker, 1983, for an example of important age changes in speech perception by older infants).

When infant and adult categorical discrimination has been directly compared, 3-month-olds' (Eimas & Miller, 1980b) and even 7- to 8-month-olds' performance differs significantly from that of adults (Aslin, Pisoni, Hennessy, & Percy; 1981; Eilers, Wilson, & Moore, 1976). Differences from adults, in fact, persist until at least 5-6 years of age (L. E. Bernstein, 1979; Garnica, 1973; Robson, Morrongiello, Best, & Clifton, 1982; Schvachkin, 1973; Simon & Fourcin, 1978; Werker & Tees, 1981; Zlatin & Koenigsknecht, 1976). Therefore, boundary effects cannot be considered innate in some absolute sense.

Second, perceptual constancy for vowels is only certain as early as a mean of 2½ months (Kuhl & Miller, 1975). Perceptual constancy for consonants has not been reported earlier than 4 months (Fodor et al., 1975) or 6 months (Holmberg et al., 1977; Kuhl, 1980). Thus, arguments for the innateness of perceptual constancy effects (e.g., Jusczyk, 1981b) should also be held in check.

The exact timing, causes, and nature (i.e., phonetic vs. psychoacoustic) of changes in speech perception cannot be inferred from this literature, however. Even if age had been systematically studied, conclusions would still be limited by the near-exclusive reliance on discrimination measures. The coincidence of an adult phoneme boundary and a peak in infant discrimination is not sufficient evidence to claim "adultlike" perception of the contrast. As argued earlier, assessment of categorical perception requires both labeling and discrimination tests. Simple discrimination cannot reveal whether the distinction was perceived as a phoneme contrast (see also Jenkins, 1980), a concern that is equally relevant to the animal research. Discrimination indicates only that some difference was detected. Since phoneme discrimination is dissociable from phoneme category identification in aphasics (Blumstein, Cooper, Zurif, & Caramazza, 1977; Riedel, 1981), it is equivocally involved in aspects of perception that are closer to the meaning of language than mere acoustic contrast detection. Because language-dependent perceptual qualities are more likely to change developmentally, discrimination is inadequate as the sole measure of development in speech perception.

It is uncertain which, if any, speech perception effects are innate, and how they might change developmentally prior to the discovery of words. More crucial for the auditory-phonetic controversy, the following questions remain unanswered: Even if perception of a phoneme contrast is innate, is that innately possessed quality phonetic or psychoacoustic in nature? And if perceptual change does occur during prelinguistic infancy, what is the nature of the change? Each side of the controversy has provided answers (see Aslin &

Pisoni, 1980; Trehub, Bull, & Schneider, 1981; Walley, Pisoni, & Aslin, 1982; cf. Eilers, 1980; Eilers, Gavin, & Wilson, 1979; Eimas, 1975a), and the auditory-phonetic choice remains unclear. One potential motive force for early perceptual changes can be eliminated, however: if they do occur, they could not have a linguistic motivation from the infant's perspective. This fact causes difficulty for both sides of the controversy, as the next section indicates.

The data on categorical perception of nonspeech, animal perception of phoneme contrasts, perceptual constancy for phonemes, and innateness of infant phoneme perception have thus failed to decide between the psychoacoustic and the phonetic explanations of infant speech perception. When an impasse as extensive as this is reached, it is important to consider whether the difficulty is not with the research but rather with the logic of the two theoretical views themselves.

3. Questioning the Auditory-Phonetic Question

As stated earlier, the tacit assumptions shared by the two sides of the dichotomy have generally been (a) that the units of speech perception are phonemes (cf. Bertoncini & Mehler, 1981; Jusczyk, 1981a); (b) that some intraperceiver interpretive process must transform acoustic properties to phonemic percepts; and (c) that the process is mediated by some specialized neural mechanism(s). All three assumptions can be questioned, particularly in relation to the infant's discovery of words. The first assumption requires that infants segment phonemes from connected speech. Phonemic segmentation has not been assessed in infants, and is not straightforward even when the perceiver is a language-user, since young children seem unable to explicitly segment phonemes (E. J. Gibson & Levin, 1975), as are illiterate adults (Morais, Cary, Alegria, & Bertelson, 1979). The notion that infants perceive phonemes can also be questioned on a linguistic level (see discussion in the next paragraph). The second assumption entails that the listener shed meaning on the presumed meaninglessness of the superficial acoustics of speech, that is, the meaning of the stimulus resides solely in the listener and not directly in the signal. According to the third assumption, the transformation is accomplished by specialized nervous system structures or information-processing stages. Thus, the psychoacoustic and the phonetic views regard speech perception as a mechanistic intraperceiver process.

We turn now to a more detailed examination of each position in the psychoacoustic-phonetic controversy. The main problem with the phonetic view of infant speech perception is that infants presumably do not have a language system, having not yet discovered words. Phoneme contrasts cannot be perceptually available to infants, since they are defined by a language system, being dependent on word meanings in that system. They represent abstract relations among speech sounds that are used by the language to convey semantic differences, as in "pat" versus "bat." Although infants, like adults, categorically distinguish between /p/ and /b/ (e.g., Eimas et al., 1971), they do not necessarily perceive such differences as phonemic contrasts (recall that discrimination is an equivocal measure of phoneme perception).

The phonetic view also has difficulty accounting for how and why infants would adjust their perceptual categories to suit the language of their environment. Presumably, the evolutionary advantage of innate phoneme categories is that they would filter out irrelevant within-category acoustic variations and thereby relieve the perceiver of having to deal with those unnecessary details. Yet young infants fail to discriminate some phoneme contrasts existing in certain languages according to the adult categories. How and why would infants learning those languages later become able to focus on those innately filtered-out details, in order to adjust their category boundaries or develop new categories? According to researchers on genetic evolution (Jacob, 1977), on nervous system function (Rose, 1976), on perceptual development (Spelke, 1979; Trevarthen, 1979), and on speech perception (Studdert-Kennedy, 1981b), the most efficient evolutionary solution for developmental adaptation to stimulus environments, such as that provided by the native language, is not an array of innate mechanisms that are tightly tuned to specific stimulus values, but instead a more flexible attunement to detect the range of stimulus values that could occur. Thus, any specialization we have for perceiving speech would have to be sufficiently flexible to adapt to the specific phoneme contrasts of one's own particular language.

A major drawback of the psychoacoustic view is the argument that the object of speech perception is intrinsically meaningless speech-neutral acoustic data. In particular, this claim has difficulty accounting for the perceived constancy of phonemes spoken by different vocal tracts (vocal-tract normalization) and spoken in different contexts of surrounding phonemes (acoustic variability). The ability to recognize the invariance of a phoneme or word spoken in different contexts and by different people is crucial for language-learning infants. The phonemes they hear occur in a variety of phonemic contexts; the vocal tracts of the older speakers they hear (and must eventually base their own vocalizations on) differ proportionally from each other, as well as from the infant's own vocal tract. Vocal-tract normalization and acoustic variability do not appear to cause perceptual difficulties for infants. Both infants and other animals apparently solve the normalization and acoustic variability problems in their discriminations among phonemes (e.g., Baru, 1975; Kuhl, 1979, 1980, 1981b; Lieberman, 1980). The psychoacoustic view does not adequately explain the infant's perceptual solution to those problems. The nontrivial acoustic variations involved have thus far defied speaker-independent and speech-neutral acoustic definitions for either phonemes or words. Thus, it cannot be assumed that the infant's solution focuses on speech-neutral acoustic information.

A problematic implication of the psychoacoustic view is that the infant must at some time move from perceiving speech in purely auditory terms to perceiving linguistic structures, such as phonemes (Jusczyk, 1981a, 1981b). What would lead the infant to take the cognitive step from meaningless acoustics to meaning at the level of either phonemes or words? One proposal would be the empiricist philosophical perspective that infants learn meanings by contextual association. However, the associationist solution is unsatisfactory on logical grounds (e.g., E. J. Gibson, 1977; J. J. Gibson, 1966; J. J. Gibson & Gibson, 1955; Jenkins, 1974). Meaning cannot emerge from meaninglessness, so some meaningful element would have to predate the infant's first association.[1] The traditional empiricist perspective is that the elements of meaning are extrinsic to the individual, introduced by sensory stimulation. However, the psychoacoustic view assumes that meaning is not intrinsic to the speech stimulus, and it has provided no argument that other stimulation is intrinsically meaningful. In other words, the psychoacoustic view gives us no reason to believe that sensory stimulation in any modality provides extrinsic elements of meaning.

If meaning cannot be assumed to derive from extrinsic stimulation, the traditional nativist perspective offers the alternative proposal that elements of meaning are innate or intrinsic to the infant. The prime candidates for innate elements of meaning in infant speech perception, of course, would be phonemes or phoneme contrasts. This is exactly the position of the phonetic viewpoint questioned previously, and is antithetical to the psychoacoustic viewpoint because it cannot be invoked for animal speech perception. A third possibility is the constructivist perspective that the infant cognitively constructs meaning for sensory stimulation that does not itself provide intrinsic meaning. But this would still depend on some mechanism that determined the nature of the meaning to be constructed, and again the most likely mechanism for speech would be phoneme-based.

A fourth, nontraditional perspective on the source of meaning in speech can be offered, however, which is not consistent with either the psychoacoustic or the phonetic viewpoints. This is the ecological perspective (e.g., J. J. Gibson, 1966) that meaning is directly available to perception in the active, adaptive relation between the perceiver and the objects/events being perceived. It will be discussed at greater length in the subsequent sections of the chapter.

The psychoacoustic position may also be troubled by its evolutionary proposal. It assumes that specialized perceptual "notches" in the sensitivity of the mammalian auditory system along certain phonemically relevant acoustic dimensions have imposed selective pressures on the phonemes that can be uttered by the human vocal tract (e.g., Aslin & Pisoni, 1980; Kuhl, 1981b; J. D. Miller, 1977; Stevens, 1972; Walley et al., 1981). An alternative proposal is that the anatomy of the human vocal tract places quantal limits on the sounds it can make (see Stevens, 1972); it is this fact that has placed selective pressures on the evolution of specialized notches in the auditory system.

Specializations of neural tissue, like any structural specializations, are naturally selected (evolve) because they have suited some purpose. Yet the psychoacoustic model fails to specify a purpose that could have selected for the speech-related perceptual notches or discontinuities in the mammalian auditory system. Certain basic properties of the auditory system probably are shared by all mammals, reflecting selective adaptation to the commonalities in their auditory environments such as the sounds of weather, vegetation, predators, and prey. However, individual mammalian species do develop more highly specialized sensitivities for certain acoustic properties that are uniquely suited to the particulars of their own ecological niche (e.g., the bat; Neuweiler, Bruns, & Schuller, 1980). For some mammals, notably humans, species-specific vocalizations are particularly important to the species' survival, and have probably placed selective pressures on the development of specialized responsivities of the auditory system. In these species, we would expect to find that specializations in auditory sensitivity have evolved to be uniquely responsive to the vocal characteristics of the species (see also Petersen, 1981; Zoloth, Petersen, Beecher, Green, Marler, Moody, & Stebbins, 1979), which is the converse of the psychoacoustic model of evolution. In the case of speech perception by humans and animals, recall that the specialized notches or discontinuities do indeed appear to be more sensitive and finely tuned in human adults than in the animals studied (see Section 2.4).

The general principle in evolution and in ontogeny of the nervous system has been that motor functions (e.g., vocalization) precede and often motivate the development of the correlated sensory functions (Bekoff, 1981; Horridge, 1968). Consistent with that principle, motor areas develop in advance of sensory areas at each level of the neuraxis (Jacobson, 1978), including the human neocortex (Marshall, 1968; Tuchmann-Duplessis, Aroux, & Haegel, 1975). Even more important, the motor-sensory precedence applies to the neural supports for human speech perception and production: the development and maturation of Broca's area (the motor speech cortex) precedes that for Wernicke's area (the receptive speech cortex) (Rabinowicz, 1979). The more likely evolutionary scenario, then, may be that the human auditory system's properties were selected for best responsiveness to the sound-producing abilities of the uniquely human vocal tract (e.g., Stevens, 1972; Studdert-Kennedy, 1981c), rather than vice versa as the psychoacoustic view suggests.

In summary, both contemporary views of infant speech perception are flawed. The chapter's introduction pinpointed four knotty perceptual problems as requisites to discovering words in the speech stream: (a) disembedding words from connected speech; (b) recognizing word pattern invariances across different utterances and vocal tracts; (c) recognizing the variations that do specify different word patterns; and (d) hearing how to imitate a pattern made by another person. Neither the psychoacoustic nor the phonetic viewpoint offers the infant adequate means for discovering words or phonemes in speech.

But what is the alternative? The remaining sections on speech source perception focus on the organization of the speech medium for an answer. The vocal tract offers a structural and dynamic meaning that is intrinsic to speech itself, since as the source of speech it determines the shape of its acoustic product (e.g., Fant, 1973). It would be more parsimonious for perceivers to attend directly and actively to the available vocal-tract information that is intrinsic to speech, than to have the perceptual process mediated with a step involving the meaningful interpretation of meaningless superficial acoustics. According to the speech source view, meaning exists in perceiver's relation to speech at its source. This alternative approach derives from the general ecological approach to perception taken by James Gibson and his followers (e.g., J. J. Gibson, 1966; Fowler & Turvey, 1978; Studdert-Kennedy, 1981, 1981d; Summerfield, 1978; Verbrugge, Rakerd, Fitch, Tuller, & Fowler, in press).

4. An Ecological Perspective on Infant Speech Perception

Perception of the speech source implies that the infant attends to intrinsically specified information about the vocal tract and the articulatory events that shaped the speech medium. This premise is consistent with the ecological argument that perceiving organisms actively seek information about distal events, which is lawfully specified in the stimulus array (J. J. Gibson, 1966; E. J. Gibson, 1977). It stands in contrast to depictions of perception as the cognitive or neural transformation of proximal sensory data, which are intrinsically meaningless and informationally impoverished with respect to the distal event. The speech medium carries many parallel messages, some of which are defined within a particular language, and thus presumably not detected by infants, who do not yet recognize that phonemes can function to distinguish word meanings. Others are human universals, such as the paralinguistic messages of emotional affect, of regional accent, and of age or gender effects on vocal-tract size and configuration.

The speech characteristics of interest for the current discussion are the structural organization and articulatory gestures of the vocal tract. According to speech source perception, to perceive these characteristics is to simultaneously apprehend the constant anatomical structure and the transforming positions of articulators in a speaking vocal tract (see also Schubert, 1974). This proposition is supported by the speed and accuracy with which adults and even young children can imitate speech sounds, in spite of the "normalization problem" (e.g., Alekin, Klass, & Chistovich, 1962; Ferguson & Farwell, 1975; Galunov & Chistovich, 1966; Kent & Forner, 1979; see Studdert-Kennedy, 1981a). However, the proposition may seem counterintuitive in two manners. First, common sense suggests that we can see the structure and movements of objects, but that we hear only "sounds." Thus, the claim that we hear structure and movements, especially the small hidden ones of vocal tracts, may seem unlikely. Second, it may seem implausible that structure and motion are captured at once in the same information, because the qualities of form and movement seem dissociable in our experience. However, an ecological appreciation of sound and hearing shows these "problems" to be false.

4.1 Ecological Acoustics and Speech

Acoustic energy is the radiation over time and space of a wave of rapid alternations in air pressure. It originates from the oscillatory motion of some object or surface that compresses and rarefies the distribution of air molecules around it. Both an object/surface and some oscillatory motion are necessary to produce acoustic energy. In fact, their contributions to sound cannot be dissociated. Structural properties constrain the sound-producing motions an object/surface can undertake; in turn those motions temporally deform the structure in a characteristic manner. It follows that the specific properties of an acoustic flow (e.g., time-varying frequencies and amplitude changes of the pressure wave) necessarily reflect those structural properties of the object and its vibratory movements that shaped the sound-production.

By the ecological view, the interdependence between structure and transformation is at the core of real events and therefore of their perception (Shaw & Pittenger, 1977). The nature of an object is revealed to the perceiver through event-determined, co-defined information about structural invariants and transformational invariants in objects/events. These terms refer, respectively, to the structural identity of an object undergoing some transformation or change (e.g., by its movement or structural deformation, or by changes in the observer's orientation to it), and the transformations or changes it partakes in. Consider, for example, one person saying three different words as opposed to three people saying a single word. In the first instance, information common to the three words reflects the structural invariance of that speaker's vocal tract. In the second case, there is a transformational invariant in the articulation of the single word by three structurally different vocal tracts. Structural and transformational invariants lawfully shape the energy medium that carries their message (i.e., acoustical or optical energy). The ecological premise is that, through their modulation of the energy medium, transformations can perceptually specify an object's structure. In support of this, infants' visual recognition of objects and their structure is enhanced by watching the objects undergo various spatial-temporal transformations (E. J. Gibson, 1980; Ruff, 1980, 1982).

From the ecological perspective, auditory perception is the co-detection of the transformations (motions) and structure of the sound source, which are veridically conveyed in the acoustic medium (see also Schubert, 1974; Warren & Verbrugge, in press). Thus, structure and motion are not only seen but heard. In the case of speech, the acoustic medium is better suited than optics for conveying the structure and transformations of the vocal tract, many of which are invisible in face-to-face communication as well as when the speaker is out of view. Articulatory gestures may be beyond the capabilities of vision in another way, since their speed and precision exceed the temporal and spatial resolution of the visually perceived manual American Sign Language (Studdert-Kennedy & Lane, 1980).

The unseen messages carried by natural speech reflect not only the origin of its acoustic energy (respiratory and laryngeal), but also the structural identity, biokinematic coupling, and specific movements of the speaker's supralaryngeal vocal tract.[2] In vocalizations and musical sounds (among others), the acoustic wave does not radiate freely from its oscillatory origin to the perceiver: the medium is also molded by the structure and transformations of an intervening resonating tube. The size and shape of a resonant cavity determine its natural resonating frequency (or frequencies), at which the air contained within its walls will oscillate when excited by a flow of air introduced from outside. If the extrinsic air flow is already oscillating (e.g., when acoustic energy from the vibrating larynx is introduced to the supralaryngeal vocal tract), then those oscillatory frequencies in the flow that match the resonant properties of the tube will be amplified in intensity; other frequencies that mismatch the resonant properties will be attenuated (filtered out). The larger a resonating cavity is, the lower its primary resonant frequency, which is also affected by the size and number and positions of openings in the cavity. As its shape deviates from perfectly spherical or cylindrical, particularly if there are corners or "side pockets," higher order resonant frequencies may be added. Surface properties, such as the smoothness and elasticity of the resonant tube's walls, largely determine the time course of intensity changes in the resonated frequencies.

In the case of speech, the critical sound-shaping properties of the resonant tube include the shape and elasticity of the cheeks, throat, lips, and tongue, and the position and rigidity of the teeth. The properties that allow the vocal tube to transform in shape are especially important for its articulatory gestures: the hinged movements of the jaw, and the moving and deforming obstructions of the tongue, lips, and velum (which opens the nasal passage to resonate for sounds like /m/ and /n/). These structural and transformational properties all shape the acoustic speech wave. The time-varying formants and other acoustic discontinuities (see Section 1.3) reflect, in particular, rapid transformations of the vocal-tract resonant configuration that are produced by movements of the tongue, lips, jaw, and velum, within the constraints posed by the enduring anatomical relations within the tract. (For more detailed discussions of the physical acoustics of speech, see Fant, 1973; Flanagan, 1973.)

Since these distal source properties determine the acoustic shape of speech, they should be available to perceivers (see also Studdert-Kennedy, 1981a, 1981d; Summerfield, 1978; Verbrugge et al., in press). Infants and even animals should be able to detect at least some of the structural and transformational invariants in speech, although evolutionary and ontogenetic history will affect how well different perceivers are attuned to pick up the various messages. As discussed earlier, although animals and human infants do show phoneme boundary effects, their performance deviates significantly from human adults' categorical perception.

The speech source perception view may be further clarified by comparison and contrast with the psychoacoustic and the phonetic views. If auditory perception is the detection of information about source structure and transformations in the acoustic medium, then the perception of speech should be abstractly similar to the perception of other sounds, in agreement with the psychoacoustic view. However, the speech source perspective disagrees with the psychoacoustic notions that the perceiver's focus is on event- or speech-neutral acoustic parameters, and that these parameters need to be transformed by the system into percepts. The speech source proposal is also consistent in an important respect with the phonetic view. That is, speech perception even by the infant is considered "special" and uniquely human; however, that special perceptual quality is not agreed to be based on phonemes for infants. Moreover, the relative emphasis on "speech" and "perception" is different. Because the speech medium conveys its source to the perceiver, acoustic cues need not be transformed into percepts, whether by codes for phoneme categories or by neuromotor codes for phoneme production.

According to the speech source view, the specialness of speech perception derives from the unique structural and transformational properties of its sound source, the human vocal tract. It is unique in its complex anatomy, its biokinematic organization, and its particular dynamic gestures (e.g., Lieberman, 1967; Liberman et al., 1971), all of which are reflected in its acoustic productions. Moreover, humans have a privileged relation to speech as the tool of human-specific language communication.

Some of the advantages of this view for several important aspects of speech perception in adults and infants will be considered next.

4.2 Speech Source Perception in Adults

The major acoustic puzzles in speech perception research have been the problems of acoustic variability and normalization (see Section 1.3). As a reminder, the acoustic variability problem refers to the sometimes quite striking variability in the acoustic properties of an invariantly perceived phoneme, which occurs primarily when it is produced in different contexts of surrounding phonemes. The variability is caused by coarticulation among phonemes as well as by the articulatory trajectories that interconnect adjacent phonemes. The acoustic properties of the phoneme are thus assimilated to the acoustic properties of its neighbors (e.g., in Figure 1, the differences among the /l/'s in "little" and "girl"). Another source of acoustic variability is the wide variety of acoustic features that can identify a given phoneme (e.g., Lisker, 1978). These sorts of acoustic variations pose a greater empirical puzzle for psychoacoustic accounts of consonant perception rather than vowel perception. They have much stronger effects on the formant trajectories and other acoustic features associated with consonants than on the formant frequencies in the nuclear portions of vowels (although the latter are also affected to considerable degree in conversational speech).

The normalization problem refers to another acoustic context effect, caused by variations in the dimensions of different vocal tracts. It causes greater difficulties for psychoacoustic explanations of vowel perception than consonant perception. The formant frequencies in vowel nuclei are more obviously affected by vocal-tract proportions than are the formant trajectory patterns associated with consonants. The different effects of these two types of acoustic variability in vowels and consonants suggest a difference in the information those two phoneme classes convey, which has been supported by several other lines of research on speech production (e.g., Fowler, 1980; Fowler et al., 1980) and perception (e.g., Ades, 1977; Crowder, 1973; Cutting, 1974; Darwin, 1971; Pisoni, 1973; Studdert-Kennedy & Shankweiler, 1970).

That adults perceive an invariant identity underlying the acoustic variations of a given consonant (e.g., acoustic differences for /d/ in "dee" versus "dah" versus "doo") is difficult for the psychoacoustic view to explain, because it requires finding a unitary acoustic principle for the invariant percept. Although several auditory solutions have been proposed for the acoustic invariance problem, for example, perceptual "templates" for consonant-specific spectral (frequency) acoustic properties (Blumstein, 1980; Searle, Jacobson, & Rayment, 1979; Stevens & Blumstein, 1978), these do not hold up well under empirical test (Blumstein, Isaacs, & Mertus, 1982; Walley et al., 1981) or logical scrutiny (e.g., Liberman, 1982; Studdert-Kennedy, 1981a, 1981d). However, the acoustic variability is actually an advantage from the speech source view that speech acoustics convey information about the structural and transformational invariants of their vocal-tract source. That is, for the variability that derives from the coarticulation of adjacent phonemes, the form of that vocal-tract transformation should clarify rather than confuse the source properties that identify both elements. As for the variety of acoustic features that can specify a given consonant, these also result from, and thus may offer equivalent information about, the vocal-tract invariants identifying that consonant.

4.2.1 Consonant perception. Research on phoneme context effects has found shifts in category boundary positions that are predictable from the coarticulatory effects of different neighboring phonemes. For example, a continuum between "s" and "sh" can be generated by varying only the center frequency of the fricative noise. But the frequency of natural fricatives is lower if the following vowel is "oo" rather than "ah." That is because the lip-rounding for "oo," which lengthens the vocal tract and therefore lowers its resonant frequencies, is coarticulated with the fricative (e.g., Bell-Berti & Harris, 1979). In support of the notion that perceivers detect coarticulatory information, the "soo-shoo" boundary occurs at a lower frication frequency than the "sah-shah" boundary (Mann & Repp, 1980). Similar coarticulatory context effects have been found for stop consonant place of articulation differences (Mann, 1980) and for stop consonant voicing boundaries (e.g., Summerfield, 1982).

Research on the perceptual unity of multiple acoustic properties for a consonant distinction offers converging support for the speech source position. The various acoustic properties are not perceived according to their acoustic differences but are perceived instead as equivalent information about the articulation of the same consonant (Best, Morrongiello, & Robson, 1981; Fitch, Halwes, Erickson, & Liberman, 1980). These perceptual equivalences and context effects are difficult to explain by speech-neutral psychoacoustic mechanisms (see Studdert-Kennedy, 1981d), and control studies have in fact failed to find analogous effects in nonspeech perception (Best et al., 1981; Mann, Madden, Russell, & Liberman, 1981; Summerfield, 1982).

4.2.2 Vowel perception. If vowel perception is accomplished by detecting the underlying vocal-tract configuration and transformations that remain invariant across the structural variations of different vocal tracts, then vocal-tract normalization would not be a problem for speech source perception. In contrast, the psychoacoustic account posits that a singular underlying neutral acoustic description of the vowel must be derivable by some formula. No such description has yet been found because proportional differences among vocal tracts, especially male versus female versus child, prevent a uniform scaling of vowel formant frequencies among speakers (Broad, 1981); that is, formant frequency ratios are speaker-specific. Nor can the problem be solved through some formula that partials out sex and age differences based on the value of some independent acoustic feature such as fundamental frequency of the voice, which is the most obvious formant-independent feature that could differentiate those speaker characteristics. The sex and age groups show considerable overlap in fundamental frequency, a laryngeal property that is imperfectly correlated with the variation in supralaryngeal configurations that affect formant properties. Furthermore, as the latter observation suggests, the perceived sex and age of a speaker depend on supralaryngeal characteristics rather than on fundamental frequency (Lehiste & Meltzer, 1973). Thus, the psychoacoustic approach is left in an untenable position: the acoustic normalization solution would appear to depend on a priori knowledge of the supralaryngeal vocal-tract properties whose influence it is trying to circumvent.

Of further relevance to the speech source view, identification performance is better for vowels spoken in CVC context than for isolated vowels, even though formant frequencies are more clearly differentiated among the isolated vowels (Strange, Verbrugge, Shankweiler, & Edman, 1976). Likewise, similarity judgments among vowels in CVCs are clearly differentiated along three dimensions, which correspond closely to vowel articulatory factors, whereas most perceivers differentiate isolated vowels along only one or two dimensions (Rakerd & Verbrugge, 1982). These contextual effects suggest that coarticulatory information aids vowel as well as consonant perception, consistent with the ecological premise that structural invariants (vocal tract configurations) are clarified by transformational (dynamic articulation) properties (see Verbrugge, Shankweiler, & Fowler, 1980). Studies with CVC syllables whose vowel nucleus has been replaced with silence, leaving only the syllable-initial and syllable-final formant transitions (in correct temporal relation), further support the ecological interpretation. Vowel identification under these conditions is remarkably well-preserved (Strange, Jenkins, & Edman, 1977), even if each remaining piece of coarticulatory information is taken from a different-sexed speaker (Verbrugge & Rakerd, 1980). These coarticulatory influences on vowel perception are no problem for the view that we detect vocal-tract source information, but would be difficult to explain via psychoacoustic mechanisms (Fowler & Shankweiler, 1978; Shankweiler, Strange, & Verbrugge, 1977; but see Howell, 1981).

These consonant and vowel findings suit the speech source interpretation of adult speech perception. Certainly, the experimental tasks often required subjects to "recover phonemes from the speech stream" (Liberman, 1982). Since speech conveys language, the detection of "pure" (nonlinguistic) structural and transformational invariants of the vocal tract may rarely if ever be ends in themselves for adults, and instead serve as means to the linguistic ends of recognizing, for example, phonemes. This would not be the case for infants, however, since phonemes can be "recovered" from the speech stream only by listeners who already know they are there (Studdert-Kennedy, 1981d). Phonemes may indeed be at the "surface of language" (Liberman, 1982) for adults, but vocal-tract source information must be at the "surface of speech" for prelinguistic infants. We now turn to some recent research with prelinguistic infants, which offers strong support for speech source perception.

4.3 Speech Source Perception by Infants

4.3.1 Infants' perceptual constancy for speech. Perceptual constancy is at the base of the acoustic variability and vocal-tract normalization puzzles. The findings just discussed suggest that the adult's solution lies in a perceptual focus on the structural and transformational invariants of a speaker's vocal tract. Even without the adult's linguistic motivations, infants also show perceptual constancy for vowels and consonants spoken by different people or in different phoneme contexts, as described earlier in the chapter (Holmberg et al., 1977; Kuhl, 1979, 1980, 1981b). Therefore, the invariant features they apprehend must exist at the surface of speech and not only at the surface of language. The speech source view suggests that the properties of relevance to the infant lie in the vocal tract and not in the speech-neutral superficial acoustics.

A psychoacoustic interpretation of infant perceptual constancy works no better than it did for adults. We cannot assume that infants are guided by knowledge about phonemes in their solution of the two acoustic puzzles, so some independent source of guidance to the invariant acoustic features of vowels and consonants would be needed. One psychoacoustic solution to the normalization problem might be that although adults do not solve it by partialing out speaker differences based on fundamental frequency, linguistically naive infants do. However, this approach does not work, because infants as well as adults appear to rely on supralaryngeal information in the formant structure of speech rather than on fundamental frequency when perceiving speaker gender (C. L. Miller, Younger, & Morse, 1982). Nor does an acoustic template model (e.g., Blumstein) appear to give an adequate psychoacoustic explanation to the acoustic invariance problem of infants' perceptual constancy for consonants across varying vowel contexts (Studdert-Kennedy, 1981d; Walley et al., 1981).

What <u>does</u> remain constant in the utterances of a vowel or consonant by different speakers or across different phoneme contexts is the underlying similarity in vocal-tract structure and articulator positioning. Speech source information would thus seem to offer a more straightforward metric than speech-neutral acoustic invariance for infant perceptual constancy, as was argued in the case of adult speech perception. Moreover, perception of speech source information would certainly be a more direct guide than speech-neutral acoustic patterns for the infant's attempts at vocal imitation of older speakers and eventual production of words provided by her native language environment.

Thus far, the central argument that the infant perceives speech source information has been oriented around the acoustic medium. However, as indicated in the next section on infants' recognition of auditory and visual commonalities in speech, this information is provided by sight as well as by sound. The intermodal perception of speech by infants provides strong support for the speech source perception view, and is particularly difficult to reconcile with the psychoacoustic and phonetic perspectives.

4.3.2 Infants' intermodal perception of speech. When adults listen and watch someone speak, the acoustic and optic information about speech is not perceptually independent. Rather, the two seemingly disparate types of information are perceptually unified, implying that a common metric underlies them. Prelinguistic infants likewise recognize underlying commonalities between audio and visual presentations of speech that cannot be described in linguistic terms for them. The speech source perception view suggests that infants perceive intermodally by attending to the underlying articulatory events that provide the auditory and visual information.

To appreciate the contribution of visual information to speech perception, recall listening to someone speaking at the front of a room. It probably seemed easier to understand what was being said if you could also keep the speaker's face in view. This intuition has recently been empirically validated with adult listeners. Under difficult listening conditions, adults perceive speech more correctly when they can watch the speaker than when they must rely on their ears alone (Binnie, Montgomery, & Jackson, 1974; Dodd, 1977; Summerfield, 1979). These findings suggest that listeners obtain information about speech not only from the acoustic signal, but also from the optical information that results from articulatory maneuvers.

Of course, the major responsibility for speech perception is carried by the auditory modality. That blind adults successfully perceive speech, whereas the deaf have serious difficulty with lipreading, would seem to imply that auditory information is both necessary and sufficient for speech perception, although visual information plays a negligible role unless listening is particulary difficult. Speech researchers accepted this logic until recently, when MacDonald and McGurk (1978) and Summerfield (1979) reported that listeners fail to recognize phonemic conflicts in concurrent auditory and visual presentations of speech, instead perceiving a unified speech event. The percepts did not veridically reflect either the acoustic or the optic signal considered in isolation. For example, when perceivers watched a face silently articulating "ga" while a voice said "ba," they heard "da." These results indicate that in face-to-face speech perception, listening is not simply supplemented by arbitrary, learned associations between vocal-tract configurations and speech sounds. Rather, at the level of the speech event itself, the information provided by the two modalities shows an intermodal articulatory equivalence.

Two opposing views have been offered for adults' perception of acoustic-optic equivalence in speech. MacDonald and McGurk (1978) have suggested that the equivalence be described linguistically, in terms of abstract features of phonemes. The other view, proposed by Summerfield (1979) and based on the Fowler et al. (1980) ecological interpretation of speech production findings, argues that the equivalence is nonlinguistic and modality-free, arising from the dynamics of articulation. The first view has been criticized, in part, because it accounts for only a limited number of the speech percepts that result from audiovisual conflict (see Summerfield, 1979, for detailed discussion of these and other criticisms).

In light of recent findings, the latter view appears to best account for how prelinguistic infants perceive speech intermodally. Infants' sensitivity to acoustic-optic equivalences in speech has been demonstrated under two conditions. Under the first condition, infants were presented with acoustic-optic speech displays in which the overall synchrony and the specific articulatory details presented in the two modalities of information were confounded. In one study, $2\frac{1}{2}$- to 4-month-old infants saw the mirrored reflection of a woman's face repeating nursery rhymes; the auditory signal was either in synchrony or delayed relative to the optic presentation by 400 milliseconds. The infants watched the reflection significantly longer when the visual and the auditory presentations were in natural synchrony (Dodd, 1979). In a second study, -month-old infants viewed two women speaking, in two adjacent video films, while the concurrent speech of one woman was played over a central loudspeaker. Infants preferred to look at the face that talked in synchrony to the audio speech presentation (Spelke & Cortelyou, 1981).

It is unclear whether the infants were responding to the general synchrony and/or the specific articulatory details of the optic and acoustic displays, however, since these two aspects of audiovisual match were confounded in both experiments. They might have only recognized the overall synchrony, for example, for syllable onsets, between the speech seen and heard. However, they might also have preferred watching the natural acoustic-optic concurrence of specific articulatory gestures. For example, infants might prefer to look at a speaker's lips being rounded and protruded for the production of the vowel "oo" as they heard an audio "oo," as opposed to looking at the speaker's lips being opened wider to produce "ah."

This prediction was recently tested experimentally (Kuhl & Meltzoff, 1982; MacKain, Studdert-Kennedy, Spieker, & Stern, 1981). Kuhl and Meltzoff (1982) presented 4- to 5-month-olds with two adjacent films of a woman's face synchronously articulating the vowels "ah" and "ee," while one of those vowels was presented auditorily and in synchrony over a central loudspeaker. The infants preferred to watch the film whose articulatory details specified the vowel presented auditorily. In the MacKain et al. study, disyllables (e.g., "mama," "lulu") were presented audiovisually, under similar experimental conditions, to 5- to 6-month-old infants. The infants looked significantly longer at the video display whose articulatory dynamics matched the acoustic presentations, for the disyllables "mama," "baby," and "zuzu." These findings indicate that young infants recognize at least some auditory-visual equivalences of articulatory gestures, and are not only sensitive to general synchrony. Moreover, this intermodal recognition was accomplished in the presumed absence of a language system, making linguistically based explanations (e.g., MacDonald & McGurk) untenable for this age group. The commonality that the infants recognized between the acoustic and optic information may best be described in nonlinguistic terms (Summerfield, 1978).

Given that the infant seems attuned to detect vocal-tract source information in speech, what may be the organization of the supporting perceptual system? A consideration of the biological basis of this attunement should move us closer to understanding the ease with which humans recognize auditory-visual equivalences in articulatory details, and the infant's apparent ease in learning to speak a first language. Research with adults suggests that the answer lies in the functional asymmetries of the left- and right-cerebral hemispheres of the human brain.

4.4 Left-Hemisphere Attunement for Articulatory Information

For the adult, the left-cerebral hemisphere shows a specialized advantage for the perception of speech, in contrast to a right-hemisphere advantage for the perception of music and certain other nonspeech sounds (e.g., Kimura, 1973). Even in young infants, the hemispheres are differentially responsive to human speech versus other sounds. Auditory evoked response asymmetries in young infants favor the left hemisphere when words or syllables are presented auditorily, whereas the right-hemisphere response is stronger when musical or other nonspeech sounds are presented (Molfese, Freeman, & Palermo, 1975). In dichotic listening tests, consistent with the adult findings previously cited, infants as young as 2½ to 3 months show a right-ear advantage (REA) in discriminating among consonants, indicating a left-hemisphere superiority. Conversely, they show a left-ear advantage (LEA) in discriminating notes played by different musical instruments, indicating a right-hemisphere superiority (Best, Hoffman, & Glanville, 1982; Entus, 1977; Glanville, Best, & Levenson, 1977).

These functional asymmetries in infants indicate an early left-hemisphere attunement to information in speech, which could be an important biological support for the infant's perceptual discovery of the articulatory patterns of spoken words. But the data do not indicate exactly the sort of information in speech to which the left hemisphere is attuned (see Molfese, Nuñez, Seibert, & Ramaniah, 1976). Two recent findings suggest that the infant's left hemisphere is apparently attuned to information about the articulatory gestures of the vocal tract.

As an additional result of their intermodal speech perception study, MacKain, Studdert-Kennedy, Spieker, and Stern (1983) found a rightward attentional bias (implying left-hemisphere activation: Kinsbourne, 1973, 1982), which facilitated the infants' recognition of acoustic-optic commonality in the articulatory details of speech. In that experiment, infants attended primarily to either the right or the left video monitor during the synchronous audio presentation. An analysis of visual preferences indicated that the infants recognized auditory-visual matches versus mismatches in articulatory properties only when they were attending to the <u>right</u> video monitor. Since intermodal perception of speech appears to entail the recognition of its vocal-tract source properties, these results indicate the infant's recognition of that information is facilitated by a left-hemisphere attentional bias toward those properties.

The results of another study (Best, 1978) may further clarify which aspects of human speech are the object of the infant's left-hemisphere attunement. Adults show a consistent left-hemisphere advantage for consonant perception, whereas isolated vowels yield a nonsignificant perceptual asymmetry (e.g., Studdert-Kennedy & Shankweiler, 1970; Weiss & House, 1973). This vowel-consonant difference in hemispheric perceptual asymmetry may depend on the earlier discussed differences in the acoustic and articulatory properties of vowels and consonants.

The aim of the infant study (Best, 1978) was to determine whether 3½ month olds show a similar hemispheric difference in perception of consonants versus vowels. The infants showed a clear left-hemisphere advantage for discriminating a set of synthetic consonants, as adults did. However, the infants also showed a clear <u>right</u>-hemisphere advantage for discriminating steady-state synthetic vowels, which differs from adult reports.

The psychoacoustic interpretation of the adult hemisphere differences is that the hemispheres are differentially specialized for processing the acoustic features that differ between consonants and vowels (e.g., Cutting, 1974;

Schwartz & Tallal, 1980). However, this interpretation confounds acoustic and articulatory differences between vowels and consonants, and must be rejected on methodological grounds (Studdert-Kennedy & Shankweiler, 1980), as well as for the general criticisms against the psychoacoustic view. A speech source interpretation would instead consider articulatory differences between consonants and vowels. Indeed, consonant and vowel productions engage different coordinations of the articulatory musculature (Fowler, 1980). Moreover, vowel-consonant differences in intermodal speech perception effects (Summerfield, 1979; Summerfield, McGrath, & Forster, 1982) suggest that such articulatory differences may be influential in perception. The pattern of the phoneme class differences in production and in intermodal perception suggests that consonant information is conveyed in rapid articulatory changes, whereas vowel information is conveyed in relatively more slowly changing configurations of the tongue, lips, and jaw.

The implication of these two recent findings on infant hemispheric asymmetries is that the left-hemisphere attunement takes the form of an attentional bias toward information about rapid articulatory transformations. In complement to that attentional bias, the infant's right hemisphere may be better attuned to information about relatively more enduring structural properties of sound-making objects, such as the structural properties of instruments that determine their musical timbre and the configurations of the articulators in the human vocal tract that determine steady-state vowel color.

In conclusion, these specializations of the cerebral hemispheres for responding to different aspects of articulatory information in speech may offer biological support for the infant's discovery and production of words. In the final section, the relation of speech source perception to the discovery of words and the broader motivation provided by the context of communicative development will be briefly discussed.

5. The Broader Context of Communicative Development

The speech medium carries a number of parallel messages (Pike, 1959), and not only the sort of vocal-tract source information we have been focusing on at the surface of speech. To learn language, the infant must discover, or learn to recognize, many or all of those other messages as they are specified by convergent information in speech and in the context of its occurrence. Some of the messages that must be discovered are linguistic, beginning with words or phrases. Although the linguistic messages are more abstract than speech source messages, their expression in the speech medium depends directly on speech source information. Words and phrases are conveyed in the medium as patterns of articulator configurations and transformations, which have invariant properties across speakers, speaking rates, and surrounding speech context. Therefore, the infant's eventual discovery of them depends on attention to informational invariants in vocal-tract shapes and gestures as they are patterned over time, and conveyed in both the acoustic and optic media.

As for phonemes, their discovery as invariant vocal patterns that convey difference in meaning appears to be <u>served by</u> the child's developing use of words rather than vice versa (Menn 1980; Menyuk & Menn, 1979). Given that the function of phonemes in a language system is determined by word meanings and contrast, it is consistent with this interpretation that maternal speech to toddlers who produce words <u>does</u> include hyperdifferentiation in the productions of some phoneme contrasts, whereas maternal speech to prelinguistic infants does <u>not</u> include such hyperdifferentiation (Malsheen, 1980).

But the recognition of invariant vocal patterns, of course, is still not support enough for the discovery of words and phonemes. In order to comprehend and use language normally, the infant must also come to recognize the speaker's communicative messages. Some of these appear at the surface of speech, where the infants can detect them. Emotional affect may be conveyed to the infant directly in the mother's speech, for example, through the level and modulations of her voice pitch and intensity (see Stern, Spieker, & MacKain, 1982). These communicative messages often gain converging support from information in the visual and even in the haptic modalities; for example, emotional affect often receives contextual support in the speaker's changing facial expressions and the way she or he touches or holds the infant. The argument here is that the infant's discovery of more abstract communicative messages, notably the referential meaning of words, requires such nonspeech contextual support. Word meanings cannot be revealed for the first time through the speech medium alone.

This observation brings us to the end of our discussion, moving as it does beyond both the prelinguistic period and the infant's perception of the surface of speech. In closing, however, it is suggested that the prelinguistic infant's ability to perceive in speech the structure and transformations of its vocal-tract source provide her or him a crucial tool for discovering the more abstract messages of language.

References

Ades, A. E. (1977). Vowels, consonants, speech, and nonspeech. *Psychological Review*, *84*, 524-530.

Alekin, R. O., Klaas, Y. A., & Chistovich, L. A. (1962). Human reaction time in the copying of aurally perceived vowels. *Soviet Physics-Acoustics*, *8*, 17-22.

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: Vol. 2. Perception* (pp. 67-96). New York: Academic Press.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effect of early experience. *Child Development*, *52*, 1135-1145.

Baru, A. V. (1975). Discrimination of synthesized vowels [a] and [i] with varying parameters (fundamental frequency, intensity, duration, and number of formants) in dog. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 91-101). New York: Academic Press.

Bekoff, A. (1981). Embryonic development of the neural circuitry underlying motor coordination. In W. M. Cowan (Ed.), *Studies in developmental neurobiology*. New York: Oxford University Press.

Bell-Berti, F., & Harris, K. S. (1979). Anticipatory coarticulation: Some implications from a study of lip rounding. *Journal of the Acoustical Society of America*, *65*, 1268-1270.

Bernstein, J. (1981). Formant-based representation of auditory similarity among vowel-like sounds. *Journal of the Acoustical Society of America*, *69*, 1132-1144.

Bernstein, L. E. (1979). Developmental differences in labeling VOT continua with varied fundamental frequency. *Journal of the Acoustical Society of America*, *65*(Suppl.), S112. (Abstract)

Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant Behavior and Development*, *4*, 247-260.

Best, C. T. (1978). *The role of consonant and vowel acoustic features in infant cerebral asymmetries for speech perception*. Unpublished doctoral dissertation, Michigan State University.

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 two acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, *29*, 191-211.

Binnie, C. A., Montgomery, A. A., & Jackson, P. L. (1974). Auditory and visual contributions to the perception of consonants. *Journal of Speech and Hearing Research*, *17*, 619-630.

Blumstein, S. E. (1980). Speech perception: An overview. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 2. Perception*. New York: Academic Press.

Blumstein, S. E., Cooper, W. E., Zurif, E. B., & Caramazza, A. (1977). The perception and production of voice-onset-time in aphasia. *Neuropsychologia*, *15*, 371-383.

Blumstein, S. E., Isaacs, E., & Mertus, J. (1982). The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, *72*, 43-50.

Broad, D. J. (1981). Piecewise planar vowel formant distributions across speakers. *Journal of the Acoustical Society of America*, *69*, 1423-1429.

Burdick, C. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Discrimination of sustained /a/ and /i/. *Journal of the Acoustical Society of America*, *58*, 415-427.

Clifton, R. K., Morrongiello, B. A., Kulig, J. W., & Dowd, J. M. (1981). Developmental changes in auditory localization in infancy. In R. N. Aslin, J. R. Alberts, & M. R. Petersen (Eds.), *Development of perception: Psychobiological perspectives* (Vol. 1). New York: Academic Press.

Cooper, F. S. (1981). Some reflections on speech research. *Haskins Laboratories Status Report on Speech Research*, *SR-65*, 201-215.

Crowder, R. G. (1973). Representation of speech sounds in precategorical acoustic storage. *Journal of Experimental Psychology*, *98*, 14-24.

Cutting, J. E. (1974). Two left-hemisphere mechanisms in speech perception. *Perception & Psychophysics*, *16*, 601-612.

Cutting, J. E., & Eimas, P. D. (1975). Phonetic feature analyzers and the processing of speech in infants. In J. F. Kavanagh & J. E. Cutting (Eds.), *The role of speech in language*. Cambridge, MA: MIT Press.

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

Darwin, C. J. (1971). Ear differences in the recall of fricatives and vowels. *Quarterly Journal of Experimental Psychology*, *23*, 46-62.

Dodd, B. (1977). The role of vision in the perception of speech. *Perception*, *6*, 31-40.

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

Dudley, H. (1940). The carrier nature of speech. *Bell System Technical Journal*, *14*, 495-515. (Reprinted in J. L. Flanagan & L. R. Rabiner (Eds.), *Speech synthesis*. Stroudsberg, PA: Dowden, Hutchinson & Ross, 1973.)

Eilers, R. E. (1980). Infant speech perception: History and mystery. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 2. Perception* (pp. 23-39). New York: Academic Press.

Eilers, R. E., & Gavin W. (1981). The evaluation of infant speech perception skills: Statistical techniques and theory development. In R. E. Stark (Ed.), *Language behavior in infancy and early childhood*. New York: Elsevier-North Holland.

Eilers, R. E., Gavin W., & Wilson, W. R. (1979). Linguistic experience and phonemic perception in infancy: A cross-linguistic study. *Child Development*, *50*, 14-18.

Eilers, R. E., & Minifie, F. D. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research*, *18*, 158-167.

Eilers, R. E., Oller, D. K., & Gavin, W. J. (1978). *A cross-linguistic study of infant speech perception*. Paper presented at the meeting of the Southeastern Conference on Human Development, Atlanta, GA.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1976). Discrimination of synthetic prevoiced labial stops by infants and adults. *Journal of the Acoustical Society of America*, *60*, S91(Suppl.). (Abstract)

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research*, *20*, 766-780.

Eimas, P. D. (1974a). Linguistic processing of speech by infants. In R. L. Schiefelbusch & L. L. Lloyd (Eds.), *Language perspectives--Acquisition, retardation ~~and~~ intervention*. Baltimore: University Park Press.

Eimas, P. D. (1974b). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, *16*, 513-521.

Eimas, P. D. (1975a). Speech perception in early infancy. In L. B. Cohen & P. Salapatek (Eds.), *Infant perception: From sensation to cognition: Vol. 2. Perception of space, speech and sound*. New York: Academic Press.

Eimas, P. D. (1975b). Auditory and phonetic coding of the cues for speech: Discrimination of the r-l distinction by young infants. *Perception & Psychophysics*, *18*, 341-347.

Eimas, P. D., & Miller, J. L. (1980a). Contextual effects in infant speech perception. *Science*, *209*, 1140-1141.

Eimas, P. D., & Miller, J. L. (1980b). Discrimination of the information for manner of articulation by young infants. *Infant Behavior and Development*, *3*, 367-375.

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

Emde, R. N., & Robinson, J. (1976). The first two months: Recent research in developmental psychobiology and the changing view of the newborn. In J. Noshpitz & J. Call (Eds.), *Basic handbook of child psychiatry*. New York: Basic Books.

Entus, A. K. (1977). Hemispheric asymmetry in processing dichotically presented speech and nonspeech stimuli by infants. In S. J. Segalowitz & F. A. Gruber (Eds.), *Language development and neurological theory* (pp. 64-73). New York: Academic Press.

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

Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, *51*, 419-439.

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

Flanagan, J. L. (1973). *Speech analysis, synthesis and perception* (2nd ed.). New York: Academic Press.

Fodor, J. A., Garrett, M. F., & Shapiro, D. B. (1970). Discriminations among phones by infants. *Quarterly Progress Report: Research Laboratory of Electronics* (MIT), *96*, 180.

Fodor, J. A., Garrett, M. F., & Brill, S. L. (1975). Pi Ka Pu: The perception of speech sounds by prelinguistic infants. *Perception & Psychophysics*, *18*, 74-78.

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

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

Fowler, C. A., & Shankweiler, D. P. (1978). Identification of vowels in speech and nonspeech contexts. *Journal of the Acoustical Society of America*, *63*, S4. (Abstract)

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

Galunov, V. I., & Chistovich, L. a. (1966). Relationship of motor theory to the general problem of speech recognition. *Soviet Physics-Acoustics*, *11*, 357-365. (Review)

Garnica, O. (1973). The development of phonetic speech perception. In T. E. Moore (Ed.), *Speech production*. New York: Academic Press.

Gibson, E. J. (1977). How perception really develops: A view from outside the network. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Potomac, MD: Erlbaum.

Gibson, E. J. (1980, Spring). *Infants' perception of structural invariants of objects*. Colloquium presented at Columbia University, Psychology Department.

Gibson, E. J., & Levin, H. (1975). *The psychology of reading*. Cambridge, MA: MIT Press.

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

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, *62*, 32-41.

Glanville, B. B., Best, C. T., & Levenson, R. (1977). A cardiac measure of cerebral asymmetries in infant auditory perception. *Developmental Psychology*, *13*, 54-59.

Goldstein, U. G. (1979). Modeling children's vocal tracts. *Journal of the Acoustical Society of America*, *65*(Suppl.1), S25. (Abstract)

Haith, M. M. (1979). Visual cognition in early infancy. In R. B. Kearsley & I. E. Sigel (Eds.), *Infants at risk: Assessment of cognitive function*. Hillsdale, NJ: Erlbaum.

Hillenbrand, J., Minifie, F. D., & Edwards, T. J. (1979). Tempo of spectrum change as a cue in speech-sound discrimination by infants. *Journal of Speech and Hearing Research*, *22*, 147-165.

Holmberg, T. L., Morgan, K. A., & Kuhl, P. K. (1977, December). *Speech perception in early infancy: Discrimination of fricative consonants*. Paper presented at the 94th Meeting of the Acoustical Society of America, Miami Beach, FL.

Horridge, G. A. (1968). *Interneurons*. London: W. H. Freeman.

Howell, P. (1981). Identification of vowels in and out of context. *Journal of the Acoustical Society of America*, *70*, 1256-1260.

Jacob, F. (1977). Evolution and tinkering. *Science*, *196*, 1161-1166.

Jacobson, M. (1978). Developmental neurobiology. New York: Plenum Press.

Jenkins, J. J. (1974). Remember that old theory of memory? Well, forget it! American Psychologist, 29, 785-795.

Jenkins, J. J. (1980). Research in child phonology: Comments, criticism, and advice. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology. Vol. 2. Perception (pp. 23-39). New York: Academic Press.

Jusczyk, P. W. (1977). Perception of syllable-final stop consonants by 2-month-old infants. Perception & Psychophysics, 21, 450-454.

Jusczyk, P. W. (1981a). Infant speech perception: A critical appraisal. In P. D. Eimas & J. L. Miller (Eds.), Perspectives in the study of speech (pp. 113-164). Hillsdale, NJ: Erlbaum.

Jusczyk, P. W. (1981b). The processing of speech and nonspeech sounds by infants: Some implications. In R. N. Aslin, J. R. Roberts, & M. R. Petersen (Eds.), Development of perception: Psychobiological perspectives (Vol. 1). New York: Academic Press.

Jusczyk, P. W., Murray, J., & Bayly, J. (1979, March). The perception of place of articulation of fricatives and stops by infants. Paper presented at the meeting of the Society for Research in Child Development, San Francisco, CA.

Jusczyk, P. W., Pisoni, D. B., Walley, A., & Murray, J. (1980). Discrimination of relative onset time of two-component tones by infants. Journal of the Acoustical Society of America, 67, 262-270.

Jusczyk, P. W., Rosner, B. S., Cutting, J. E., Foard, C. L., & Smith, L. B. (1977). Categorical perception of nonspeech sounds by 2-month-old infants. Perception & Psychophysics, 21(1), 50-54.

Jusczyk, P. W., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. Perception & Psychophysics, 23, 105-109.

Kaye, K. (1980). Why we don't talk "baby talk" to babies. Journal of Child Language, 7, 70-78.

Kent, R. D., & Forner, L. L. (1979). Developmental study of vowel formant frequencies in an imitation task. Journal of the Acoustical Society of America, 65, 208-217.

Kimura, D. (1973). The asymmetry of the human brain. Scientific American, 228, 70-78.

Kinsbourne, M. (1973). The control of attention by interaction between the hemispheres. In S. Kornblum (Ed.), Attention and performance IV. New York: Academic Press.

Kinsbourne, M. (1982). Hemispheric specialization and the growth of human understanding. American Psychologist, 37, 411-420.

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

Kuhl, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology: Vol. 2. Perception (pp. 41-66). New York: Academic Press.

Kuhl, P. K. (1981a). Auditory category formation and development speech perception. In R. E. Stark (Ed.), Language behavior in infancy and early childhood. New York: Elsevier-North Holland.

Kuhl, P. K. (1981b). 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). Bimodal speech perception in early infancy. *Journal of the Acoustical Society of America*, *71*(Suppl.), S77-78. (Abstract)

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, *190*, 69-72.

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.

Lehiste, I., & Meltzer, D. (1973). Vowel and speaker identification in natural and synthetic speech. *Language and Speech*, *16*, 356-364.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

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.

Lieberman, P. (1967). *Intonation, perception and language*. Research Monograph No. 38. Cambridge: MIT Press.

Lieberman, P. (1980). On the development of vowel production in young children. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology*: *Vol. 1*. *Production* (pp. 113-142). New York: Academic Press.

Lieberman, P., Harris, K. S., Wolff, P., & Russell, L. H. (1972). Newborn infant cry and nonhuman primate vocalizations. *Journal of Speech and Hearing Research*, *14*, 718-727.

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.

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. (1981, August). *Cross-modal coordination in infants' perception of speech*. Paper presented at the meeting of the 2nd International Conference on Child Psychology, Vancouver.

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

Malsheen, B. J. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology*: *Vol. 2*. *Perception*. New York: Academic Press.

Mann, V. A. (1980). Influence of preceding liquid on stop consonant perception. *Perception & Psychophysics*, *28*, 407-412.

Mann, V. A., Madden, Russell, J. M., & Liberman, A. M. (1981). Further investigation the influence of preceding liquids on stop consonant perception. *Journal of the Acoustical Society of America*, *69*(Suppl.), S91. (Abstract)

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

Marshall, W. A. (1968). *The development of the brain.* London: Oliver & Boyd.

McCulloch, W. S. (1965). *The embodiments of mind.* Cambridge, MA: MIT Press.

Mehler, J., & Bertoncini, J. (1979). Infants' perception of speech and other acoustic stimuli. In J. Morton & J. C. Marshall (Eds.), *Psycholinguistics 2: Structures and processes* (pp. 45-66). Cambridge, MA: MIT Press.

Menn, L. (1980). Phonological theory and child phonology. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson, (Eds.), *Child phonology: Vol. 1. Production* (pp. 23-41). New York: Academic Press.

Menyuk, P., & Menn, L. (1979). Early strategies for the perception and production of words and sounds. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 49-70). Cambridge: Cambridge University Press.

Miller, C. L., & Morse, P. A. (1976). The "heart" of categorical speech discrimination in young infants. *Journal of Speech and Hearing Research, 19,* 578-589.

Miller, C. L., Morse, P. A., & Dorman, M. F. (1977). Cardiac indices of infant speech perception: Orienting and burst discrimination. *Quarterly Journal of Experimental Psychology, 29,* 533-545.

Miller, C. L., Younger, B. A., & Morse, P. A. (1982). The categorization of male and female voices in infancy. *Infant Behavior and Development, 5,* 143-159.

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.

Miller, J. D., Henderson, B. C., Sullivan, H. T., & Rigden, G. K. (1978). Speech perception by the chinchilla: Learning functions for pairs of synthetic stimuli from various points along a VOT continuum. *Journal of the Acoustical Society of America, 64*(Suppl. No. 1), S18.

Miller, J. D. Wier, C. C., Pastore, R., Kelly, W. J., & Dooling, R. J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America, 60,* 410-417.

Miller, J. L., & Eimas, P. D. (1981, April). *Contextual perception of voicing by infants.* Paper presented at the meeting of the Society for Research in Child Development, Boston, MA.

Moffit, A. R. (1971). Consonant cue perception by twenty- to twenty-four week old infants. *Child Development, 42,* 717-731.

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.

Molfese, D. L., & Molfese, V. J. (1979). VOT distinctions in infants: Learned or innate? In H. A. Whitaker & H. Whitaker (Eds.), *Studies in neurolinguistics* (Vol. 4). New York: Academic Press.

Molfese, D. L., Nuñez, V., Seibert, S. M., & Ramaniah, N. V. (1976). Cerebral asymmetry: Changes in factors affecting its development. *Annals of the New York Academy of Sciences, 280,* 821-833.

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

Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology, 14,* 477-492.

Morse, P. A. (1978). Infant speech perception: Origins, processes, and Alpha Centauri. *Infant Development Laboratory Research Status Report, 2*, 89-144.

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

Morse, P. A., & Snowdon, C. T. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perception & Psychophysics, 17*, 9-16.

Neuweiler, G., Bruns, V., & Schuller, G. (1980). Ears adapted for the detection of motion, or how echolocating bats have exploited the capacities of the mammalian auditory system. *Journal of the Acoustical Society of America, 68*, 741-753.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 1. Production* (pp. 93-112). New York: Academic Press.

Paget, R. (1930). *Human speech*. London: Routledge and Kegan Paul.

Petersen, M. R. (1981). Perception of acoustic communication signals by animals: Developmental perspectives and implications. In R. N. Aslin, J. R. Alberts, & M. R. Petersen (Eds.), *Development of perception: Psychobiological perspectives* (Vol. 1). New York: Academic Press.

Pike, K. L. (1959). Language as particle, wave, and field. *The Texas Quarterly, 2*, 37-55.

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

Pisoni, D. B. (1975). The role of auditory short-term memory in vowel perception. *Memory & Cognition, 3*, 7-18.

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. (1978). Speech perception. In W. K. Estes (Ed.), *Handbook of cognitive processes: Vol. 6. Linguistic functions in cognitive theory*. Hillsdale, NJ: Erlbaum.

Rabinowicz, T. (1979). The differentiate maturation of the human cerebral cortex. In F. Falkner & J. M. Tanner (Eds.), *Human growth: Vol. 3. Neurobiology and nutrition*. New York: Plenum Press.

Rakerd, B., & Verbrugge, R. R. (1982). Individual differences in the perception of isolated vowels and vowels in a consonantal context. *Journal of the Acoustical Society of America, 71*, S76. (Abstract)

Repp, B. H. (1982). Categorical perception. *Haskins Laboratories Status Report on Speech Research, SR-70*, 99-183.

Riedel, K. (1981). *Durational factors in the speech perception of aphasics*. Unpublished doctoral dissertation, City University of New York Graduate Center.

Robson, R. C., Morrongiello, B. A., Best, C. T., & Clifton, R. K. (1982). Trading relations in the perception of speech by five-year-old children. *Haskins Laboratories Status Report on Speech Research, SR-70*, 255-274.

Rose, S. R. (1976). *The conscious brain*. New York: Vintage Books.

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

Ruff, H. A. (1980). the development of perception and recognition of objects. *Child Development, 51*, 981-992.

Ruff, H. A. (1982). Effect of object movement on infants' detection of object structure. *Developmental Psychology, 18*, 462-472.

Schubert, E. D. (1974). The role of auditory perception in language processing. In D. D. Duane & M. B. Rawson (Eds.), *Reading, perception and language*. Baltimore: York Press.

Schvachkin, N. K. (1973). The development of phonemic speech perception in early childhood. In C. A. Ferguson & D. I. Slobin (Eds.), Studies of child language development. New York: Holt, Rhinehart, and Winston.

Schwartz, J., & Tallal, P. (1980). Rate of acoustic change may underlie hemispheric specialization for speech perception. Science, 207, 1380-1381.

Searle, C. J., Jacobson, J. Z., & Rayment, S. G. (1979). Stop consonant discrimination based on human audition. Journal of the Acoustical Society of America, 65, 799-809.

Shankweiler, D. P., Strange, W., & Verbrugge, R. R. (1977). Speech and the problem of perceptual constancy. In R. E. Shaw & J. D. Branford (Eds.), Perceiving, acting, and knowing: Toward an ecological psychology. Hillsdale NJ: Erlbaum.

Shaw, R. E., & Pittenger, J. (1977). Perceiving the face of change in changing faces: Implications for a theory of object perception. In R. E. Shaw & J. Bransford (Eds.), Perceiving, acting, and knowing: Toward an ecological psychology. Hillsdale, NJ: Erlbaum.

Simon, C., & Fourcin, A. J. (1978). Cross-language study of speech-pattern learning. Journal of the Acoustical Society of America, 63, 925-935.

Sinnott, J. M., Beecher, M. D., Moody, D. B., & Stebbins, W. C. (1976). Speech sound discrimination by monkeys and humans. Journal of the Acoustical Society of America, 60, 687-695.

Spelke, E. (1979). Exploring audible and visible events in infancy. In A. D. Pick (Ed.), Perception and its development: A tribute to Eleanor J. Gibson. Hillsdale, NJ: Erlbaum.

Spelke, E., & Cortelyou, A. (1981). Perceptual aspects of knowing: Looking and listening in infancy. In M. E. Lamb & L. R. Sherrod (Eds.), Infant social cognition. Hillsdale, NJ: Erlbaum.

Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology: Vol. 1. Production (pp. 73-92). New York: Academic Press.

Stern, D., Spieker, S., & MacKain, K. S. (1982). Intonation contours as signals in maternal speech to infants. Developmental Psychology, 18, 717-735.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. Davids, Jr. & P. B. Denes (Eds.), Human communication: A unified view (pp. 51-66). New York: McGraw-Hill.

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

Strange, W., Jenkins, J. J., & Edman, T. R. (1977). Journal of the Acoustical Society of America, 61, S39. (Abstract)

Streeter, L. A. (1976). Language perception of 2-month old infants shows effects of both innate mechanisms and experience. Nature, 259, 39-41.

Studdert-Kennedy, M. (1981a). The emergence of phonetic structure. Cognition, 10, 301-306.

Studdert-Kennedy, M. (1981b). A note on the biology of speech perception. Haskins Laboratories Status Report on Speech Research, SR-65, 223-232.

Studdert-Kennedy, M. (1981d). Cerebral hemispheres: Specialized for the analysis of what? The Behavioral and Brain Sciences, 4, 76-77.

Studdert-Kennedy, M. (1981). Perceiving phonetic segments. In T. Myers, J. Laver, & J. Anderson (Eds.), The cognitive representation of speech. Amsterdam: North Holland.

Studdert-Kennedy, M., & Lane, H. (1980). The structuring of language: Clues from the differences between signed and spoken languages. *Haskins Laboratories Status Report on Speech Research, SR-63/64*, 1-8.

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

Studdert-Kennedy, M., & Shankweiler, D. P. (1970). Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America, 48*, 579-594.

Summerfield, Q. (1978). Perceptual learning and phonetic perception. In *Interrelations of the communicative senses: Proceedings of the National Science Foundation conference at Asilomar, California, September-October 1978*. Washington, DC: National Science Foundation Publications.

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

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., McGrath, M., & Forster, J. (1982, April). *Detection and resolution of audio-visual conflict in the perception of vowels*. Paper presented at the meeting of the Acoustical Society of America, Chicago, IL.

Swoboda, P., Kass, J., Morse, P., & Leavitt, L. (1978). Memory factors in infant vowel discrimination of normal and at risk infants. *Child Development, 49*, 332-339.

Swoboda, P., Morse, P., & Leavitt, L. (1976). Continous vowel discrimination in normal and at-risk infants. *Child Development, 47*, 459-465.

Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology, 9*, 91-96.

Trehub, S. E., Bull, D., & Schneider, B. A. (1981). Infant speech and nonspeech perception: A review and reevaluation. In R. L. Schiefelbusch & D. D. Bricker (Eds.), *Early language acquisition and intervention: Language intervention series*. Baltimore, MD: University Park Press.

Trehub, S. E., & Rabinovitch, M. S. (1972). Auditory-linguistic sensitivity in early childhood. *Developmental Psychology, 6*, 74-77.

Trevarthen, C. (1979). Neuroembryology and development of perception. In F. Falkner & J. M. Tanner (Eds.), *Human growth: Neurobiology and nutrition*. New York: Plenum Press.

Tuchmann-Duplessis, H., Aroux, M., & Haegel, P. (1975). *Illustrated human embryology: Vol. 3. Nervous system and endocrine glands*. New York: Springer Verlag.

Verbrugge, R. R., & Rakerd, B. (1980). Talker-independent information for vowel identity. *Haskins Laboratories Status Report on Speech Research, SR-62*, 205-215.

Verbrugge, R. R., Rakerd, B., Fitch, H., Tuller, B., & Fowler, C. A. (in press). The perception of events: An ecological perspective. In R. E. Shaw & W. Mace (Eds.), *Event perception*. Hillsdale, NJ: Erlbaum.

Verbrugge, R. R., Shankweiler, D. P., & Fowler, C. A. (1980). Context-conditioned specification of vowel identity. *Haskins Laboratories Status Report on Speech Research, SR-62*, 217-224.

Walley, A. C., Pisoni, D. B., & Aslin, R. N. (1981). The role of early experience in the development of speech perception. In R. N. Aslin, J. Alberts, & M. R. Peterson (Eds.), *Sensory and perceptual development*. New York: Academic Press.

Warren, W. H., & Verbrugge, R. R. (in press). Toward an ecological acoustics. In R. E. Shaw & W. Mace (Eds.), *Event perception*. Hillsdale, NJ: Erlbaum.

Waters, R. S., & Wilson, W. A., Jr. (1976). Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. *Perception & Psychophysics*, *19*, 285-289.

Weiss, M. S., & House, A. S. (1973). Perception of dichotically presented vowels. *Journal of the Acoustical Society of America*, *53*, 51-53.

Werker, J. F. (1983, April). *Language-specific developmental changes in the speech perception of prelinguistic infants*. Paper presented at the meeting of the Society for Research in Child Development, Detroit, MI.

Werker, J. F., & Tees, R. C (1981, June). *Developmental changes in the perception of non-native speech sounds*. Paper presented at the meeting of the Canadian Psychological Association, Toronto.

Werker, J. F., & Tees, R. C. (1982, June). *Evidence for perceptual reorganization during the first year of life*. Paper presented at the meeting of the Canadian Psychological Association, Montreal.

Williams, L., & Bush, M (1978). Discrimination by young infants of voiced stop consonants . and without release bursts. *Journal of the Acoustical Society of America*, *63*, 1223-1226.

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

Zlatin, M., & Koenigsknecht, R. (1976). Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research*, *19*, 78-92.

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.

Footnotes

[1] The general underlying issue of meaning is complex and certainly cannot be settled here. The source of meaning for our knowledge of the world is at the heart of a centuries-long debate in epistemology between phenomenologists and realists. Psychology has taken up this debate. No satisfactory solution, acceptable to all, has been reached in either field. In the context of this chapter, the discussion reflects the author's view on the relation of the topic to how infants perceive speech.

[2] Natural speech also identifies the speaker as a member of the human species. Synthetic speech, insofar as it "works" perceptually, must capture necessary information about human vocal-tract dynamics, and usually also about the structure of a generic vocal tract, often appropriate for an adult male of indeterminate age. Rarely, however, does synthetic speech capture sufficient "textural" detail about a natural vocal tract to sound like a live human speaker, even an unknown one.

VOWEL-TO-VOWEL COARTICULATION IN CATALAN VCV SEQUENCES*

Daniel Recasens

Abstract. Electropalatographic and acoustical data on V-to-V coarticulatory effects were obtained for Catalan VCV sequences, with the consonants representing different degrees of tongue-dorsum contact (dorsopalatal approximant [j], alveolo-palatal nasal [ɲ], alveolo-palatal lateral [ʎ] and alveolar nasal [n]). Results show that the degree of V-to-V coarticulation in linguopalatal fronting and F2 frequency varies monotonically and inversely with the degree of tongue-dorsum contact, for larger carryover effects than anticipatory effects. The temporal extent of coarticulation also varies with the degree of tongue-dorsum contact, much more so for anticipatory effects than for carryover effects. Overall, results indicate that V-to-V coarticulation in VCV sequences is dependent on the mechanical constraints imposed on the tongue dorsum to achieve dorsopalatal closure during the production of the intervening consonant. Moreover, anticipatory effects but not carryover effects involve articulatory preprogramming.

Introduction

Studies on coarticulation address the question of how the phonemic string is produced in running speech. The failure to discover a one-to-one mapping between phonemes and articulatory targets suggests that the production units involve patterns of spatial and temporal coordination among several articulators (see, for example, Bell-Berti & Harris, 1981). Fowler (1980) and Fowler, Rubin, Remez, and Turvey (1980) have proposed that coarticulation results naturally from such coordinated patterns of articulatory activity. According to these researchers, the process of speech production is executed by means of coordinative structures, namely, muscle groupings organized functionally to actualize linguistic units in fluent speech. The constraints on articulatory movement imposed by the coordinative structure define those articulatory dimensions along which adjustment to context may take place. Thus, in light of this approach, coarticulatory effects ought to be predictable from constraints on articulatory displacement. On these grounds, evidence is presented in this

---

*Journal of the Acoustical Society of America, 1984, 76, 1624-1635.

Acknowledgment. I wish to thank Arthur S. Abramson, Ignatius G. Mattingly, Katherine S. Harris and Alvin M. Liberman for their helpful assistance at all stages of this investigation. Bruno Repp and Thomas Baer made very useful comments on a previous version of this paper. I also thank two anonymous reviewers and Nancy O'Brien for contributing significantly to the improvement of the manuscript. This research was supported by NICHD Grant HD-01994 and NINCDS Grant NS-13617 to Haskins Laboratories, and by a fellowship from the government of Catalonia (Spain) (CIRIT Commission) to the author.

study for systematic variability in transconsonantal vowel-to-vowel coarticulatory effects as a function of the degree of tongue-dorsum contact for the intervening consonant.

Öhman (1966) has proposed a model to account for coarticulation for bilabial, alveolar, and velar stops in VCV sequences. In this model, VCV coarticulatory effects are interpreted as reflecting an underlying V-to-V tongue movement with a superimposed consonantal constriction, which is actualized by commands directed towards different regions of the tongue. Öhman distinguishes at least three separate tongue regions that can be independently controlled: regions that shape the whole tongue body (as used for the production of vowels), the apical region (as used for the production of alveolars), and the dorsal region (as used for the production of velars). Tongue regions left uncontrolled by these consonantal commands can conform to the underlying diphthongal gesture, thus allowing for V-to-V coarticulation.

Öhman's interpretation has the interesting implication that degree of coarticulation should vary with the constraints exerted upon the kinematics of the different tongue dimensions under control. Thus, for instance, it could be that the production of place categories other than bilabial, alveolar, and velar imposes restrictions upon tongue activity so severe as to almost prevent V-to-V coarticulation from occurring. In fact, there is evidence from the literature that palatal articulations block V-to-V coarticulation to a large extent. Thus, it has been found for Russian palatalized consonants (produced with a primary constriction plus some raising of the tongue dorsum towards the palate) that formant transitions are barely influenced by the quality of the transconsonantal vowel (Öhman, 1966; Purcell, 1979). Also, data on V-to-C coarticulation show that English [j] (Lehiste, 1964; Stevens & House, 1964) and Italian [ʎ] (Bladon & Carbonaro, 1978) are highly resistant to effects from the surrounding vocalic environment.

The prediction tested in the present study was that the degree of V-to-V coarticulation in VCV sequences varies monotonically and inversely with the degree of tongue-dorsum contact required for the production of the consonant. Thus, for consonants produced with varying degrees of constraint on tongue-dorsum displacement towards the palate, more tongue-dorsum contact ought to allow less transconsonantal coarticulation, and less tongue-dorsum contact, larger transconsonantal coarticulatory effects. Moreover, degrees of tongue-dorsum contact and degrees of transconsonantal coarticulation ought to vary in similar amounts.

The dorsopalatal approximant [j], alveolo-palatal nasal [ɲ], alveolo-palatal lateral [ʎ] and alveolar nasal [n] in Catalan (a Romance language spoken in Catalonia, Spain) were chosen for analysis. The degree of tongue-dorsum contact associated with these consonants varies in the order [j]>[ɲ]>[ʎ]>[n], as traditionally described and according to a survey of palatographic recordings from the literature across different Romance languages and contextual conditions (e.g., Haden, 1938; Rousselot, 1924-1925). Thus, in a language with this set of consonants, [j], [ɲ], [ʎ] and [n] ought to show increasing degrees of V-to-V coarticulation. This hypothesis is based on the assumption that articulatory control during the production of [j], [ɲ] and [ʎ] is primarily exerted upon tongue-dorsum raising towards the hard palate. On these grounds, the dorso-palatal [j] ought to show maximum degree of tongue-dorsum constraint and minimum degree of V-to-V coarticulation since all [illegible] activity is directed towards this gesture; less tongue-dorsum constraint and more V-to-V coarticulation occurs for alveolo-palatals (more so for [ʎ] than for [ɲ]) since muscular activity is directed simultaneously towards tongue-blade contact and tongue-dorsum contact.

Another purpose of this investigation was to analyze: 1) the relative salience of V1-to-V2 (carryover) vs. V2-to-V1 (anticipatory) effects; 2) the temporal extent of coarticulatory effects.

Data on coarticulation in asymmetrical VCV sequences (mainly English) with consonants involving lingual closure show large anticipatory and carryover effects during closure and along the VC and CV transitions (see, for review, Parush, Ostry, & Munhall, 1983). However, small and asystematic anticipatory (English: Kent & Moll, 1972; German: Butcher & Weiher, 1976; and carryover (English: Gay, 1974) V-to-V effects have been reported at the steady-state vowel period. Several studies show that carryover effects are larger than anticipatory effects for English (Bell-Berti & Harris, 1976; Gay, 1974). In this study, the relative salience of transconsonantal anticipatory vs. carryover effects at the formant transitions and at the steady-state vowel are investigated for Catalan.

If the anticipatory process reflects articulatory preprogramming and the carryover process is primarily due to mechanical inertia constraints, anticipatory effects should be more sensitive than carryover effects to the temporal aspects of coarticulation. Recent evidence shows that this is the case for English (Parush et al., 1983). In the present study, this issue is also investigated for Catalan, as well as the extent to which V-to-V temporal effects are dependent on or independent of the degree of dorsal contact required for the production of the consonant.

I. Method

A. Articulatory Analysis

Electropalatographic (EPG) data were collected for the Catalan consonants [j], [ɲ], [ʎ] and [n] in all possible VCV combinations for V=[i], [a], [u]. All combinations can occur in running speech in Catalan. The utterances were embedded in a Catalan frame sentence "Sap ___ poc," meaning "He knows ___ just a little." A single speaker of Catalan (speaker Re, the author), also fluent in Spanish, English, and French, repeated all utterances 10 times with the artificial palate in place while the electropalatographic signal and the corresponding acoustic signal were recorded on tape for later analysis.

A mouth cast for speaker Re was used to build the artificial palate. The artificial palate is a device, 2-mm thick, made of acrylic resin, equipped with 63 small gold electrodes evenly distributed over its surface (shown in Figure 1). Patterns of linguopalatal contact were tracked over time (1 frame= 15.6 ms) on a display panel with an array of 63 lamps in an analogous configuration; as the tongue touches an electrode, the corresponding lamp lights up. Electrodes were reproduced on the panel in a two-dimensional display (as in Figure 1), which does not account for the vaulting of the subject's palate. Detailed information about this palatographic system (Rion Electropalatograph Model DP-01) is available in Shibata (1968) and Shibata et al. (1978).

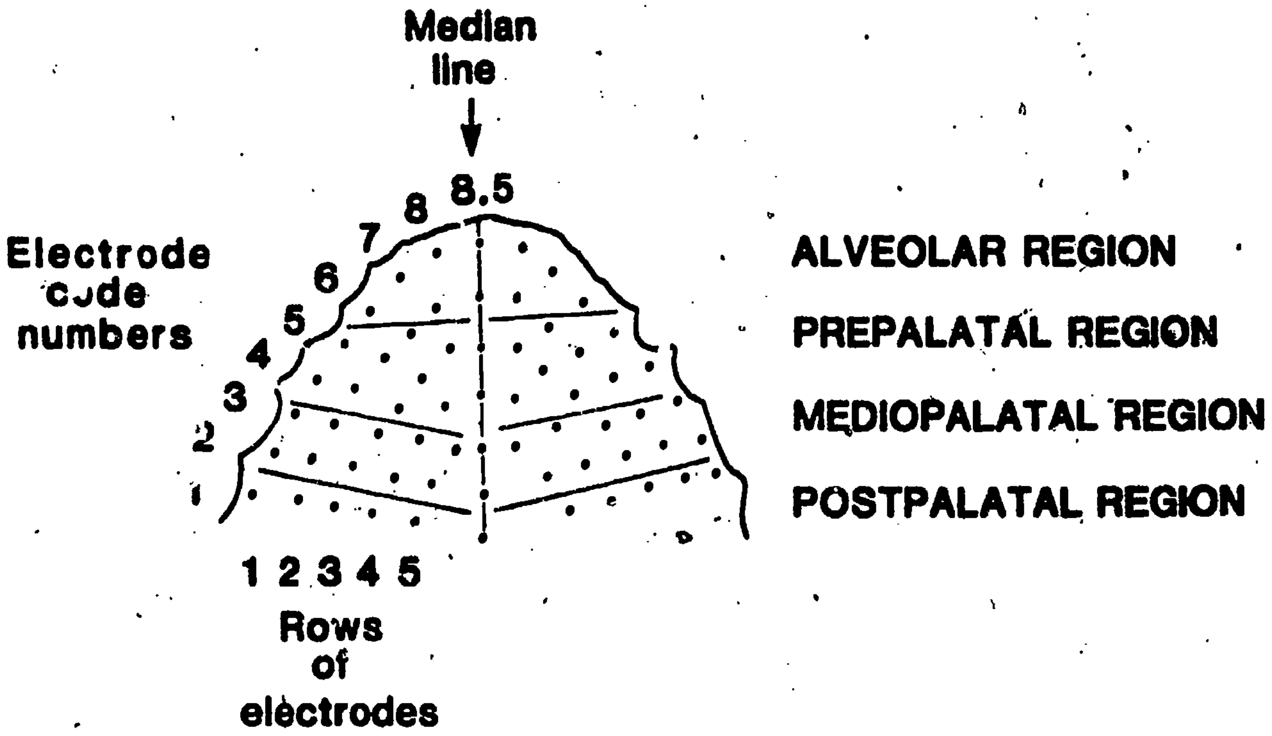


Figure 1. Electropalate.

The electrodes are arranged in five semicircular rows. For purposes of data interpretation, they are grouped in articulatory regions and sides, taking advantage of their equidistant arrangement in parallel curved rows on the artificial palate. As shown in Figure 1, the surface of the palate was divided into four articulatory regions (alveolar, prepalatal, mediopalatal, and postpalatal) and into two symmetrical sides (right and left) by a median line traced along the central range of electrodes. This division into articulatory areas on the palatal surface is based on anatomical considerations (Catford, 1977).

For each VCV utterance, data were tabulated from onset to offset of palatal contact, for a variable number of on-electrodes on each side of the palate. To tabulate the placement of on-electrodes frame by frame, every electrode was given a code number on each semicircular row for each side of the palate starting from the backmost electrode (1) up to the frontmost electrode (8.5 for row 1, 7.5 for row 2, and so on). Electrodes placed on the median line were assigned to both sides; thus, row 1 had 8.5 electrodes, row 2 had 7.5, and so on (see Figure 1). Given the fact that contacts were always made first at the rear of the palate (except for the sequence [uʎu], as discussed in the Results section) and that back electrodes stayed on during the entire production, the number of on-electrodes on each row was equivalent to the code number of the frontmost on-electrode. Therefore, a recording of each code number for each row in each frame simultaneously indicated the amount of linguopalatal contact and the degree of linguopalatal fronting for that row at that moment in time. For data interpretation, means were obtained by averaging the number of on-electrodes on each row frame by frame across repetitions of the same sequence lined up according to the point of maximum contact (PMC). PMC for a token was considered to be at the frame that presented the highest number of on-electrodes.

B. Acoustical Analysis

Four repetitions of all VCV combinations from this and two other Catalan speakers (Bo and Ca), also fluent in Spanish, were recorded for acoustical analysis. They were digitized at a sampling rate of 10 kHz, after preemphasis and low-pass filtering. An LPC (linear predictive coding) program included in the ILS (Interactive Laboratory System) package available at Haskins Laboratories was used for spectral analysis. Dynamic trajectories for the three lowest spectral peaks from onset to offset of voicing as detected on the waveform displays were reproduced on tracing paper and averaged across repetitions of the same sequence lined up according to PMC. To identify PMC on the acoustic wave for speaker Re, EPG data were also digitized at a sampling rate of 20 kHz, with no previous preemphasis or filtering. Labeling procedures were executed by means of WENDY (Haskins Laboratories Wave Editing and Display system). For speakers Bo and Ca, for whom no EPG data were available, PMC was estimated by visually identifying the F1 frequency minimum in the transition from the first vowel to the consonant. This procedure was chosen on the grounds that, of all the spectral characteristics present in the acoustical display of the utterances under study, such a point was found empirically to match PMC for speaker Re.

For each consonant, articulatory and acoustical data are presented as a function of time, considering first the general production characteristics in symmetrical VCV environments, and subsequently V-to-V coarticulatory effects in asymmetrical VCV environments. In the articulatory domain, patterns of contact in the palatal region (mediopalate and postpalate) that reflect tongue-dorsum activity are of particular concern; in the acoustic domain, F2 frequencies that, for palatal and alveolar consonants, reflect changes in the size of the back cavity behind the primary constriction and in degree of palatal constriction (Fant, 1960) are emphasized.

II. Results

A. General Production Characteristics

VCV utterances involving [j], [ɲ], [ʎ], and [n] were found to exhibit different patterns of linguopalatal contact and contrasting F2 patterns. Such patterns were correlated with different degrees of tongue-dorsum contact required for the production of each consonant, with [j]>[ɲ]>[ʎ]>[n].

1. Articulatory Data

Trajectories of linguopalatal contact in VCV symmetrical environments for [j], [ɲ], [ʎ], and [n] with V=[i], [a], [u] are displayed in the top panels of Figures 2, 3, 4, and 5. Each trajectory represents an average over 10 repetitions. Each panel provides data on linguopalatal contact (vertical axis) over time (horizontal axis). Data have been displayed for three rows of electrodes on the right side of the palate, namely, row 1 (contact with the tongue sides), row 3 (contact with the region between the tongue sides and the center of the tongue dorsum) and row 5 (contact with the center of the tongue dorsum). Linguopalatal contact has been plotted in terms of the code numbers for any on-electrodes on each row starting from the backmost electrode (1) up to the frontmost electrode (8.5 for row 1, 6.5 for row 3, 3.5 for row 5). As explained in the Method section, the plot of each trajectory over time represents the frontmost contacted electrode. For all consonants, contact is cumulative from back to front such that the frontmost electrode is also a good representation of total amount of contact. For [uʎu] (see Figure 4), the tongue did not always make contact with electrode 1 on row 1 of the artificial palate presumably because of the positioning of the tongue sides required to allow lateral airflow. Time has been measured in ms frame by frame. The line-up point for VCV sequences with [ɲ], [ʎ] and [n] is at PMC. Line-up procedures for [VjV] sequences were handled differently. EPG data for [aja] and [uju] showed a single frame (PMC) with maximum contact all over the palatal surface; however, maximum contact for [iji] was found to last for six frames (=95 ms). To account for this contrast, PMC for [aja] and [uju] was lined up with the midpoint of the period of maximum contact for [iji].

For all consonants in all sequences (see Figures 2 through 5), onset of contact occurs earlier at the tongue sides (row 1) and intermediate tongue regions (row 3) than at the center of the tongue dorsum (row 5); analogously, offset of contact occurs later on rows 1 and 3 than on row 5. Displacement along the vertical axis for any row indicates degree of linguopalatal fronting. All sequences show that the degree of linguopalatal fronting increases during the VC period from onset to PMC and decreases during the CV period from PMC to offset. This pattern of displacement over time has been reported in the literature for dorsal articulations such as [j] and velar consonants (Kent & Moll, 1972).

Trajectories for all consonants on row 5 show that tongue-dorsum contact decreases for [j], [ɲ]>[ʎ]>[n] (see also Introduction). Thus, the number of on-electrodes at PMC on row 5 adding across different vocalic conditions for each consonant varies for [j] (5.3), [ɲ] (5.3), [ʎ] (2.9), [n] (1.3). Moreover, vowel [a] shows contact with [j] but not with [ɲ] nor with [ʎ] and [n]. Vowel [u] shows no contact with [n]. Differences in degree of tongue-dorsum contact for alveolo-palatals [ɲ] and [ʎ] vs. alveolar [n] are related to the fact that, while the two categories of place of articulation involve alveolar contact, alveolo-palatals but not alveolars are produced with simultaneous raising of the tongue dorsum towards the palatal vault resulting in lingual contact at the center of the mediopalatal and postpalatal regions.

Table 1 shows maximum and minimum onset and offset contact values with respect to PMC at the tongue sides (row 1) and at the center of the tongue dorsum (row 5), as derived from figures 2 through 5. These values give a good estimate of the duration of the VC period (from onset to PMC) and the CV period (from PMC to offset) of linguopalatal contact. Onset values across rows show the pattern [j]>[ɲ]>[ʎ], [n]; offset values across rows show the pattern [j]>[ɲ]>[ʎ]>[n]. Thus, VC, CV, and VCV contact durations decrease as the degree of tongue-dorsum contact for the consonant decreases.

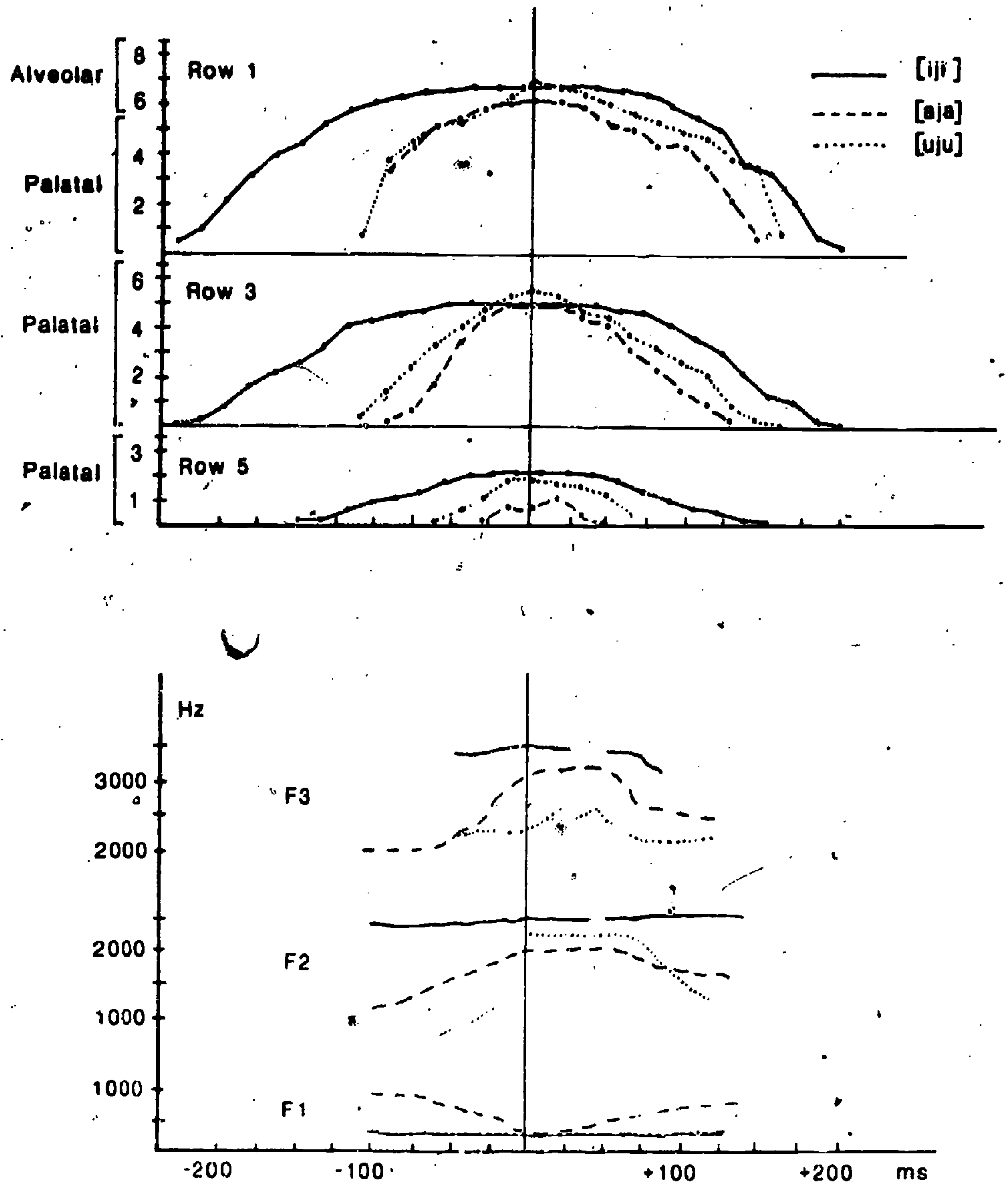


Figure 2. Trajectories of articulatory dynamics over time (in ms) for [j] in contrasting symmetrical environments lined up at PMC (for [aja] and [uju]) and at the MC midpoint (for [iji]) (speaker Re). Top: trajectories for the frontmost contacted electrode on rows 1 (tongue sides), 3 (between the tongue sides and center of the tongue dorsum) and 5 (center of the tongue dorsum) of the right side of the palate; electrode code numbers and articulatory regions have been given for each row. Bottom: trajectories for F1, F2 and F3 in Hz.

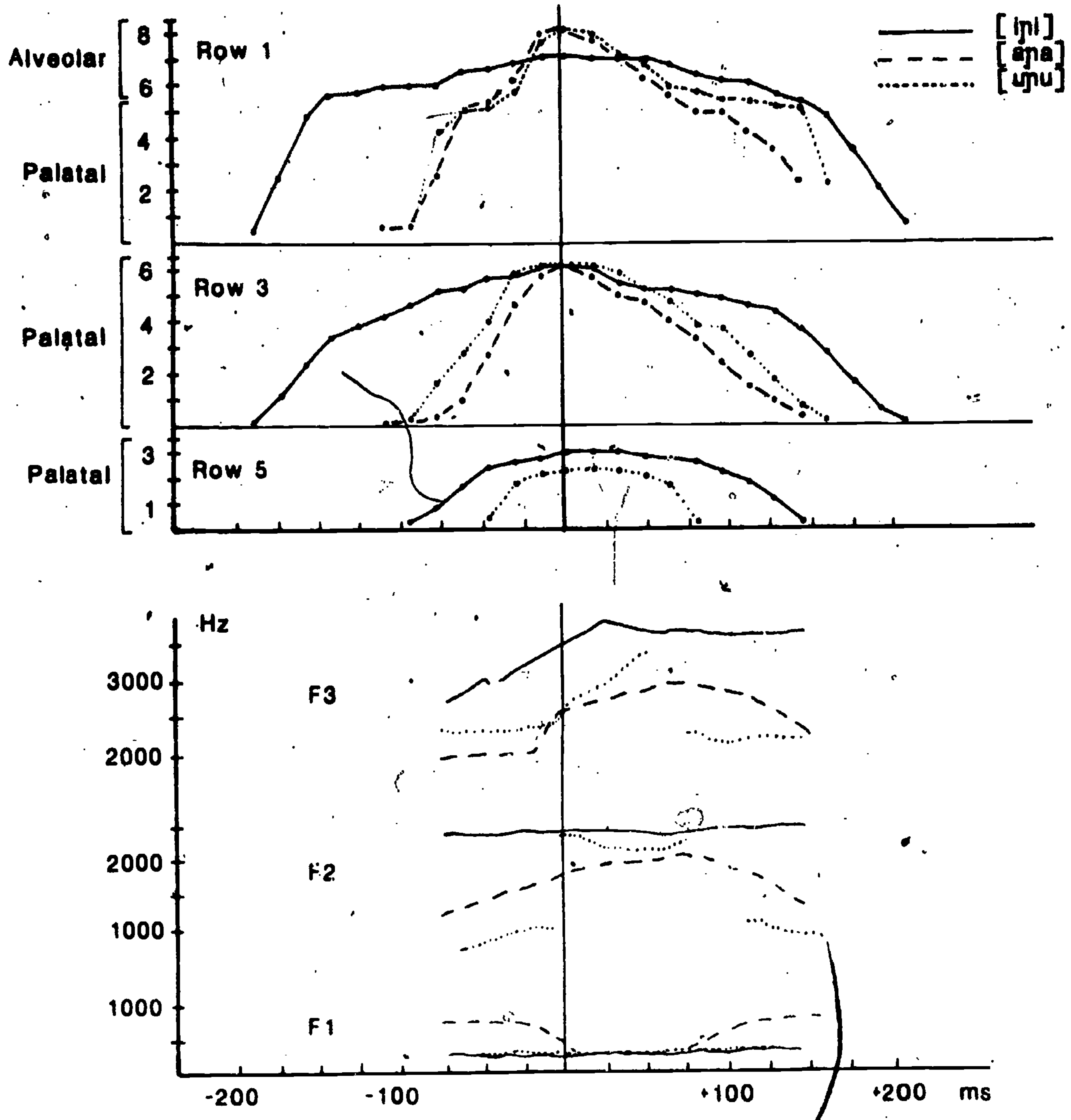


Figure 3. Trajectories of articulatory dynamics over time for [ɲ] in contrasting symmetrical environments lined up at PMC (speaker Re). Top: EPG data; bottom: acoustical data. See Figure 2 for details about the displays.

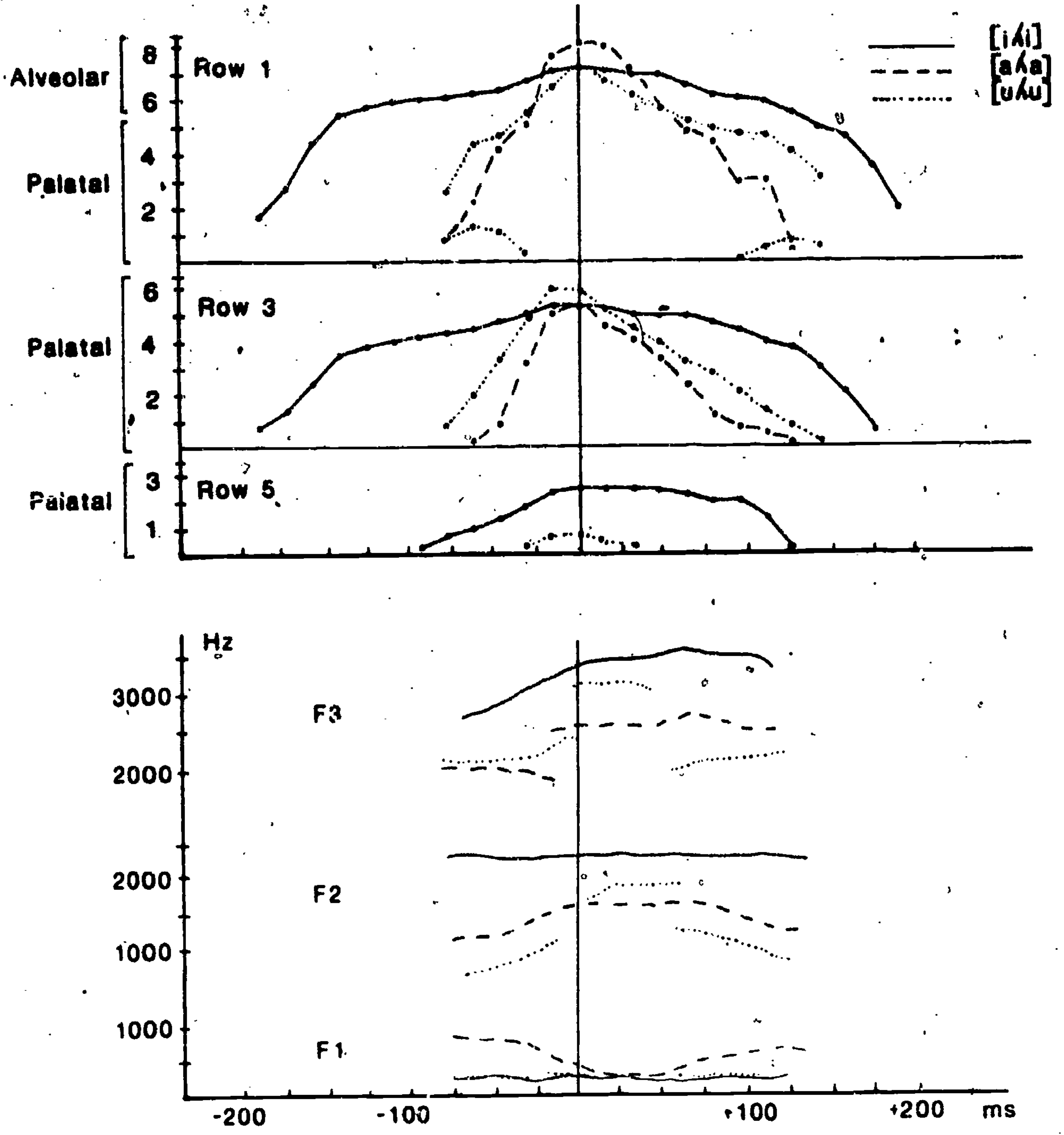


Figure 4. Trajectories of articulatory dynamics over time for [ʎ] in contrasting symmetrical environments lined up at PMC (speaker Re). Top: EPG data; bottom: acoustical data. See Figure 2 for details about the displays.

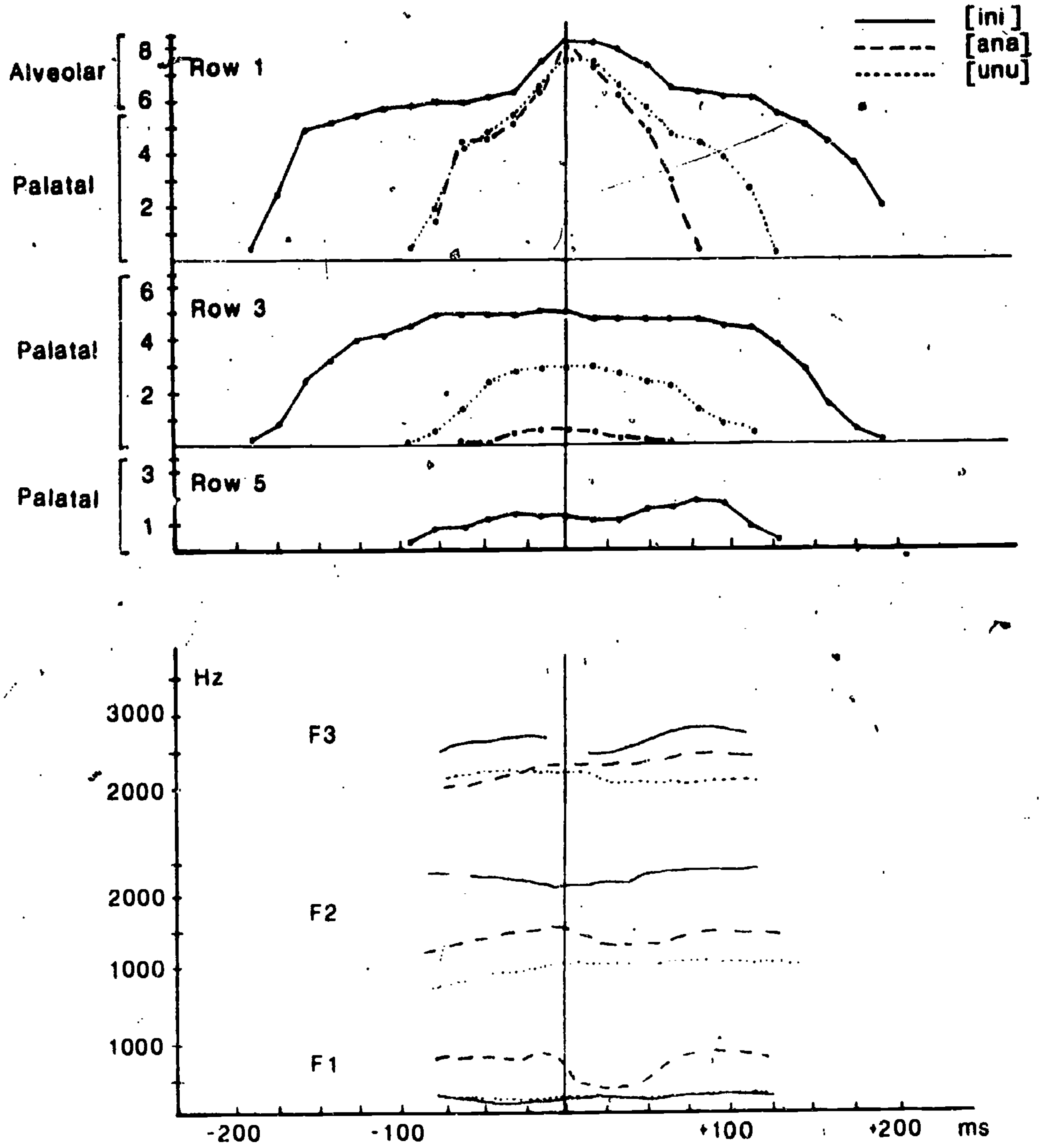


Figure 5. Trajectories of articulatory dynamics over time for [n] in contrasting symmetrical environments lined up at PMC (speaker Re). Top: EPG data; bottom: acoustical data. See Figure 2 for details about the displays.

Table 1

Maximum and minimum values of onset and offset time (in ms) for linguopalatal contact at the tongue sides (row 1) and at the center of the tongue dorsum (row 5) for consonants [j], [ɲ], [ʎ], and [n] in symmetrical VCV environments. Data are from one Catalan speaker (Re).

	Onset values		Offset values	
	Row 1	Row 5	Row 1	Row 5
[j]	-220/-95	-140/-30	+200/+140	+155/+45
[ɲ]	-185/-95	-95/ 0	+200/+140	+140/ 0
[ʎ]	-185/-80	-95/ 0	+185/+125	+125/ 0
[n]	-185/-80	-95/ 0	+185/ +80	+125/ 0

Table 2

Maximum and minimum F2 values (in Hz) at PMC for consonants [j], [ɲ], [ʎ], and [n] in symmetrical VCV environments. Data are from three Catalan speakers (Re, Bo, and Ca).

	Speaker Re	Speaker Bo	Speaker Ca
[j]	2350 - 1925	2450 - 2150	2150 - 1925
[ɲ]	2350 - 1775	2450 - 2000	2250 - 1575
[ʎ]	2275 - 1600	2400 - 1850	2000 - 1600
[n]	2210 - 1075	2350 - 1150	2075 - 1100

The trajectories give information about average velocity of linguopalatal displacement over time. Average velocity of movement can be obtained by dividing the total displacement for a given movement by the time required to execute the movement (Kuehn & Moll, 1976). Accordingly, velocity increases as duration decreases with degree of displacement remaining constant. Trajectories for the sequences [aCa] and [uCu] in Figures 3, 4 and 5 show that, for consonants involving fronted alveolar contact ([ɲ], [ʎ], and [n]) and, thus, similar degree of tongue-tip and/or tongue-blade displacement, time to achieve and release alveolar constriction (see row 1) decreases for [ɲ]>[ʎ]>[n]. Therefore, velocity of displacement for this articulatory gesture increases in that progression, as the degree of tongue-dorsum contact for the consonant decreases. On the other hand, velocity decreases as degree of displacement decreases and duration increases. This is the case for [j] vs. all other consonants; thus, [j] is articulated with a lesser degree of alveolar fronting and involves greater VCV contact duration. Overall, at the tongue sides, the velocity of displacement appears to be inversely related to the degree of tongue-dorsum contact.

In summary, several patterns of fronting, duration, and velocity of linguopalatal contact vary with the degree of tongue-dorsum contact for [j]>[ɲ]>[ʎ]>[n].

2. Acoustic Measurements

Table 2 reports maximum and minimum F2 values for [j], [ɲ], [ʎ], and [n] at PMC in symmetrical environments for speakers Re, Bo and Ca. As the table shows, F2 values for all speakers decrease for [j]>[ɲ]>[ʎ]>[n], namely, as the degree of tongue-dorsum contact decreases. Moreover, since F2 is dependent on the back cavity behind the place of constriction for [j] and [ʎ], its frequency decreases (for [j]>[ʎ]) as that cavity becomes larger (for [ʎ]>[j]). F2 for nasal consonants [ɲ] and [n] is presumably pharynx-cavity dependent, as indicated by an F2 continuation from V1 into the consonant in Figures 3 and 5 (bottom) and given the fact that, for nasal consonants, the mouth cavity behind the constriction acts as a shunting cavity. On these grounds, the pharynx-cavity size for [ɲ] and [n] ought to be smaller than the whole mouth-pharynx system for [j] and [ʎ] and, thus, cause a higher F2; however, acoustical data reported in Table 2 show that this is not the case. Also, if, as revealed by the area functions of Russian [ɳ] and [n] reported by Fant (1960), Catalan [ɲ] and [n] are produced with similar pharynx-cavity size, these two consonants ought to show similar F2 frequencies; however, acoustical data reported in Table 2 show a much higher F2 for [ɲ] than for [n]. In the absence of X-ray data on vocal tract configurations for these two Catalan consonants, it can only be stated with confidence that F2 differences for [j], [ɲ], [ʎ] and [n] are inversely correlated with differences in degree of tongue-dorsum contact.

The same F2 relationship holds during the consonantal steady-state period for all speakers. This is exemplified by the displays of formant trajectories for speaker Re, lined up at PMC with the EPG data in Figures 2, 3, 4, and 5 (bottom). The steady-state period for [j], [ɲ], [ʎ], and [n] lasts roughly from PMC up to +75 ms. Thus, little change in F2 frequencies appears to be taking place during the consonantal steady-state period and, presumably, in degree of tongue-dorsum contact.

In summary, as for the EPG data, F2 trajectories show frequency values that vary with the degree of tongue-dorsum contact for [j]>[ɲ]>[ʎ]>[n].

B. Coarticulation

It was also found that the extent of V-to-V coarticulatory effects for [j]<[ɲ]<[ʎ]<[n] is inversely related to the different degrees of tongue-dorsum contact.

1. Articulatory Data

Trajectories of linguopalatal contact were plotted for contrasting V2 to study anticipatory coarticulation and for contrasting V1 to study carryover coarticulation. All VCV sequences except [ijV] and [Vji] were lined up according to PMC. For [ijV] sequences, in which the period of maximum contact lasts for several frames, the onset of the period of maximum contact was taken as the line-up point in measuring anticipatory effects; for [Vji] sequences, for the same reasons, the offset of the period of maximum contact was taken as the line-up point in measuring carryover coarticulation. Coarticulation was considered to occur when an observable difference between two vowels in fronting of linguopalatal contact caused an analogous difference to occur on the other side of the line-up point and such difference was found to be significant at some moment in time. Since the main concern was to measure the correlation between degree of tongue-dorsum contact and degree of transconsonantal coarticulation, only data from rows 3 and 5 were selected for analysis in view of the fact that those rows show contact at the palatal region exclusively. The analysis procedure chosen to study V-to-V coarticulatory effects is described below.

Figure 6 shows anticipatory effects for [uʎV] (top, two upper panels) and carryover effects for [Vʎu] (bottom, two upper panels) on rows 3 and 5 at the right side of the palate. For the anticipatory condition, differences in linguopalatal fronting can be observed during V2 on rows 3 and 5 as [i]>[u]>[a]. On row 3 such differences cancel out during the period of maximum contact but appear between V1 onset and 15 ms before PMC as V2= [i]>[u], [a]; two-tailed *t*-tests show that anticipatory effects for V2= [i]>[u] (but not for V2= [i]>[a]) are significant ($p < .05$) between -60 and -30 ms. No significant anticipatory effects occur on row 5.

For the carryover condition, differences in linguopalatal fronting can be observed during V1 on rows 3 and 5 as [i]>[u]>[a]. Differences cancel out during the period of maximum contact (on row 3 but not on row 5) but appear during V2 (as V1= [i], [a]>[u] on row 3 and as V1= [i]>[u]>[a] on row 5). They were found to be significant for V1=[i]>[u] on rows 3 and 5 ($p < .01$) and for V1= [i]>[a] on row 5 ($p < .001$) between PMC and V2 offset.

This procedure was used to analyze effects between all possible VCV pairs in all coarticulatory conditions for all consonants reported in this study. Any possible coarticulatory effect on rows 3 and 5 on both sides of the palate, as determined visually from the plottings for all contextual combinations of VCV pairs, was tested frame by frame by means of the *t*-test procedure. Transconsonantal anticipatory effects (Table 3) and carryover effects (Table 4) are reported for all consonants in all vocalic environments. For each VC context (Table 3) and CV context (Table 4), the size of the significant coarticulatory effects and the onset and offset times of such effects

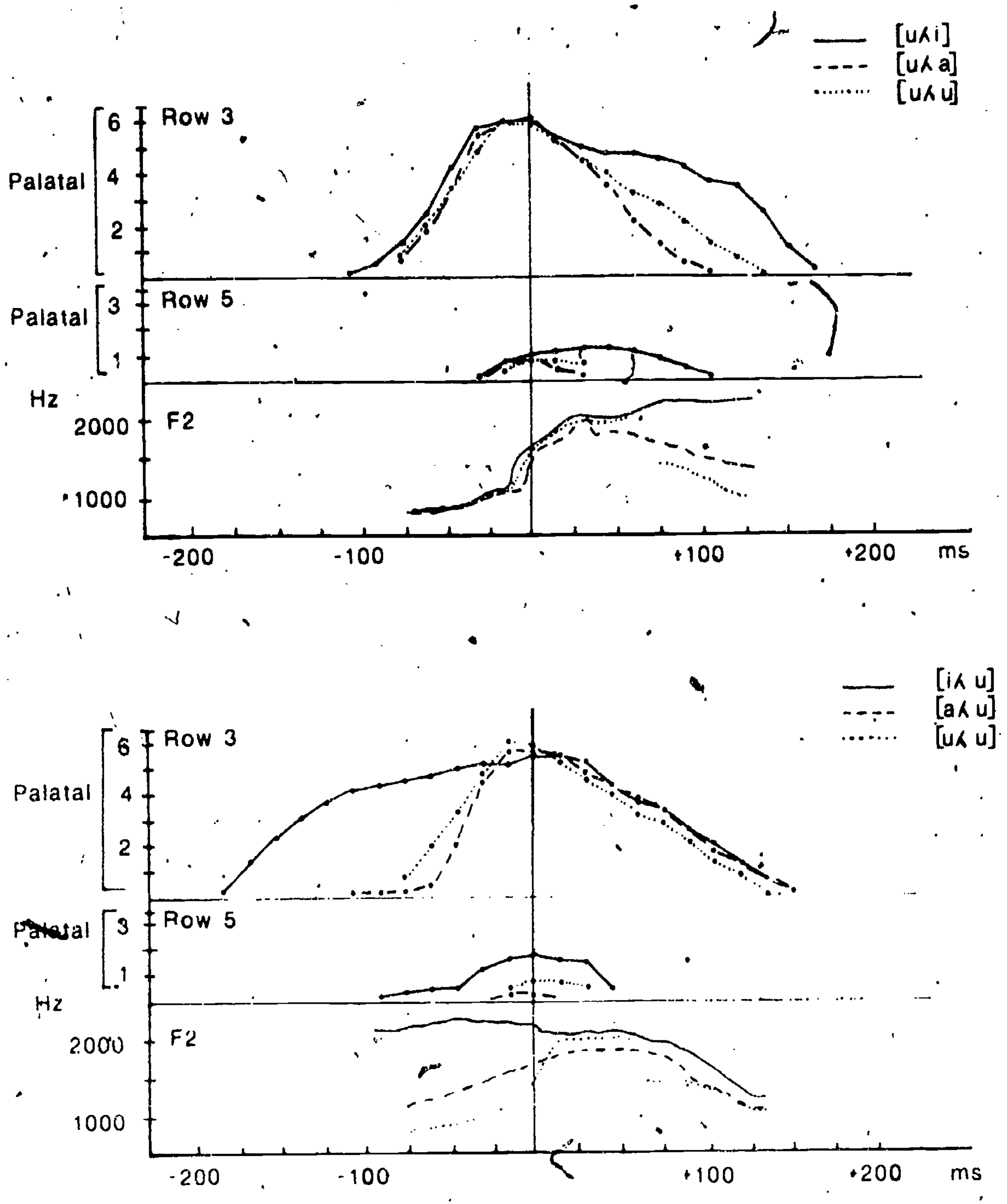


Figure 6. Top: EPG data (two upper panels) and F2 data (lower panel) on anticipatory effects over time for sequences [uʎV] lined up at PMC. Bottom: EPG data (two upper panels) and F2 data (lower panel) on carryover effects over time for sequences [Vʎu] lined up at PMC. Data correspond to speaker Re; EPG data correspond to rows 3 and 5 on the right side of the palate.

Table 3

EPG data on significant anticipatory (V2-to-V1) effects for different degrees of tongue-dorsum fronting. The magnitude of the effects is plotted for different levels of significance (*p<.05; **p<.01; ***p<.001). Onset of coarticulation (in ms before PMC) occurs at the number preceding "/" or at V1 onset if no number is present; offset of coarticulation (in ms before PMC) occurs at the number following "/" or at PMC if no number is present.

	V1=[i]			V1=[a]			V1=[u]		
	Ci>Ca	Ci>Cu	Cu>Ca	Ci>Ca	Ci>Cu	Cu>Ca	Ci>Ca	Ci>Cu	Cu>Ca
[j]	/-140*	—	—	/***	/***	—	/-30***	/30**	—
[ɲ]	—	—	/-75*	—	—	/-30*	—	—	—
[ʎ]	/-140**	—	—	/-30*	—	/-30*	—	-60/-30*	—
[n]	/-140**	/-140**	—	/***	/***	-45/***	-60/***	-60/***	—

Table 4

EPG data on significant carryover (V1-to-V2) effects for different degrees of tongue-dorsum fronting. The magnitude of the effects is plotted for different levels of significance (*p<.05; **p<.01; ***p<.001). Onset of coarticulation (in ms after PMC) occurs at the number preceding "/" or at PMC if no number is present; offset of coarticulation (in ms after PMC) occurs at the number following "/" or at V2 offset if no number is present.

	V2=[i]			V2=[a]			V2=[u]		
	iC>aC	iC>uC	uC>aC	iC>aC	iC>uC	uC>aC	iC>aC	iC>uC	uC>aC
[j]	—	—	—	/***	—	/***	—	—	—
[ɲ]	/+30*	—	+95/ +125***	/***	—	/***	/***	—	+30/***
[ʎ]	/***	/***	—	/***	/+65*	/***	/***	/**	—
[n]	+95/**	/***	—	/***	/***	/***	/+95***	/***	/***

are given for pairs of V2 (Table 3) and V1 (Table 4) differing in degree of linguopalatal fronting. The magnitude of the significant coarticulatory effects is plotted for different levels of significance (* $p<.05$; ** $p<.01$; *** $p<.001$); data correspond to the largest significant effects on rows 3 and 5 on any of the two sides of the palate. Onset and offset times of the significant coarticulatory effects are plotted in ms. The number of ms before "/" indicates onset time (as for -60 in the case of [uʎi] vs. [uʎu]) and the number of ms after "/" indicates offset time (as for -30 in the case of the same pair of sequences). No number before "/" indicates onset of coarticulation at V1 onset (anticipatory effects) and at PMC (carryover effects), and no number after "/" indicates offset of coarticulation at PMC (anticipatory effects) and at V2 offset (carryover effects).

a. Anticipatory effects (Table 3). Data show that, as expected, the instances and the degree of significance of the anticipatory effects decrease for [n]>[j] and for [n]>[ʎ]>[ɲ] as tongue-dorsum contact increases. However, [j] shows more contact and more (not less, as expected) coarticulation than [ɲ] and [ʎ]. According to Table 3, large anticipatory effects for [j] occur for [aCV] and [uCV] when V2= [i] vs. [a], [u]; given that [i] and [j] are produced with a highly similar articulatory configuration, it could be that, during V1, the speaker shows greater fronting for V2= [i] vs. [a], [u] to make sure that the upcoming consonant will be produced with a very salient gesture so that it can be articulatorily and perceptually distinguishable from V2=[i]. Data on F2 (see section IIB.2a) suggest strongly that the same effects occur for this and other speakers when [j] is preceded by V1=[i]. This strategy supports the view that the process of anticipatory coarticulation is regulated by articulatory preprogramming.

Onset time of the anticipatory effects occurs always at V1 onset for consonants that involve large tongue-dorsum contact ([j] and [ɲ]) and at different moments in time before PMC (at V1 onset, -60 ms and -45 ms) for consonants that involve less tongue-dorsum contact ([n] and, to a lesser extent, [ʎ]). Offset time values show that anticipatory effects can cancel out about the period of closure for articulations that show large tongue-dorsum contact (at 30 ms before PMC for [j], [ɲ] and [ʎ]; at the onset of V1=[i] about -140 ms) but not for articulations that show small tongue-dorsum contact ([n]).

Overall, it appears that consonants that involve low requirements on tongue-dorsum activity (as for [n]) show a small degree of tongue-dorsum contact and allow large transconsonantal anticipatory effects with different onset times before PMC. On the other hand, as tongue-dorsum contact increases (as for [ʎ] and [ɲ] but not for [j], for reasons stated above), consonants allow small transconsonantal anticipatory effects with a fixed onset time at V1 onset.

These coarticulatory phenomena suggest that the magnitude and the temporal extent of transconsonantal anticipatory coarticulation in VCV sequences is controlled by anticipatory preprogramming with reference to the mechanical constraints on articulatory activity required for the production of the consonant.

b. Carryover effects (Table 4). The instances and the degree of significance of the carryover effects decrease (for [n]>[ʎ]>[ɲ]>[j]) inversely and monotonically with the degree of tongue-dorsum contact (for [j]>[ɲ]>[ʎ]>[n]).

For all consonants (except for [j], for reasons indicated in section IIB.1a), carryover effects are larger than anticipatory effects; also, while anticipatory and carryover effects occur from front vs. back vowels, carryover effects among back vowels are much larger than anticipatory effects.

Contrary to anticipatory coarticulation, no contrasting timing effects are found among different consonants; thus, carryover coarticulation extends generally from PMC up to V2 offset.

These coarticulatory phenomena suggest that transconsonantal carryover coarticulation in VCV sequences results from mechanical inertia constraints on articulatory activity required for the production of the consonant and involves no articulatory programming.

2. <u>Acoustic Measurements</u>

F2 trajectories for pairs of VCV sequences for speakers Re, Bo and Ca were lined up according to the same procedure used to study coarticulatory effects for EPG data. Carryover effects for [j] could be measured only for speaker Re since no reference point was available for line-up procedures for speakers Bo and Ca. To detect transconsonantal effects, a procedure analogous to that for EPG data was used; thus, effects were considered to occur when observable frequency differences between two vowels caused analogous differences to occur at some moment in time on the other side of the line-up point. This method of analysis is exemplified below.

Figure 6 shows anticipatory effects for [uʎV] (top, lower panel) and carryover effects for [Vʎu] (bottom, lower panel). Data correspond to speaker Re and have been lined up at PMC with EPG data for the same speaker. For the anticipatory condition, differences in F2 frequency can be observed during V2 as [i]>[a]>[u]. Anticipatory effects for V2= [i]>[a] and V2= [i]>[u] occur between -15 ms and PMC; they never exceed 250 Hz. No differences for V2= [a]>[u] take place before PMC.

For the carryover condition, differences in F2 frequency can be observed during V1 as [i]>[a]>[u]. Carryover effects for V1= [i]>[a] and V1= [i]>[u] occur between PMC and V2 offset; the largest magnitude for the two effects is found between 250 and 500 Hz. No effects take place after PMC for V1= [a]>[u].

Coarticulatory effects on F2, as determined visually from the plottings for all contextual combinations of VCV pairs, are reported on Tables 5 (anticipatory effects) and 6 (carryover effects) analogously to Tables 3 and 4. The magnitude of the coarticulatory effects is plotted for different magnitude levels (* 0-250 Hz; ** 250-500 Hz; *** more than 500 Hz); onset and offset times are reported in ms.

a. <u>Anticipatory effects</u> <u>(Table 5)</u>. Similarly to the EPG data, the magnitude of the coarticulatory effects increases slightly as tongue-dorsum decreases for [ɲ]>[ʎ]>[n], while [j] shows similar effects to [ʎ]. As for the EPG data, speaker Re (but not speakers Bo and Ca) shows large anticipatory effects for [Vji] vs. [Vja], [Vju] when V1= [a] and [u] as a result of tongue-dorsum reinforcement before PMC. Moreover, the same strategy is used by all speakers for all palatal consonants when V1=[i], including speaker Re who did not show anticipatory effects during the articulation of this V1 on the surface of the palate.

Table 5

Acoustical data on significant anticipatory (V2-to-V1) effects for differences in F2 frequency. The magnitude of the effects is plotted for different magnitude levels (*0-250 Hz; **250-500 Hz; ***more than 500 Hz). Temporal effects are indicated as for the EPG data (see Table 3).

	V1=[i]			V1=[a]			V1=[u]		
	Ci>Ca	Ci>Cu	Ca>Cu	Ci>Ca	Ci>Cu	Ca>Cu	Ci>Ca	Ci>Cu	Ca>Cu
[j]									
Re	/*	/*	———	/*	/*	/*	/**	/**	———
Bo	/*	/*	———	-15/*	-15/*	———	———	———	———
Ca	/*	/*	———	———	———	———	———	———	———
[ɲ]									
Re	/*	/*	———	———	———	———	/*	/*	———
Bo	/-25*	———	———	-70/-25*	———	———	———	———	———
Ca	/*	/*	———	———	/*	/*	———	———	———
[ʎ]									
Re	/*	/*	———	/*	/*	———	-15/*	-15/*	———
Bo	-15/*	-40/**	———	/*	/*	———	———	-25/*	———
Ca	/-15*	/-15*	———	-15/*	———	———	———	———	———
[n]									
Re	———	———	-55/*	-40/*	/**	/*	-50/**	-50/-10*	———
Bo	———	———	———	/*	-45/*	———	/*	/*	———
Ca	/*	/*	———	-45/*	-55/*	———	———	-25/*	———

Timing effects can be explained analogously to the EPG data. For consonants involving a large degree of tongue-dorsum contact ([j] and [ɲ]), onset time of the anticipatory effects occurs at V1 onset; for consonants involving less tongue-dorsum contact ([n] and, to a large extent, [ʎ]), onset of anticipatory coarticulation occurs at different moments in time before PMC (mainly at V1 onset and between -55 and -25 ms). Contrary to the EPG data, as for [n], anticipatory effects for palatal consonants last until PMC.

Overall, anticipatory effects on F2 are in agreement with anticipatory effects over the surface of the palate (see section IIB.1a) and, therefore, are due to articulatory preprogramming. In some instances, different coarticulatory trends take place in different regions of the vocal tract for the same speaker; thus, articulations that require large degrees of tongue-dorsum contact block coarticulatory effects about the period of closure at the surface of the palate but not at other regions of the vocal tract.

b. <u>Carryover effects</u> (<u>Table 6</u>). As for the EPG data, the magnitude of the coarticulatory effects decreases (for [n]>[ʎ]>[ɲ]>[j]) inversely and monotonically with the degree of tongue-dorsum contact (for [j]>[ɲ]>[ʎ]>[n]). As for the EPG data, for all consonants (except for [j], for reasons indicated in section IIB.2a), carryover effects are larger than anticipatory effects; also, while effects from front vs. back vowels occur in the anticipatory and carryover conditions, carryover effects for [n] are much larger than anticipatory effects.

There is some indication that, in the carryover condition, timing may be controlled with reference to the degree of tongue-dorsum constraint required for the production of the consonant. Thus, carryover effects end at different moments in time after PMC for consonants produced with small degrees of tongue-dorsum contact ([n] and, to a lesser extent, [ʎ]) and last until V2 offset for consonants produced with large degrees of tongue-dorsum contact ([ɲ]). However, consistent to EPG data on carryover coarticulation, speaker Re does not show this trend with respect to coarticulatory effects common to all the consonants. Onset of carryover effects at +65/+80 ms after PMC for [ɲ] results from the cancellation of differences in F2 frequency for V1=[i] vs. [u] during closure due to oronasal coupling.

Overall, carryover effects on F2 are in agreement with carryover effects on the surface of the palate (see section IIB.1b); they are due to mechanical inertia constraints on articulatory activity and involve little or no articulatory programming.

III. Summary and Conclusions

The articulatory mechanisms that underlie the relationship between degrees of tongue-dorsum contact and degrees of coarticulatory activity are discussed. The dorsopalatal [j] is produced with one articulatory command for the raising of the tongue dorsum while the tongue blade makes no contact over the palate. For alveolo-palatals, two commands are actualized simultaneously: tongue-blade occlusion and tongue-dorsum raising. As a result of this synergistic activity, a large degree of contact is obtained for alveolo-palatals over the entire surface of the palate. Thus, the production of [j] vs. alveolo-palatals results in more tongue-dorsum constraint since all muscu-

Table 6

Acoustical data on significant carryover (V1-to-V2) effects for differences in F2 frequency. The magnitude of the effects is plotted for different magnitude levels (*0-250Hz; **250-500 Hz; ***more than 500 Hz). Temporal effects are indicated as for the EPG data (see Table 4).

	V2=[i]			V2=[a]			V2=[u]		
	iC>aC	iC>uC	aC>uC	iC>aC	iC>uC	aC>uC	iC>aC	iC>uC	aC>uC
[j]									
Re	/*	/*	___	/+75*	___	___	/+50*	/+50*	___
Bo	___	___	___	___	___	___	___	___	___
Ca	___	___	___	___	___	___	___	___	___
[ɲ]									
Re	/**	/*	___	/**	+80/**	+80/*	+80/**	+80/**	___
Bo	/+50*	___	___	/**	+65/**	___	/+25*	___	___
Ca	/**	___	___	/**	+80/*	___	/**	+70/**	___
[ʎ]									
Re	/+85**	/+85***	___	/***	/***	___	/**	/**	___
Bo	/+40**	___	___	/**	+50/*	___	/+50*	/+50**	___
Ca	/+65*	___	___	/**	/*	___	/***	/***	___
[n]									
Re	___	___	___	/***	/***	/+65**	/***	/***	/+50**
Bo	/+55***	/+45***	/+45**	/***	/***	/+40*	/+75***	/+75***	/+75**
Ca	/**	/+65***	/+50*	/+125**	/+125*	___	/+125**	/***	/**

lar activity is directed exclusively towards tongue-dorsum raising; for alveolo-palatals, on the other hand, less tongue-dorsum constraint results from the fact that muscular activity towards tongue-dorsum raising is accompanied by tongue-blade activity to make contact at the front of the palatal surface (more so for [ʎ] than for [ɲ]). Alveolar [n] is produced with a command to the tongue tip against the alveolar region and involves no constraint on the tongue dorsum to achieve dorsopalatal contact; in fact, contact with the tongue dorsum for [n] occurs only at the sides of the palate. Data on the degree of linguopalatal fronting and F2 frequency reported in this study show that V-to-V coarticulatory effects for [j], [ɲ], [ʎ], and [n] can be predicted from these differences in constraint on the tongue dorsum to achieve dorsopalatal contact. Thus, effects have been found to vary monotonically and inversely as a function of such differences in tongue-dorsum constraint.

Transconsonantal effects extended back to V1 onset (anticipatory effects) and up to V2 offset (carryover effects), more so on the surface of the palate (80% of anticipatory and carryover effects for the EPG data) than at other regions of the vocal tract (60% of anticipatory and carryover effects for the F2 data). There is evidence from the literature that acoustical measurements can be less sensitive than articulatory measurements to coarticulatory effects (Gay, 1974, 1977).

Directionality of coarticulatory effects has been taken to be inherent in the programming of speech sequences (Kent, 1976). Several findings reported in this study speak to this issue. Carryover effects were found to be larger than anticipatory effects in the light of articulatory and acoustical data for most consonants and speakers. From the present study, it can be concluded that this finding reflects a language-specific property of how articulatory programming is organized in Catalan. Evidence for a similar trend has been found for English (Bell-Berti & Harris, 1976; Gay, 1974; MacNeilage & DeClerk, 1969).

The temporal extent of coarticulation was found to vary with differences in tongue-dorsum contact, much more so for anticipatory effects than for carryover effects. Thus, EPG data and acoustical data show that onset time of anticipatory coarticulation is determined with higher precision (at V1 onset vs. different times before PMC) as the degree of tongue-dorsum contact for the consonant increases; on the other hand, EPG data and, less so, acoustical data show that carryover effects extend up to V2 offset independent of the degree of tongue-dorsum contact for the consonant. This finding is consistent with the view that anticipatory coarticulation results from articulatory preprogramming which, for the set of consonants investigated in this study, operates with reference to the mechanical constraints involved during the production of the consonant. Preprogramming also results in the reinforcement of tongue-dorsum activity during V1 to anticipate the articulatory and perceptual differentiation of a palatal consonant followed by V2=[i] vs. [a], [u].

Data reported in this study support the view that the speech production mechanism involves independent control of different regions of the vocal tract. Thus, anticipatory effects on the surface of the palate (EPG data) but not at other regions of the vocal tract (F2 data) are blocked about the period of closure for articulations that involve large tongue-dorsum contact.

With respect to alternative models that have been proposed in the literature to explain coarticulation in VCV utterances (syllabic or V-to-C model and cross-syllabic or V-to-V model; see Gay, 1978, for discussion), data reported here show that coarticulation is a context-dependent process in the sense that the extent to which V-to-V effects occur depends on the articulatory mechanisms involved in the production of the entire VCV sequence. The finding that coarticulation can be largely predicted from the degree of constraint involved during the activity of specific articulators suggests that the process of speech production is organized around precisely controlled patterns of articulatory activity. Thus, it has been shown that, independent of whether the primary constriction takes place at the palatal and/or alveolar regions, the degree of tongue-dorsum contact needs to be specified with accuracy. These observations are consistent with the view that linguistic units are actualized in running speech by muscle groupings that are synergistically controlled.

References

Bell-Berti, F., & Harris, K. S. (1976). Some aspects of coarticulation. *Haskins Laboratories Status Report on Speech Research*, *SR-45/46*, 1-204.

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

Bladon, R. A. W., & Carbonaro, E. (1978). Lateral consonants in Italian. *Journal of Italian Linguistics*, *3*, 43-54.

Butcher, A., & Weiher, E. (1976). An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics*, *4*, 59-64.

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

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

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

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.

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

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

Gay, T. (1978). Articulatory units: Segments and syllables. In A. Bell & J. Hooper (Eds.), *Syllables and segments* (pp. 121-131). New York: North Holland.

Haden, E. F. (1938). *The physiology of French consonant changes*. Supplement to *Language*, *14*, Dissertation no. 26.

Kent, R. (1976). Models of speech production. In N. J. Lass (Ed.), *Contemporary issues of experimental phonetics* (pp. 79-104). New York: Academic.

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

Kuehn, D. P., & Moll, K. L. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics*, *4*, 303-320.

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

MacNeilage, P. F., & DeClerk, J. L. (1969). On the motor control of coarticulation in CVC monosyllables. *Journal of the Acoustical Society*

of America, 45, 1217-1233.
Ohman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. Journal of the Acoustical Society of America, 39, 151-168.
Parush, A., Ostry, D. J., & Munhall, K. G. (1983). A kinematic study of lingual coarticulation in VCV sequences. Journal of the Acoustical Society of America, 74, 1115-1125.
Purcell, E. T. (1979). Formant frequency patterns in Russian VCV utterances. Journal of the Acoustical Society of America, 66, 1691-1702.
Rousselot, A. P. (1924-1925). Principes de phonétique expérimentale (Vols. I and II). Paris: Didier.
Shibata, S. (1968). A study of dynamic palatography. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo), 2, 28-36.
Shibata, S., Ino, A., Yamashita, S., Hiki, S., Kiritani, S., & Sawashima, M. (1978). A new portable type unit for electropalatography. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo), 12, 5-10.
Stevens, K. N., & House, A. S. (1964). Analysis of formant transitions for consonants. Quarterly Progress Report of the Research Laboratory of Electronics (Massachusetts Institute of Technology), 73, 165-166.

CATEGORICAL TRENDS IN VOWEL IMITATION: PRELIMINARY OBSERVATIONS FROM A REPLICATION EXPERIMENT*

Bruno H. Repp and David R. Williams†

Abstract. The question whether isolated stationary vowels are imitated in a continuous or categorical fashion--raised by Chistovich, Fant, de Serpa-Leitão, and Tjernlund (1966), and followed up primarily by Kent (1973)--was pursued further by replicating Kent's study with some slight modifications. The subjects (the two authors) imitated synthetic stimuli from [u]-[i] and [i]-[æ] continua at three different temporal delays. Acoustic analysis of the response vowels revealed very similar patterns across delays. Both subjects showed clear evidence of nonlinearities in the stimulus-response mapping and of preferred response formant frequencies, though strictly categorical responses were generally absent. The origin of these nonlinearities and their relation to the phonemic vowel categories of English are not fully understood at present. Vowel imitation responses presumably reflect the joint influences of perceptual and articulatory factors that need to be disentangled in future research.

1. Introduction

Almost two decades ago, Chistovich and her colleagues (Chistovich, Fant, de Serpa-Leitão, & Tjernlund, 1966) raised the interesting possibility that (isolated, stationary) vowels might have an internal representation intermediate between the auditory pattern, which presumably is a continuous function of the input, and the discrete categories of the vowel phonemes in the language. To warrant a separate existence in a theoretical model of speech processing, such an intermediate level of representation must have a noncontinuous structure distinct from that found at the phonemic level.

To test for this possible intermediate level, Chistovich, Fant, de Serpa-Leitão, and Tjernlund (1966) used a vowel imitation task. There is reason to believe that oral reproduction, particularly when it occurs as rapidly as possible ("shadowing"), may bypass the level at which familiar phonemic categories are associated with the input. This may be so even when the imitation response occurs at some delay ("mimicking"):[1] Latencies of shadowing and mimicking responses do not increase for phonemically ambiguous vowel tokens, but the latencies of written responses do (Chistovich, de Serpa-Leitão, &

---

*Speech Communication, in press.

†Also Department of Psychology, University of Connecticut.

Acknowledgment. This research was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. We thank Michael Studdert-Kennedy and Ignatius Mattingly for their helpful comments.

Tjernlund, 1966). The increase presumably reflects the time needed to decide among several categorical response alternatives. The hypothetical intermediate stage thus may serve to translate auditory information directly into motor instructions, as well as to store such information over time periods exceeding the life span of the raw auditory trace. Indeed, this intermediate representation may be thought of as motor in nature, thus anchoring it firmly in one of the three essential components of any model of speech communication (sensory, motor, and central processing). The question posed by Chistovich et al. thus concerns the continuous versus discrete motor representation of (isolated, stationary) vowels.

To answer this question, Chistovich, Fant, de Serpa-Leitão, and Tjernlund (1966) constructed a series of 12 synthetic vowel stimuli forming a continuum between the Russian vowel categories [i]-[ε]-[a]. The formant values of the endpoint stimuli were closely matched to natural productions of the single subject (L.C.), a female speaker of Russian. Her imitation responses in two conditions (shadowing--average latency of 190 ms; mimicking--average latency of 900 ms) were analyzed in three ways: (1) The formant frequencies (F1, F2, F3) of the responses, obtained from spectrograms, were plotted as a function of the formant frequencies of the synthetic stimuli. (2) The standard deviations of the formant frequencies across multiple imitations of the same stimulus were plotted in a similar fashion. (3) Histograms of the frequencies of each formant across all responses were constructed.

For the mimicking task, the formant frequency plots revealed stepwise changes in F2 suggesting at least four distinct categories. This impression was supported by the presence of three peaks in the standard deviation plot, corresponding to the boundaries between those categories. Finally, the F2 histograms also seemed to represent four distinct distributions of frequency values. The shadowing data seemed to agree, although only the first type of analysis was presented and some additional arguments were required to establish the similarity. Phonetic transcriptions by the (sophisticated) subject of her own responses also suggested four categories--corresponding to /i/, /ε/ (or /e/), /ö/, and /a/--and possibly a fifth category (/I/). Further support for this division came from a vowel matching task that required long-term memory for a fixed standard, using subject L.C. and the same set of stimuli. Chistovich et al. concluded that there is a discrete representation of vowels in terms of categories whose number exceeds that of the relevant phonemic categories in the language, and that this representation is used both to guide articulation and to retain vowel sounds in long-term memory.

Although these results are suggestive and challenging, they are not without weaknesses: (1) There was convincing evidence for only one category (/ö/) beyond the three phonemic categories that, *a priori*, might have been expected to play a role. That extra vowel category, moreover, is functional in Swedish, the language of the country where Chistovich et al. conducted their study. Thus, the results are not inconsistent with a two-stage model, in which the outcome of a rapid phonemic decision (prior to the stage of response selection) constrains the motor program for imitation, without any intermediate stage. (2) The conclusions are based entirely on the pattern of F2 frequencies. Although there appeared to be some correlated trends in F1 and F3, these data were not discussed by Chistovich et al. (3) The analysis of shadowing responses is incomplete, and not totally conclusive. (4) All the analyses are qualitative; no statistical criterion for, e.g., detecting steps in a function was specified. (5) There was only a single, highly sophisticated subject.

Some preliminary follow-up data were reported by Chistovich, Fant, and de Serpa-Leitão (1966). Several subjects (including L.C.) imitated noise-excited synthetic two-formant vowels lying along arbitrary trajectories in the F1-F2 space. Again, some step-wise changes in response formant frequencies were observed, but the data show some remarkable irregularities and are suggestive at best. Some of the stimuli may have exceeded the range of formant values subjects were able to produce.

A full-scale replication, with some modifications, was attempted by Kent (1973). He used two 11-member vowel continua, ranging from [u] to [i] and [i] to [æ], respectively. The formant values of the endpoint stimuli were derived from Peterson and Barney (1952). Four English-speaking subjects who had some phonetic training imitated each vowel 10 times. There was no time pressure; subjects had up to 7 s to respond.

The [i]-[æ] continuum roughly corresponds to two-thirds of the [i]-[a] continuum used by Chistovich, Fant, de Serpa-Leitão, and Tjernlund (1966). As Kent points out, this continuum is "phonemically rich" in English, spanning as many as five categories (/i/, /I/, /e/, /ε/, /æ/). The [u]-[i] continuum, on the other hand, contains no familiar vowel categories between the endpoints, despite covering a much larger range of F2 frequencies (though with F1 nearly constant). The question of interest, then, was whether categorical tendencies in vowel imitations--if replicated--would be restricted to the [i]-[æ] continuum.

Kent interpreted his data as providing an affirmative answer to this question. Response F2 frequencies along the [u]-[i] continuum, while not a strictly linear function of stimulus F2, did not exhibit any clear steps for any of the four subjects. The standard deviations, however, showed pronounced peaks--a single central peak for subjects 1 and 3, and twin peaks for subjects 2 and 4. Moreover, the F2 histograms tended to have several modes for each subject--two for subjects 1 and 3, and three for subjects 2 and 4. Finally, comparison of these patterns with the F2 frequency plots shows that increased standard deviations and histogram valleys correspond to regions of increased slope on the F2 frequency plots. Kent is very conservative in interpreting these data, conceding only that responses are more accurate at the [u] and [i] ends of the continuum. In fact, his data offer some support for the presence of three categories in two of the four subjects. It is true, however, that these categories correspond only to regions of reduced discrimination, never to a true constancy in imitation. No labeling data were collected for these stimuli, nor was F3 analyzed.

The results for the [i]-[æ] continuum were presented by Kent in an incomplete and somewhat confusing manner, which makes it difficult to object to his conclusion that these data were hard to interpret. The response locations in F1-F2 space (F3 was not analyzed) showed some pronounced discontinuities, but they occurred in different places for different subjects and agreed only partially with the patterns of standard deviations and frequency histograms. For only one subject could histogram peaks be tentatively aligned with the response categories used in labeling the [i]-[æ] stimuli. From these results, Kent drew the "very tentative conclusion that the stimuli of the /i/-/æ/ series are represented in memory in a more classificatory fashion than are the stimuli of the /u/-/i/ series" (Kent, 1973, p. 16). The emphasis on memory is explained by the lack of time constraints in Kent's task.

The available data on categorical tendencies in vowel imitation must therefore be considered as merely suggestive. Both Kent (1973) in his concluding paragraph and, quite recently, Chistovich (1984) acknowledge that further data on this issue are needed. Their initial studies do not seem to have been followed up.[2] Kent went on to conduct a series of interesting vowel imitation studies, but none of these addresses directly the issue of categorical tendencies (Kent, 1974, 1978, 1979; Kent & Forner, 1979). A relevant study by Schouten (1977), in which stimuli from an [i]-[æ] continuum were presented to Dutch-English bilingual subjects, yielded some evidence of categorical imitation, but the data were complex and were analyzed by quite different procedures, such as cluster analysis. Thus, the ambiguity of the earlier results remains.

We decided to continue where Kent (1973) left off, and to begin with a straightforward replication of his study. In this preliminary report we present data obtained for the two authors as subjects. There were a few methodological differences between our study and Kent's. The most important of these was that we employed three different response timing conditions, in an attempt to incorporate the "shadowing" versus "mimicking" comparison (Chistovich, Fant, de Serpa-Leitão, & Tjernlund, 1966) into Kent's design. Subjects were required to imitate each vowel (a) immediately, (b) following a short (750 ms) delay, and (c) following a longer (3 s) delay. Other procedural changes were relatively minor: First, the stimuli were presented over a loudspeaker rather than over earphones. Second, to prevent overlap of shadowing responses with the end of the stimulus, the synthetic vowels were shortened to 150 ms (versus 250 ms [Kent, 1973] and 300 ms [Chistovich, Fant, de Serpa-Leitão, & Tjernlund, 1966]). Extrapolating from perceptual findings (Pisoni, 1973) shortening of vowel stimuli should, if anything, result in a more categorical response. Finally, there were 12 (rather than Kent's 11) stimuli per vowel continuum.

2. Methods

2.1. Subjects

The two authors served as subjects in this pilot study. BR is a native speaker of Southern German who has been speaking English almost exclusively for over 15 years. He also has smatterings of French, Italian, and Swedish, but no professional phonetic training. DW is a native speaker of American English from California. Although not a fluent speaker of any foreign languages, he has studied French, German, Japanese, and Latin, and has had several years of phonetic training.

2.2. Stimuli

Five-formant vowel stimuli were generated on the software serial synthesizer at Haskins Laboratories. The center frequencies of the fourth and fifth formants were fixed at 3500 and 4500 Hz, respectively. All stimuli were 150 ms in duration and all formants were stationary with amplitude rise-fall times of 30 ms. Bandwidths of the five formants were set at 50, 80, 110, 150, and 200 Hz, respectively. For all stimuli, the fundamental frequency fell linearly from an initial value of 120 Hz to a final value of 105 Hz.

Two stimulus continua--one from [u] to [i] and the other from [i] to [æ]--were employed in the experiment. Frequency values for F1, F2, and F3 of the endpoint stimulus vowels corresponded to the mean values reported by Peterson and Barney (1952) for their male talkers. Two 12-stimulus vowel series were constructed by interpolating in equal frequency steps between the endpoint values for the first three formants. The formant frequency values of all stimuli are listed in Table 1. Twelve randomized blocks of the twelve stimuli were arranged to form a stimulus sequence. The order of items in the sequence was adjusted until it was almost perfectly balanced; i.e., each stimulus was preceded once by every other stimulus, with a few exceptions. For each vowel continuum, three such stimulus sequences were recorded on tape for presentation in the imitation task.

Table 1

Formant Frequencies of the Stimuli

[u]-[i] continuum

	F1	F2	F3
1	300	870	2240
2	297	999	2310
3	295	1128	2380
4	292	1257	2450
5	289	1386	2520
6	286	1515	2590
7	284	1644	2660
8	281	1773	2730
9	278	1902	2800
10	276	2031	2870
11	273	2161	2940
12	270	2290	3010

[i]-[æ] continuum

	F1	F2	F3
1	270	2290	3010
2	305	2238	2955
3	341	2186	2901
4	376	2134	2846
5	412	2082	2792
6	447	2030	2737
7	483	1979	2683
8	518	1927	2628
9	554	1876	2574
10	589	1824	2519
11	625	1772	2465
12	660	1720	2410

2.3. Imitation Task

Each of the 144 trials in a stimulus sequence required six seconds. Within that interval, the subject was presented with a 150 ms stimulus vowel that was either preceded or followed by a 1000 Hz tone, 100 ms in duration. The position of the tone in a trial depended on the condition. In the immediate imitation condition (A), a trial began with the tone, which served as a warning signal; its onset preceded that of the stimulus vowel by 500 ms. Subjects were asked to produce a response vowel as soon as possible after the onset of the stimulus vowel. In the delayed (B) and deferred (C) imitation conditions, a trial began with a stimulus vowel, the onset of which preceded that of the tone by 750 ms an 3000 ms, respectively. Subjects were asked in these conditions to respond immediately upon the onset of the tone, which served as a "go" signal here. All responses were initiated from a closed mouth position; that is, phonation was always preceded by a silent opening gesture.

Each of the three stimulus sequences was divided into three parts of 48 trials each, separated by pauses and assigned to the three imitation conditions as follows: (1) A, B, C; (2) B, C, A; (3) C, A, B. Before hearing the stimulus sequences, subjects listened to one 12-item block of trials from each of the imitation conditions for practice. The [u]-[i] session preceded the [i]-[æ] session for both subjects.

The stimuli were presented in a sound-insulated (IAC) booth over a Realistic loudspeaker placed about 20 degrees to the right of center at a distance of 3 feet from the subject. A Sennheiser MKH415T microphone was placed directly in front of the subject approximately eighteen inches from his lips. Both the stimuli and the subject's responses were recorded on a second tape recorder.

2.4. Identification Tasks

In addition to the imitation responses, each subject provided written phonemic and numeric identifications of the stimuli. For the phonemic identification task, the stimulus vowels from the [i]-[æ] series were recorded twelve times in random sequence with ISIs of 3 s. The subjects used the labels /i,I,e,ɛ,æ/. This identification task followed the imitation task by several weeks. Phonemic identification of the [u]-[i] stimuli was not attempted, since both authors found it difficult to think of appropriate response categories.

Several weeks later, the subjects identified the stimuli from both continua on a numerical scale ranging from 1 to 12. To facilitate this absolute identification task, the stimuli were presented in an AXB format, with the A and B stimuli representing the continuum endpoints, in either order. A total of 12 x 12 = 144 triads were presented for each continuum, with ISIs of 1 s within triads and 3 s between triads.

Each identification task was preceded by a familiarization sequence in which the stimulus series was presented four times, twice in forward order and twice in reverse order.

2.5. Analysis

The imitation recordings were digitized at a sampling rate of 10 kHz. With a waveform editor, the time interval between the onset of the warning tone and the onset of the response vowel was measured. Each response vowel was then isolated, labeled, and stored in a disk file. The center frequencies of the first three spectral peaks were obtained using LPC analysis, with a 20-ms window moving in 10-ms steps. For each formant, three separate estimates were obtained: (a) the mean formant frequency over the whole vowel, including its standard deviation across the vowel token; (b) its value at the onset of the response vowel; (c) its value two-thirds into the vowel (the measurement point used by both Chistovich, Fant, de Serpa-Leitão, and Tjernlund [1966] and Kent [1973]). These values were further averaged across the 12 responses to each stimulus token, and the standard deviation for each stimulus token was computed. Formant frequency histograms for each delay condition and for all delay conditions combined were obtained for each stimulus in a series using Program 5D of the BMDP package.

3. Results and Discussion

3.1. Latencies

Chistovich, Fant, de Serpa-Leitão, and Tjernlund (1966) reported essentially constant response latencies for subject L.C. across their vowel continuum, both in the rapid shadowing condition (average latency: 190 ms) and in the slower mimicking condition (average latency: 900 ms). Kent (1973) did not measure reaction times. The average latencies for the vowels along the two continua used in the present study are plotted in Figure 1 as a function of stimulus number, separately for the two subjects and for the three imitation conditions. Although there is more variability here than in the data of Chistovich et al., the functions are either flat or vary in a nonsystematic fashion. The standard deviations were rather large--about 60 ms on the average--which must be taken into account when interpreting the latency functions. Despite occasional peaks and valleys, these functions do not provide any strong evidence of systematic variation in latency across the vowel continua. Thus, the conclusion of Chistovich et al. that imitation bypasses conscious response selection, regardless of delay, is not contradicted.

Some striking differences among conditions are evident in Figure 1. Both subjects showed consistently longer reaction times for immediate than for delayed (+750 ms) imitations. (Note that the delayed-imitation latencies were measured from the onset of the "go" signal; to obtain stimulus-response latencies, 750 ms must be added to the times shown.) This is not unexpected: After all, the subjects knew exactly what to produce when the "go" signal occurred in the delayed condition, whereas in the immediate condition they had no such advance knowledge, and therefore had to wait until at least part of the stimulus vowel had been processed. It should also be noted that the latencies of subjects DW and BR were much longer than those of L.C. Neither of the present subjects was capable of the very close shadowing evinced by L.C. (Chistovich, Fant, de Serpa-Laitão, 1966). The fact that all responses were initiated from a closed mouth position may have something to do with this; L.C.'s articulatory resting position was not reported.

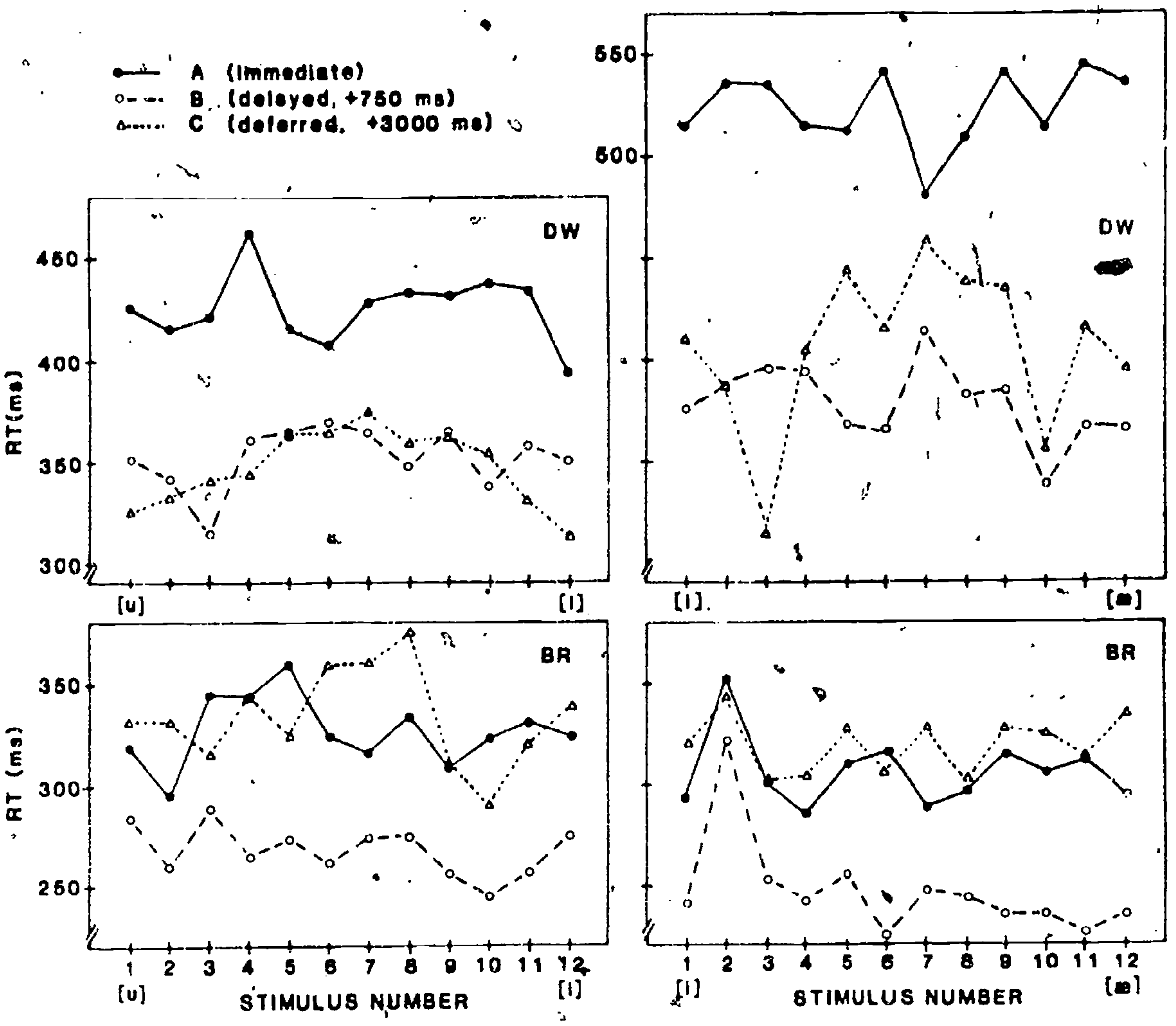


Figure 1. Response latencies (averaged over 12 responses) as a function of stimulus number, imitation condition, vowel continuum, and subject.

In the deferred imitation condition, there was increased temporal uncertainty about the moment of occurrence of the "go" signal; hence, an increase in reaction times might have been expected relative to the delayed imitation condition. Such an increase was shown by BR but not (or to a much lesser extent) by DW. The reason for this difference between subjects is not clear. Another difference of uncertain origin is DW's slower response to the [i]-[æ] continuum.

3.2. Formant Frequencies

Presentation of formant frequency data is simplified by the finding that two factors seemed to play only a very minor role. First, formant frequencies remained roughly constant or fell slightly over the course of the response vowels. Frequency measures at response vowel onset and two thirds into the vowel followed patterns almost identical to those of the mean formant frequencies across the whole vowel; therefore, the latter were chosen as the primary dependent variable. Second, formant frequencies were virtually identical across the three imitation conditions. Therefore, mean formant frequencies will be presented averaged across conditions (n=36 per stimulus).

Figure 2 plots the formant frequencies of the two subjects' responses to the 12 members of each vowel continuum. The formant frequencies of the stimulus vowels are indicated by the dashed lines. These plots are modeled after Figure I.-A-4 of Chistovich et al. [2], which showed several steps, especially in the function for F2. No such steps (i.e., true plateaus) can readily be discerned in Figure 2, except at the [u] end of the [u]-[i] continuum for DW. The response formant frequencies are not a linear function of the stimulus formant frequencies, however. Changes in the slope of the F2 function are evident especially along the [u]-[i] continuum, and also at the [i] end of the [i]-[æ] continuum for DW. Note also that the [i] stimulus was not well matched to the subjects' vocal capabilities, even though it was based on Peterson and Barney's (1952) male norms. For both DW and BR, F2 and especially F3 frequencies in the response vowels were much lower than in [i]-like stimuli.

A closer examination of these stimulus-response relationships is possible in Figures 3 and 4, which present two-dimensional formant frequency plots. Figure 3 shows the data for the [u]-[i] continuum in F2-F3 space; F1 varied very little across this series. Each stimulus vowel is connected by a line to its corresponding response. There is an obvious similarity between the two subjects' data in that the linear stimulus continuum is mapped onto a compressed, curvilinear contour in F2-F3 space. It is also evident that there are local expansion and compression effects that are quite different for the two subjects. DW did not distinguish among stimuli 1-3; this is the only clear instance of a categorical response in the present data. His responses to stimuli 3-6 were spaced widely apart, while stimuli 6-9 were mapped onto a compressed response space; responses to stimuli 9-12 roughly matched the distances between stimuli. Stimuli 6-9 could be interpreted as forming a second category on this continuum, although DW clearly was capable of responding discriminatively within that category. The stimulus-response mapping for BR is quite different. Major gaps appear between responses to stimuli 2-3, 4-5, and especially 9-10. Thus, four clusters of responses may be distinguished, perhaps corresponding to categories of some kind. Clearly, however, BR was able to respond distinctively to stimuli within each of these categories.

It might be added that the stimulus-response functions for F1, which are difficult to discern in Figure 2, were quite systematic and different for the two subjects, even though F1 changed over only a 30 Hz range. For DW, F1 was practically constant for responses to stimuli 1-10 and then decreased rapidly. For BR, on the other hand, F1 was roughly constant for stimuli 1-4 and then decreased almost continuously, although a local increase for stimulus 9 might lead one to consider stimuli 5-9 as a second grouping. Such a grouping would be consistent with the patterning seen in the lower panel of Figure 3.

The data for the [i]-[æ] continuum are shown in Figure 4; they are plotted in F1-F2 space, disregarding F3. Again, the response trajectories in this space are similar for the two subjects; they are very nearly linear, parallel to the stimulus continuum, and compressed, although more so for BR. The linearity reflects the fact that, as in the stimuli, changes in F1 and F2 were correlated in the responses. The correlation (computed between the F1

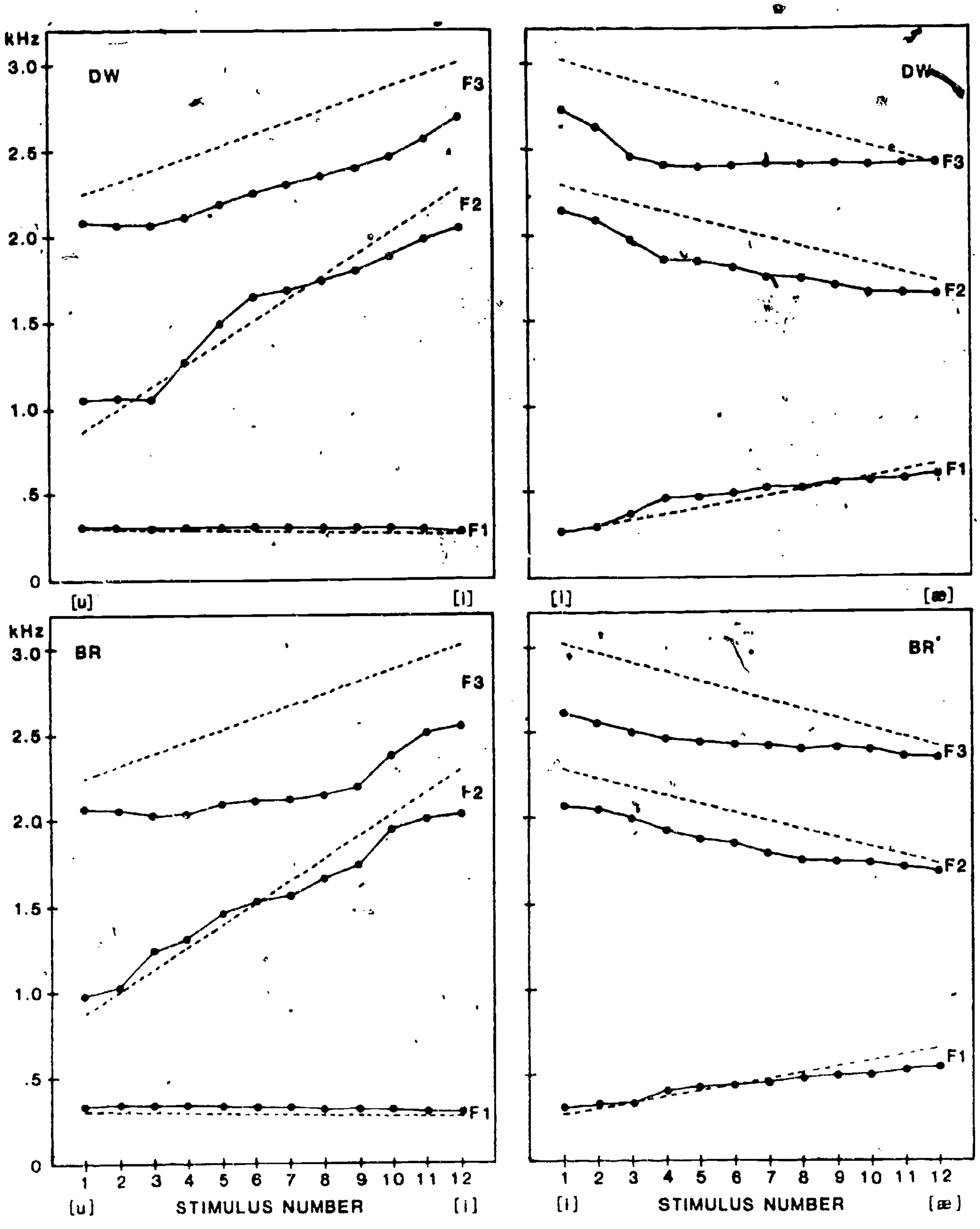


Figure 2. Response formant frequencies (averaged over 36 responses) as a function of stimulus number, vowel continuum, and subject. The dashed lines represent the stimulus formant frequencies.

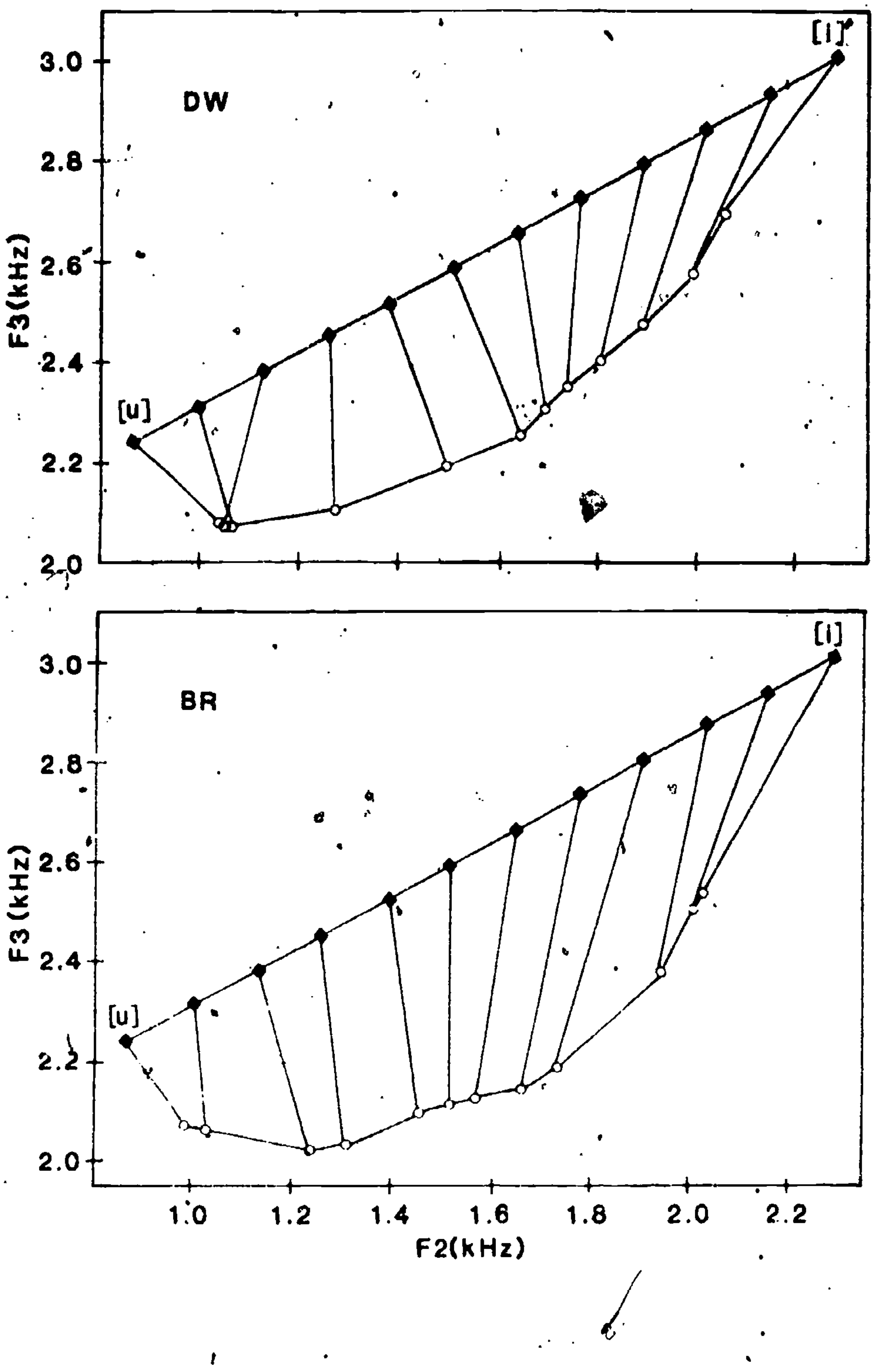


Figure 3. Average responses (circles) to stimuli from the [u]-[i] continuum (diamonds) in F2-F3 space. Corresponding stimuli and responses are connected by straight lines.

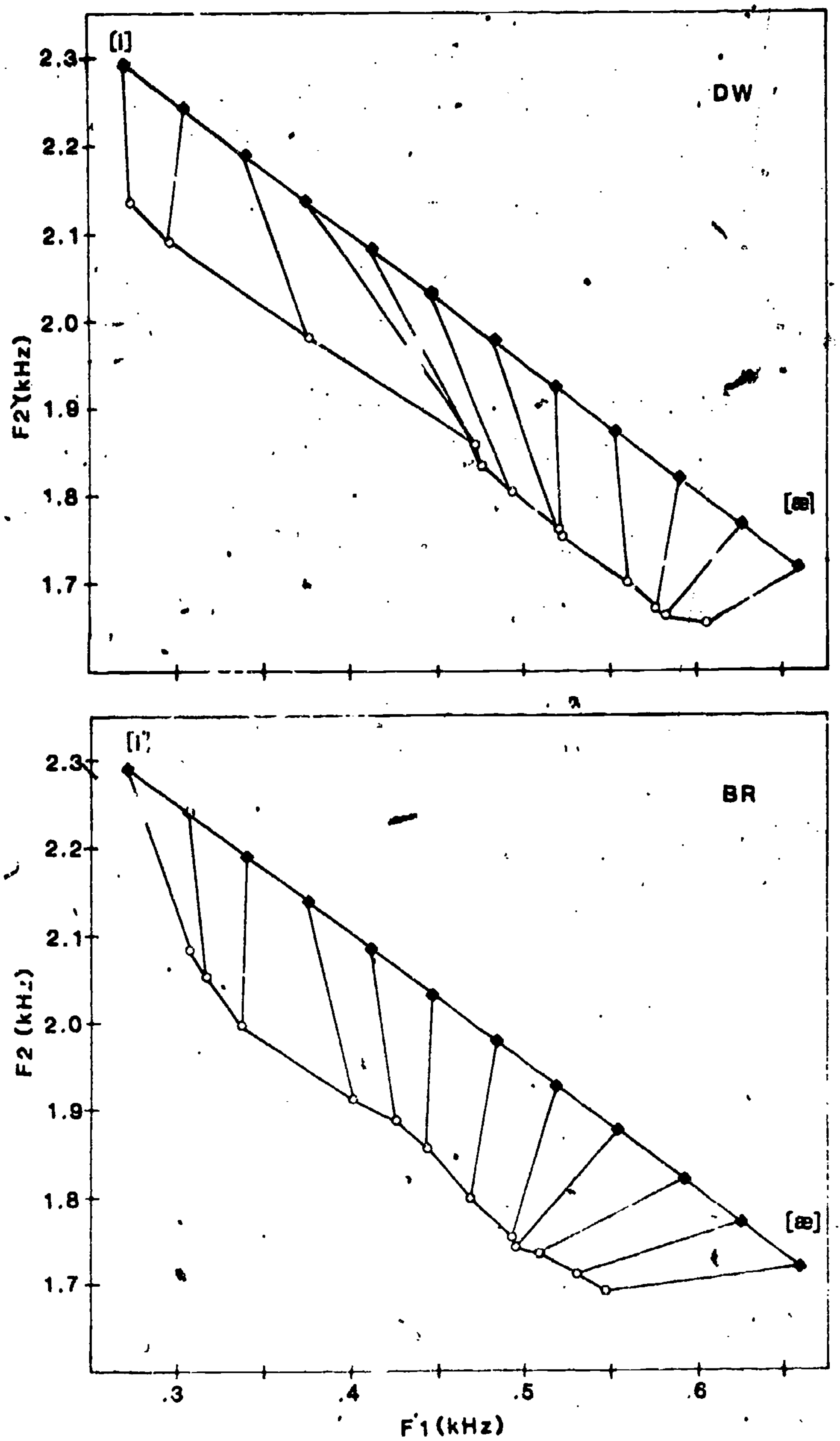


Figure 4. Average responses (circles) to stimuli from the [i]-[æ] continuum (diamonds) in F1-F2 space.

and F2 frequency differences of responses to adjacent stimuli on the continuum) was 0.97 for DW and 0.81 for BR ($p < .001$). The exact stimulus-response mapping, however, again shows individual differences. DW has some striking gaps between responses to stimuli 2-3-4; responses to stimuli 4-5, 7-8, and 10-11, on the other hand, are very similar. Clearly, there are strong distortions in this mapping, but they do not reveal a clear categorical structure. The same can be said for BR's data, although the distortions are less strong here, with the largest gap occurring between responses to stimuli 3-4. Neither subject's F3 values lead to different conclusions, as is evident from Figure 2: For DW, F3 changed rapidly in response to stimuli 1-4 and then remained completely insensitive to stimulus variations; for BR, similarly, F3 decreased more rapidly at first and then decreased more gradually from stimulus 4 on. There are no indications of any local categories in these F3 values.

To summarize, these formant frequency data offer ample evidence for nonlinearities in stimulus-response mapping, but little evidence for response categories in the strict sense. Moreover, they offer little support for Kent's (1973) tentative conclusion that responses to the [i]-[æ] continuum are more categorical than those to the [u]-[i] continuum. We turn now to an examination of the formant frequency standard deviations.

3.3. Standard Deviations

To simplify presentation, the standard deviations, like the formant frequencies, will be presented averaged across the three imitation conditions. These average within-condition standard deviations do not include between-condition variability, but this variability was small, as noted above. The absolute magnitude of response variability was comparable across conditions. The patterns of standard deviations, on the other hand, showed considerable differences among conditions. These differences must be viewed with caution, however, because of the similarity in mean formant frequencies across conditions; moreover, we have no estimate of measurement error (i.e., of the random variability of standard deviations). Some differences will be mentioned below.

The standard deviations for all three formants are displayed in Figure 5, separately for the two subjects and the two stimulus continua. Some general observations may be made at the outset: For both subjects, the standard deviations of F1 are larger on the [i]-[æ] continuum than on the [u]-[i] continuum, perhaps because only the former requires active control of degree of jaw opening. The standard deviations of F2, on the other hand (and, to some extent, those of F3), are much larger and more variable on the [u]-[i] continuum. This probably reflects the relative unfamiliarity of the vowel sounds along that continuum (cf. Kent, 1973). The fact that F3 variability is often lower than F2 variability may also be noted.

If there is a quasi-categorical structure underlying the imitation responses, then the standard deviations should increase whenever the slope of the formant frequency function increases (i.e., in the "between-category" regions). The striking peak in the F2 function for DW on the [u]-[i] continuum indeed coincides with the rapid change in response F2 frequency between stimuli 3 and 6 (see Fig. 2). The peaks at stimuli 4 and 9 in the F2 function for BR, on the other hand, have no such clear correlate in the pattern of mean F2 frequencies (Fig. 2). A finding not shown in Figure 5 is that, for both sub-

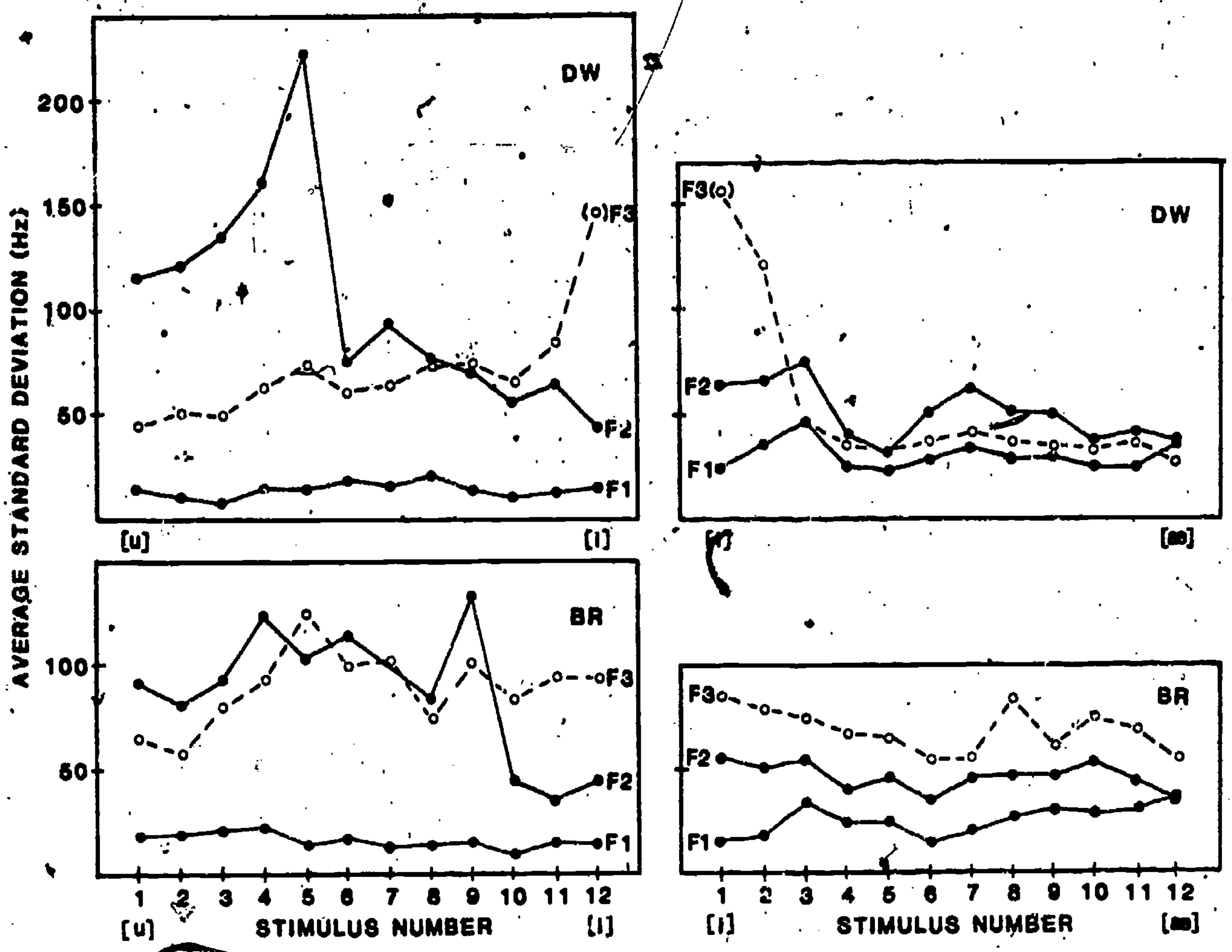


Figure 5. Formant frequency standard deviations (averaged over imitation conditions) as a function of stimulus number, vowel continuum, and subject.

jects, the F2 standard deviation peaks along the [u]-[i] continuum were most pronounced in the immediate imitation condition. The patterns of F1 and F3 standard deviations show only occasional correspondence to the pattern of F2 standard deviations, and they also varied considerably across conditions. (The large standard deviation of F3 for the [i] endpoint stimulus on both stimulus continua for subject DW was entirely due to the delayed imitation condition, for unknown reasons; hence the parentheses in Figure 5.)

Along the [i]-[æ] continuum, a correlation of F1 and F2 standard deviations can be seen for DW ($r = 0.70$, $p < .01$) but not for BR ($r = -0.03$). Standard deviation peaks for F1 seemed most pronounced in the immediate imitation condition. The pattern of F1 and F2 standard deviations finds some correspondences in slope changes in the formant frequency functions (Fig. 2). Correlations between the standard deviations and the absolute formant frequency differences between responses to stimuli n and n+1 on the [i]-[æ] continuum were significant for subject DW, both for F1 ($r = 0.80$, $p < .01$) and for F2 ($r = 0.74$, $p < .01$). For BR, on the other hand, the standard deviations were not significantly correlated with the pattern of response formant frequencies.

3.4. Frequency Distributions

As might have been expected from the similarity in mean formant frequencies across imitation conditions, the formant frequency distributions for each continuum were highly similar across conditions also. Therefore, histogram envelopes will be presented for responses in all three conditions combined (n = 432 per continuum). They are shown in Figures 6 and 7.

The data for the [u]-[i] continuum appear in Figure 6. The F1 distributions at the bottom are strongly unimodal and reflect the general upward shift of F1 in the subjects' responses, relative to the stimuli. The F2 distributions in the middle panels, on the other hand, cover the stimulus range rather well and show evidence of trimodality for both subjects. The first and third peaks of both subjects are located similarly near the endpoints of the continuum and presumably represent /u/ and /i/ categories, respectively. The middle peak is located differently: closer to [i] for DW but in the center of the continuum for BR. The clear trimodality of these distributions is surprising after the somewhat ambiguous patterns of the formant frequency and standard deviation curves. It indicates definite response preferences (or avoidances?) on the part of the subjects, although it should be noted that the speakers were able to produce any F2 frequency along the continuum, except for that corresponding to the [i] endpoint stimulus. The F2 response distributions for individual stimuli (not shown here) were also examined for bimodality. Although there were several distributions with two peaks (e.g., stimuli 4 and 5 for DW, stimuli 3 and 9 for BR), these peaks generally did not coincide with the major peaks seen in the overall histogram. They might indicate tendencies to avoid certain F2 values, or else they reflect just random variability. The F3 distributions (top panels) are strongly skewed toward low frequencies, and considerable "undershoot" is present, as noted earlier. In addition to the major peak reflecting the constancy of F3 over part of the continuum (cf. Fig. 2), two minor peaks may be distinguished for each subject; however, their locations are only partially consistent with those of the F2 peaks.

The histogram envelopes for the [i]-[æ] continuum are plotted in Figure 7. The F1 distributions at the bottom show good coverage of the stimulus range as well as very pronounced peaks, four for DW and three for BR. The peaks correspond to different stimuli for the two subjects. The absolute frequency locations of the three major peaks, however, are remarkably similar: 325, 475, and 550 Hz (±12 Hz) for DW; 324, 468, and 530 Hz (±7 Hz) for BR. There is much less correspondence in the F2 distributions for the two subjects (middle panels). For DW, the F2 histogram is clearly trimodal: One mode is in the [i]-[I] region and two modes are in the [ε]-[æ] region. The distribution for BR is less clear, having five peaks. Note the general asymmetry of these distributions; there seemed to be a strong "pull" toward lower F2 frequencies in the subjects' responses, apart from the absolute downward shift in F2. The F2 distributions for some individual stimuli (not shown) showed signs of bimodality (e.g., stimuli 3 and 8 for DW, stimuli 3 and 9 for BR), but not so strongly as to suggest discrete response categories. As on the [u]-[i] continuum, the whole range of F2 frequencies was represented in the responses. The F3 distributions (top panels) were unimodal and were centered at the lower end of the stimulus range for both subjects (cf. Fig. 2).

These data may be summarized by stating that (1) subjects show pronounced preferences for particular formant frequencies in their responses to both stimulus continua, and (2) the two subjects' responses are rather similar with

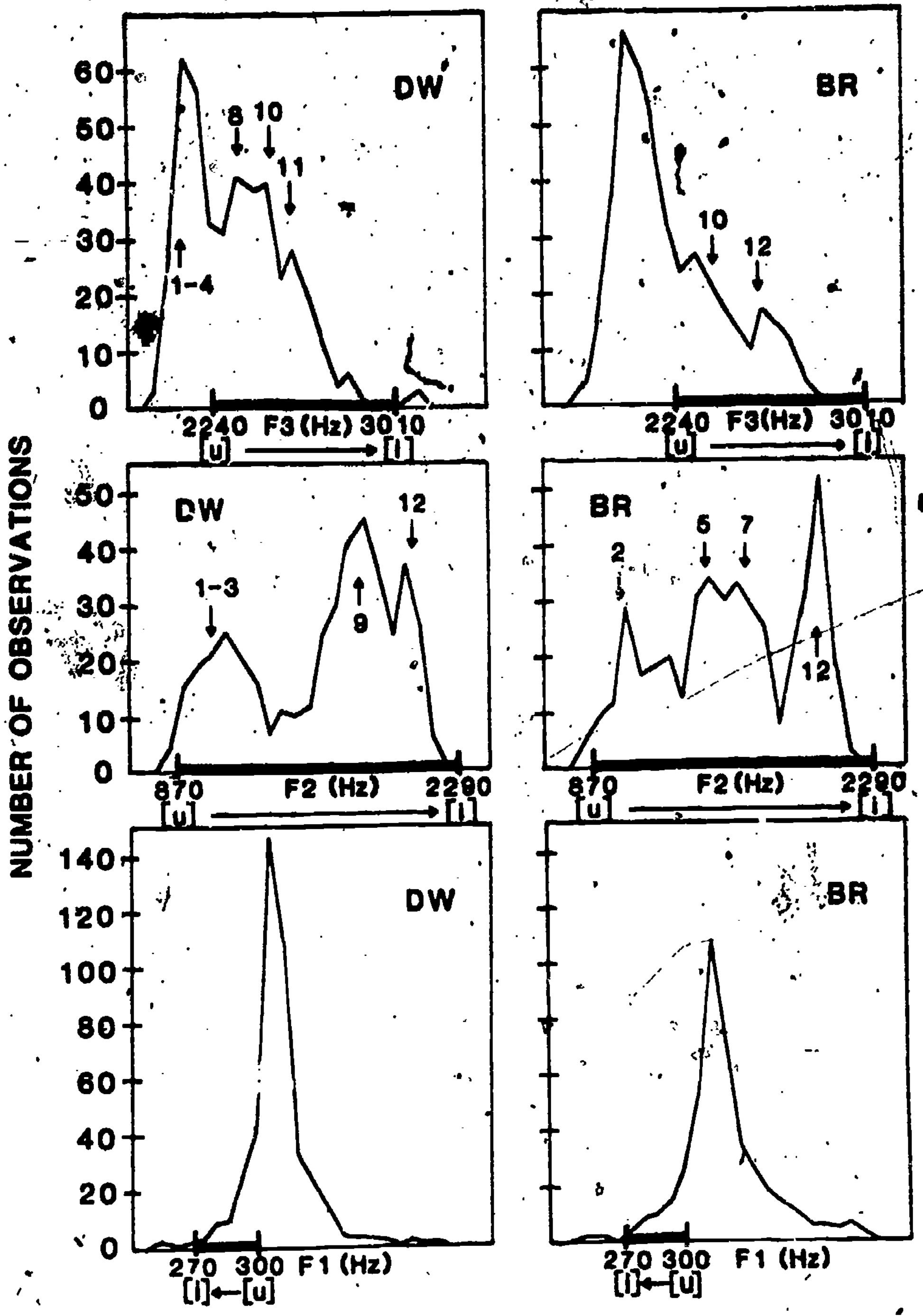


Figure 6. Formant frequency distributions for responses to stimuli along the [u]-[i] continuum. The heavy line at the bottom of each graph indicates the range of formant values along the stimulus continuum. Numbered arrows indicate stimuli for which the mean formant frequencies of the responses fell in the vicinity of peaks in the histogram. Note that the distributions for the different formants are not aligned with respect to each other or across subjects; rather, they are spread out over a constant number of histogram bins.

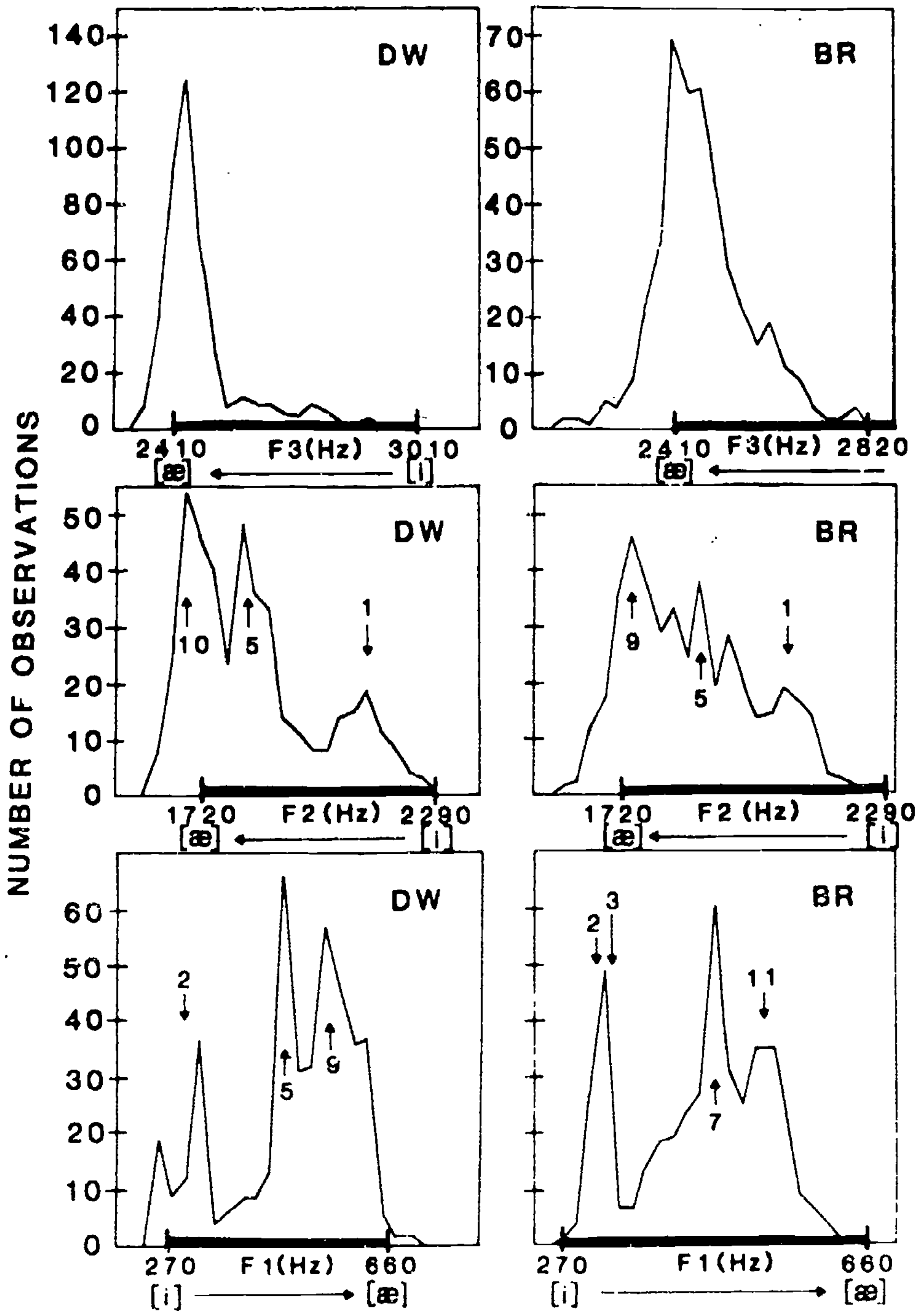


Figure 7. Formant frequency distributions for responses to stimuli along the [i]-[æ] continuum.

regard to F1 and F3, with individual differences residing primarily in the F2 distributions, particularly in responses to stimuli from the centers of the continua.

3.5. Written Identification

The subjects' phonemic labeling responses to the stimuli on the [i]-[æ] continuum are shown in Figure 8. It is evident that DW and BR applied somewhat different criteria, possibly because of their different language backgrounds. DW divided the continuum fairly consistently into five categories: /i/ (stimuli 1-4), /I/ (5), /e/ (6), /ɛ/ (8-11), and /æ/ (12). Stimulus 7 was ambiguous (i.e., less than 75 percent responses in any category). BR, on the other hand, applied only four categories consistently: /i/ (1-2), /e/ (4-5), /ɛ/ (7-9), and /æ/ (11-12). Stimuli 3, 6, and 10 were ambiguous to this listener. The /I/ category was not consistently applied by him to any stimulus; stimuli 2-7 received a few responses in that category. Essentially, for BR the /e/ category occupied the place of DW's /I/ and /e/ categories. A prominent role of /e/ might be expected in a native speaker of German, which has a monophthongal /e/ phoneme; however, the /I/ phoneme is also distinctive in German, not to speak of BR's long exposure to English. It will also be noted that BR's phoneme boundaries are all shifted toward the lower end of the continuum relative to DW's boundaries. The reason for this shift is not immediately evident.

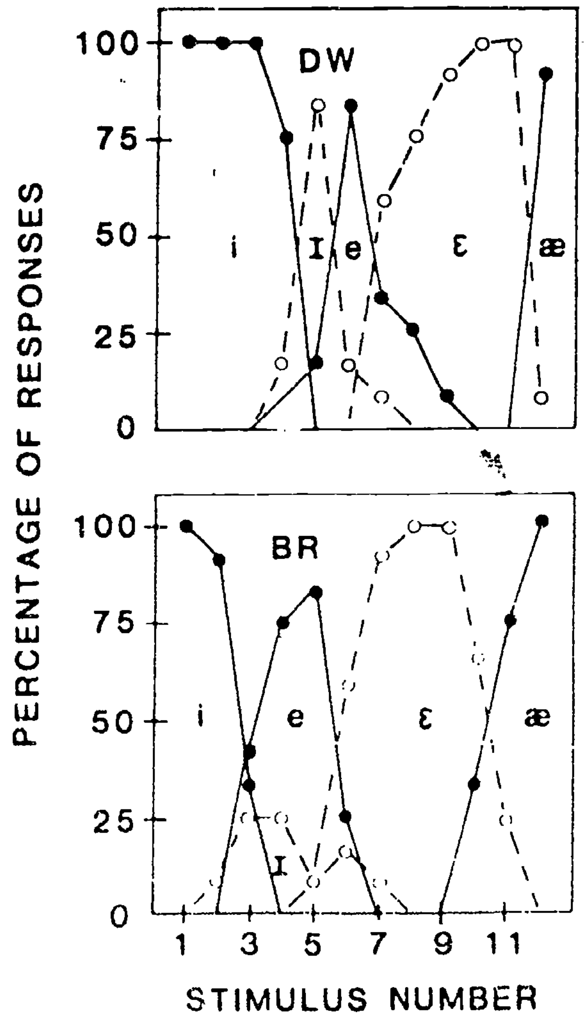


Figure 8. Phonemic identification responses to stimuli along the [i]-[æ] continuum.

The patterns of these labeling responses may be compared with the histogram peaks in Figure 7. For DW, the three major peaks in the F1 distribution (bottom left) may be identified with the /i/, /I/, and /ɛ/ categories, respectively. There are no peaks corresponding to /e/ and /æ/. The three peaks in the F2 distribution for DW (center left) correspond to /i/, /I/, and /ɛ/ again. The three peaks in BR's F1 distribution (bottom right) correspond somewhat less clearly to /i/, /ɛ/, and /æ/, while the three major peaks in his F2 distribution (center right) reflect more unambiguously /i/, /e/, and /æ/. There are no peaks for /e/ in F1 and for /æ/ in F2. For DW, and to some extent also for BR, it appears that the number of functional categories on this continuum is smaller in production than in perception, contrary to the conclusions of Chistovich, Fant, de Serpa-Leitão, and Tjernlund (1966).

This should not be taken to imply, however, that the subjects were somehow less sensitive to stimulus differences in the imitation task than in the phonemic labeling task. On the contrary, Figure 4 shows clearly that DW, for example, imitated quite differently the three stimuli (1-3) uniformly labeled as /i/. In fact, the clustering of modal responses in F1-F2 space (Fig. 4) is difficult to relate to the phonemic category boundaries (Fig. 8), which leaves the issue of the origin of the categorical tendencies in production unresolved.

Phonemic labeling of vowels does not reflect the extent of the subjects' perceptual sensitivity, as is abundantly clear from many earlier categorical perception studies (see Repp, 1984, for a review). This is also demonstrated by the absolute identification responses of the present two subjects, which are graphed in Figure 9. These responses are a monotonic function of stimulus

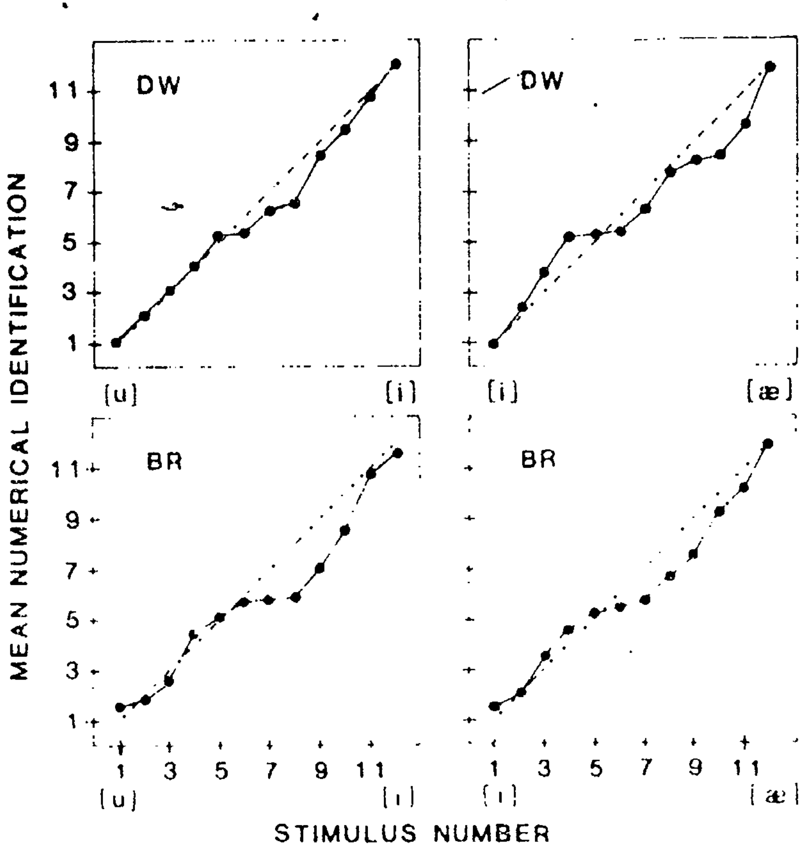


Figure 9. Average numerical identification responses to stimuli along both vowel continua. The dashed lines have a slope of 1.

number and show good discrimination of most adjacent stimuli, especially toward the ends of the continua. Thus, for example, it is clear that DW not only imitated differentially stimuli 1-3 on the [i]-[æ] continuum (all labeled /i/) but also was able to discriminate them perceptually without special training. More interesting, perhaps, is the observation that he also discriminated stimuli 1-3 on the [u]-[i] continuum, which he had imitated in identical fashion (cf. Fig. 3). Whether this implies a limitation on production or a perceptual limitation caused by the higher stimulus uncertainty in the imitation task remains to be seen. The poorer discrimination in the centers of the continua may be a consequence of the AXB paradigm, which provided endpoint anchors that facilitated discrimination of stimuli in their vicinity. The steps in the functions bear only a vague correspondence to the compression regions in the response formant space (Figs. 3 and 4).

4. Summary and Conclusions

The data presented here are preliminary and do not permit any strong conclusions, especially since their pattern exhibits some of the ambiguities observed by Kent (1973), after whose study the present experiment was modeled. A few tentative observations can be made, however, which should help guide future research on this topic.

(1) It appears from the present data that the pattern of vowel imitation responses is essentially insensitive to the delay between stimulus and response (from 300 to 3000 ms). This agrees with the conclusions of Chistovich, Fant, de Serpa-Leitão, ~~and~~ Tjernlund (1966), although it should be noted that the present subjects did not shadow as rapidly as subject L.C. Whatever internal representation of the stimulus mediates vocal imitation, it seems to be both (virtually) immediately available and relatively long-lasting in memory. A trend toward more pronounced patterns of formant variability in immediate imitation was observed.

(2) Vocal response latencies do not seem to vary much across a vowel continuum, as also observed by Chistovich, Fant, de Serpa-Leitão, and Tjernlund (1966). Thus, imitation seems to bypass conscious response selection, regardless of delay.

(3) In contrast to Chistovich, Fant, de Serpa-Leitão, and Tjernlund (1966), and, to some extent, in contrast to Kent (1973), we found no discrete steps in the response formant frequency functions. There was ample evidence, however, for nonlinearities in the stimulus-response relationships. The pattern of response variability resembled that of the formant frequencies for one subject only; basically, however, changes in mean formant frequencies and standard deviations can be assumed to reflect the same underlying tendencies. The absence of a strictly categorical structure in responses to isolated, stationary vowels is in agreement with the categorical perception literature (see Repp, 1984). It is not clear whether the stimulus-response nonlinearities should be interpreted in terms of perceptual categories, especially since they were difficult to relate to the phonemic labeling responses. The hypothesis needs to be pursued that these distortions have an independent origin in the production system.

(4) Response frequency histograms for individual formants showed very pronounced peaks and valleys. Thus the subjects had definite response preferences, which seemed to correspond to some of their phonemic categories. Not

all phoneme categories were represented however, and some categories had a peak only in a single formant. The relation of the histogram peaks to the stimulus-response nonlinearities is not easily characterized, although they are, of course, not independent. Generally, peaks correspond to regions of compression in the response formant space, and valleys correspond to regions of expansion. The response histograms, however, seem to provide a much clearer indication of underlying categories than do the formant frequency plots.

(5) We have found little support for Kent's (1973) tentative claim that responses to the [u]-[i] continuum, which does not harbor familiar phonemic categories apart from the endpoints, are less categorical than responses to the [i]-[æ] continuum, which spans five phonemic categories in English. The F2 histograms show three major peaks on both continua. It is true, however, that on the [i]-[æ] continuum F1 histograms provide additional striking information about response preferences, whereas F1 along the [u]-[i] continuum is too restricted in range to be informative. Response variability, especially in F2, was also much larger on the [u]-[i] continuum, presumably due to the unfamiliarity of the vowel sounds on this continuum.

(6) There were considerable individual differences between the two subjects, which might be related to their different language experiences. Some instances of congruity were also noted. Individual differences seemed most pronounced in the pattern of F2 frequencies.

(7) Chistovich, Fant, de Serpa-Leitão, and Tjernlund (1966) hypothesized the existence of an intermediate stage of representation, characterized by a number of categories exceeding (but including) the functional categories in the subjects' language. With regard to the [u]-[i] continuum, the hypothesis is supported, since both subjects seemed to have an additional category between the two familiar endpoints. The hypothesis is not supported for the [i]-[æ] continuum, however, where the number of categories in production (i.e., peaks in the formant histograms) was smaller than that of relevant vowel phonemes in the language. It is possible that this continuum was not sampled finely enough to reveal its full categorical structure in vocal reproduction.

In conclusion, the present results leave unresolved the issue of whether an intermediate mental representation needs to be postulated to account for nonlinearities in vowel imitation responses. These nonlinearities exist, however, and the search for their origin should continue. We plan to extend our research in several directions: by testing monolingual speakers of different languages to examine the role of linguistic experience; by matching the stimuli more closely to the subjects' vocal capabilities (as done originally by Chistovich, Fant, de Serpa-Leitão, & Tjernlund, 1966) so as to eliminate distortions that may arise in perceptual normalization; and by studying in more detail the possibility that the observed nonlinearities have their origin in articulation itself. Eventually, we also want to use more realistic, time-varying speech stimuli. In general, the thrust of our research will be to disentangle the perceptual and articulatory factors that jointly constrain vocal imitation. This enterprise seems interesting and worthwhile, and it provides but one example of the immense stimulus to speech research provided by the pioneering work of Ludmilla Chistovich and her colleagues.

References

Chistovich, L. A. (1984). Relation between speech production and speech perception. In M. P. R. van den Broecke & A. Cohen (Eds.), Proceedings of the Tenth International Congress of Phonetic Sciences (pp. 55-58). Dordrecht: Foris Publications.

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

Chistovich, L. A., Fant, G., & de Serpa-Leitão, A. (1966). Mimicking and perception of synthetic vowels, part II. Quarterly Progress and Status Report (Royal Technical University, Speech Transmission Laboratory, Stockholm), 3, 1-8.

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

Kent, R. D. (1974). Auditory-motor formant tracking: A study of speech imitation. Journal of Speech and Hearing Research, 17, 203-222.

Kent, R. D. (1978). Imitation of synthesized vowels by pre-school children. Journal of the Acoustical Society of America, 63, 1193-1198.

Kent, R. D. (1979). The imitation of synthesized English and non-English vowels by children and adults. Journal of Psychological Research, 8, 43-60.

Kent, R. D., & Forner, L. L. (1979). Developmental study of vowel formant frequencies in an imitation task. Journal of the Acoustical Society of America, 65, 208-217.

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

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

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, pp. 243-335). New York: Academic Press.

Schouten, M. E. H. (1977). Imitation of synthetic vowels by bilinguals. Journal of Phonetics, 273-283.

Footnotes

[1] The term "mimicking," used by Chistovich, Fant, de Serpa-Leitão, and Tjernlund (1966), suggests that the subject's response must match the stimulus in every respect. This was nearly true in their study, where the stimuli were modeled after vowels produced by the subject. In discussing our own data and those of Kent (1973), we will use the more general term, "imitation," which allows for stimulus-response differences in various irrelevant properties (duration, fundamental frequency, etc.) as well as for differences in formant frequencies caused by mismatches between the vocal tract implied by the stimulus and that of the imitator.

[2] We yet have to conduct a full survey of the relevant literature, especially that published in Russian.

INFLUENCE OF FOLLOWING CONTEXT ON PERCEPTION OF THE VOICED-VOICELESS DISTINCTION IN SYLLABLE-FINAL STOP CONSONANTS*

Bruno H. Repp and David R. Williams†

Abstract. This paper reports acoustic measurements and results from a series of perceptual experiments on the voiced-voiceless distinction for syllable-final stop consonants in absolute final position and in the context of a following syllable beginning with a different stop consonant. The focus is on temporal cues to the distinction, with vowel duration and silent closure duration as the primary and secondary dimensions, respectively. The main results are that adding a second syllable to a monosyllable increases the number of voiced stop consonant responses, as does shortening of the closure duration in disyllables. Both of these effects are consistent with temporal regularities in speech production: Vowel durations are shorter in the first syllable of disyllables than in monosyllables, and closure durations are shorter for voiced than for voiceless stops in disyllabic utterances of this type. While the perceptual effects thus may derive from two separate sources of tacit phonetic knowledge available to listeners, the data are also consistent with an interpretation in terms of a single effect, one of temporal proximity of following context.

Introduction

Acoustic cues to the perception of the phonological voiced-voiceless distinction in American English syllable-final stop consonants have been investigated quite intensively in recent years. One important cue is "vowel duration" (i.e., the duration of the periodic stimulus portion taken to correspond to the vowel--see, e.g., Raphael, 1972; Raphael, Dorman, & Liberman, 1980), which is consistent with the commonly observed longer duration of vowels preceding voiced consonants in speech production (House & Fairbanks, 1953; Peterson & Lehiste, 1960). Other relevant perceptual cues include the offset characteristics (i.e., formant transitions and amplitude envelope) of the "vowel" (Wang, 1959; Wolf, 1978), its fundamental frequency contour (Gruenenfelder & Pisoni, 1980; Lehiste, 1976), and--if the stop consonant is released--the acoustic properties of the release (Malécot, 1958; Wolf, 1978), as well as the duration and voicing of the closure interval (Hogan & Rozsypal, 1980; Raphael, 1981). All these cues are also relevant for stop consonants

---

*Journal of the Acoustical Society of America, in press.

†Also Department of Psychology, University of Connecticut.

Acknowledgment. This research was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. Results from Experiments 6 and 7 were reported at the 107th Meeting of the Acoustical Society of America in Norfolk, VA, May 1984.

in intervocalic position, where additional voicing information may be contained in the vocalic portion following the release burst (Lisker, 1978).

Although vowel duration is not always the most salient voicing cue (e.g., Wardrip-Fruin, 1982), it is nevertheless an acoustic dimension that has consistently been found to influence the perception of phonological stop consonant voicing in English. As a purely temporal cue, it has attracted researchers' attention because it offers an opportunity to study the sensitivity of phonetic perception to local and global changes in speaking rate--a topic of much theoretical interest (Miller, 1981; Port, 1981). A second temporal cue to the voicing distinction--the duration of the closure interval--is available in intervocalic and released final stops. Provided that closure voicing, an often overriding cue (e.g., Lisker, 1981), is eliminated, closure duration provides important voicing information for intervocalic stops (Lisker, 1957), though it is less salient in released utterance-final stops (Raphael, 1981). Port and Dalby (1982) have proposed that the joint perceptual influence of the two temporal variables, vowel duration and (silent) closure duration, is best expressed by a constant ratio rule (however, see Massaro & Cohen, 1983a). In production, too, the ratio of the two interval durations (the C/V ratio) seems to be fairly constant across changes in global speaking rate (Barry, 1979; Port, 1981). The ratio varies, however, across different utterance positions and as a function of other voicing cues (Barry, 1979).

The present study investigates the perception of phonological stop consonant voicing in a context that has not been studied previously, viz., when a syllable-final stop is followed by another syllable beginning with stop consonant having a different place of articulation. Although the basic voicing cues are likely to be those already studied extensively in connection with stop consonants in intervocalic or absolute final position, sequences of two different stop consonants have several peculiar properties that warrant detailed investigation.

One consideration is that the total closure period for sequences of two nonhomorganic stop consonants is about twice as long as that for a single intervocalic stop (Repp, 1982; Westbury, 1977). When the first stop is unreleased, as is frequently the case (Henderson & Repp, 1982), there is no acoustic or perceptual basis for subdividing the total closure interval into portions pertaining to the two consecutive (perhaps overlapping) stop closures. This raises the question of whether closure duration is a salient cue for the perception of voicing in this context. Certainly, if it has any effect at all, one would expect to find different critical C/V ratios than for single intervocalic stops.

Whe[illegible]st stop is released, the situation is similar to that for released [illegible] absolute final position, except that the release bursts of stops fo[illegible]y another stop are generally much weaker (Henderson & Repp, 1982). [illegible] raises the question of whether such weak bursts can serve as a perceptual marker delimiting the closure interval pertaining to the syllable-final stop, thereby possibly changing the relative salience of the closure duration cue and with it the critical ratio to vowel duration. It has been shown that these weak release bursts carry considerable place of articulation information (Repp, 1983b), so it is not unreasonable to expect that they might have some influence on voicing perception as well.

Another prediction that could be made is that, in sequences of two nonhomorganic stop consonants, the onset characteristics of the second syllable will have less of an effect on the perceived voicing of the preceding syllable-final stop than they have in the case of single intervocalic stops. For one thing, the temporal separation of pre- and post-closure cues is greater because of the extended closure interval, which makes perceptual integration more difficult. In addition, the two stops have different places of articulation, which may result in a perceptual segregation of the respective voicing cues, syllable-initial cues pertaining only to the syllable-initial consonant. On the other hand, it has been observed in such VC_1C_2V stimuli that the perception of the place of articulation of one stop is influenced by that of the other (Repp, 1983a); so it may be asked whether there are similar (contrastive) interactions with regard to voicing perception. Repp (1983a) linked apparent perceptual contrast effects in place-of-articulation perception to listeners' intrinsic knowledge of systematic variations in closure duration in natural speech. However, apart from an unpublished study by Westbury (1977), little is known about the acoustic consequences of phonological voicing in nonhomorganic stop sequences. Therefore, the present study reports acoustic as well as perceptual data.

A final important goal of the present investigation was to demonstrate an effect of following context on voicing perception by comparing the absolute vowel durations required to change voiceless to voiced final stop percepts in monosyllables and in disyllables. Since it is known that, in production, the duration of a syllable decreases when a second syllable is added to form a disyllabic word (Klatt, 1973; Lehiste, 1972), it was predicted that addition of a second syllable would considerably reduce the absolute vowel duration at the voiced-voiceless boundary in the first syllable, while maintaining the perceptual relevance of the vowel duration cue. Indeed, an analogous effect on the perception of phonological vowel length was shown long ago by Nooteboom (1973). Given such a contextual effect, it might be asked further whether the effect is dependent on whether or not the listener considers the two syllables as parts of the same word, and by how much the temporal separation between the two syllables can be increased before the voicing boundary for the final stop of the first syllable approaches that for the final stop in an isolated monosyllable (cf. Nooteboom & Doodeman, 1980).

For the present experiments, two pairs of monosyllabic English words were sought, such that one ended and the other began with either a voiced or a voiceless stop consonant (in terms of spelling, at least), and that made sense in all four possible disyllabic combinations as well as in isolation. Such a set does not exist. Rather than using nonsense materials, an approximation was devised by using the words LAB/LAP and GOAT/COAT, which in combination yield the real word LABCOAT, as well as the novel but potentially meaningful compounds LAPCOAT, LABGOAT, and LAPGOAT. This set was considered more attractive to listeners than complete nonsense, although the possibility of semantic bias in phonetic perception (cf. Ganong, 1980) must be considered. As will be seen, however, it is unlikely that such a bias influenced the results in any significant way.

I. Acoustic Measurements

Acoustic measurements were obtained to get a general idea of the acoustic consequences of phonological stop consonant voicing in sequences of two nonhomorganic stop consonants, and particularly in the kinds of utterances used also in the perceptual experiments. The only relevant previous data were reported by Westbury (1977) who measured three speakers' productions of isolated CVC$_1$C$_2$VC nonsense utterances, in which C$_1$ and C$_2$ were stop consonants differing in both voicing and place of articulation. Westbury's acoustic measurements showed that, in voiced-voiceless stop sequences, the vowel in the first syllable was about 10 ms longer and the closure interval about 20 ms shorter than in voiceless-voiced sequences. Whether these differences were due to the voicing characteristics of the first or the second stop, or both, cannot be determined from Westbury's data. Also, C$_1$ in these utterances was either consistently unreleased, or Westbury ignored the C$_1$ release burst in his measurements.

In the present study, too, speakers produced isolated utterances. While these productions are not representative of fluent speech, and a certain amount of deliberate enhancement of phonetic differences may be expected, the data are appropriate for comparisons with perceptual responses to similarly isolated utterances, as collected in the subsequent experiments.

A. Method

1. Subjects. Four native speakers of American English, three females (CG, JM, AB) and one male (DW, the second author), served as talkers. CG and JM grew up in New York, AB in the Midwest, and DW in California.

2. Utterances. The utterances were LAB, LAP, GOAT, COAT, LABGOAT, LABCOAT, LAPGOAT, and LAPCOAT. Ten different random orders of these eight words were concatenated and printed on a sheet of paper in standard English spelling.

3. Recording procedure. Each talker read from the list after practicing for a few minutes. The instructions were to read at a steady rate, pausing after each word, and to speak clearly but naturally. The utterances were recorded in a sound-insulated booth using high-quality equipment.

4. Measurement procedures. Temporal properties of the utterances were measured from magnified CRT waveform displays. Durations to the nearest tenth of a millisecond were obtained between the following acoustic landmarks (described here with reference to a disyllabic utterance): (a) the onset of significant energy; (b) the point of change from [l] to [æ], as determined visually by a noticeable change in the waveform of the glottal cycle; (c) the beginning of the closure interval, as indicated by a significant damping or cessation of voicing pulses; (d) the onset of the C$_1$ release burst, if present; (e) the end of the C$_1$ burst; and (f) the onset of the C$_2$ release burst.

In this way, the durations of the following acoustic segments were obtained: (1) [l] resonance, (2) [æ] resonance ("vowel")[1], and (3) total closure interval. If C$_1$ was released, (3) could be subdivided into (3a) C$_1$ closure, (3b) C$_1$ release burst, and (3c) C$_2$ closure. Monosyllabic LAB and LAP were always released by three talkers; talker CG did not release 8 out of 20 utterances. In disyllabic context, labial release bursts were produced in all tokens by talkers JM and CG; there were 11/40 unreleased tokens for AB and 5/40 for DW.

For each talker, means and standard deviations of these acoustic segment durations were calculated from the 10 repetitions of each utterance. True outliers were omitted; there were not more than a few for each talker. Analyses of variance were conducted separately for each dependent variable, using a repeated-measures design on the mean durations.

B. Results and Discussion

One effect of interest is the shortening of the first syllable in a disyllabic word, as compared to its production in isolation. Although such shortening has been described previously (e.g., Klatt, 1973; Lehiste, 1972; Nooteboom, 1973), the added syllables in these studies were unstressed, while the present disyllables had a spondaic stress pattern. Nevertheless, shortening of the first syllable was exhibited consistently by all talkers. Table 1 compares the durations of [l], [æ], C_1 closure, and C_1 release burst in mono- and disyllables, averaging over the four talkers and over voicing distinctions (rows labeled "mean"). It is evident that most of the shortening took place during the [æ] portion--an average reduction of 73 ms (30 percent), which was highly significant, $F(1,3) = 44.6$, $p < .007$. By contrast, the C_1 closure changed by only 10 ms (10 percent), $F(1,3) = 16.5$, $p < .03$, and the [l] portion did not change significantly. In addition, a dramatic difference in C_1 release burst durations is evident, $F(1,3) = 71.5$, $p < .004$: While utterance-final labial release bursts contained significant amounts of aspiration or voicing, those in disyllabic utterances basically represented only the brief noise generated by the parting of the lips (cf. Henderson & Repp, 1982).

The amount of vowel shortening is comparable to that observed for trochaic words (Klatt, 1973). Klatt also observed about twice as much shortening when the first syllable ended in a voiced consonant than when it ended in a voiceless consonant. Such a trend was also found in the present data (32 percent for LAB versus 26 percent for LAP), though it was much smaller, primarily due to relatively less shortening of LAB than would be predicted from Klatt's data.

The second comparison of interest is that between phonologically voiced and voiceless syllable-final stop consonants (LAB versus LAP in Table 1). It is evident that the [æ] portion was longer, $F(1,3) = 39.4$, $p < .009$, and the C_1 closure duration was shorter, $F(1,3) = 5.0$, $p < .12$, in LAB than in LAP. (The difference in C_1 closure duration was shown by all four subjects but varied considerably in magnitude; hence the low level of significance.) C_1 release bursts tended to be longer when the stop was voiceless; however, this difference was shown by only two talkers. The C_2 closure was also affected by C_1 voicing, being shorter for LAB than for LAP, $F(1,3) = 22.5$, $p < .02$. The duration of the initial [l], on the other hand, was completely unaffected by stop consonant voicing. Another difference, not shown in Table 1, was that the C_1 closure of LAB usually contained low-amplitude voicing while that of LAP did not. The voicing usually ceased before the end of the C_1 closure.

The average difference in [æ] duration between LAB and LAP was 98 ms (33 percent) in monosyllables and 55 ms (27 percent) in disyllables. For C_1 closure, the difference in duration was -25 ms (30 percent) in monosyllables and -16 ms (21 percent) in disyllables. The C/V ratios in mono- and disyllables, respectively, were 0.28 and 0.39 for LAB, and 0.55 and 0.65 for LAP. This comparison shows that the addition of a second syllable increased the C/V ratio, which thus was not invariant.

Table 1

Comparison of acoustic segment durations (ms) in monosyllables and in disyllables, averaged across talkers and tokens.

	[l]	[æ]	C_1 closure	C_1 burst	C_2 closure	Total closure
Monosyllables						
LAB	82	295	83	117		
LAP	83	197	108	136		
Mean	83	246	96	126		
Disyllables						
LAB-	78	201	78	14	84	176
LAP-	77	146	94	21	117	232
Mean	78	173	86	17	101	204

Table 2

Comparison of average acoustic segment durations (ms) as a function of following GOAT or COAT in disyllables.

	[l]	[æ]	C_1 closure	C_1 burst	C_2 closure	Total closure
-GOAT	78	177	87	18	110	215
-COAT	78	170	85	17	91	193

A third comparison relevant to the perceptual experiments concerns the effect of the voicing of C_2 (GOAT vs. COAT) on acoustic properties of the preceding syllable. Table 2 lists the relevant data, averaged over all other factors. It is evident that there was very little effect: Only the [æ] vowel was slightly shorter preceding COAT than preceding GOAT--a small difference that was almost unreasonably consistent across talkers, $F(1,3) = 358.9$, $p = .0003$. In addition, the C_2 closure seemed to be affected by C_2 voicing, being longer for GOAT than for COAT. However, this difference, which runs counter to the common finding of longer closures for voiceless than for voiced stops, was exhibited by only two talkers and hence was nonsignificant. Westbury's (1977) observation that the total closure was shorter in voiced-voiceless than in voiceless-voiced C_1-C_2 sequences is thus confirmed by the present data, even though closure durations exhibited large individual differences.

To summarize: "Vowel duration" in LAB/LAP is substantially reduced when a second syllable is added; it is shorter in LAP than in LAB; and it is slightly shorter preceding GOAT than preceding COAT. C_1 closure duration tends to be shorter for LAB than for LAP, and it is shortened somewhat when a second syllable is added. C_2 closure may be longer preceding GOAT than preceding COAT. C_1 release bursts are substantially reduced in intersyllabic position as compared to absolute final position.

II. Perception Experiments

The perceptual studies were carried out in two stages. An initial five-part study (Exps. 1-5) was followed by a two-part replication experiment conducted several years later with a new set of stimuli (Exps. 6 and 7).

A. General Method: Experiments 1-5

1. Subjects. Nine paid student volunteers served as subjects. They were all native speakers of American English and reported having no speech or hearing problems.

2. Stimuli. Selected utterances of one female talker (CG) were used for stimulus construction. A continuum from LAP to LAB was constructed from the first syllable of a representative token of LABGOAT. The original duration of that syllable (not including the closure interval) was 320 ms. A seven-member continuum, not including the original syllable, was constructed by deleting pitch pulses from the interior of the [æ] portion. The initial [l] portion, approximately 62 ms long, was left undisturbed. The members of the LAP/LAB continuum had vowel durations ranging from 118 to 238 ms in 20-ms steps. A second LAP/LAB continuum, used only in Experiment 1, was constructed from the first syllable of a good token of LAPCOAT whose original duration was 226 ms (51 ms for [l]), by either deleting or duplicating pitch pulses in the [æ] portion. The vowel durations of the seven members of that continuum ranged from 134 to 255 ms in 20-ms steps.[2] These disyllable-derived stimuli were acceptable as monosyllables with a neutral intonation, whereas stimuli fashioned from LAB or LAP produced in isolation would have been unacceptable in the context of a disyllabic word because of their falling intonation. A strongly falling intonation contour would also have made construction of a syllable duration continuum problematic.

In some conditions, the stimuli from the LAP/LAB continuum were followed by one of two C_1 release bursts. These bursts and the surrounding closure interval were derived from tokens of LABCOAT and LAPGOAT, respectively. Any closure voicing present was replaced with silence, so as to eliminate a potentially overriding (Lisker, 1981) nontemporal cue. The C_1 closure durations were 65 and 88 ms, respectively, the release burst durations were 9 and 13 ms, and the total closure durations were 170 and 243 ms. These durations were representative of those observed in the sample of utterances recorded and analyzed (see Table 1). Note that closure duration and origin of release burst were confounded.

A good token of GOAT, with a voice onset time (VOT) of 22 ms and a total duration of 473 ms (including the final [t] release burst), was excerpted from LABGOAT. Since it was desirable to use tokens of GOAT and COAT that differed only in their initial VOT, the initial 66 ms of the second syllable of LABCOAT, representing the aperiodic portion (i.e., the VOT) of COAT, was substituted for the initial 66 ms of GOAT to yield an acceptable token of COAT.[3]

The stimuli for each experiment were recorded on audio tape in five randomized blocks with intertrial intervals of 2.5 s.

3. Procedure. Each subject participated in two sessions. In the first session, the stimulus tapes representing Experiments 1, 2, and one additional test (see Footnote 3) were presented. In the second session, Experiments 5, 3, and 4 were administered, always in that order. Subjects listened over TDH-39 earphones in a quiet room. Their task was to identify in writing the final stop consonant of the first syllable ("b" or "p") and, when a second syllable followed, its initial stop consonant as well ("g" or "c"), even when it was constant (except for Exp. 4).

B. Experiment 1

The purpose of this first test was twofold: to assess the influence of a final release burst on the LAP/LAB distinction in monosyllabic stimuli, and to compare perception of two LAP/LAB continua, one derived from an original utterance of LAB-, the other from LAP-. Several earlier investigators have noted that it is easier to change an originally voiced syllable-final stop consonant into a voiceless one by manipulating vowel and/or silent closure duration than vice versa (Hogan & Rozsypal, 1980; Price & Lisker, 1979). At the very least, a "trading relation" between vowel duration and differential vowel offset cues was expected in Experiment 1. As to the perceptual contribution of the C_1 closure and release burst, previous investigations (e.g., Hillenbrand, Ingrisano, Smith, & Flege, 1984; Malécot, 1958; Wolf, 1978) employing release bursts appropriate for utterance-final stops generally found only small effects. Although the disyllable-derived release bursts in the present stimuli were acoustically weaker, the relative ambiguity created by varying vowel duration was expected to enhance any effects of secondary voicing cues.

1. Method. For the seven stimuli from each LAP/LAB continuum, a final release burst (with associated C_1 closure) was either present or absent and, if present, derived from either LAB- or LAP-. Stimuli from a GOAT/COAT continuum (see Footnote 3) were interspersed as fillers.

2. Results and discussion. The results are shown in Figure 1. It can be seen, first, that the average percentage of "b" responses increased as vowel duration increased. Thus, vowel duration was an effective cue to the voiced-voiceless distinction, as intended. Second, "b" responses were much more frequent to the LAB-derived stimuli than to the LAP-derived stimuli, which reflects additional important cues presumably located at the offset of the periodic stimulus portion. This difference was highly significant in a repeated-measures analysis of variance, $F(1,8) = 33.1$, $p < .001$. Third, the presence of a release burst (and of an associated C_1 closure interval) did make a difference, $F(2,16) = 11.9$, $p < .001$. The effect was generally one of reducing "b" responses, even when the burst and closure derived from LAB-, at least in the case of the LAP-derived continuum. A separate analysis of variance was conducted on stimuli with bursts only. Bursts derived from LAP- led to fewer "b" responses than did bursts derived from LAB-, $F(1,8) = 8.8$, $p < .02$, and this difference was more pronounced for the LAB-derived continuum, $F(1,8) = 5.8$, $p < .05$ for the interaction.

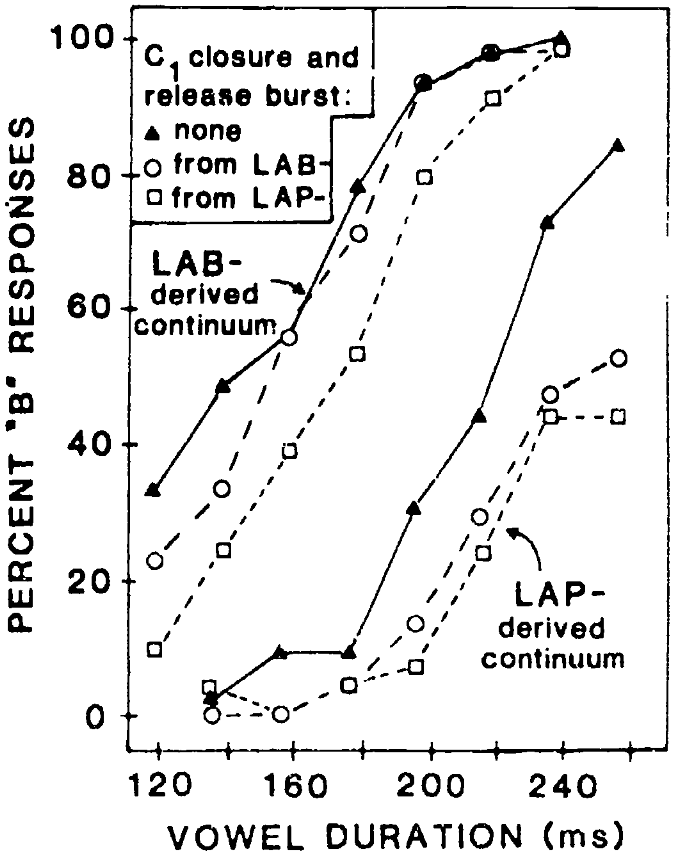


Figure 1. Effects of vowel offset cues (LAB- vs. LAP-derived continuum) and C_1 closure/release burst on the voiced-voiceless distinction for syllable-final stops along a vowel duration continuum (Experiment 1).

These results confirm the relative salience of vowel duration as a voicing cue. Although, within the range of durations used here, a complete change of perceived category was not achieved in either continuum, vowel duration was about equally effective in changing LAB to LAP, and LAP to LAB. This is in contrast to some earlier studies that have found it difficult to change final voiceless stops into voiced ones (Hogan & Rozsypal, 1980; Price & Lisker, 1979). The lack of such an asymmetry in the present stimuli may be due to their having been produced in the context of a disyllabic word, which perhaps made the vowel offset cues somewhat less pronounced. Still, they were quite strong, being "worth" roughly 80 ms of vowel duration at the point of maximal ambiguity.

The finding that release bursts reduced "b" responses regardless of the burst's origin may be attributed to the absence of closure voicing: Presence of a release burst defined a closure interval that was always silent and thus more appropriate for "p" than for "b". The differential effect of LAB- and LAP-derived bursts may have been due either to properties of the release bursts themselves or to C_1 closure duration (or both). Why this effect was more pronounced with the LAB-derived continuum is not clear.

C. Experiment 2

The purpose of Experiment 2 was fourfold. First, the LAP/LAB stimuli were now presented in a disyllabic context (i.e., followed by GOAT or COAT), and a shortening of the absolute vowel duration necessary to cue the LAP/LAB distinction was expected relative to Experiment 1, by analogy to the findings of Nooteboom (1973; Nooteboom & Doodeman, 1980). Second, Experiment 2 investigated the perceptual contribution of (total) closure duration in the disyllabic context. Third, the effect of presence versus absence of a C_1 release burst was also studied in this new context. Finally, possible perceptual contrast effects due to the voicing category of the initial consonant of the second syllable (GOAT or COAT) were assessed.

1. Methods. Only the LAB-derived LAP/LAB continuum was used. These syllables were followed by one of four closure intervals and by either GOAT or COAT. Two of the closure intervals were those also used in Experiment 1, which contained a release burst derived from either a voiced or a voiceless syllable-final stop. In contrast to Experiment 1, however, where only the C_1 closure was defined (65 or 88 ms), here the C_2 closure was defined as well by the onset of the GOAT or COAT syllable. Two additional conditions resulted from substituting silence for the release bursts so that the closure interval consisted of either 170 or 243 ms of pure silence.

2. Results and discussion. Analysis of subjects' responses revealed no influence of the GOAT/COAT contrast on the LAP/LAB distinction, $F(1,8) = 1.7$. Therefore, Figure 2 shows the results collapsed over this factor, as a function of total closure duration and presence versus absence of a release burst. It is evident from the figure and from the statistical analysis that the release burst had no systematic effect, $F(1,8) = 1.5$. However, total closure duration did have an influence: Fewer "b" responses were obtained with the longer closure duration, $F(1,8) = 20.9$ $p < .002$. Finally, contrary to expectations, the LAP/LAB boundaries were located at about the same point as in Experiment 1, revealing no influence of the addition of a second syllable.

The absence of this expected contextual effect must be interpreted with caution because of the different stimulus ensembles used in Experiments 1 and 2. In Experiment 1 the inclusion of LAP-derived stimuli in the test sequence may have had a contrastive effect that pushed the boundary for the LAB-derived stimuli toward shorter vowel durations. That this was the case is suggested by the results of Experiments 4 and 6, which directly compared LAP/LAB stimuli

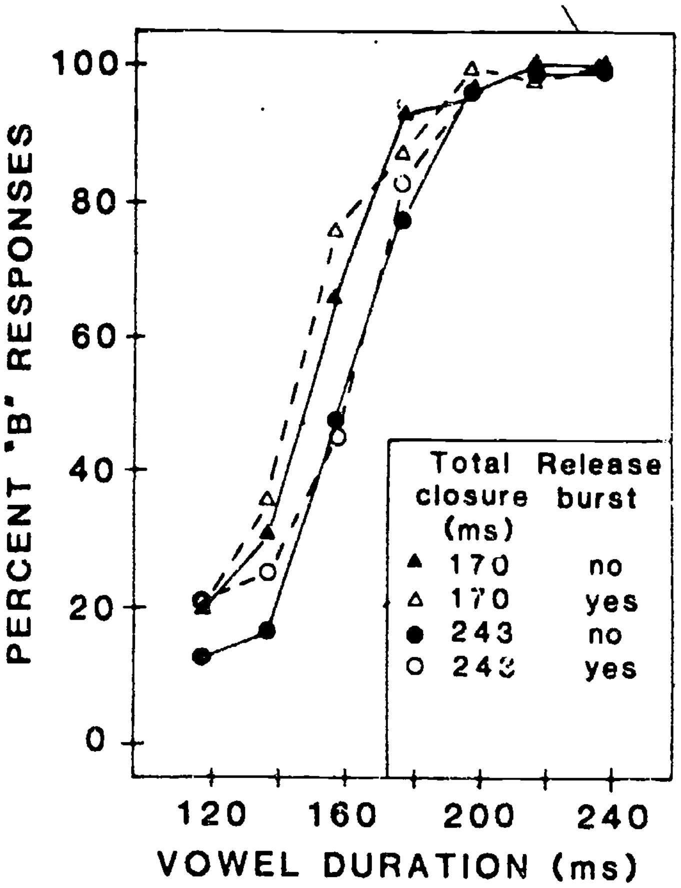


Figure 2. Effects of total closure duration and release burst on the voiced-voiceless distinction for syllable-final stops in disyllables (Experiment 2).

in isolation and in disyllabic context, and obtained a reliable difference (see below).

The absence of an effect of GOAT versus COAT on perception of the LAP/LAB contrast indirectly supports Repp's (1983a) conclusion that there are no perceptual contrast effects between syllable-final and syllable-initial stop consonants (see also Ades, 1974; Samuel, Kat, & Tartter, 1984). What seemed like a contrast effect (for place of articulation) in Repp's study was ultimately attributed to perceptual information conveyed by closure duration. The present acoustic measurements showed that GOAT versus COAT had only a negligible influence on the duration of the preceding closure, so the absence of any perceptual effect on the LAP/LAB distinction is consistent with speech production. Incidentally, the absence of any response preference for the real word LABCOAT over the disyllabic pseudowords suggests that semantic biases played no role in the present experiment.

The absence of any effect due to the release burst was somewhat surprising in view of the fairly large effect obtained in Experiment 1. While that effect could have been due to either C_1 closure duration or properties of the release bursts themselves, neither variable was effective in the disyllabic context. One possible explanation is that the following syllable had a masking effect on the weak release burst, making it difficult to detect (cf. Henderson & Repp, 1982).

The only significant effect obtained in Experiment 2, that of total closure duration, is consistent with the acoustic measurements that showed total closure duration to be longer following a voiceless syllable-final stop.

D. Experiment 3

Experiment 3 investigated further the potential role of the release burst in disyllabic context by manipulating its position within the closure interval.

1. Methods. Stimuli from the LAB-derived continuum were always followed by COAT. The single release burst used was the one originally taken from LAP-COAT. The silent intervals surrounding this release burst were modified as follows: The "total closure" duration was made either short (120 ms) or long (200 ms), and the release burst was placed 40 or 80 ms (in the short interval) or 40, 80, 120 or 160 ms (in the long interval) after the beginning of the closure, defining corresponding C_1 closure durations.

2. Results and discussion. The effect of total closure duration was again obtained, $F(1,8) = 12.2$, $p < .01$; the results resembled those obtained in Experiment 2 (see Fig. 2). Effects of release burst position (i.e., C_1 closure duration), on the other hand, were small and apparent only in the longer closure interval: $F(3,24) = 3.7$, $p < .03$, in a separate analysis. The effect was not monotonic: "b" responses decreased slightly as C_1 closure duration increased from 40 to 80 to 120 ms--which is in the expected direction--but increased again for a C_1 closure of 160 ms. Perhaps, this very late-occurring release burst effectively suggested a shortening of the total closure interval. Alternatively, the burst may have been masked by the onset of the second syllable in that condition, which restored a "long C_1 closure" to a neutral value.

E. Experiment 4

In this test, isolated LAP/LAB stimuli were directly compared with disyllabic stimuli, either with or without release bursts. Thus, the issue of whether LAP/LAB syllables from a vowel duration continuum exhibit a shorter category boundary in disyllabic context than in isolation was re-examined. Recall that the Experiment 1 vs. Experiment 2 comparison yielded a negative result, but this was attributed to possible stimulus range effects. In addition, a possible influence of the rate of production of the second syllable on the perception of the LAP/LAB contrast was investigated (cf. Port & Dalby, 1982).

1. Methods. The LAB-derived continuum was used in conjunction with the LAB-derived closure interval (total duration 170 ms) and release burst, which was either present or absent. There were two versions of the second syllable: the GOAT used previously, which had been produced in a disyllabic context and was 473 ms in total duration, and another token of GOAT from the same speaker, which had been produced in isolation and measured 610 ms. In this test, the subjects identified only the syllable-final stop.

2. *Results and discussion.* Presence versus absence of a release burst made no significant difference, $F(1,8) = 1.5$. In hindsight, this is not surprising in view of the fact that the burst used happened to be the one that had little effect in Experiment 1 (see the two left-most functions in Fig. 1). The average data did suggest an effect in the expected direction for isolated LAP/LAB syllables, but the relevant interaction was not nearly significant, $F(1,8) = 1.2$.

Collapsing over this factor, Figure 3 compares the labeling functions for LAP/LAB syllables in isolation and when followed by either version of GOAT. In the disyllables, there were somewhat more "b" responses when the long GOAT followed than when the short GOAT followed. This effect was very small but reached significance, $F(1,8) = 7.8$, $p < .03$. A similar effect of the rate of production of the second syllable on the DIGGER-DICKER distinction was reported by Port and Dalby (1982), although in their stimuli closure duration, not vowel duration, was the primary voicing cue. In addition, a shift in the boundary for isolated syllables relative to disyllables can be seen in Figure 3, $F(1,8) = 9.5$, $p < .02$. This is the hypothesized effect of syllabic context, which failed to emerge in a comparison of Experiments 1 and 2. Thus, the suspicion that this earlier comparison was invalid because of differences in stimulus ensemble tends to be confirmed by the present data. Another replication of this context effect was sought in Experiment 6.

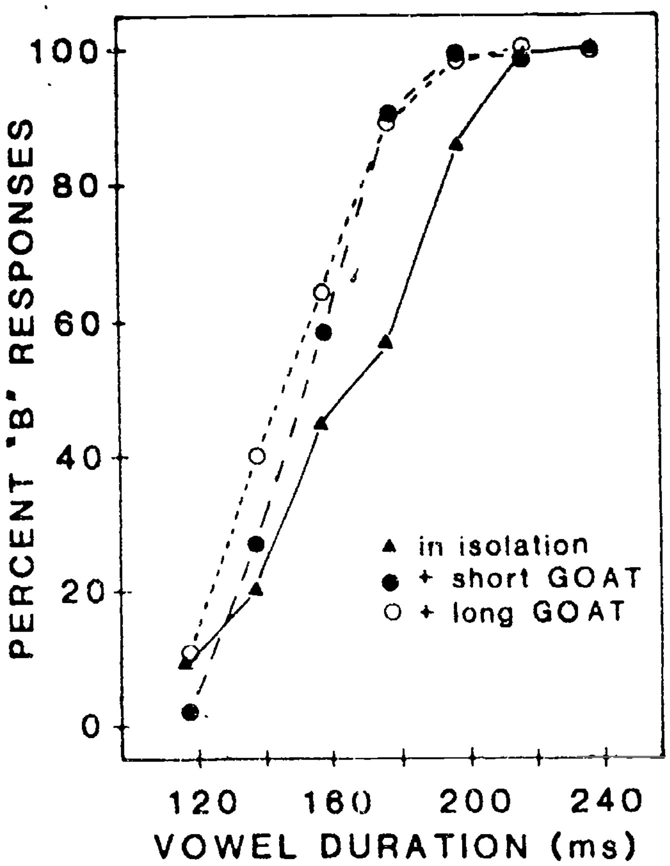


Figure 3. Effect of adding a second syllable, and of the rate of production of that syllable, on the voiced-voiceless distinction for syllable final stops (Experiment 4).

F. Experiment 5

One result has emerged clearly from Experiments 2 and 3: In disyllables, the number of "b" responses decreases as total closure duration increases. Presumably, this indicates that phonological voicing information is conveyed by closure duration, in agreement with the acoustic measurements. Experiment 4 also suggests that the LAP/LAB boundary for isolated syllables is indeed at a longer vowel duration than that for the same syllables in disyllabic context. This difference may be attributed to a perceptual compensation for the expected shortening of the first syllable in disyllabic context (cf. Nooteboom, 1973). Note, however, that the direction of the boundary shift when a second syllable is added (viz., the increase in "b" responses) is the same as results from a shortening of the closure duration in disyllables. Thus, it is conceivable that the effects of closure duration and of syllabic context are one and the same, reflecting temporal proximity of following context.

The precise time course of the change in the LAP/LAB boundary in disyllables as a function of a wide range of closure durations may provide relevant information. Certainly, as the closure duration is increased to very long values, the influence of the second syllable on the LAP/LAB boundary should cease, and the boundary should equal that for isolated monosyllables. Experiment 5 sought to determine the temporal separation (closure duration) at which this asymptote is reached, as well as the shape of the function relating the LAP/LAB boundary to closure duration. If there is only a single factor involved--temporal proximity of following context--this function should be monotonically increasing until the asymptote is reached. On the other hand, if there are two factors--closure duration acting as a voicing cue and presence/absence of syllabic context making an independent contribution--then closure durations typical of voiceless stops (i.e., around 230 ms, cf. Table 1) should lead to a relative <u>decrease</u> in "b" responses counteracting the effect of syllabic context, which increases "b" responses. Thus, depending on the relative strengths of the two opposing effects, the function may either be nonmonotonic, or have an early asymptote, or exhibit a change in slope around the point where the cue value of closure duration changes polarity (i.e., around 200 ms).

The subjects in Experiment 5 were also asked to judge on each trial whether they thought the two syllables formed a single compound (pseudo-) word or whether they sounded like two unrelated monosyllables. It was of interest to determine whether the "one word"-"two words" boundary (1/2 boundary, for short) would coincide with the intersyllabic temporal separation (i.e., closure duration) at which the LAP/LAB boundary function reached its asymptote. Such a finding might suggest a top-down influence on phonetic perception, or at least a common factor influencing both types of judgment.

1. <u>Methods</u>. The stimuli from the LAB-derived LAP/LAB continuum were followed by the standard GOAT at each of 8 temporal separations (closure intervals) ranging from 150 to 500 ms in 50-ms steps. The closure intervals were completely silent; release bursts were not included in this test.

In addition to identifying both stop consonants, subjects were asked to indicate for each two-syllable sequence whether it sounded like a single disyllabic word (LABGOAT or LAPGOAT) or like two unrelated monosyllabic words (LAB, GOAT or LAP, GOAT). They indicated the latter judgment by placing a comma between the two consonant responses ("b,g" or "p,g"). The results of one subject were discarded because she gave no "p" responses. (Reasons unknown.)

2. Results and discussion. The average category boundary on the LAP/LAB continuum was determined by linear interpolation between the two data points straddling the 50-percent cross-over, separately for each closure duration. The solid line in Figure 4 shows this vowel duration boundary as a function of closure duration. It is evident that the boundary shifted to longer vowel durations (i.e., "b" responses decreased) as closure duration increased from 150 to 250 ms. The overall effect of closure duration was highly significant, $F(7,49) = 8.6$, $p < .0001$. However, the boundary remained fixed at closure durations beyond 250 ms, $F(5,35) = 1.4$. This is just slightly beyond the typical closure duration following voiceless stops. No nonmonotonic trend is evident.

"Two words" responses increased monotonically as closure duration increased, as expected. The 1/2 boundary, expressed in terms of closure duration, was calculated separately for each vowel duration and is shown as the dotted line in Figure 4. This boundary decreased strongly as vowel duration increased, $F(6,42) = 30.4$, $p < .0001$, except at the shortest vowel durations. The slope of this function is not far from −1 (a dashed line with this slope is drawn in Figure 4); in other words, the sum of vowel and closure durations at the 1/2 boundary tended to be constant. Thus, it appears that the subjects based their judgments not on the silent intersyllable interval but on the interval between the onsets of the first and second syllables. Note also that the boundary continued to decrease at the longest vowel durations where the syllable-final consonant was almost uniformly labeled "b". Therefore, these judgments did not seem to be contingent on identification of the stop consonant, although the steepest slope of the 1/2 boundary function did occur in the region of the LAP/LAB boundary. In addition, it may be noted that the 1/2 boundary function intersects the LAP/LAB boundary function at 350 ms of closure silence, i.e., 100 ms beyond the point at which closure duration loses its effectiveness as a voicing cue. Thus, whatever the process responsible for the effect of closure duration on voicing judgments, it does not seem to be a direct consequence or a common determinant of perceiving the two syllables as part of a single word.

These results give no reason to consider the effect of adding a second syllable as different, in principle, from the effect of shortening the closure interval in a disyllable. Indeed, it appears that adding a second syllable has an effect only when the resulting closure interval is sufficiently short. Relative to LAP produced in isolation, shortening of the vowel in LAP was observed even when a closure interval averaging 232 ms intervened before the second syllable (cf. Table 1). Yet, the perceptual effect of adding a syllable after that long a closure duration was almost nil (cf. Fig. 4).[*] This may mean that closure duration should be viewed as an asymmetric cue: Short closures are a cue for the category "voiced", but long closures are devoid of any perceptual cue value. That is, a stop never sounds "more voiceless" in disyllabic context than in isolation. Alternatively, a tendency to hear LAP at closure durations characteristic of voiceless stops (200-250 ms) may have been cancelled by an opposing tendency to hear LAB because of temporal recalibration (i.e., a subjective stretching of the vowel) in the presence of a second syllable. Although this two-process model is not implausible, a single-process explanation must be preferred on grounds of parsimony. A summary of this argument is presented schematically in Figure 5. (However, see the General Discussion.)

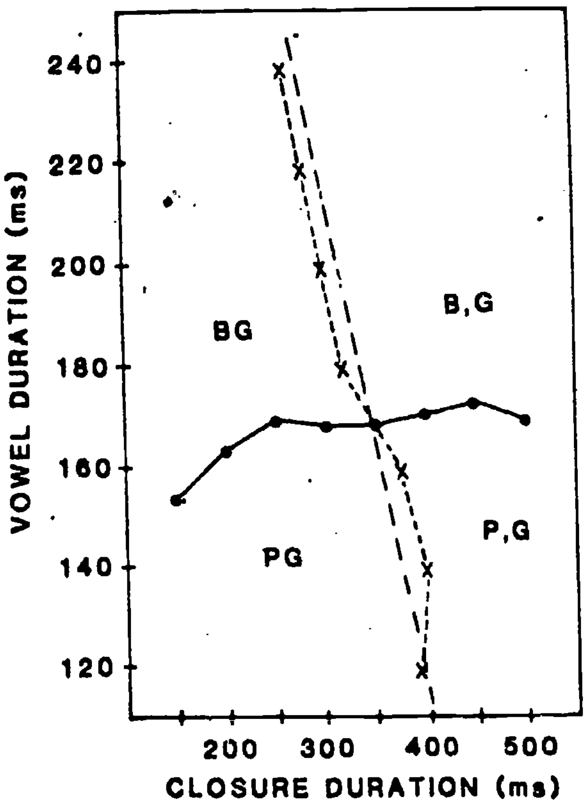


Figure 4. The voiced-voiceless boundary (solid line) and the 1/2 boundary (dotted line) as a function of closure duration and vowel duration in disyllables (Experiment 5). The dashed line represents a slope of -1, i.e., a constant sum of vowel and closure durations.

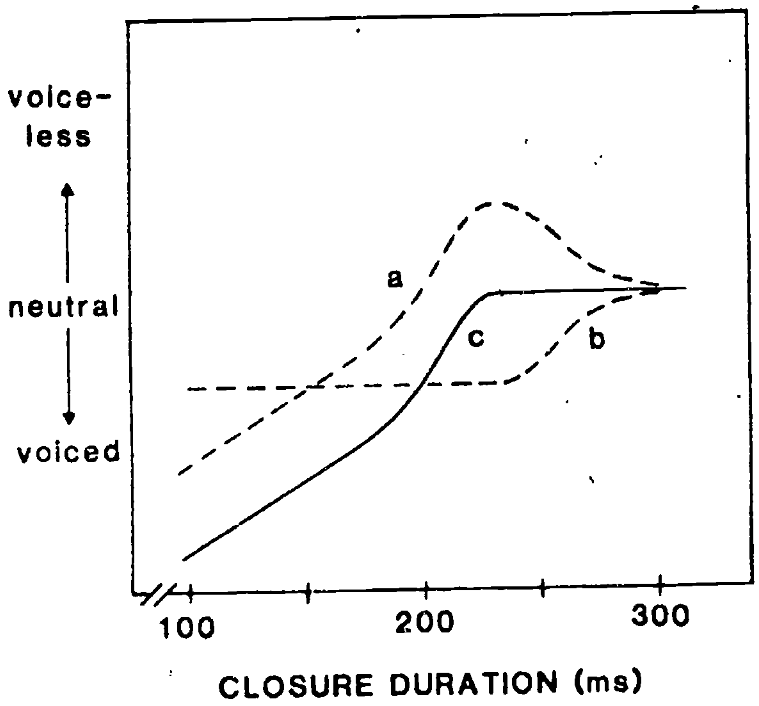


Figure 5. Schematic illustration of a two-process model of the perceptual effect of closure duration. Closure durations of less than 200 ms are assumed to cue voicedness, whereas closure durations of about 200-300 ms are assumed to cue voicelessness (function a). At the same time, a constant temporal compensation due to the presence of a second syllable is assumed to occur, as long as the closure duration does not exceed about 250 ms (function b). The resultant (function c) shows only an increase in voiced percepts because functions a and b cancel beyond 230 ms or so.

G. Summary of Experiments 1-5

In summary, these studies show:

(1) that addition of a second syllable shifts the LAP/LAB boundary toward shorter vowel durations, as long as the temporal separation (closure) is less than about 250 ms;
(2) that total closure duration (in the range below 250 ms) is a cue to the LAP/LAB distinction, with shorter closures leading to more "b" responses;
(3) that the effect of syllabic context is not a direct consequence of hearing the two syllables as part of a single word, and that it may indeed be identical with (2);
(4) that judgments of "two words" increase almost linearly with the duration of the first syllable and thus seem to rest on the perceived separation of syllable onsets;
(5) and that C_1 release burst and properties of the second syllable (VOT, overall duration) play at best a minor role in the perception of the LAP/LAB distinction in disyllabic context.

Experiments 6 and 7 attempted to replicate findings (1)-(4) with a new set of stimuli and a new group of subjects.

H. General Methods: Experiments 6-7

1. Subjects. Sixteen undergraduate students enrolled in an introductory psychology course at the University of Connecticut participated in the experiment for course credit. All subjects were native speakers of American English with no history of hearing impairment.

2. Stimuli. Three representative disyllabic utterances of the male talker (DW), digitized at 10 kHz, served as bases for the LAB, GOAT, and COAT stimuli used in the present experiments. The LAB stimulus was excerpted from a good LABCOAT. Tokens of LABGOAT and a second LABCOAT yielded the GOAT and COAT stimuli.

A LAP/LAB continuum was constructed by successively deleting every other pitch pulse from the vocalic region of the LAB stimulus. The first deleted pitch pulse began 55 ms into the syllable (following the formant transitions for /l/) and the last, one pitch pulse prior to closure for /b/. In all, eight pitch pulses were excised yielding a series of nine stimuli. The longest stimulus was the original LAB (220 ms); the shortest ("LAP") was 135 ms in duration. The vowel durations thus ranged from 80 to 165 ms across the LAP/LAB continuum. Note that these durations are considerably shorter than those employed in Experiments 1-5, reflecting differences in the speaking rates of talkers DW and CG.

When GOAT or COAT was appended to the LAP/LAB stimuli to form disyllables, the closure was completely silent; no release bursts of the syllable-final stop consonant were included in Experiments 6 and 7. The duration of the COAT stimulus (487 ms) was somewhat greater than that of the GOAT stimulus (439 ms), although this difference was located mainly in the final release burst.

3. Procedure. Two stimulus tapes (one for each experiment) were prepared with intertrial intervals of 2.5 s and presented to subjects binaurally over TDH-39 headphones. Both tapes were heard during the 75-minute session, with Experiment 7 always following Experiment 6.

J. Experiment 6

In this study, voicing judgments for disyllables were compared with those for LAP/LAB monosyllables when presented in a separate test and when included in the disyllable test. We also wished to replicate the effect of closure duration on the voicing boundary, using (silent) closure durations that were appropriate for the present, shorter stimuli. In addition, the possible influence of the natural GOAT and COAT on voicing judgments was re-assessed.

1. Method. Two different stimulus sequences were presented. In the first, only the nine stimuli from the monosyllabic LAP/LAB series were included. Following five repetitions of the endpoint stimuli, subjects listened to ten randomized blocks of the nine stimuli. In the second sequence, each block included two occurrences of the monosyllabic stimuli interspersed among single occurrences of stimuli from four disyllabic continua. The disyllabic stimuli were constructed by appending the GOAT or COAT stimulus to each member of the LAP/LAB series following either a 120 ms or a 170 ms silent interval. These intervals corresponded to speaker DW's average closure durations in his utterances of LAB- and LAP- disyllables, respectively. There were five blocks of 54 stimuli. Appropriate responses to the monosyllabic stimuli were "b" and "p". For the disyllables, subjects responded with "bg", "pg", "bc", or "pc" depending on whether LABGOAT, LAPGOAT, LABCOAT, or LAPCOAT was heard.

2. Results and discussion. In the disyllables, there was a small but significant effect of the identity of the following syllable, GOAT vs. COAT, $F(1,15) = 14.9$, $p < .002$. Voiced responses were more frequent preceding COAT, which is consistent with three alternative explanations: (1) If closure durations were longer preceding COAT in production, then listeners' tacit knowledge of that regularity might lead them to shift their perceptual criterion in favor of voiced responses in that context. (2) The effect may be due to second syllable duration, COAT being longer than GOAT in the present experiment. (3) The effect may represent response contrast between the voicing categories of C_1 and C_2. Considering that (1) is not supported by our acoustic measurements (see Table 2) and that (3) was not obtained in Experiment 2, the most likely explanation seems to be (2), in accordance with the effect of second-syllable duration obtained in Experiment 4.

Figure 6 presents the results for monosyllables (dashed lines) and for disyllables collapsed over the GOAT/COAT factor (solid lines). As expected, increasing the closure duration in disyllables had the effect of shifting the voicing boundary toward longer vowel durations. Significantly fewer "b" responses were made to disyllables with long closures than to those with short closures, $F(1,15) = 21.7$, $p < .0003$. It is also evident that subjects gave significantly fewer "b" responses to the monosyllables than to the disyllables, $F(1,15) = 73.0$, $p < .0001$. Thus, these results replicate for the present set of stimuli the finding (Experiment 4) that the LAP/LAB boundary is shifted toward shorter vowel durations in disyllabic context.

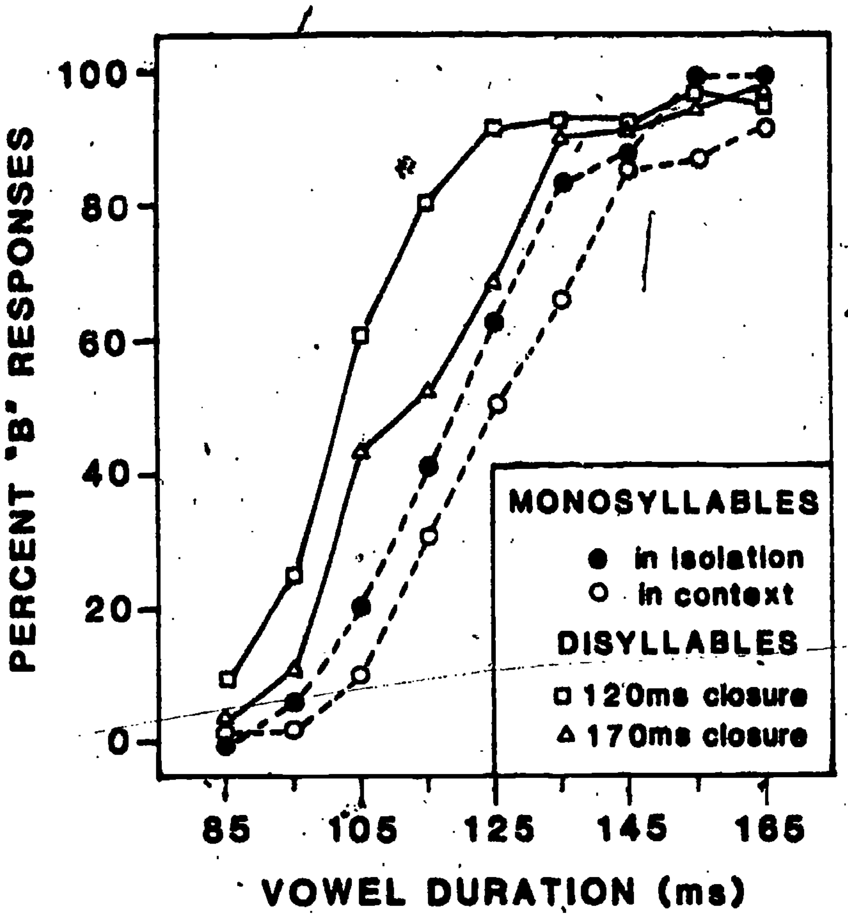


Figure 6. Percent voiced responses as a function of vowel duration for isolated monosyllables, monosyllables interspersed among disyllables, and disyllables with two closure durations (Experiment 6).

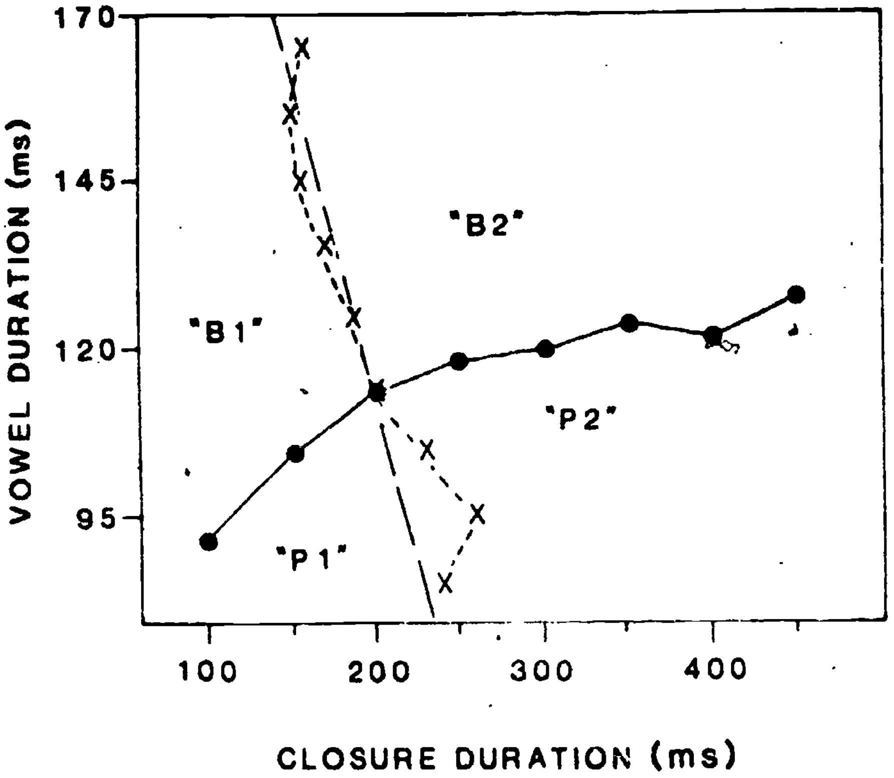


Figure 7. The voiced-voiceless boundary (solid line) and the 1/2 boundary (dotted line) in disyllables as a joint function of closure duration and vowel duration (Experiment 7). The dashed line represents a slope of -1, i.e., a constant sum of vowel and closure durations.

Figure 6 further shows that subjects gave fewer "b" responses to the monosyllables that were interspersed among the disyllables than to those that were presented alone, $F(1,15) = 13.2$, $p = .003$. This represents a stimulus range effect: Because the syllable-final stops sounded relatively more voiced in disyllables, they presumably sounded relatively more voiceless in the interspersed monosyllables. The finding of such an effect lends further credence to the earlier argument proferred in connection with Experiments 1 and 2, where the apparent absence of an effect of syllabic context was attributed to the presence of stimulus range effects.

K. Experiment 7

This experiment replicated Experiment 5 using the new set of stimuli.

1. Method. The GOAT stimulus was appended to the stimuli from the LAP/LAB series at each of eight temporal separations (closure durations) ranging from 100 to 450 ms in 50 ms steps. These stimuli were recorded in five randomized blocks. Subjects were asked to decide on each trial whether they heard /b/ or /p/, and whether they heard one two-syllable word or two one-syllable words, using the responses B1, P1, B2 or P2.

2. Results and discussion. As in Experiment 5, the average category boundary on the LAP/LAB continuum was determined for each closure duration by means of linear interpolation. These boundaries are plotted as a function of closure duration in Figure 7 (solid line). As expected, the overall effect of closure duration was highly significant, $F(7,105) = 50.4$, $p < .0001$: Increasing closure duration shifted the LAP/LAB boundary toward longer vowel durations. The effect leveled off around closure durations of 200-250 ms. But, in contrast to Experiment 5, there was a small increase in the boundary even beyond 300 ms, $F(3,45) = 3.7$, $p < .02$. Again, the boundary function is monotonic, with an asymptote close to the boundary for monosyllables in Experiment 6.

The 1/2 boundary was calculated separately for each vowel duration and is plotted as a function of vowel duration in Figure 7 (dotted line). As in Experiment 5, it is evident that the boundary decreases sharply as vowel duration increases, $F(8,120) = 29.7$, $p < .0001$ (except perhaps at the shortest and longest vowel durations). Again, the slope of the function is close to −1, indicating that the sum of vowel and closure duration at the 1/2 boundary tends to be constant. Thus, the data confirm that these judgments were based on the interval between the onsets of the first and second syllables.

Because of the shorter syllable durations, the subjects in Experiment 7 were probably more inclined to consider the two syllables as separate words than were the subjects in Experiment 5.[5] Thus, the "one word"/"two words" boundary function intersects the voicing boundary function at 200 ms of closure duration (versus 350 ms in Experiment 5). While this would be consistent with the hypothesis that syllabic context (or closure duration) has its effect contingent on perception of the two syllables as part of the same utterance, this hypothesis was rejected on the basis of the results of Experiment 5. That is, the coincidence of the 1/2 boundary with the leveling off of the syllabic context (or closure duration) effect may be just that, a coincidence. At any rate, the data of Experiment 7 are again consistent with a single-process explanation of the effect of closure duration: Short closures increase voiced responses.

244

L. Summary of Experiments 6-7

The results of these replication experiments affirm all the major conclusions of Experiments 1-5, as summarized above. There are two minor discrepancies: the absence of a stable asymptote of the LAP/LAB boundary function, and of a clear dissociation of voicing and 1/2 judgments. The earlier data were clearer in these regards, but the replication must nevertheless be considered useful in view of the stability of the major findings across different stimulus materials and subject groups.

III. General Discussion

The principal aim of the present series of studies was to demonstrate two effects in the perception of the voicing category of syllable-final stop consonants, one being due to the addition of a second syllable beginning with a different stop, and the other reflecting the perceptual contribution of the closure duration cue in disyllables. An effect of closure duration was consistently obtained: The shorter the closure, the more likely subjects were to report a voiced stop consonant. This is in agreement with our measurements of closure durations in natural speech and suggests that listeners have incorporated tacit knowledge about these temporal regularities into their perceptual criteria for the voiced-voiceless distinction. By the same token, one should expect that this knowledge includes the fact, well-known from earlier speech production studies and substantiated by our acoustic measurements, that a syllable contracts when a second syllable is added to it (or, equivalently, that a syllable is lengthened in utterance-final position). Nooteboom (1973) has demonstrated such perceptual compensation in a task requiring judgments of phonological vowel length. The present data are consistent with such a temporal compensation mechanism, which operates in addition to an independent perceptual effect of closure duration as a voicing cue (see Fig. 5).

Although the results are compatible with such a two-process model, they do not provide compelling evidence in its favor. A more parsimonious interpretation of the results, at least when considered in isolation from other findings in speech perception research, is that there is only a single effect, that of closure duration. That effect, moreover, is unidirectional: A final stop consonant sounds increasingly "voiced" as closure duration decreases, but even at closure durations that are optimal for voiceless stops in disyllables subjects do not give more voiceless responses than they give to monosyllables that lack any closure duration information. In other words, the closure duration cue apparently contributes only to the perception of stops as voiced, not as voiceless. This is not unreasonable: Even though "voicing" may be used as an abstract cover term for a variety of acoustic manifestations of a phonological distinction, it may also be understood more narrowly as designating the common acoustic feature of "presence of low-frequency energy" within a certain time span (Stevens, Keyser, & Kawasaki, 1985). Absence of low-frequency energy is the neutral state; that is, voicing is an acoustically "marked" feature. The problem in applying this view to the present data lies in the finding that it made little difference whether the second syllable began with a voiced or a voiceless (aspirated) stop. If the decisive factor was presence of low-frequency energy within a certain interval following the offset of the first syllable, aspiration following the closure should not have increased voiced responses as much as did a voiced (actually, weakly aspirated) signal.

A specific auditory mechanism that has been discussed in connection with speech is backward recognition masking (e.g., Massaro, 1975). Although the ineffectiveness of C_1 release bursts in disyllables may be due to auditory backward masking, the perceptual effect of decreasing closure duration is not compatible with such an explanation: Masking of either the vowel offset cues (which favored voiced percepts, since the stimuli were derived from an original utterance of LAB) or of the perceived duration of the first syllable (leading to a reduction in subjective vowel duration--see Massaro & Idson, 1978) should have decreased, not increased voiced responses. It seems, therefore, that proximity of following context had its effect without interfering with or altering those auditory properties of the first syllable that feed into phonetic decisions.

These hypotheses surely do not exhaust the possible mechanisms that may underlie a unidimensional, unidirectional effect of closure duration. The failure of two specific accounts, however, raises doubt about whether the isolated parsimony of a single-process model is indeed preferable to a two-process model, particularly one that ties in with a multitude of related observations suggesting that listeners make phonetic decisions in accord with criteria that reflect the phonetic regularities of the language (Nooteboom, 1973; Nooteboom & Doodeman, 1980; Repp, 1982, 1983c).

A comment is in order concerning the relationship between vowel duration and closure duration at the voiced/voiceless boundary. For single intervocalic post-stressed stops, as in DIGGER/DICKER (Port & Dalby, 1982), the ratio of closure to vowel durations (C/V) at the voicing boundary remains approximately constant as one of the two temporal variables is manipulated. Port and Dalby varied vowel duration while determining the boundary on a closure duration continuum; we manipulated closure duration while determining the boundary on a vowel duration continuum. In the context of two-stop sequences, closure duration is a much less salient voicing cue than it is for single intervocalic stops. Because of the longer closure durations, larger absolute C/V ratios were to be expected; the question was whether they would remain constant in the region where closure duration influenced the boundary on the vowel duration continuum. As can be seen in Figure 8, the answer is negative. The average C/V ratios from both Experiments 5 and 7 increase as a nearly linear function of closure duration over the whole range. Thus, a constant-ratio rule does not hold for these stimuli; such a rule may be restricted to the specific utterance types considered by Port and Dalby (1982).

Another type of constancy was found in the present data, however: The sum of vowel duration and closure duration at the 1/2 boundary was approximately constant. That is, subjects based their "one word"/"two words" judgments on the onset-to onset interval between the two syllables and not on the separation (the silent closure) between them. Perception of the syllable-final stop as voiced or voiceless seemed to be independent of these timing judgments. This is in agreement with the recent findings of Miller, Aibel, and Green (1984), who showed that perception of a particular temporally-cued phonetic contrast was independent of explicit judgments of perceived speaking rate. It appears that the speech signal supports a variety of independent judgments that do not interact, though they may combine at higher levels of organization (cf. Ganong, 1980; Massaro & Cohen, 1983b).

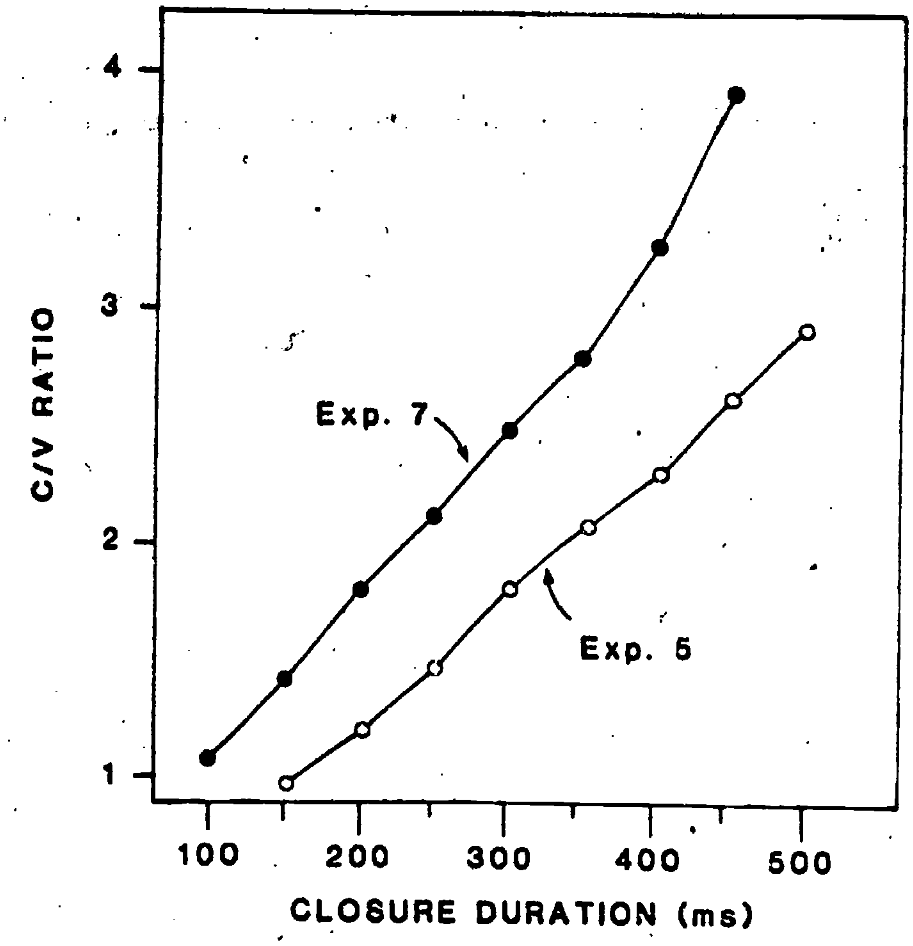


Figure 8. The average C/V ratio as a function of closure duration (Experiments 5 and 7). The ratios were computed from the average data shown in Figures 4 and 7, respectively.

On the whole, then, the present results are consistent with the general notion that human listeners behave as if they knew all the detailed acoustic consequences of articulation, including context-conditioned and position-specific variation. The perceptual effects of various acoustic cues can almost always be rationalized by reference to the systematic patterns that emerge in the acoustic analysis of speech, although, considering the many factors that play a role in perception, a precise prediction of experimental results from acoustic regularities is rarely possible. The future development of a more economic description of speech in terms of dynamic articulatory processes may ease the burden on the perception theorist.

References

Ades, A. E. (1974). How phonetic is selective adaptation? Experiments on syllable position and vowel environment. *Perception & Psychophysics*, *16*, 61-66.

Barry, W. J. (1979). Complex encoding in word-final voiced and voiceless stops. *Phonetica*, *36*, 361-372.

Ganong, W. F., III. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 110-125.

Gruenenfelder, T. M., & Pisoni, D. B. (1980). Fundamental frequency as a cue to post-vocalic consonantal voicing: Some data from speech perception and production. Perception & Psychophysics, 28, 514-520.

Henderson, J. B., & Repp, B. H. (1982). Is a stop consonant released when followed by another stop consonant? Phonetica, 39, 71-82.

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., & Flege, J. E. (1984). Perception of the voiced-voiceless contrast in syllable-final stops. Journal of the Acoustical Society of America, 76, 18-26.

Hogan, J. T., & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. Journal of the Acoustical Society of America, 67, 1764-1771.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. Journal of the Acoustical Society of America, 25, 105-133.

Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. Journal of the Acoustical Society of America, 54, 1102-1104.

Lehiste, I. (1972). Timing of utterances and linguistic boundaries. Journal of the Acoustical Society of America, 51, 2018-2024.

Lehiste, I. (1976). Influence of fundamental frequency pattern on the perception of duration. Journal of Phonetics, 4, 113-117.

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

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. (1981). On generalizing the Rabid-Rapid distinction based on silent gap duration. Haskins Laboratories Status Report on Speech Research, SR-65, 251-259.

Malécot, A. (1958). The role of releases in the identification of released final stops. Language, 34, 370-380.

Massaro, D. W. (1975). Perceptual images, processing times, and perceptual units in speech perception. In D. W. Massaro (Ed.), Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics (pp. 125-150). New York: Academic Press.

Massaro, D. W., & Cohen, M. M. (1983a). Consonant/vowel ratio: An improbable cue in speech. Perception & Psychophysics, 33, 501-505.

Massaro, D. W., & Cohen, M. M. (1983b). Phonological context in speech perception. Perception & Psychophysics, 34, 338-348.

Massaro, D. W., & Idson, W. L. (1978). Temporal course of perceived vowel duration. Journal of Speech and Hearing Research, 21, 37-55.

Miller, J. L. (1981). The effect of speaking rate on segmental distinctions: Acoustic variation and perceptual compensation. In P. D. Eimas & J. L. Miller (Eds.), Perspectives in the study of speech. Hillsdale, NJ: Erlbaum.

Miller, J. L., Aibel, I. L., & Green, K. (1984). On the nature of rate-dependent processing during phonetic perception. Perception & Psychophysics, 35, 5-15.

Nooteboom, S. G. (1973). The perceptual reality of some prosodic durations. Journal of Phonetics, 1, 25-45.

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.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. Journal of the Acoustical Society of America, 32, 693-703.

Port, R. F. (1981). Linguistic timing factors in combination. Journal of the Acoustical Society of America, 69, 262-274.

Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics*, *32*, 141-152.

Price, P. J., & Lisker, L. (1979). (/b/->/p/) but (/p/->/b/). In J. Wolf & D. Klatt (Eds.), *Speech communications papers: 97th Meeting of the Acoustical Society of America* (pp. 49-52). New York: Acoustical Society of America.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of voicing in word-final consonants in American English. *Journal of the Acoustical Society of America*, *51*, 1296-1303.

Raphael, L. J. (1981). Durations and contexts as cues to word-final cognate opposition in English. *Phonetica*, *69*, 1462-1464.

Raphael, L. J., Dorman, M. F., & Liberman, A. M. (1980). On defining the vowel duration that cues voicing in final position. *Language and Speech*, *23*, 297-308.

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

Repp, B. H. (1983a). Bidirectional contrast effects in the perception of VC-CV sequences. *Perception & Psychophysics*, *33*, 147-155.

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

Repp, B. H. (1983c). Trading relations among acoustic cues in speech perception: Speech-specific but not special. *Haskins Laboratories Status Report on Speech Research*, *SR-76*, 129-132.

Samuel, A. G., Kat, D., & Tartter, V. C. (1984). Which syllable does an intervocalic stop belong to? A selective adaptation study. *Journal of the Acoustical Society of America*, *76*, 1652-1663.

Stevens, K. N., Keyser, S. J., & Kawasaki, H. (1985). Toward a phonetic and phonological theory of redundant features. In J. S. Perkell and D. H. Klatt (Eds.), *Invariance and variability of speech processes*. Hillsdale, NJ: Erlbaum.

Wang, W. S.-Y. (1959). Transition and release as perceptual cues for final plosives. *Journal of Speech and Hearing Research*, *2*, 66-73.

Wardrip-Fruin, C. (1982). On the status of phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants. *Journal of the Acoustical Society of America*, *71*, 187-195.

Westbury, J. R. (1977, June). *Temporal control of medial stop consonant clusters in English*. Paper presented at the 93rd meeting of the Acoustical Society of America, State College, Pennsylvania.

Wolf, C. G. (1978). Voicing cues in English final stops. *Journal of Phonetics*, *6*, 299-309.

Footnotes

[1] Authors such as Klatt (1973) and Port (1981) took the midpoint of the formant transitions in spectrograms to be the acoustic liquid-vowel boundary. The oscillographic criterion used here allots most of the transition portion to the vowel, which seems justified on the basis of perceptual data (Raphael et al., 1980).

[2] The syllable durations were similar across the two continua, as intended, except for a small difference caused by the discreteness of the pitch pulses. The difference in vowel durations is mainly a consequence of the difference in [l] duration. The decision to use vowel duration, rather than

syllable duration, in the presentation of results was made at a rather late stage.

[3] In addition, a GOAT-COAT continuum was constructed by substituting aspiration noise from COAT for successive pitch periods in GOAT (see the appendix in Ganong, 1980, for a description of this procedure). This continuum was used in another test administered to the present subjects at the end of the first session. That condition investigated the influence of preceding LAB or LAP, closure duration, and C1 release bursts on perception of the GOAT/COAT distinction. There were no systematic effects of any of these variables; thus, VOT appeared to be the only salient cue to the GOAT/COAT distinction in these stimuli.

[4] The assumption here is that the constant asymptotic vowel duration boundary of about 170 ms matches that for isolated LAP/LAB syllables. This is supported by a comparison with the results of Experiment 4 (cf. Fig. 3).

[5] Note that the 1/2 boundary did not just bisect the range of closure durations but was clearly to the left of the center of the stimulus range. This indicates that 1/2 judgments were not arbitrary and rested on some pre-established internal criteria.

DIFFERENTIAL PROCESSING OF PHONOGRAPHIC AND LOGOGRAPHIC SINGLE-DIGIT NUMBERS BY THE TWO HEMISPHERES[*]

Daniel Holender[†] and Ronald Peereman[†]

In planning to include the topic of this chapter in a book devoted to acquired mathematical disabilities, the editors must have assumed not only that the theoretical problem is interesting but also that the literature contains enough relevant data to discuss the issues. A full assessment of the differential roles played by each hemisphere in dealing with different surface forms of numbers would require the availability of results from a variety of tasks comparing different kinds of number representation. Ideally, each task should also have been investigated in the conditions resulting from the combination between different number surface forms with left and right hemifield presentations, or with left and right brain injuries. However, such is certainly not the case, and many conclusions will have to be drawn from experiments in which the requisite conditions are only partially met. The existing data further impose two restrictions on the scope of the review: Only single-digit numbers (hereafter referred to as "numbers," unless otherwise specified) will be considered and only nonmathematical tasks will be dealt with.

Before speculating on cognitive processes and mental representations of numbers, we should have a good description and classification of what is represented in the stimulus. In spite of the restriction of this chapter to the differential processing of single-digit numbers according to their surface form, some space will also be devoted to specifying the notational principles that underlie multidigit number writing. The fact that the resulting classification of symbols that will emerge is different for single-digit and multidigit numbers may highlight what we expect to find, and what has already

---

[*] To appear in G. Deloche & X. Seron (Eds.), Mathematical disabilities: A cognitive neuropsychological perspective. Hillsdale, NJ: Lawrence Erlbaum Associates, in press.

[†] Université Libre de Bruxelles, Brussels, Belgium

Acknowledgment. This chapter was written while the first author was a guest at Haskins Laboratories, New Haven, CT. We are grateful to Michael Studdert-Kennedy for his thorough comments and his suggestions for style improvement of an earlier version. Thanks are also due to Noriko Kobayashi for providing information about number writing in Japan. This work has been partially supported by NICHD grant HD-01994 to Haskins Laboratories, by the Belgian "Fonds de la Recherche Fondamentale Collective" (FRFC) under convention 2.4505.80 (with P. Bertelson), and by the Research Council of the University (Project: "Processus cognitifs dans la lecture"). The sojourn of the first author in the United States was supported by the University and partially subsidized by a NATO grant for scientific research.

been found, when these symbols are considered from the vantage point of the cognitive neuropsychology of number processing. The first of the five sections comprising the chapter is therefore devoted to an analysis of different types of number representations. This is followed by three sections reviewing and discussing the data from (a) numerical size comparison tasks, (b) lateral hemifield presentations, and (c) the performances of brain damaged patients. The fifth section summarizes the main conclusions.

Number Representations

The Arabic numeral 5, the Roman numeral V, the English written word five, and the corresponding Chinese character are different arbitrary symbols that denote the same abstract concept: the number five. In addition to having different surface forms, these various symbols also belong to different notational systems when they are used as components of multidigit numbers. Two distinctions have to be made. One concerns the difference between numerals and number names and the other, the difference between logographic and phonographic number representations. The first distinction is better captured by characterizing multidigit number notational systems and the second is better illustrated by specifying the surface form of single digit symbols.

Notational Systems

The first important distinction to bear in mind concerns the difference between numerals and number names. Numerals are special symbols for representing numbers visually. In many written languages they coexist with number names, which are translations of the spoken form, according to the writing system of the language. The only numerals extensively used now are Arabic numerals. The universality of Arabic numerals contrasts with the language specificity of number names, but the main reason for distinguishing between numerals and number names lies elsewhere: Only number names allow for a term-by-term translation of the spoken miltidigit numbers. In other words, we may write and say "two hundred and thirty three," but we do not usually say "two-three-three" or "hundred-hundred-ten-ten-ten-one-one-one" when we are confronted with 233 and CCXXXIII. Hence, the rules governing the way in which numbers are transcribed differ according to notational systems.

These rules are better illustrated by Chinese number-writing instead of by English or some other alphabetically written language, and by hieroglyphic Egyptian instead of Roman numerals. This allows us to capture the essence of the underlying notational principle without having to deal with irrelevant and confusing features, such as the use of special words to denote the multiple of 10 in English number-naming or the incorporation of a subtractive principle in the Roman numeral system (e.g., IV instead of IIII). For the few following examples, let us represent the ranks of the units, tens, and hundreds by U,T, and H, respectively.

In the form of hieroglyphic Egyptian used for lapidary inscriptions, one symbol denoted the unit and a different symbol denoted each of the successive powers of 10 up to 10,000. The number 543 was written in the form, HHHHHTTTTUUU. The only important aspect of this representation is that it is based on an additive principle. The conventional grouping of the units of the same rank and the usual order of writing was irrelevant to the understanding of the number. On a stone monument of ancient Egypt, the number would have been written right-to-left (instead of left-to-right as here) and the symbols would probably have been displayed on more than one line, but our sequence would have been unequivocally understood, even if it had been written TTHHUHTTHUHU. The same commutative principle applies to Roman numerals, except that elements entering into a subtractive relation must be kept together in their conventional order.

The Chinese number-naming system is also based on an underlying additive principle, but a supplementary multiplicative principle allows for suppression of the cumbersome repetitions of the symbols belonging to the same rank. This entails a different symbol for each unit (u1, u2...u9). The Chinese 543 is therefore written in the form, u5Hu4Tu3, using five different symbols instead of the three needed in hieroglyphic Egyptian. Here too, provided the symbols entering into a multiplicative relation are kept together, permuting the terms would not transform one number into another. In this case, however, the psychological impact of doing so would be stronger, because although the order of the elements plays no intrinsic role in the representation, their usual order corresponds to their order of utterance in the spoken number. Similarly, it may be unusual to write "twenty eight and four hundred" but it clearly means 428. The Chinese number writing principle has been called a "named place-value" notation by Menninger (1969), as opposed to the "abstract place-value" notation realized with Arabic numerals. English number-name writing is also a named place-value and it should be clear from what we have said that it is a pseudopositional system.

The only true positional number-writing system still in use was developed some time in the first half of the sixth century A.D. in India, whence it spread more or less rapidly to the whole world. The system uses only 10 symbols, named Arabic numerals after their first principal propagators rather than after their creators. In this system the rank of the units is abstractly symbolized by the position occupied by these units in the written number. Permutations of terms are no longer allowed without changing the value of the number and the whole system works only because of the great intellectual accomplishment of symbolizing nothing by something; namely, by using zero to fill in the positions of the unemployed ranks (compare the English three thousand and twenty, in which nothing stands for the unused ranks of the hundreds and the units, with 3020). Aside from the Greeks' ephemeral use of a complete abstract place-value system including a zero, the only known independent invention of such a notation took place in Mesoamerica. The extent to which the Mayas really grasped the concept of zero is likely to remain controversial forever, but they undoubtedly used a symbol functionally equivalent to zero in their place-value notation of numbers (e.g., Kelley, 1976).

In Europe, widespread use of the Arabic numeral place-value notation began toward the end of the fifteenth century, rapidly supplanting the Roman numerals from then on. The most important consequence of this event is that calculation, mainly realized by means of counting boards and quite independent of number writing before this date, now became intimately bound to the Arabic numeral notational system.

Much of what precedes can be found in the extensive and insightful coverage of the topic by Menninger (1969; see Flegg, 1983, for a condensed account). From an information-processing point of view, this rapid survey of the number notational systems still in use reveals three important points.

1. The Arabic numerals stand alone in being the only symbols that enter into an abstract place-value notation, an inherently positional system, and in being used for purpose of calculation, a highly specialized cognitive activity of symbol manipulation.

2. Number names, whether written in an alphabetic script, such as English, or in a logographic script (see below) such as Chinese, constitute a different notational system whose purpose is mainly to provide a visual term-by-term translation of spoken numbers.

3. Unlike Arabic numerals and number names, Roman numerals are more concrete representations of numbers, combining some properties of tally counts with simple additive and subtractive rules. They are quite easily decoded, but they are no longer widely used, and they have never been considered an efficient medium for calculation.

Symbol Surface Forms

Number names are represented according to the writing systems in use for general writing purposes. We shall therefore distinguish between logographic and phonographic systems. In a logographic system the written symbols represent linguistic units of meaning; namely, morphemes. In phonographic systems the linguistic units represented by each symbol are phonological, being either syllables in syllabic systems or phonemes in alphabetic systems (see Gelb, 1963, for a history and description of the writing systems. For discussions of the psycholinguistic aspects of the written symbols and their consequences for the analysis of mental processes involved in reading, see Gleitman & Rozin, 1977; Henderson, 1984; Liberman, Liberman, Mattingly, & Shankweiler, 1980; Mattingly, 1984; Rozin & Gleitman, 1977).

In Chinese writing, the most complete logographic system ever designed and still in use today, each symbol represents one morpheme. Each morpheme is also a word, although many words are composed of more than one morpheme, and are therefore written with more than one character. Chinese characters are often called ideograms, but this terminology is misleading because few characters are actually designed on a truly ideographic principle. We shall call the characters logograms to fit the linguistic description of the unit they represent.

As already mentioned, nine symbols represent the numbers one to nine in Chinese. The first three consist of one, two, and three horizontal strokes and the others are arbitrary symbols. Thus the first three symbols are built on an ideographic, or even a pictographic, principle, representing the beginning of a stick count. Knowing that they stand for numbers, someone who cannot read Chinese at all would be able to interpret them correctly; but this is not the case with the symbols for the numbers four to nine. It is nonetheless clear that the two horizontal strokes stand for the monomorphemic word meaning *two* in Chinese and that the arbitrary symbol representing the number "six" stands for the monomorphemic word meaning *six* in Chinese. Hence, the exact nature of any of these symbols is certainly better captured by the term logogram than by any other term.

In Japan, many Chinese characters, called Kanji characters, have been borrowed to be used conjointly with a syllabary. The simple syllabic structure of Japanese allows any word of the language to be written by using only the symbols of the syllabary. These symbols are called Kana and they exist in two forms: Hiragana and Katakana. In a normal text the content morphemes (mainly nouns, verbs, and adjectives) are usually written in Kanji and the grammatical morphemes are written in Hiragana; foreign loan words are written exclusively in Katakana.

Japanese number names are represented in Kanji, the characters being exactly those used in China. Like any other words they can also be written in Hiragana and Katakana, but this seldom, if ever, occurs in daily life. This point should be kept in mind in interpreting the results of experiments that have exploited this possibility.

The most important point of this entire discussion is that although the 10 Arabic numerals can be considered as logographic representations of the numbers zero to nine, the nine Chinese (or Japanese) logograms (zero not being represented) are not numerals, but number names. This fact has not always been correctly evaluated, either in the recent psychological literature, or by Menninger (1969) who was struck by the fact that the Chinese number symbols realize a perfect synthesis, being both numerals and number names. That this position is incorrect can be appreciated from the fact that throughout history, Chinese number names have coexisted with genuine autochthonous numerals (incorporating the Indian zero, but not the other symbols, in the thirteenth century). These have now been replaced by Arabic numerals. Hence, the relation between Chinese characters (or Japanese Kanji), denoting single-digit numbers, and Arabic numerals is exactly the same as that between the corresponding English alphabetically written words and these very same Arabic numerals.

This is, of course, the conclusion we reached in our discussion about number notational systems. It is clear that symbols do not lose their identity as number names or as numerals when they denote single-digit numbers. Nevertheless, in dealing with single-digit rather than multidigit numbers, processing operations should be more dependent on the surface form of the symbols than on the notational system to which these symbols belong. Therefore, in investigating the processing of single-digit numbers considered as lexical units, it is a priori more natural to regroup the symbols with respect to their surface forms irrespective of the notational system. Accordingly, in what follows, Arabic numerals and Chinese or Japanese Kanji number names are subsumed under the logographic category, while the generic term phonographic is applied to number names written alphabetically or in Hiragana (hereafter simply referred to as Kana because the Katakana form has not yet been used).

Roman numerals are part of a different notational system, but their surface form can be considered as logographic.

Numerical Size Comparison Judgments

A common experimental task calls on subjects to judge which of two simultaneously presented Arabic numerals is the larger (less often, the smaller) numerically, with response latency as the dependent variable. Such experiments have provided a rich pattern of results revealing at least four different effects: symbolic distance, serial position, semantic congruity, and size congruity. This abundance of effects (not confined to the comparison of number numerical sizes, but apparent also in many other comparative judgment tasks) has recently been the subject of much theorizing (see Banks, 1977; Moyer & Dumais, 1978, for reviews). For our purposes, the main point of interest is the possibility of observing different configurations of results as a function of the surface form of the numbers. In what follows each effect will be briefly characterized and studies contrasting different types of number representations will be reviewed and discussed.

Symbolic Distance Effect

The latency of the comparative judgment is an inverse function of the subtractive difference between the two numbers; for example, subjects are faster in judging that 7 is the larger in a pair like 2-7 than in a pair like 5-7 (Aiken & Williams, 1968; Banks, Fujii, & Kayra-Stuart, 1976; Buckley & Gillman, 1974; Duncan & McFarland, 1980; Moyer & Landauer, 1967; Parkman, 1971; Sekuler & Mierkiewicz, 1977). The effect is also observed when numbers are symbolized by patterns of dots (Buckley & Gillman, 1974) or with numbers written in Kana and in Kanji (Takahashi & Green 1983). In the latter study, distances of 1, 3, and 5 were compared; the general trend was the same for both kinds of script, but the detailed pattern of results was slightly different in each case. With Kana stimuli there was a relatively small decrease in reaction time between distances 1 and 3 and a relatively large decrease between distances 3 and 5, whereas with Kanji the opposite configuration was observed, a large decrease between distances 1 and 3 and a small one between distances 3 and 5. Since only 12 out of the 36 possible pairs were studied, the effect could have arisen from an interaction between the relative coding difficulty of the pairs, symbolic distance, and type of script, rather than from a different comparison process taking place with each kind of script. This is a likely possibility in view of the absence of interaction between symbolic distance and type of script (Arabic numerals vs. alphabetic number names) in the experiment of Foltz, Poltrock, and Potts (1984, Experiment 2). In this case, the complete set of 36 pairs was used.

Serial Position Effect

In the present framework serial position refers to the position of each member of a pair of numbers relative to the boundaries of the ordered sequence of single-digit numbers. For a given symbolic distance, pairs composed of small numbers (e.g., 1-3, 2-4) are compared more rapidly than pairs composed of large numbers (e.g., 6-8, 7-9). The effect, often expressed as an increase in reaction time as a function of the increase in the smaller member of each pair, has been consistently observed with Arabic numerals (Aiken & Williams, 1968; Buckley & Gillman, 1974; Parkman, 1971). As for symbolic distance, the serial position effect was also obtained with numbers symbolized by patterns of dots (Buckley & Gillman, 1974), and there was no interaction between the serial position effect and the type of script (Arabic numerals vs. alphabetically written names) in the study of Foltz et al. (1984, Experiment 2).

Semantic Congruity Effect

This effect was identified by Banks, Clark, and Lucy (1975). It results from an interaction between the way the instructions are formulated with respect to the boundaries of the ordered set of numbers and the position of the pair of numbers with respect to these boundaries. With small numbers (e.g., 2-4) subjects make their comparisons more rapidly under the instruction "choose the smaller" than under the instruction "choose the larger." Conversely, with larger numbers (e.g., 6-8) decisions are reached more rapidly under the instruction "choose the larger" than under the instruction "choose the smaller." The semantic congruity effect has been observed twice with Arabic numerals (Banks et al., 1976; Duncan & McFarland, 1980). Although the effect has not yet been investigated with other number representations, it is unlikely that the outcome of such a study would show differential effects according to the surface form of the numbers. One reason for this is that in judging the size of two objects or the intelligence of two animals, the semantic congruity effect has been found to be independent of the representation of the referents as pictures or as alphabetically written names (Banks & Flora, 1977).

At present, the picture that emerges from contrasting logographic and phonographic representations of numbers in numerical comparison judgments is incomplete, but quite consistent. As regards the symbolic distance and serial position effects there is no evidence that the task is performed differentially according to the surface form of the stimuli, and with respect to the semantic congruity effect the relevant information is not yet available. For the size congruity effect, to be described next, the results are more contradictory; this is also the case for experiments using lateral hemifield presentations in numerical size comparisons. In order to draw some tentative conclusions from these data a more detailed analysis will be necessary than has sufficed for the three effects discussed before.

Size Congruity Effect

This effect, labeled by Banks and Flora (1977), was first observed by Paivio (1975) in a size comparison task involving objects represented either by pictures or by words. It appeared in a Stroop-like situation in which an irrelevant dimension, the relative physical size of each member of a pair of stimuli, was combined orthogonally with the relative real sizes of the referents. In a congruent trial the stimulus referring to the larger object was also physically larger than the other. In an incongruent trial the stimulus referring to the larger object was physically smaller. Neutral trials in which both members of the pairs of stimuli were the same physical size were also included. Paivio observed a size congruity effect with pictures, the mean response latency being 89 ms faster for congruent than for incongruent trials. The most striking result was that there was no congruity effect at all when the same referents were represented by words instead of pictures. As regards number comparisons, this Stroop-like task was first used by Besner and Coltheart (1979) who obtained results parallel to those of Paivio; namely, a large size congruity effect with Arabic numerals and no effect at all with the alphabetical representations of the numbers. Subsequent experiments confirmed the result with Arabic numerals, but were discrepant with the initial study in showing a large size congruity effect with alphabetical number names as well (Besner, Davelaar, Alcott, & Parry, 1984; Foltz et al., 1984; Peereman & Holender, 1984). The size congruity effect was also observed with numbers written in Kanji whereas Kana numbers showed ambiguous results (Takahashi & Green, 1983).

Table 1 summarizes the main results of the experiments published so far, except for some forthcoming data of the second author (Peereman, in preparation). In addition to presenting the mean reaction time for each type of trial (congruent, neutral, and incongruent), the table also splits the congruity effect into facilitation and interference effects. The facilitation effect is obtained by subtracting the mean latency of congruent trials from the mean latency of neutral trials, while subtracting the latter from the mean latency of incongruent trials yields the interference effect. A few more procedural details are worth describing before we discuss the results. In some experiments the numbers were presented side by side, to left and right of a fixation point, and responses were made on a left and right response key (Besner & Coltheart, 1979, logographic condition; Foltz et al., 1984; Henik

Table 1

Size Congruity Effect in Numerical Size Comparison Judgments[a]

Authors	Experiment or condition	Logographic C	Logographic N	Logographic I	Logographic N - C	Logographic I - N	Phonographic C	Phonographic N	Phonographic I	Phonographic N - C	Phonographic I - N
Henik & Tzelgov, 1983[b,c]	Exp. 1	588	624	696	36	72	--	--	--		
Besner & Coltheart, 1979[b]		531	542	586	11	44	800	--	800		
Foltz et al., 1984	Exp. 2	564	585	641	21	56	749	762	795	13	33
Peereman & Holender, 1984	Central field	472	500	561	28	61	719	724	756	5	32
	Left field	481	523	577	42	54	717	755	777	38	22
	Right field	472	528	577	56	49	704	745	759	41	14
Peereman, in preparation	Manual response	520	552	622	32	70	749	784	805	35	21
	Vocal response	568	602	694	34	92	751	768	795	17	27
Takahashi & Green, 1983b,d		752	790	855	38	65	1076	1054	1095	-22	41

[a] C=congruent, N=neutral, I=incongruent. [b] Data estimated from a graph. [c] First session only. [d] Data pooled over sessions 1 and 2.

& Tzelgov, 1982). The other experiments used numbers displayed above and below a fixation point, and responses were made either on two vertically aligned response keys (Besner & Coltheart, 1979, alphabetic condition; Takahashi & Green, 1983) or by activating a forward-backward switch (Peereman & Holender, 1984; Peereman, in preparation,). Only one study used the complete set of 36 pairs generated by using the numbers 1 to 9 (Foltz et al., 1984), whereas the others used only a small subset of these pairs, from 4 to 12 according to the experiment. In addition to central presentations, Peereman and Holender (1984) also included lateral ones and Peereman (in preparation) contrasted the usual manual response with a vocal response, the naming of the larger number.

The left side of Table 1 shows the results with logographic scripts, i.e., Kanji numbers in the experiment of Takahashi and Green (1983), and Arabic numerals in all the other cases. The main results can be summarized as follows.

1. There is a large overall size congruity effect (sum of the facilitation and interference effect in Table 1) in each experiment. The magnitude of the effect tends to increase with the increase in the absolute level of performance.

2. Both the facilitation and the interference effects are substantial in each experiment (except for a very small facilitation effect in the experiment of Besner and Coltheart). With central presentations, the magnitude of the facilitation effect is in the range of 20 to 60% of the magnitude of the interference effect. With lateral presentations (Peereman & Holender, 1984), the ratio of the two effects is closer to 1.

3. The Kanji numbers used by Takahashi and Green (1983) behave in pretty much the same way as the Arabic numerals used in the other experiments, except that response latencies are much longer than with Arabic numerals, probably because Kanji numbers are not widely used.

The right side of Table 1 shows the results for phonographic scripts, the syllabic Kana writing in Takahashi and Green's report and alphabetic writing in all the other cases. The most prominent aspects of the results are the following.

1. Overall response latencies are in the range of 200 to 250 ms longer than with logographic numbers. The absence of a congruity effect, reported by Besner and Coltheart (1979), is not confirmed in subsequent experiments, although the effect tends to be a bit smaller than with logographic numbers. There is no systematic relation between the absolute level of performance and the magnitude of the size congruity effect.

2. With central presentations, much of the size congruity effect is due to the interference caused by incongruent trials, congruent trials provoking almost no facilitation or even a detrimental effect (Takahashi & Green, 1983). With lateral presentations, the opposite tendency is observed; that is, strong facilitation effects and weak interference effects (Peereman & Holender, 1984).

3. Kana numbers (Takahashi & Green, 1983) are responded to much slower than alphabetic numbers, but this form of representation is almost never used outside the laboratory. There is also a reversal in the facilitation effect.

The most important point to discuss is the discrepancy between the absence of a congruity effect with alphabetic numbers in the experiment of Besner and Coltheart (1979) and the presence of such an effect in the two other experiments (Foltz et al., 1984; Peereman & Holender, 1984). Foltz et al. interpreted the difference between their results and those of Besner and Coltheart as due to their use of a repeated-set design instead of the fixed-pair design of the conflicting experiment. In a repeated-set design each item (the number 1 to 9) is paired equally often with each other item, whereas only a small subset of these pairs is used repeatedly in a fixed-pair design (12 pairs repeated 20 times and 9 pairs repeated 10 times in the logographic and alphabetic conditions of Besner and Coltheart, respectively) and each item is paired with only a few other items (one, two, or three in Besner and Coltheart's experiment). It is argued that there is an increasing probability of bypassing the comparison stage as a function of the increase in the number of repetitions in the fixed-pair design: Subjects may respond on the basis of specific response-pair associations established during the experiment. This accounts for the lack of a size congruity effect in Besner and Coltheart's fixed-pair design and the presence of such an effect in Foltz et al.'s repeated-set design. Moreover, the prediction was nicely supported in a study using names of objects (Experiment 1 of Foltz et al.) where in the fixed-pair design of Paivio (1975, six pairs repeated eight times, each item being paired with only one other item), no size congruity effect was observed. However, with an infinite-set design in which 48 different pairs were presented only once, as if they were drawn from an infinite set of pairs, a strong 115-ms size congruity effect was obtained, which reduced to 49 ms after three further presentations of the set. Recently, Besner et al. (1984, p. 127) also alluded to the observation of a size congruity effect in using a larger set of alphabetic numbers than in the original experiment of Besner and Coltheart (1979). There is, however, one result that is clearly at odds with this interpretation. In our alphabetical condition (Peereman & Holender, 1984), only four different pairs were repeated 72 times, each of four numbers being paired with only two other numbers. This should have maximized the chances of bypassing the comparison stage, thereby suppressing the size-congruity effect, but this did not happen.

A further assumption is needed to account for the fact that the repeated-set design does not suppress the size congruity effect when Arabic numerals are used instead of alphabetic number names. Foltz et al. (1984) suggested that, because pictures or Arabic numerals provide much shorter latencies than their spelled names, retrieving and comparing the size information could be faster than retrieving the appropriate previously learned response in the former than in the latter case. This is a completely ad hoc interpretation. In addition, it cannot explain why, in a fixed-pair design, Takahashi and Green (1983) observed a very strong size congruity effect with Kanji numbers in spite of the fact that the absolute level of performance was equivalent to that of Besner and Coltheart (1979) in the alphabetic condition (see Table 1). In such a case, according to Foltz et al.'s interpretation, the retrieval of previously associated responses should have been faster than the size retrieval and comparison process, leading to no size congruity effect.

For other tendencies revealed in Table 1, such as the smaller congruity effect with phonographic than with logographic script and the different ratios between the facilitation and the interference effect with each kind of script, no unequivocal conclusion can be drawn at present. The problem is that the situation is a little too complicated. Several confounding factors whose role are not well understood could be responsible for these effects. Moreover, none of them might give any interesting hint toward a possible differential role of the surface form of the stimuli in the operations needed to perform the numerical size comparison judgment. Let us mention two such confounding factors.

1. The relative salience of the irrelevant dimension affects the magnitude of its influence on the decision about the relevant dimension (Besner & Coltheart, 1976; Dixon & Just, 1978). In the present context, the salience of the irrelevant dimension may well be influenced by the factors affecting the judgment of dissimilarity between rectangles, because, roughly speaking, the areas occupied by Arabic numerals or by upper case number names are rectangular in shape. The psychophysics of dissimilarity judgments between rectangles varying in shape and area (Krantz & Tversky, 1975; Wender, 1971; Wiener-Ehrlich, 1978) is surprisingly complex, no simple dimensional structure emerging from the data. There are two ways in which the data discussed in this section could be affected by these psychophysical factors. First, the difference between the physical size of two Arabic numerals can simply be more conspicuous than that between two multiletter words, leading to a stronger size congruity effect in the former than in the latter case. Second, the speed with which a dissimilarity judgment can be made, or for our purposes, the speed with which the difference in size becomes compelling, should depend on the magnitude of the physical difference, at least within a certain range. This could be responsible for subtle differences between the magnitude of the interference and facilitation effects according to type of script.

2. From our experience with the task, we know that the magnitude of the congruity effect and the relative magnitudes of the facilitation and interference effects vary considerably between different pairs of numbers, especially with the alphabetical representation. Having used only a small subset of pairs in our experiments, it is hard to find any systematic factor underlying either the intra-surface form or the inter-surface form variability. We nevertheless suspect that some pairs are more easily encoded than others, thus affecting the time at which the information becomes available for performing the comparison operation. This could, of course, generate different patterns of results between experiments using different subsets of pairs

These two confounding factors emphasize the role that the relative time course of processing both the relevant and irrelevant aspects of the pairs of stimuli might play in the determination of the size congruity effect, independent of the comparison process itself. Of course, this could be systematically studied, but we then run the risk of completely losing sight of the real goal of this research, which is precisely to investigate whether or not the surface form of the stimuli affects the numerical size comparison operations, not to untangle the complexity of Stroop-like situations.

Hemifield Presentations

The rationale for using hemifield presentations of stimuli will be explained in the next main section of the paper. Suffice it to say here that a relatively better performance for stimuli displayed in one hemifield than in the other is generally interpreted in terms of a contralateral hemispheric superiority for a particular class of stimuli or for a particular experimental task. The investigation of lateral presentations of numbers for comparison of their numerical magnitudes has led to a perplexing picture, because every possible outcome has been reported. Katz (1980, 1981) found a left visual field (LVF) advantage; Besner, Grimsell, and Davis (1979), a right visual field (RVF) advantage; and Peereman and Holender (1984), no difference between fields.

The opposite field advantages of Katz (1980, 1981) and Besner et al. (1979) can be explained by the difference in the exposure durations that were used. A short exposure duration, 50 ms in Katz's experiments, could engender a RVF advantage that has little to do either with the specific material presented or the specific task performed, but is determined rather by the nature of the available visual information (see Sergent, 1983a, 1983b). According to Sergent, the right hemisphere is more efficient than the left in extracting the relevant information from low spatial frequencies than from high spatial frequencies, and vice versa for the left hemisphere. Physical parameters such as very short exposure duration, large stimulus size, and large eccentricity should favor processing on the basis of low spatial frequencies, therefore increasing the odds of finding a LVF advantage whatever the type of stimulus. On the other hand, long exposure durations, such as the 150 ms used by Besner et al. (1979), generally lead to a RVF, which was indeed observed in this particular study. Notice, however, that the authors strongly favored an interpretation of their field advantage in terms of a left hemispheric superiority for performing the comparison process rather than for encoding the stimuli.

Why then, using a relatively long exposure duration of 120 ms, did Peereman and Holender (1984) fail to show any laterality effect? There is no ready interpretation for the discrepancy between their results and those of Besner et al. (1979). However, some tentative suggestions can be made.

The combination of left and right presentations with responses that are also spatialized along the left-right dimension may generate the compatibility effect first reported by Simon (Craft & Simon, 1970; Simon & Rudell, 1967). Asking their subjects to press a right key at the sound of a high tone and a left key at the sound of a low tone (Simon & Rudell, 1967), or to associate the right key with a red bulb and the left key with a green bulb (Craft & Simon, 1970), Simon and his collaborators observed that the right side response was made faster if the stimulus was presented in the right hemispace rather than in the left hemispace, and conversely for the left side response. This compatibility effect has been described as a tendency to react toward the source of stimulation. It is genuinely a semantic congruity effect similar to that of Banks et al. (1976), discussed earlier, because the coding of responses in terms of left and right entails an unavoidable influence of the coding of stimulus location in the same terms, thereby facilitating or interfering with the response according to the congruency or incongruency of stimulus and response positions. This compatibility effect has also been observed with lateralized presentations of pairs of numbers. Besner et al. (1979) found that right-index responses were shorter for displays presented in the RVF than in the LVF and vice versa for left-index responses. The same was true for the relation between the rightmost or leftmost finger and the visual field when two fingers of the same hand were used to make the response (Katz, 1981). However, with bimanual responses Katz (1980, Experiment 1) failed to find the effect, a surprising outcome in view of the usual robustness of the phenomenon.

Figure 1 illustrates what happens when the side of presentation affects each response in an opposite direction through one factor--compatibility--and in the same direction through another factor--the presumed hemispheric superiority. In Besner et al.'s experiment, the RVF advantage, which is very strong for the right response, gives way to a small (nonsignificant) LVF advantage for the left response. Similarly, in Katz's (1981) experiment, the LVF advantage, which is very strong for the left response, is considerably reduced for the right response. The explanation of this phenomenon goes as follows, taking Besner et al.'s results as the basis for the reasoning. For the right response, compatibility and the assumed left hemispheric superiority add their effects to enhance performance with RVF presentations and to impede performance with LVF presentations, thereby inducing a large field difference. For the left response, the advantage of the stimulus being presented in the LVF due to compatibility is counteracted by the disadvantage of being first channeled to the wrong hemisphere, whereas the benefit of the RVF presentation due to left hemispheric superiority is counterbalanced by the cost of being on the wrong side in terms of compatibility, thereby reducing, and even reversing the field advantage.

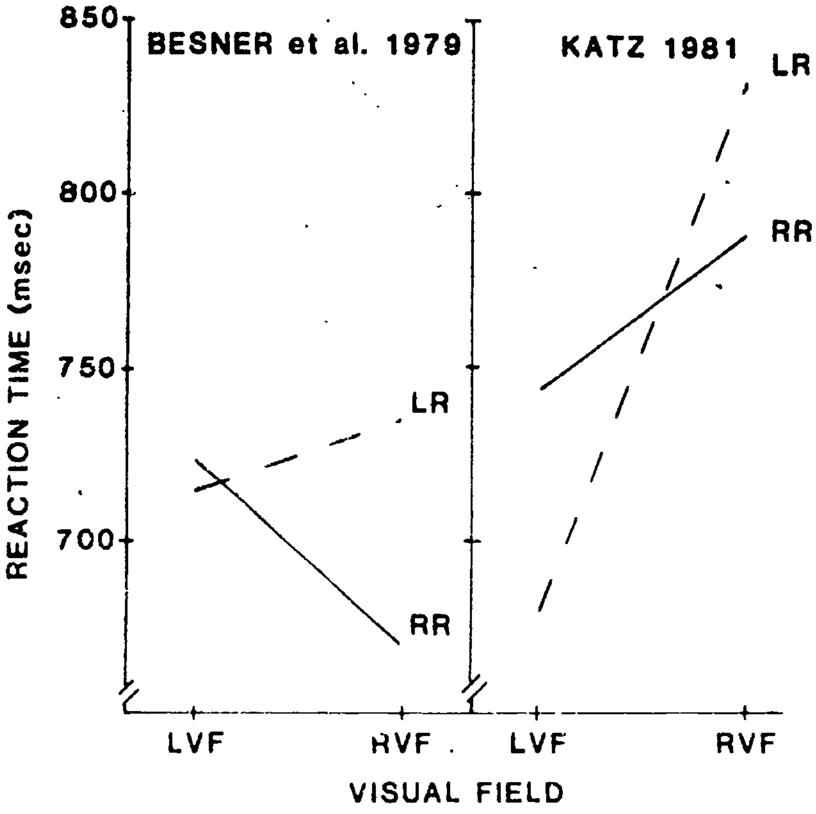


Figure 1. Mean reaction time for judging which is the larger of two numbers as a function of visual field (LVF=left visual field, RVF=right visual field) and response side (LR=left response, RR=right response) in the experiments of Besner et al. (1979) and Katz (1981). The data were estimated from graphs.

The existence of a compatibility effect that interacts in a complex manner with a laterality effect, presumably linked to hemispheric superiority, is an obstacle to the study of this phenomenon per se. It would therefore be better to get rid of the compatibility effect, by suppressing the left-right polarity of the response, presenting the pairs of numbers vertically and replacing the left-right responses with forward-backward responses, exactly the procedure used by Peereman and Holender (1984). If the left hemisphere is really better than the right in performing the comparison process, as Besner et al. (1979) believed, it should be so whatever the spatial disposition of the numbers in the pair and the ensuing RVF should show up uncontaminated by the compatibility effect. However, as we have already pointed out, the LVF advantage disappeared altogether, when we followed this procedure. Thus, the factor that combined with compatibility to determine the pattern of results found by Besner et al. (see Figure 1) was not a left hemisphere superiority for the task, but something else.

What else? We do not know, but we suggest looking to other forms of compatibility that almost certainly play a role when left-right polarized displays and responses are involved in the comparison of the numerical magnitude of numbers. For instance, in deciding which number is the larger, the response is faster if the larger number is on the right side of the pair (e.g., 3-7) than with the opposite configuration (7-3). This effect was as large as 30 ms in the experiment of Aiken and Williams (1968), using 18 pairs among the 36 possible, and 20 ms in Experiment 2 of Banks et al. (1976), using 21 pairs. However, the effect was null in Banks et al.'s Experiment 1 involving only six pairs, suggesting possible interactions with specific characteristics of the pairs.

A last point should be stressed. Peereman and Holender's (1984) experiment is the only one fulfilling the requirements of this chapter for numerical size comparisons; that is to say, it is the only study that combines factorially the type of script (Arabic numerals and their French alphabetic names) with the side of presentation. It is clear from Table 1 that there is no field advantage whatever the type of script, a conclusion that can probably be safely accepted. Whether there is evidence for a differential influence of the type of script on the comparison process cannot be answered on the basis of these data because, as remarked in the preceding subsection concerning the size congruity effect, several possible confounding factors must be controlled before any reliable conclusion can be reached.

Lateral Hemifield Presentations

Rationale Underlying the Approach

From the standpoint of understanding how numbers represented logographically or phonographically are processed, the study of laterality should be considered as one of the tools for analysis of processing operations into components. However, the extent to which the method succeeds in doing so depends on a number of difficult, unsettled issues.

In vision, provided gaze fixation is controlled, it is a matter of anatomical fact that a stimulus displayed laterally in the LVF or in the RVF is first channeled to the contralateral hemisphere and that its access to the homolateral hemisphere depends on its transit through the interhemispheric commissures. The most common interpretation of a better performance in one hemifield than in the other is in terms of a greater ability of the contralateral hemisphere to perform the task. In other words, a given hemifield advantage is almost automatically translated into a contralateral hemispheric superiority. Two points are worth stressing. First, even if a task is fully lateralized (i.e., can be accomplished by only one hemisphere), this need not entail better performance in terms of response latency or accuracy for contralaterally displayed stimuli (see G. Cohen, 1982, for an excellent discussion of this point). Second, besides hemispheric superiority, a number of factors can determine a hemifield advantage. This point has been repeatedly stressed by Bryden (1978, 1982; see also Bertelson, 1982, for a similar case). Our venture at interpreting the contradiction between the results of Besner et al. (1979) and Peereman and Holender (1984) as resulting from a combination of different compatibility effects is a good example of such an alternative approach.

Be that as it may, the logic underlying the study of lateralized presentations of different types of script implies that the results of an experiment should show some kind of interaction between hemifield and type of script. Three different interactive patterns could emerge: (a) opposite visual field advantages for each type of script, (b) no field advantage for one type of script and a field advantage for the other, (c) different degrees of field advantages in the same direction for both scripts. The first pattern is called a nonordinal interaction because each level of one factor (RVF vs. LVF) has an opposite effect on each level of the other factor (type of script). The third pattern is an ordinal interaction because the laterality effect has the same direction for each level of the other factor. The second pattern, in which there is no significant field advantage for one type of script, is a special case of either the first or the third pattern. Among these three possible interactive patterns, the first is certainly the most appealing because it takes the form of a double dissociation between the field advantages and the two kinds of stimuli, and because a nonordinal interaction cannot be removed by a nonlinear transformation of the dependent measure. The existence of such an interaction is therefore relatively independent of the choice of the dependent measure.

Claims for opposite field advantages in processing phonographic and logographic scripts arose from the initial observation of a RVF advantage in the identification of Kana words (Hatta, 1978) or nonwords ([illegible] Shimizu, & Hori, 1978; Sasanuma, Itoh, Mori, & Kobayashi, 1977), and of a LVF advantage in the identification of Kanji words (Hatta, 1977a, 1977b, 1978). Since then, the RVF advantage for processing Kana words has been clearly confirmed. Moreover, Kanji words composed of more than one character are also better processed in the RVF. For single Chinese or Kanji logograms the results are more contradictory because all possible outcomes--RVF, LVF, or no field advantages--have been reported. In spite of this, there is still a widespread tendency to consider that the bulk of the evidence favors the hypothesis of a right hemispheric superiority for the processing of single characters (see Coltheart, 1983, for a recent example). From our reading of that literature, we believe that too many confounding factors could have flawed most of these results for the existing data to be conclusive. If a conclusion is nonetheless to be drawn, we would argue that a right hemisphere superiority for logographic processing is extremely unlikely (see Peereman & Holender, 1985; Holender & Peereman, in preparation).

Review of the Data

The best way to characterize the investigation of lateral differences in the processing of numbers is to say that the data are scarce, the procedures diverse, and the results quite consistent. Most experiments have been concerned exclusively with Arabic numerals, although two have included alphabetically written numbers. Let us distinguish between those studies using response latency and those relying on response accuracy as the dependent variable, reviewing the latter first.

Hines, Satz, Schell, and Schmidlin (1969, Experiment 3) inaugurated a series of experiments in which three pairs of numbers were successively presented: One member of each pair was displayed at fixation point, the other either 3° to the left or 3° to the right of fixation. In any particular trial, the lateral member of each pair was always on the same side. (A fourth centrally placed number was temporally interpolated between the third pair and recall in two subsequent studies [Hines & Satz, 1971, 1974].) The task of the subjects was first to recall all the central numbers, and then to recall the lateral ones, only trials with 100% correct central identification being taken into account. The results always showed an overall better recall for right than for left numbers. Further examination showed the RVF advantage to be confined to the first two pairs of a trial, the last pair showing no field advantage. These data are generally disqualified on the ground that the central task in itself can generate a RVF independent of the nature of the stimuli (see Bryden, 1982, for a discussion of this long-standing debate). These data also involve a mixture between perceptual and memory processes without allowing us to disentangle their respective contributions to the RVF advantage, if any.

However, if both members of each pair of stimuli are laterally displayed one in the LVF simultaneously with the other in the RVF (rather than one laterally, the other centrally), a weak LVF advantage may be observed. This effect was not significant in Experiments 1 and 2 of Hines et al. (1969), but reached significance in Experiment 4 of Hirata and Osaka (1967). This LVF advantage could result from the strategy of report, rather than from nature of the stimuli, the left member of each pair being generally reported before the right one.

Carmon, Nachshon, and Starinsky (1976) reported a higher percentage of recall for two- or four-digit numbers (represented by Arabic numerals) in the RVF than in the LVF with fifth- and seventh-grade children. First- and third-grade children were tested only with two-digit numbers, and showed no field advantage. Hatta and Dimond (1980) also reported better RVF recognition of six-digit numbers with adult Japanese and English subjects. However, this RVF advantage might be caused by the combinatorial process involved in forming multidigit numbers rather than by the logographic nature of the representation.

Yet, Besner, Daniels, and Slade (1982, Experiment 1) obtained a very large RVF advantage with single-digit Arabic numerals, right presentations leading to 80% correct responses and left presentations to only 40%. In their second experiment, they tested Japanese and Chinese subjects with both Kanji numbers and Arabic numerals. This time the 14% RVF advantage for Arabic numerals was less pronounced than in Experiment 1. Overall performance with Kanji numbers was much lower than with Arabic numerals, but a 16% RVF advantage was again observed. It is a pity the authors limited their material to the numbers 4 to 9. Remember that in Kanji, the first three numbers are concrete representations of the quantity they denote, being composed of one, two, or three horizontal strokes, whereas the other numbers are arbitrary symbolic representations, like Arabic numerals. It would have been interesting to compare the laterality effect in the two cases.

We should point out that the extent to which the huge laterality effect obtained in these experiments was caused by physical characteristics of the displays is not known. Given the importance of visual parameters in determining visual field advantages (Sergent 1983a, 1983b), this point is worth stressing. Most studies resort to stimuli physically smaller than the Arabic numerals, subtending 5.9° x 8.5°, 4.6° x 5.7°, and 2.0° x 3.3°; in Besner's et al.'s Experiment 1 and than the 10.6° x 10.6° stimuli of their Experiment 2. These stimuli were centered 8.8° to the left or right of fixation. The exposure duration was individually adjusted to yield an overall performance of 50 to 60% correct responses, mean durations being 32, 44, and 56 ms for small, medium, and large stimuli in Experiment 1, and 54 ms in Experiment 2. Finally, a 50-ms patterned mask immediately followed stimulus presentation, which is also unusual.

We now turn to studies in which response latency was the dependent variable. Naming latencies for Arabic numerals showed no field advantage in the experiment of Gordon and Carmon (1976) and a small, but significant, 10-ms RVF advantage in Experiment 3 of Geffen, Bradshaw, and Wallace (1971). Procedural differences between experiments inspire no special comments. The main parameters of the task were, for Gordon and Carmon (1976) and Geffen et al. (1971), respectively: 7 and 4 different stimuli, exposure durations of 100 and 160 ms, stimulus visual angles of 2° and 0.5°, and eccentricities of 3° and 4°.

With two-choice manual response tasks involving only two Arabic numerals, a significant 13-ms RVF advantage was found by Geffen et al. (1971, Experiment 5) and a similar but not significant 14-ms advantage was reported by G. Cohen (1975) in her cued condition. Cohen mixed three different representations of the numbers 4 and 5: Arabic numerals, their English names presented vertically, and the corresponding patterns of dots found on a die. Subjects were either cued or not cued about the specific representation to be used on each trial. Under precuing, number names yielded a slightly greater RVF advantage (20 ms) than Arabic numerals and dots showed a nonsignificant 12-ms LVF advantage. Without cuing, there was no field difference, whatever the type of stimuli.

Classification tasks also yield a small RVF advantage with Arabic numerals. Geffen, Bradshaw, and Nettleton (1973) used a many-to-one stimulus response mapping in a go-no go task involving four numbers and one vocal response. Two Arabic numerals called for the response "bong" and two others required no response. This yielded a 16-ms RVF advantage. In a number-nonnumber classification modeled on the classical lexical decision task, Peereman and Holender (1985) showed a significant 13-ms RVF advantage and a significant 26-ms advantage in the same direction for alphabetically written number names, the interaction between visual field and type of script being nonsignificant.

Tasks involving more complex decisions than those just described have been almost exclusively concerned with numerical size comparison judgments (Besner et al., 1979; Katz, 1980, 1981; Peereman & Holender, 1984); they were reviewed and discussed in the preceding main section. There is only one more study to mention. Hatta (1983, Experiment 1) orthogonally varied the numerical size and the physical size of each member of laterally displayed pairs of Arabic numerals, asking his Japanese subjects to perform a congruity judgment. An overall 29-ms RVF advantage ensued. A 47-ms RVF advantage also showed up when the same task was performed with the Kanji logograms denoting million units (Experiment 3). By contrast, in judging the congruity between the relative physical size of pairs of logograms and the relative physical sizes of the referents (Experiment 2), the subjects showed a 26-ms LVF advantage. The author interpreted his results as evidence that the comparative judgment is based on different types of mental representation in dealing with Kanji object names and with numbers, but that in this latter case, the surface form of the stimuli (Arabic numerals vs. Kanji words) is immaterial.

Interpreting the Results

There is nothing to indicate that opposite visual field advantages for each kind of script, the first possible pattern of results mentioned above, will ever be found in contrasting numbers written logographically and phonographically. On the contrary, both surface forms lead to RVF advantages. If the LVF sometimes reported with single Chinese logograms is valid, then Arabic numerals belong to a small class of logograms behaving differently as regards laterality, as is also suggested by the results of Hatta (1983).

To date, there has been no report of a significant ordinal interaction between visual field and type of script, but the prospect of finding one is quite good. With Arabic numerals and simple tasks like naming or categorizing, a RVF of 10 to 15 ms is typically found; this is the lower bound for the effect to be statistically significant. On the other hand, Peereman and Holender (1985) pointed out that the magnitude of their RVF advantage (26 ms) for numbers written alphabetically was more substantial and well within the range of the large RVF advantages typically reported in lexical decisions involving larger classes of words. Hence, there would be nothing very unexpected if a statistically more powerful study in the future came up with a significant ordinal interaction indicating a larger RVF advantage for alphabetic number names than for Arabic numerals, both RVF advantages being significant.

Let us assume that the ordinal interaction has indeed been found. What, and how much information would then have been gained regarding logographic and alphabetic number processing? To answer this question, we will be obliged to integrate laterality research into the broader framework of mainstream information processing analysis--a highly desirable, but so far unfulfilled accomplishment (Allen, 1983; Bertelson 1982). Bertelson optimistically closed his recent analysis of laterality research with the words "Progress can be expected, provided laterality research is conducted as an integral part of the study of human cognition" (1982, p. 203). Taking a few steps in this direction, in search of an answer to the question asked at the outset of this paragraph, we came up with a more distressing conclusion (Peereman & Holender, 1985). An analysis similar to that leading to this conclusion will now be presented.

288

The aim of the analysis is to show how the ordinal and nonordinal interactions described in the rationale for the approach can be interpreted in the relatively constrained framework of a stage analysis of reaction time. The two basic assumptions are as follows:

1. Response latency can be decomposed into a series of additive component durations corresponding to different stages of processing. For additivity to hold, the processing stages should be strictly serial, each stage starting only when the preceding stage has provided an output. Under these constraints, any modification of the duration of one particular stage under the influence of any factor (i.e., hemifield of presentation) should be reflected in the response latencies. This is the one of the assumptions underlying Sternberg's (1969, 1984) additive factor method, one of the most popular methods of analysing processing into components.

2. Hemispheric specialization is relative rather than absolute: Each hemisphere can perform the task, but one is more efficient than the other. This is more reasonable than the alternative assumption of absolute hemispheric specialization (only one hemisphere can perform the task), which would imply that the difference in latency between visual fields is due to the time needed to transfer information from one hemisphere to the other when the stimulus is displayed on the wrong side. This alternative is unlikely because it would entail a relative constancy across experiments in the magnitudes of the difference between response latencies in the two fields, which is hardly the case (G. Cohen, 1982).

Within this framework, the simplest possible account for the presence of an interaction, either ordinal or nonordinal, between visual field and type of script requires the addition of two specific assumptions. (1) All processing stages are neutral with respect to laterality save one, or at maximum two--let us call them Stages A and B--which can be either neutral or lateralized according to circumstances. In the neutral state of a stage the operations performed during that period take the same mean amount of time in each hemisphere. If a stage is lateralized, the corresponding operations are performed faster in one hemisphere than in the other one. (2) We cannot exclude a priori the possibility that (a) both Stage A and Stage B are lateralized on the same side for both types of stimulus, or (b) that each stage is neutral for one type of stimulus and lateralized for the other, or (c) that the two stages are lateralized in opposite directions for each type of stimulus. Within these constraints, each pattern of interaction can be realized in three extreme ways according to the following principles. In each of the three cases, the ordinal interaction is labelled 1 and the nonordinal, 2:

A. Only Stage A is lateralized, Stage B is neutral.

1. Stage A is left-lateralized for both kinds of number representations, but the magnitude of the RVF advantage depends on the surface form of the stimuli, being larger for alphabetical number names than for Arabic numerals.

2. Stage A is left-lateralized for alphabetical number names and right-lateralized for Arabic numerals.

B. Stage A is left-lateralized for both scripts and produces the same degree of RVF for both scripts.

1. Stage B is neutral for Arabic numerals and left-lateralized for alphabetical number names; hence, Stage B adds its RVF advantage to that of Stage A, producing the required interaction.

2. Stage B is neutral for alphabetical number names and right-lateralized for Arabic numerals. Stage B produces a LVF advantage sufficient to supersede the RVF caused by Stage A.

C. Stage A is neutral for alphabetical number names and Stage B is neutral for Arabic numerals.

1. Both stages are left-lateralized when dealing with their specific stimulus type. It just happens that the RVF advantage due to Stage B is larger than that produced by Stage A.

2. Stage A is right-lateralized for Arabic numerals and Stage B is left-lateralized for alphabetical number names.

Within the constraints defined above, one can easily find the different possible interpretations corresponding to the third interactive pattern, in which only one type of script shows a visual field advantage. Similarly, the different possibilities corresponding to an absence of interaction can also be worked out.

In performing a similar analysis (Peereman & Holender, 1985), we showed that a significant ordinal interaction is no more informative than a nonsignificant interaction. We now extend this conclusion by showing that the favorite nonordinal interaction is no more informative than the ordinal. Using the simplest possible model for the organization of processing operations, and looking at the interaction between visual field and stimulus type, we always come up with three different possible interpretations. In other words, laterality as a tool for analysing processing into components simply fails to do its job. One can, of course, retort that nobody ever pretended to disentangle these various alternatives by using the laterality approach. The point would be well taken, but then what is the purpose of presenting all these beautiful phonograms and logograms in the left or in the right visual field? To avoid such criticisms researchers using the laterality methodology should be much more explicit about their goals than they usually are.

Number Processing After Brain Injury

The discussion of the data provided by brain damaged people is divided into two parts. In the first, all patients have lesions affecting different language areas of the left hemisphere. These patients display a variety of aphasic troubles, including alexia with agraphia. Potentially, the investigation of these patients can teach us something about the way different notational processing systems can break down, but the respective roles played by each hemisphere in determining the preserved aspects of performance cannot be ascertained. In the second part of the discussion, data concern the partial or total disconnection of the right hemisphere from the left. These data can potentially tell us something about the competence of the right hemisphere in dealing with different representations of single-digit numbers.

Number Processing with Lesions Located in the Language Areas

What is clear from the fragmentary information available is that the ability to read single and multidigit numbers represented by Arabic numerals can be somewhat preserved in patients unable to read letters and words in the alphabetic code. From the anatomo-clinical study of 183 retrorolandic brain injured patients, Hécaen, Angelergues, and Houillier (1961) concluded that the frequency with which letter or digit reading breaks down is different according to the site of lesion; this could indicate that partially different functional subsystems are indeed involved in each case. The same authors also mentioned 16 patients who, as a group, showed a relatively stronger inability to identify mathematical signs than Arabic numerals. A fully selective loss of this competence has been reported in two patients by Ferro and Botelho (1980). Unfortunately, they did not investigate the patients' ability to identify the written names of the mathematical signs.

We have found one patient-group study in which the ability to process different number surface forms was investigated (Dahmen, Hartje, Büssing, & Sturm, 1982). These authors selected three groups of 20 patients, each group corresponding to a different pathology--Wernicke's aphasia, Broca's aphasia, and right-sided retrorolandic lesion. These groups varied in their identification performance for numbers (chosen in the set 1 to 25), but showed no difference according to the type of representation (Arabic numerals or their German names). The mean numbers of correct identifications (out of [illegible]) for Arabic numerals and number names were: 13.3 and 12.0, 16.2 and 14.8, and 19.7 and 18.0, for Wernicke's aphasics, Broca's aphasics, and patients with a right-side lesion, respectively. The same was true in a numerical size comparison task in which the patients had to point to the larger number in pairs of numbers. For Arabic numerals and number names the mean number of correct responses were, 8.7 and 8.6, 14.2 and 14.3, and 16.2 and 15.9, for Wernicke's aphasics, Broca's aphasics, and patients with a right-side lesion, respectively. Three points should be stressed. First, the difference in performance between Broca's and Wernicke's aphasics is in the direction expected on the basis of the overall differences exhibited by these patients in terms of language comprehension. Second, the results of the comparison task confirm the trend we observed with normal subjects in being unaffected by the surface form of the numbers. Third, unlike the identification process, the comparison process seems to require the integrity of the right hemisphere as indicated by the difference in performance in each task in the group of patients suffering from a right-side lesion.

Most of the data reviewed so far are based on the comparison of the mean performances of groups of patients. Caramazza (1984) has recently pointed out that such an approach is ill-suited for addressing the issue of the analysis of cognitive processes because the patients included in a given group could differ greatly in terms of the mechanisms underlying their performance. The remaining data come either from single cases or from very small, relatively homogeneous groups of patients, whose individual symptomatology is generally available.

The few single case studies to mention in closing this subsection all concern Japanese patients who offer the additional interest of being able to show a dissociation between the processing of two forms of logographic script (Kanji words and Arabic numerals). One such aphasic patient was described by Sasanuma and Monoi (1975). He was severely impaired in language comprehension, whether spoken or written. His most prominent symptom was his greater ability to read aloud Kana than Kanji words, though with little comprehension in either case. This dissociation is extremely rare in Japanese aphasics, most of whom show a differential ability to process each kind of script, being better with Kanji than with Kana words (Sasanuma, 1974a, 1975). Aside from that, the patient was able to carry out arithmetical operations, and he could read and understand Arabic numerals.

The other Japanese patients are alexics with agraphia. Strictly speaking, the syndrome consists of a selective impairment in reading and writing, unaccompanied by any trouble in spoken language comprehension and expression. However, this ideal definition almost always overstates the true state of spoken language performance. It would be closer to reality to say that the most prominent syndrome coexists with mild aphasic troubles (e.g., Hécaen & Kremin, 1976). In these cases, reading impairment is always stronger in Kana than in Kanji. Yamadori (1975) reported one such patient who was severely impaired in calculation and in number reading. Yamadori also summarized two other reports published in Japanese (Kotani, 1935; Ohashi, 1965, cited by Yamadori, 1975) concerning two other cases of alexia with agraphia accompanied by strong calculation impairments.

Sasanuma (1974b) described a case of alexia with transient agraphia. The patient's reading in both Kana and Kanji was strongly deficient, performance in Kanji being a little better than in Kana. "Reading of digits, both Arabic and Chinese was impaired also" (Sasanuma, 1974b, p. 93), but less than for Kana and Kanji. The patient was good at mental calculation, but written calculation was hampered by his reading problem. Six months later almost all symptoms other than alexia had disappeared, but nothing more specific was stated.

Because of his preservation of mental calculation, the patient of Sasanuma (1974b) is sometimes considered as a counter-example to the observations of Yamadori (1975). Such cannot be the case because it is extremely unlikely that the two patients suffered from the same pathology. Yamadori's patient was alexic with agraphia, whereas the patient of Sasanuma presented all the symptoms of an alexia without agraphia (see next subsection) in which preservation of mental calculation is typical (e.g., Geschwind, 1965; Symonds, 1953).

To sum up, most patients with lesions affecting the language areas of the left hemisphere show various degrees of disintegration of their mathematical abilities and a poor ability to read Arabic numerals.

Number Processing by the Disconnected Right Hemisphere.

For theoretical reasons that will become apparent as we proceed, it is convenient to examine successively the data from patients having one of the following characteristics: (a) alexia without agraphia, (b) section of the splenium (posterior part) of the corpus callosum, (c) commissurotomy, and (d) hemispherectomy.

Alexia without Agraphia. An ideal patient with alexia without agraphia cannot read, but can write spontaneously and to dictation, without being able to reread what he or she has written. Such a patient has no trouble in spoken language expression or comprehension, but has some difficulties in visual object naming and a strong impairment in color naming; mental calculation is preserved. The classical account of the syndrome by the pioneer neurologists, as revived and specified by Geschwind (1965), describes isolation of the intact left angular gyrus from each occipital visual cortex. This condition is caused by (a) destruction of the left visual cortex or of the connections between the left visual cortex and the left angular gyrus and (b) destruction of the splenium of the corpus callosum, which cuts off from the left hemisphere the visual information reaching the right intact hemisphere. The essence of the trouble is, therefore, the disconnection of intact language zones from the visual world (but not from the auditory or somatosensory world). Logically, the lesions should entail an incapacity to name any visual scene; this is not the case, although it is partially realized by some difficulty in object naming and a very poor ability to name colors. The supplementary hypothesis needed to account for the preserved ability to name objects is that objects can be recognized (but not named) in the right hemisphere and that it is this interpreted information, not the visual information, that is transmitted to the left hemispheric language areas through the intact anterior portion of the corpus callosum. We assume that the right hemisphere is unable to provide verbal responses (see below). By extension, any reading performance preserved (e.g., for Arabic numerals) should reflect right hemisphere competence in processing the information. The same rationale is used by Coltheart (1980, 1983) in his attempt to account for deep dyslexics' preserved reading competence (the etiology of this syndrome is different from that of alexia without agraphia).

The recent literature has usually described four of the six alexic patients of Hécaen and Kremin (1976) as displaying the symptomatology required to fit the ideal model. They all performed better in dealing with Arabic numerals than with single letters or single words. Close examination of the constellation of symptoms displayed by these patients reveals that their classification is problematic: Part of their deficit could well be due to some lesion in the language area of the left hemisphere as well. Lack of space precludes any full analysis of this very complex question here. Only a brief account sufficient to make the point will be presented, but it should be kept in mind that including of a patient in one of the subgroups of Table 2 is often tentative because we generally lack the decisive anatomo-clinical data to remove the uncertainty. For example, the inclusion of Stengel, Vienna, and Edin's (1948) two patients in the group consisting of close to ideal cases would be disputed by Oxbury, Oxbury, and Humphrey (1969).

Table 2 includes many of the cases of alexia without agraphia reported in the literature published in English between 1948 and 1976. All the tabulated cases are bad at reading words, and most of them are relatively better at reading Arabic numerals than letters. They can be further differentiated on the basis of several features, among which four have been selected for the present discussion. These features are (a) presence or absence of a right hemianopsia, (b) color naming performance, (c) spelling performance, and (d) mental calculation performance. Spelling is evaluated either by the ability of a patient to spell and to recognize orally spelled words or by his use of a spelling strategy in attempting to read word. A good mental calculation performance indicates that very simple arithmetic operations can be performed. Here follows the description of the groups.

Table 2

Tentative classification of cases of alexia without agraphia reported between 1948 and 1976[a]

Authors	Patients	Hemi-anopsia	Spelling	Mental calcul.	Color naming	Word naming	Letter naming	Single digit naming
				Group 1: Close to ideal cases with minimal additional deficits				
Stengel et al., 1948	1	yes	good	good	bad	bad	medium	good
	2	yes	good	good	?	bad	good	good
Holmes, 1950		yes	good	good	bad	bad	bad	good
Kriendler & Ionăşescu, 1961		yes	?	good	bad	bad	medium	good
Geschwind & Fusillo, 1966		yes	good	good	bad	bad	bad	good
Oxbury et al., 1969	1	yes	?	?	bad	bad	medium	good
Cumming et al., 1970		yes	good	good	good	bad	bad	good
Benson et al., 1971	1	yes	good	good	bad	bad	medium	good
	2	yes	good	?	bad	bad	medium	?
	3	yes	good	good	bad	bad	good	?
Sasanuma, 1974b		yes	good[b]	good	medium	bad	--	medium
				Group 2: Close to ideal cases with potential additional deficits				
Ajax, 1967	1	transient	good	good	good	medium	good	good
Goldstein et al., 1971		no	?	good	good	bad	bad	medium
Heilman et al., 1971	1	no	good	medium	bad	bad	bad	?
Greenblatt, 1973		no	?	good	good	bad	good	good
				Group 3: Nonideal cases with attested additional lesions				
Warrington & Zangwill, 1957		yes	good	bad	bad	bad	medium	good
Ajax, 1964	1	yes	good	good	good	medium	good	good
Kinsbourne & Warrington, 1964		no	bad	?	bad	bad	good	good
Caplan & Hedley-Whyte, 1974		yes	good	bad	bad	bad	bad	bad
L. N. Cohen et al., 1976		yes	?	bad	medium	bad	good	?
				Group 4: Hécaen and Kremin's (1976) presumably nonideal cases				
		[c]	[c]	[c]	[c]	[d]	[d]	[d]
Hécaen & Kremin, 1976	TRO	yes	good	bad	good	[illegible]	[illegible]	100%
	DEL	yes	good	bad	medium	[illegible]	[illegible]	100%
	SAL	yes	good	medium	good	12%	70%	[illegible]
	CLI[e]	yes	medium	bad	medium	0%	[illegible]	[illegible]
	BLA[f]	Quadr. sup.	?	medium	good	[illegible]	[illegible]	[illegible]
	MAG[f]	Quadr. sup.	good	good	good	[illegible]	[illegible]	[illegible]

---

[a] An ideal case is characterized by a) a destruction of the left visual cortex or of the connections between the left visual cortex and the left angular gyrus and b) a destruction of the splenium of the corpus callosum. [b] Spelling of Kana characters. [c] From Table [illegible] [d] Percent correct from Figures 4 to 9 (pp. [illegible]-308). [e] Mild agraphia. [f] [illegible]

Group 1: These patients are close to the ideal model in showing an anatomically verified (Cumming, Hurwitz, & Perl, 1970; Geschwind & Fusillo, 1966) or presumed brain infarction (Benson, Brown, & Tomlinson, 1971, Cases 1 to 3; Holmes, 1950; Kreindler & Ionășescu, 1961; Oxbury et al., 1969, Case 1; Sasanuma 1974b; Stengel et al., 1948). The infarction is of the left occipital lobe (responsible for the hemianopsia) and of the splenium, both caused by some pathology of the left posterior cerebral artery. The 11 patients show a very consistent pattern of results. They all exhibit the right hemianopsia expected from their lesion in the left visual cortex. When investigated, spelling and mental calculation are always good and color naming is always bad except in one case (Cumming et al., 1970). Even the last case does not cast doubt on the homogeneity of this group of patients because the dorsal part of the splenium was preserved in this patient; this could allow for a transfer of the visual color information from the right occipital lobe to the left angular gyrus (see Greenblatt, 1973, for further discussion of this point). It is clear that the case of Sasanuma (1974b) reviewed in the preceding subsection fits perfectly well in this group of patients. could be tentatively concluded that the right hemisphere of these patients has a much better ability to identify Arabic numerals than letters or phonographically written words.

Group 2: The four cases included in this group are remarkable for their lack of right hemianopsia, indicating an intact left visual cortex. This can be related to the etiology of their trouble, which is different from that of patients in Group 1. The cause of the alexia was either a surgical removal of a vascular anomaly (Ajax, 1967, Case 1), a carbon monoxide intoxication (Goldstein, Joynt, & Goldblatt, 1971), a head trauma (Heilman, Safran, & Geschwind, 1971, Case 1), or a tumor (Greenblatt, 1973). In the last case anatomical analysis of the brain showed that the tumor had destroyed the splenium of the corpus callosum and part, but not all, of the connections between the intact left visual cortex and the left angular gyrus. One can tentatively hypothesize that the connections needed to transmit color information to the left angular gyrus were also preserved in two of the other patients. If this were indeed the case, these four patients, showing good spelling and good mental calculation, could be considered examples of the ideal model of alexia without agraphia as good as those of Group 1. However, it seems unlikely, in view of the etiologies of these alexias, that the brain damage was really so selectively localized. We cannot preclude the possibility that the language areas have been more or less affected as well, rendering those cases potentially less conclusive than those of Group 1 with respect to the assessment of right hemisphere competence in number identification.

Group 3: The patients in this group are certainly the least appropriate for our purposes because their left occipital lesion extended to the parietal lobe as well and because we generally do not know whether the splenium was lesioned or not (Ajax, 1964, Case 1; D. N. Cohen, Salanga, Hully, Steinberg, & Hardy, 1976; Kinsbourne & Warrington, 1964; Warrington & Zangwill, 1957). The patient of Caplan and Hedley-Whyte (1974) had an anatomically verified lesion of the left occipital cortex and of the splenium, but she suffered from additional small left parietal lesions. This can explain why this patient had dyscalculia, finger agnosia, and left-right confusion. The heterogeneity of performance in this group contrasts with the homogeneity of performance of Group 1 patients, which could support the idea either of a pathology different from that of alexia without agraphia, or, at least, of the existence of supplementary problems compared with the ideal model. Therefore, these data cannot safely be used to infer anything about right hemisphere competence.

Group 4: This group includes the six alexics studied by Hécaen and Kremin (1976). It is immediately apparent that all six patients would fall in our third group, even though three of them are generally considered as ideal cases of alexia without agraphia (CRO, DEL, SAL). CLI, who is slightly agraphic, is generally assimilated to the nonagraphic patients, whereas BLA and MAG are dissociated from them on the basis of their strong agraphia. It is clear that none of these patients presents the profile of those of Group 1 in having good spelling and mental calculation and poor color naming. Within the group of four patients with a right hemianopsia, only SAL is attested as suffering from a lesion of vascular origin sufficiently selective to affect only the occipital lobe. Although his performances depart from these of patients in group 1, he is the only one whose inclusion in this group might be defended.

In summary, we can probably safely rely on the patients of Group 1 in attempting to assess the right-hemisphere ability to process visual symbols semantically. Some other cases are probably valid as well, but we have enough patients in Group 1 to adopt a conservative position, excluding all others from further discussion.

Section of the splenium of the corpus callosum. The logic of the interpretation of alexia without agraphia implies that, under LVF tachistoscopic presentations, a patient whose only lesion is a section of the splenium of the corpus callosum should exhibit exactly the same reading performance as an alexic without agraphia. To our knowledge, only six such cases have been reported, three of them being examined at a time at which the hemifield presentation technique was not well developed. All six cases had their splenium severed in the process of removing a subcortical small tumor. The patient of Trescher and Ford (1937) could not recognize letters presented in the LVF (for what duration?), but other symptoms, such as a left hand astereognosis, did not guarantee that the splenium section was the only damage suffered by the patient. The two patients studied by Maspes (1948) did not present any hand astereognosis (for wooden letters). Letters and Arabic numerals presented for 1 or 0.5 sec were very well recognized in the RVF, but not in the LVF, as expected. Three similar Japanese cases have been reported in the recent literature (Sugishita, Iwata, Toyokura, Yoshioka, & Yamada 1978). With RVF brief presentations (66 ms), oral reading and comprehension of both Kana and Kanji words were almost perfect. With LVF presentations, performance was poorer; being at chance level for Kana, but somewhat better than chance for Kanji. Moreover, performance improved relatively more with Kanji than with Kana when the same material was retested two or three times at intervals of several months. Unfortunately, numbers were not tested.

Sugishita et al. (1978) interpreted their results as showing that the right hemisphere can understand the logographic Kanji better than the phonographic Kana; this is consonant with the better recognition of Arabic numerals than of letters or alphabetically written words by alexics without agraphia. It is unfortunate that Maspes (1948) did not systematically investigate the difference in performance for letters and Arabic numerals. Sugishita et al. also assumed that the vocal response was given by the left hemisphere, not by the right one. This implies that, however accurately identified the LVF stimuli were, naming could not have been achieved at all if the corpus callosum were completely sectioned. This brings us to the next stage in our review.

Commissurotomy. The initial investigation of patients having undergone a complete section of the interhemispheric commissures revealed the very poor language competence of the right hemisphere in most cases (see Gazzaniga, 1970). However, two patients of the California series (L.B. and N.G.) and three patients of the East Coast series (P.S., J.W., and V.P.) show a considerable right hemisphere language comprehension (both spoken and written). In addition, within two years following commissurotomy, P.S. and V.P. have developed the ability to access speech from the right hemisphere (see Gazzaniga, 1983). All these patients have a complete section of the corpus callosum and the hippocampal commissure. Most patients of the California series, including L.B. and N.G., also have a section of the anterior commissure, whereas most patients of the East Coast series, including P.S., J.W., and V.P., do not.

Recognition performance for letters and Arabic numerals was investigated in six patients of the California series by Teng and Sperry (1973). In one condition, pairs of letters or of Arabic numerals were presented either in the LVF or in the RVF, calling for a verbal report. With RVF presentations 86% of the letters and 80% of the digits were named correctly, whereas with LVF presentations these scores dropped to 13% and 35%, respectively. Notice that one patient, N.W., made 100% errors with LVF presentations of both letters and numerals and that L.B. reported only 22% of letters, but 80% of Arabic numerals from the LVF. In another experiment involving fewer numerals (Gazzaniga & Hillyard, 1971), L.B. described a strategy of enumerating the numbers and stopping when the response popped out, which the authors found compatible with the idea that the response was actually generated in the left hemisphere through cross-cuing with the right hemisphere. This strategy should be easier to use with single-digit numbers than with letters because the set is smaller in the former than in the latter case. Whether L.B. or the other often tested patients used in Teng and Sperry's (1973) experiment were using a similar strategy is not known, but it cannot be ruled out. Hence, these data are not strong enough either to challenge the hypothesis of the muteness of the right hemisphere, or to provide unequivocal evidence of a greater intrinsic ability of this hemisphere to deal with Arabic numerals than with letters.

Gazzaniga and Smylie (1984) tested two of the right-hemisphere language-proficient patients of the East Coast series, V.P. and J.W. Both patients showed errorless performance in multiple choice pointing to numbers presented to the right hemisphere (LVF). V.P. was also able to read these numbers aloud perfectly well, whereas J.W. was completely unable to do so. Both patients showed extremely poor performance in carrying out simple arithmetic operations with the right hemisphere.

The data of Gazzaniga and Smylie (1984) are compatible with the idea that the left hemisphere normally subserves calculation (see preceding subsection) and that the right hemisphere can identify numbers. However, multidigit number identification is probably better in these two patients than in most patients showing the ideal symptomatology of alexia without agraphia. The extent to which these data can be generalized to the entire population of commissurotomized patients, and, a fortiori, to normal people, is debatable (see Gazzaniga, 1983, and Zaidel, 1983, for somewhat opposite views on this question).

Hemispherectomy. The last logical step in this story is to assess the number processing ability of a completely isolated right hemisphere, the left hemisphere having been removed completely (actually the left cortex, the left subcortical structures being almost entirely preserved).

Gott (1973) described such a patient who underwent a left hemispherectomy at the age of 10 years, because of malignancy. She had already undergone brain surgery at the age of eight for removal of a tumor in the left ventricule. When she was tested two years after the hemispherectomy, she showed good comprehension of spoken language but very poor verbal expression (mainly single words or short stereotyped sentences) and very poor reading of single words. She was unable to name a single letter presented visually or to choose above chance (30% correct) which of four visually displayed letters was the one just spoken by the experimenter, but she performed much better in this task when Arabic numerals were used instead of letters (80% correct). When Zaidel (1976) tested her one year later, using a similar procedure, she was a little better in pointing to a spoken multidigit number (out of six) than to a spoken letter.

The description given by Hillier (1954) of the performance of his patient is more anecdotal. After three surgical interventions in the left hemisphere during the preceding 15 months, a complete hemispherectomy was finally performed. The patient was 14 years old at the time of the first intervention. Each intervention left him with severe aphasic troubles, indicating that language functions were subserved by his left hemisphere. However, after hemispherectomy, he was described as having good comprehension of spoken language and an ability to say some words and to read single letters.

However poor the verbal performance of these two patients may appear, it is nevertheless much better than would have been expected if the right hemisphere were completely unable to subserve any linguistic function. Due to the youth of the patients, no generalization of this conclusion is allowed because the plasticity of the nervous system is probably still important at that age. This plasticity is now well documented in patients who have undergone hemidecortication because of infantile hemiplegia, accompanied by intractable seizures. It is clear that if the illness starts before the age of one year the healthy hemisphere, whether right or left, subserves all the functions normally shared between two hemispheres (McFie, 1961). In these cases it requires subtle testing with tasks varying in complexity to show that patients retaining their left hemispheres are relatively better at complex syntax comprehension than those retaining their right hemispheres (Dennis, 1980a; Dennis & Kohn, 1975), whereas the opposite relation between relative levels of performance holds for complex spatial tasks (Kohn & Dennis, 1974). Hence, behind the tremendous plasticity showed by each hemisphere in developing functions for which it is usually less proficient, there seems to be an irreducible difference in processing ability as well.

At the other extreme, two adult left hemispherectomies, performed to remove tumors developed during adulthood, reveal extremely poor verbal ability, but not its complete lack (McFie, 1961). Between the age of one year and some unknown upper limit, the brain seems to keep some of its initial plasticity, allowing each hemisphere to develop abilities for which it is normally not very proficient (McFie, 1961). The two patients just described (Gott, 1973; Hillier, 1954) were probably still in this phase.

A thorough examination of the linguistic abilities of the left and right hemispheres has been undertaken by Dennis and her colleagues (Dennis, 1980b; Dennis, Lovett, & Wiegel-Crump, 1981; Dennis & Whitaker, 1976) on one case of right hemidecortication and two cases of left hemidecortication, performed before the age of five months. The examinations published to date took place when the children were between 9 and 14 years old. One fascinating finding of these studies is that equal performances in decoding written words can be mediated by different mental representations. The child retaining his left hemisphere shows a good awareness of the phonological structure of language: His reading draws on morphophonological properties of English orthography, and he reveals a tacit knowledge of rules that map writing onto speech when he reads new or unfamiliar words. None of these abilities is displayed by the two children retaining their right hemispheres. Yet with known words, their reading performance is equivalent to that of the child retaining his left hemisphere. Only with unknown words does their performance disintegrate; this shows that their word knowledge is not based on a morphophonological representation, so that they cannot exploit English orthographical principles to decode new words. These findings are remarkably well in line with the ideas developed by Mattingly (1972, 1984) concerning the relation between proficient reading and the availability of morphophonological representations of words in the mental lexicon. This author has also stressed that spoken language comprehension is probably less dependent on the existence of such representations than is reading (Mattingly, 1984). This claim is supported by the failure of Dennis and Whitaker (1976) to demonstrate differences in the abilities of left and right hemidecorticate children in their ability to deal with the phonemic and semantic aspects of spoken language. We may also note that a capacity for syntactic processing was much greater in the child retaining his left hemisphere than in the other two children. This is consonant with other data mentioned earlier (Dennis, 1980a; Dennis & Kohn, 1975).

Conclusions

The most important point of this section is the contrast between the disintegration of calculation and Arabic numeral reading caused by lesions in the language areas of the left hemisphere and the relative preservation of these abilities by patients showing a disconnection between intact language areas and the visual cortex.

The only point left for discussion is the interpretation of the better identification of Arabic numerals than of letters by the 11 alexics without agraphia of Group 1, those for whom the presumption of a pure disconnection syndrome is most likely to be correct. All these patients had their language functions located in the left hemisphere, they did not display any known cerebral brain disorder before their alexia; and the syndrome was caused by brain lesions in adulthood. Hence, these patients are the best suited for the assessment of the ability of the right hemisphere to deal with visual symbols.

A first possible explanation of the better performance with Arabic numerals than with letters is that it is simply easier to discriminate one visual symbol out of 10 possible visual configurations than one out of 26 possibilities. We find this extremely unlikely. Both sets of symbols have evolved from the need to allow efficient reading. They incontestably succeed in doing so, especially under the temporally unlimited viewing conditions typical of the neuropsychological examination.

Far more likely is that performance is determined by the extent to which the right hemisphere can process the meaning of different stimuli. In essence, this amounts to proposing exactly the same schema of interpretation as that used by Geschwind (1965) to explain the differential ability to name objects and colors. Remember that the two basic assumptions are that (a) only the left hemisphere can generate a naming response, and (b) although visual information reaching the right hemisphere cannot be transmitted to the left hemisphere because the splenial route is sectioned, the information can be transfered, once it has been given a semantic interpretation, through the anterior intact portion of the corpus callosum. Hence, the solution of the problem should be sought by analysing the nature of the semantic information conveyed by letters and Arabic numerals.

The meaning of a letter is determined by the phonological unit of the spoken language to which it refers and by its relation with other similar units. In other words, the meaning of a letter is defined in terms of properties that the right hemisphere is unable to process, even when it has developed an idiosyncratic language competence, due to complete loss of the left hemisphere (Dennis et al., 1981). A fortiori, a right hemisphere that has never faced the problem of associating sounds to letters should be even less able to extract their meaning. This entails that, beyond the untransmittable visual information, there is simply no other form of information that can be conveyed to the left hemisphere. In this vein, the very poor performance of alexics without agraphia in reading and understanding phonographically written words argues for the hypothesis that word recognition is mediated by letter or syllable (Kana) recognition.

By contrast, Arabic numerals have a meaning in a symbolic system that has nothing to do with phonology. There is therefore no reason why the right hemisphere could not generate a semantic representation of the digit and transfer it to the other side, a task it seems able to perform with objects as well. This explanation is consonant with the better ability of the right hemisphere to interpret Kanji logograms than Kana phonograms (Sasanuma, 1974b; Sugishita et al., 1978).

In concluding this section it is worth specifying the exact scope of the interpretation of the right hemisphere's better performance with Arabic numerals than with letters. Arabic-numeral reading in alexia without agraphia is not always perfect, and this points to the fact that, though feasible, the task is nevertheless strained. The inefficiency of the procedure is also demonstrated by the fact that the reading of multidigit numbers is rarely preserved. This would not be the case if transmission of the component numerals to the left hemisphere were more efficient. At present, we do not know whether the poor naming performance is caused by inadequacy of the semantic representation generated in the right hemisphere, or by the poor ability of the corpus callosum in transmitting interpreted rather than raw sensory information, or both. A final point worth emphasizing is that, while we may infer from the performance of brain damaged patients that the right hemisphere has some ability to process Arabic numerals, we may not infer that it is superior to the left hemisphere in doing so.

Summary and Conclusions

Let us take the different points in the reverse order to their presentation in the chapter.

1. With respect to brain injured patients, the fact that (a) number processing is strongly, but perhaps not fully, dependent on the integrity of the language areas of the left hemisphere and (b) these areas can be disconnected from the visual information reaching the right hemisphere, allows us to assess the differential ability of this hemisphere to deal with various surface forms of numbers and of other types of visual information. Hemifield presentation of stimuli is a useful technique in this framework. An understanding of how the information is processed will require both the general progress of the analytical power of cognitive psychology as a whole and the comprehension of the basic modes of processing of each hemisphere in particular.

2. As for lateral hemifield presentation of stimuli to normal subjects, we may doubt whether the technique will help us to achieve either or both of the requirements just mentioned, at least insofar as one adopts a multicomponential view of processing. The analysis of the problem presented at the end of the third section is, of course, not the only one possible. However, it is based on the simplest and most tractable view of processing we have, and this casts serious doubt on the ability of the approach to fare better in more complex theoretical frameworks. The results show a RVF advantage for both logographic and phonographic number representations. Whether this pattern of results should be considered at odds with claims for a LVF advantage in the processing of logograms in general depends on the validity of this assertion, which is still controversial.

3. As regards numerical size comparison judgments, two of the basic effects--symbolic distance and serial position--were found to be independent of the surface form of the stimuli in experiments published to date. By inference from related data we hypothesized that such will also prove to be the case for a third effect--semantic congruity--for which the information is still lacking. The Stroop-like task leading to a size congruity effect has been judged too complicated to provide useful, nonparadigm-bound information, a conclusion that extends to hemifield presentations for the reasons just invoked. The strategy of research illustrated in this approach could, of course, be extended to cover a variety of questions about the basic knowledge associated with single-digit numbers. One can, for instance, use the same paradigm in comparison judgments related to the odd vs. even, prime vs. nonprime, multiple of two vs. nonmultiple of two questions. The task need not be an explicit comparison between two numbers; it can also take the form of judging whether a single number possesses the property under investigation.

4. Three points should be made about the discussion of number representations.

First, concerning multidigit number processing, it seems appropriate to distinguish between Arabic numerals and number names irrespective of the surface form of the latter. The facts that Arabic numerals belong to a different, more abstract notational system and that they are also intimately bound to mathematical activities make them a priori distinct from number names. The irrelevance of the surface form can be further emphasized by predicting that Japanese (or Chinese) aphasics would show the same difficulty in transcoding multidigit numbers written in one logographic form into another logographic form (Arabic numerals into Kanji words and vice versa) as occidental aphasics have in transcoding these same numbers represented logographically (Arabic numerals) into alphabetically written number names, and vice versa (Deloche & Seron, 1982; Seron & Deloche, 1984).

Second, as soon as one focuses on the processing of single-digit numbers, the characteristics of the notational system of which the number representations are the elements ceases to play a prominent role, whereas the nature of the surface form of the numbers now becomes the important variable. We should expect that access to the stored knowledge associated with these elements would be influenced by factors affecting the reading of any kind of word. From this point of view it is, therefore, appropriate to regroup the symbols into a logographic and a phonographic category, irrespective of the underlying notational system. With normal subjects, we expect the surface form of the number to affect the speed with which their conceptual knowledge is accessed, but we expect the characteristics of this knowledge, as revealed by the pattern of interactions between different variables, to be the same irrespective of the surface form of the numbers. As far as the available evidence goes, this belief is not yet contradicted (cf. conclusion 3).

Third, single-digit number processing by adult brain-injured people could lead to a more complex picture. One should consider two cases. If, on the one hand, the language areas of the left hemisphere are intact, but disconnected from the visual cortex (ideal cases of alexia without agraphia or LVF presentations with section of the splenium of the corpus callosum), the right intact hemisphere could translate the visual numbers into interpreted representations transmissible to the left hemisphere, provided the numbers are represented logographically (one should also allow for the possibility that the right hemisphere may learn to process the small set of phonographic numbers as if they were logograms). In this case, a task could be performed according to the normal synergic activity of the hemispheres of an intact brain, leading to a performance qualitatively equivalent to that of normal subjects, save for some eventual loss in efficiency. If, on the other hand, the language areas of the left hemisphere are injured and if the task can be performed at all, then performance should be at least partially determined by right-hemisphere competence in dealing with numbers. A performance qualitatively different from that of normal subjects could then be considered an index of the idiosyncrasy of the right hemisphere's knowledge of numbers. The nature of the right hemisphere competence could then be studied in commissurotomized patients, provided the cognitive capacities of the right hemisphere of these severely epileptic people could be considered representative of those of normal subjects.

References

Ajax, E. T. (1964). Acquired dyslexia. Archives of Neurology, 11, 66-72.

Ajax, E. T. (1967). Dyslexia without agraphia. Archives of Neurology, 17, 645-652.

Aiken, L. R., & Williams, F. N. (1968). Three variables related to reaction time to compare single-digit numbers. Perceptual and Motor Skills, 27, 197-206.

Allen, M. (1983). Models of hemispheric specialization. Psychological Bulletin, 93, 73-104.

Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgments. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 11, pp. 101-159). New York: Academic Press.

Banks, W. P., Clark, H. H., & Lucy, P. (1975). The locus of the semantic congruity effect in comparative judgments. Journal of Experimental Psychology: Human Perception and Performance, 1, 35-47.

Banks, W. P., & Flora, J. (1977). Semantic and perceptual processes in symbolic comparisons. Journal of Experimental Psychology: Human Perception and Performance, 3, 278-290.

Banks, W. P., Fujii, M., & Kayra-Stuart, F. (1976). Semantic congruity effects in comparative judgments of magnitudes of digits. Journal of Experimental Psychology: Human Perception and Performance, 2, 435-447.

Benson, D. F., Brown, J., & Tomlinson, E. B. (1971). Varieties of alexia. Neurology, 21, 951-957.

Bertelson, P. (1982). Lateral differences in normal man and lateralization of brain functions. International Journal of Psychology, 17, 173-210.

Besner, D., & Coltheart, M. (1976). Mental size scaling examined. Memory & Cognition, 4, 525-531.

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. Neuropsychologia, 17, 467-472.

Besner, D., Daniels, S., & Slade, C. (1982). Ideogram reading and right hemisphere language. British Journal of Psychology, 73, 21-28.

Besner, D., Davelaar, E., Alcott, D., & Parry, P. (1984). Holistic reading of alphabetic print: Evidence from the FDM and FBI. In L. Henderson (Ed.), Orthographies and reading (pp. 121-135). London: Erlbaum.

Besner, D., Grimsell, D., & Davis, R. (1979). The mind's eye and the comparative judgment of number. Neuropsychologia, 17, 373-380.

Bryden, M. P. (1978). Strategy effects in the assessment of hemispheric asymmetry. In G. Underwood (Ed.), Strategies of information processing (pp. 117-149). London: Academic Press.

Bryden, M. P. (1982). Laterality: Functional asymmetry in the intact brain. New York: Academic Press.

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. Journal of Experimental Psychology, 103, 1131-1136.

Caplan, L. R., & Hedley-Whyte, T. (1974). Cuing and memory dysfunction in alexia without agraphia: A case report. Brain, 97, 251-262.

Caramazza, A. (1984). The neuropsychological research and the problem of patient classification in aphasia. Brain and Language, 21, 9-20.

Carmon, A., Nachshon, I., & Starinsky, R. (1976). Developmental aspects of visual hemifield differences in perception of verbal material. Brain and Language, 3, 463-469.

Cohen, D. N., Salanga, V. D., Hully, W., Steinberg, M. C., & Hardy, R. W. (1976). Alexia without agraphia. Neurology, 26, 455-459.

Cohen, G. (1975). Hemisphere differences in the effects of cuing in visual recognition tasks. Journal of Experimental Psychology: Human Perception and Performance, 4, 366-373.

Cohen, G. (1982). Theoretical interpretations of lateral asymmetries. In J. G. Beaumont (Ed.), Divided visual field studies of cerebral organisation (pp. 87-111). New York: Academic Press.

Coltheart, M. (1980). Deep dyslexia: a right hemisphere hypothesis. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), Deep dyslexia (pp. 326-380). London: Routledge and Kegan Paul.

Coltheart, M. (1983). The right hemisphere and disorders of reading. In A. Young (Ed.), Functions of the right cerebral hemisphere (pp. 171-201). London: Academic Press.

Craft, J. L., & Simon, J. R. (1970). Processing symbolic information from a visual display. Interference from an irrelevant directional cue. Journal of Experimental Psychology, 83, 415-420.

Cumming, W. J. K., Hurwitz, L. J., & Perl, N. T. (1970). A study of a patient who had alexia without agraphia. Journal of Neurology, Neurosurgery, and Psychiatry, 33, 34-39.

Dahmen, W., Hartje, W., Büssing, A., & Sturm, W. (1982). Disorders of calculation in aphasic patients--Spatial and verbal components. Neuropsychologia, 20, 145-153.

Deloche, G., & Seron, X. (1982). From one to 1: An analysis of a transcoding process by means of neuropsychological data. Cognition, 12, 119-149.

Dennis, M. (1980a). Capacity and strategy for syntactic comprehension after left or right hemidecortication. Brain and Language, 10, 287-317.

Dennis, M. (1980b). Language acquisition in a single hemisphere: Semantic organization. In D. Caplan (Ed.), Biological studies of mental processes (pp. 159-185). Cambridge, MA: The MIT Press.

Dennis, M., & Kohn, B. (1975). Comprehension of syntax in infantile hemiplegics after cerebral hemidecortication: Left-hemisphere superiority. Brain and Language, 2, 472-482.

Dennis, M., Lovett, M., & Wiegel-Crump, C. A. (1981). Written language acquisition after left or right hemidecortication in infancy. Brain and Language, 12, 54-91.

Dennis, M., & Whitaker, H. A. (1976). Language acquisition following hemidecortication: Linguistic superiority of the left over the right hemisphere. Brain and Language, 3, 404-433.

Dixon, P., & Just, M. A. (1978). Normalization of irrelevant dimensions in stimulus comparisons. Journal of Experimental Psychology: Human Perception and Performance, 4, 36-46.

Duncan, E. M., & McFarland, C. E., Jr. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments: An additive-factors analysis. Memory and Cognition, 8, 612-622.

Endo, M., Shimizu, A., & Hori, T. (1978). Functional asymmetry of visual fields for Japanese words in Kana (syllable-based) writing and randon shape-recognition in Japanese subjects. Neuropsychologia, 16, 291-297.

Ferro, J. M., & Botelho, M. A. S. (1980). Alexia for arithmetical signs: A cause of disturbed calculation. Cortex, 16, 175-180.

Flegg, G. (1983). Numbers: Their history and meaning. New York: Schocken Books.

Foltz, G. S., Poltrock, S. E., & Potts, G. R. (1984). Mental comparison of size and magnitude: Size congruity effects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 442-453.

Gazzaniga, M. S. (1970). The bisected brain. New York: Appleton-Century-Crofts.

Gazzaniga, M. S. (1983). Right hemisphere language following brain bisection. American Psychologist, 38, 525-537.

Gazzaniga, M. S., & Hillyard, S. A. (1971). Language and speech capacity of the right hemisphere. Neuropsychologia, 9, 273-280.

Gazzaniga, M. S., & Smylie, C. S. (1984). Dissociation of language and cognition. Brain, 107, 145-153.

Geffen, G., Bradshaw, J. L., & Nettleton, N. C. (1973). Attention and hemispheric differences in reaction time during simultaneous audio-visual tasks. Quarterly Journal of Experimental Psychology, 25, 404-412.

Geffen, G., Bradshaw, J. L., & Wallace, G. (1971). Interhemispheric effects on reaction time to verbal and nonverbal visual stimuli. Journal of Experimental Psychology, 87, 415-422.

Gelb, I. J. (1963). A study of writing. Chicago: University of Chicago Press.

Geschwind, N. (1965). Disconnexion syndromes in animal and man: Part I. Brain, 85, 237-294.

Geschwind, N., & Fusillo, M. (1966). Color-naming defects in association with alexia. Archives of Neurology, 15, 137-146.

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: The proceedings of the CUNY conference (pp. 1-53). Hillsdale, NJ: Lawrence Erlbaum Associates.

Goldstein, M. N., Joynt, R. J., & Goldblatt, D. (1971). Word blindness with intact central visual field. Neurology, 21, 873-876.

Gordon, H. W., & Carmon, A. (1976). Transfer of dominance in speed of verbal response to visually presented stimuli from right to left hemisphere. Perceptual and Motor Skills, 42, 1091-1100.

Gott, P. S. (1973). Language after dominant hemispherectomy. Journal of Neurology, Neurosurgery, and Psychiatry, 36, 1082-1088.

Greenblatt, S. H. (1973). Alexia without agraphia or hemianopsia. Brain, 96, 307-316.

Hatta, T. (1977a). Lateral recognition of abstract and concrete Kanji in Japanese. Perceptual and Motor Skills, 45, 731-734.

Hatta, T. (1977b). Recognition of Japanese Kanji in the left and the right visual fields. Neuropsychologia, 15, 685-688.

Hatta, T. (1978). Recognition of Japanese Kanji and Hirakana in the left and the right visual fields. The Japanese Psychological Research, 20, 51-59.

Hatta, T. (1983). Visual field differences in semantic comparative judgments with digits and Kanji stimulus materials. Neuropsychologia, 21, 669-678.

Hatta, T., & Dimond, S. J. (1980). Comparison of lateral differences for digit and random form recognition in Japanese and Westerners. Journal of Experimental Psychology: Human Perception and Performance, 6, 368-374.

Hécaen, H., Angelergues, R., & Houillier, S. (1961). Les variétés cliniques des acalculies au cours des lésions rétrorolandiques: Approche statistique du problème. Revue Neurologique, 105, 86-103.

Hécaen, H., & Kremin, H. (1976). Neurolinguistic research on reading disorders resulting from left hemisphere lesions: Aphasic and "pure" alexias. In H. Whitaker & H. A. Whitaker (Eds.), Studies in neurolinguistics (Vol. 2, pp. 269-329). New York: Academic Press.

Heilman, K. M., Safran, A. & Geschwind, N. (1971). Closed head trauma and aphasia. Journal of Neurology, Neurosurgery, and Psychiatry, 34, 265-259.

Henderson, L. (Ed.). (1984). Orthographies and reading. Hillsdale, NJ: Lawrence Erlbaum Associates.

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. Memory & Cognition, 10, 389-395.

Hillier, W. F. (1954). Total left cerebral hemispherectomy for malignant glioma. Neurology, 4, 718-721

Hines, D., & Satz, P. (1971). Superiority of right visual half-fields in right-handers for recall of digits presented at varying rates. Neuropsychologia, 9, 21-25.

Hines, D., & Satz, P. (1974). Cross-modal asymmetries in perception related to asymmetry in cerebral functions. Neuropsychologia, 12, 239-247.

Hines, D., Satz, P., Schell, B., & Schmidlin, S. (1969). Differential recall of digits in the left and right visual half-fields under free and fixed order of recall. Neuropsychologia, 7, 13-22.

Hirata, K., & Osaka, R. (1967). Tachistoscopic recognition of Japanese letter materials in left and right visual fields. Psychologia, 10, 7-18.

Holender, D., & Peereman, R. (in preparation). Is there a right hemispheric superiority specific to logographic processing?

Holmes, G. (1950). Pure word blindness. Folia Psychiatrica, Neurologica et Neurochirurgica Neerlandica, 53, 279-288.

Katz, A. N. (1980). Cognitive arithmetic: Evidence for right hemispheric mediation in an elementary component stage. Quarterly Journal of Experimental Psychology, 32, 69-84.

Katz, A. N. (1981). Spatial compatibility effects with hemifield presentation in a unimanual two-finger task. Canadian Journal of Psychology, 35, 63-68.

Kelley, D. H. (1976). Deciphering the Maya script. Austin, TX: University of Texas Press.

Kinsbourne, M., & Warrington, E. K. (1964). Observation on colour agnosia. Journal of Neurology, Neurosurgery, and Psychiatry, 27, 296-299.

Kohn, B., & Dennis, M. (1974). Selective impairments of visuo-spatial abilities in infantile hemiplegics after right cerebral hemidecortication. Neuropsychologia, 12, 505-512.

Krantz, D. H., & Tversky, A. (1975). Similarity of rectangles: An analysis of subjective dimensions. Journal of Mathematical Psychology, 12, 4-34.

Kreindler, A., & Ionășescu, V. (1961). A case of 'pure' word blindness. Journal of Neurology, Neurosurgery, and Psychiatry, 24, 275-280.

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 (pp. 137-153). Baltimore, MD: University Park Press.

Maspes, P. E. (1948). Le syndrome expérimental chez l'homme de la section du splénium du corps calleux: Alexie visuelle pure hémianopsique. Revue Neurologique, 80, 100-113.

Mattingly, I. G. (1972). Reading, the linguistic process and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), Reading by ear and by eye (pp. 133-147). 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 (pp. 9-25). New York: Springer-Verlag.

McFie, J. (1961). The effects of hemispherectomy on intellectual functioning in cases of infantile hemiplegia. Journal of Neurology, Neurosurgery, and Psychiatry, 24, 240-249.

Menninger, K. (1969). Number words and number symbols: A cultural history of numbers. Cambridge, MA: MIT Press.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. Nature, 215, 1519-1520.

Moyer, R. S., & Dumais, S. T. (1978). Mental comparison. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 12, pp. 117-155). New York: Academic Press.

Oxbury, J. M., Oxbury, S. M., & Humphrey, N. K. (1969). Varieties of colour anomia. Brain, 92, 847-860.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. Memory & Cognition, 3, 635-647.

Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. Journal of Experimental Psychology, 91, 191-205.

Peereman, R. Size congruity effects for alphabetically and ideographically written numbers. Submitted for publication.

Peereman, R., & Holender, D. (1984). Relation entre taille physique et taille numérique dans la comparaison de chiffres écrits alphabétiquement et idéographiquement. *Psychologica Belgica*, *24*, 147-264.

Peereman, R., & Holender, D. (1985). Visual-field differences for a number-nonnumber classification of alphabetic and ideographic stimuli. *Quarterly Journal of Experimental Psychology*, to appear.

Rozin, P., & Gleitman, L. R. (1977). The structure and acquisition of reading II: The reading process and the acquisition of the alphabetic principle. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading: The proceedings of the CUNY conference* (pp. 55-141). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sasanuma, S. (1974a). Impairment of written language in Japanese aphasics: Kana versus Kanji processing. *Journal of Chinese Linguistics*, *2*, 141-158.

Sasanuma, S. (1974b). Kanji versus Kana processing in alexia with transient agraphia: A case report. *Cortex*, *10*, 89-97.

Sasanuma, S. (1975). Kana and Kanji processing in Japanese aphasics. *Brain and Language*, *2*, 369-383.

Sasanuma, S., Itoh, M., Mori, K., & Kobayashi, Y. (1977). Tachistoscopic recognition of Kana and Kanji words. *Neuropsychologia*, *15*, 547-553.

Sasanuma, S., & Monoi, H. (1975). The syndrome of Gogi (word-meaning) aphasia: Selective impairment of Kanji processing. *Neurology*, *25*, 627-632.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development*, *48*, 630-633.

Sergent, J. (1983a). Role of the input in visual hemispheric asymmetries. *Psychological Bulletin*, *93*, 481-512.

Sergent, J. (1983b). The effect of sensory limitations on hemispheric processing. *Canadian Journal of Psychology*, *37*, 345-366.

Seron, X., & Deloche, G. (1984). From 2 to two: An analysis of a transcoding process by means of neuropsychological evidence. *Journal of Psycholinguistic Research*, *13*, 215-236.

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*, 300-304.

Stengel, E., Vienna, M. D., & Edin, L. R. C. P. (1948). The syndrome of visual alexia with colour agnosia. *Journal of Mental Science*, *94*, 46-58.

Sternberg, S. (1969). The discovery of processing stages: Extension of Donder's method. *Acta Psychologica*, *30*, 276-315.

Sternberg, S. (1984). Stage models of mental processing and the additive-factor method. *Behavioral and Brain Sciences*, *7*, 82-84

Sugishita, M., Iwata, M., Toyokura, Y., Yoshioka, M., & Yamada, R. (1978). Reading of ideograms and phonograms in Japanese patients after partial commissurotomy. *Neuropsychologia*, *16*, 417-426.

Symonds, C. (1953). Aphasia. *Journal of Neurology, Neurosurgery, and Psychiatry*, *16*, 1-6.

Takahashi, A., & Green, D. (1983). Numerical judgments with Kanji and Kana. *Neuropsychologia*, *21*, 259-263.

Teng, E. L., & Sperry, R. W. (1973). Interhemispheric interaction during simultaneous bilateral presentation of letters or digits in commissurotomized patients. *Neuropsychologia*, *11*, 131-140.

Trescher, J. H., & Ford, F. R. (1937). Colloid cyst of the third ventricule. *Archives of Neurology and Psychiatry*, *37*, 959-973.

Warrington, E., & Zangwill, O. L. (1957). A study of dyslexia. *Journal of Neurology, Neurosurgery, and Psychiatry*, *20*, 208-215.

Wender, K. (1971). A test of independence of dimensions in multidimensional scaling. Perception & Psychophysics, 10, 30-32.

Wiener-Ehrlich, W. K. (1978). Dimensional and metric structures in multidimensional stimuli. Perception & Psychophysics, 24, 399-414.

Yamadori, A. (1975). Ideogram reading in alexia. Brain, 98, 231-238.

Zaidel, E. (1976). Auditory vocabulary of the right hemisphere following brain bisection or hemidecortication. Cortex, 12, 191-211.

Zaidel, E. (1983). A response to Gazzaniga: Language in the right hemisphere, convergent perspectives. American Psychologist, 37, 542-546.

A NOTE ON PROCESSING KINEMATIC DATA: SAMPLING, FILTERING, AND DIFFERENTIATION

B. A. Kay,† K. G. Munhall, E. V.-Bateson,†† and J. A. S. Kelso†

I. Introduction

Until quite recently, the predominant dependent measures used in the field of motor behavior have been product scores such as absolute, constant, and variable error, or discrete temporal measures such as reaction time and movement time. Much current work, on the other hand, has emphasized the detailed analysis of the movement itself, particularly kinematic patterns of displacement, velocity, and acceleration over time. In addition, several recent models of voluntary movement control use kinematic measures as their major data base. For example, Nelson (1983) describes a number of constraints that could govern the "economy" or "efficiency" of skilled movements. Optimization criteria such as minimum-force, minimum-impulse, and minimum-energy are offered as possible constraints that can, in principle, be distinguished by subtle differences in the movement's velocity pattern. Similarly, Hogan (1984) has proposed "minimum-jerk" as an organizing principle for voluntary movement, a model that requires careful measurement of the first derivative of acceleration ("jerk") for its evaluation. Finally, several investigators have used kinematic relations (1) to uncover putative invariants of limb (e.g., Jeannerod, 1984; Kelso, Southard, & Goodman, 1979; Soechting & Lacquaniti, 1981) and speech articulator motions (e.g., Tuller & Kelso, 1984) that persist in the face of systematic changes in the task; and (2) as a window into a movement's dynamic control structure (e.g., Kelso, V.-Bateson, Saltzman, & Kay, 1985; Ostry & Munhall, 1985).

From the above sampling of movement research, the desirability of obtaining accurate estimates of kinematic variables is clear, particularly in light of models that demand sensitive measurement procedures for their evaluation. Such considerations have led us to re-evaluate very carefully the manner in which we process movement data--the subject of the present note.

Among the many factors affecting accurate kinematic measures are the presence of noise in the transducer signals, the conversion of analog signals to digital form, and digital differentiation. In this paper we evaluate some techniques used to minimize these effects with the aim of assuring ourselves that the measurements are, in fact, valid. We concentrate mainly on

---

†Also University of Connecticut.
††Also Indiana University.

Acknowledgment. Work on this paper was supported by NINCDS Grant NS-13617 and Contract No. N0014-83-K-0083 from the U. S. Office of Naval Research. K. Munhall was supported by a Natural Science and Engineering Research Council of Canada Postdoctoral Fellowship. The authors wish to thank T. Baer and R. McGowan for very helpful comments on an earlier version of the paper.

optoelectronic movement transduction data because of its common usage in our laboratory and elsewhere. First we outline the processing sequence involved, from movement transduction to final analyses. Then we describe in detail the characteristics of each of these steps, in particular, those required to minimize the effects of noise and those that produce reliable differentiated data. We also provide evaluations of these two key steps. Our intent here is not to provide a comprehensive survey of physiological signal processing (see Kaiser & Reed, 1981; Winter, 1979; Woltring, 1984, for many more details). Rather it is to ensure that our particular measurement and analysis system provides us with accurate and interpretable data so that quantitative statements can be made about the values of, and relations among, kinematic variables. This discussion may also prove useful to others who are beginning to use kinematic data extensively in their experimental investigations.

II. The Signal Processing Sequence

In a typical experiment, articulator displacement data are collected using a position-sensitive optoelectronic transduction system, similar to the commonly-used Selcom SELSPOT system. The output signals provide the bases on which instantaneous velocity and acceleration are derived for more in-depth kinematic analysis. The general sequence of operations used to obtain movement data is as follows: During the experimental session, the analog outputs of the transduction system are fed through a bank of amplifiers and recorded on FM tape. Later, the FM signals are played back, converted to digital form, and transferred to disk. Next, the digitized displacement data are low-pass filtered (smoothed) to reduce high-frequency noise components. The filtered displacement data are then differentiated using a central-difference algorithm. These derived velocity data are low-pass filtered and then differentiated using the central-difference algorithm, and finally, these derived acceleration data are low-pass filtered. Table 1 provides a summary description of the methods we use at each of these steps. The filtered displacement, velocity, and acceleration waveforms (movement trajectories) are subsequently analyzed with our Waveform EditiNg and DisplaY program (WENDY; Szubowicz, 1977) and many other waveform analysis programs.[1]

III. Movement Transduction, Recording, and Analog-to-Digital Conversion

The outputs of our optoelectronic movement transduction system (Weyrich-Tarbox model 400; see Kelso et al., 1985, for a description) are analog DC signals. Since movement trajectory signals contain significant spectral components down to 0 Hz, we use a frequency-modulation (FM) tape deck (SE model 7000) to produce an archival record of the experimental session, rather than an amplitude modulation deck, which would not adequately record such low frequencies. In order to use the full dynamic range of the recorder (signal-to-noise ratio = 48 dB for a voltage range of +/- 4 V), transducer signals are routed through a bank of DC amplifiers before they are recorded.

The analog-to-digital (A/D) step is performed by a Datel model ST-PDP A/D converter, the digitized result being stored on disk. The entire process is controlled by the in-house Physiological Signal Processing (PSP) software system (Gulisano, 1982) on a DEC PDP 11/45. Two parameters are of interest here. First, the resolution of the converter is twelve bits, that is, it produces 2^{12} or 4096 discrete levels ("machine units") over an input voltage range of +/- 10 V. The same bank of DC amplifiers used in the recording step allows us to exploit the full dynamic range of the converter on playback. Twelve bits

Table 1

Steps performed in the signal processing sequence

1) Movement transduction and recording:

-- Transducers typically used:

- optoelectronic system (position-sensitive)
- potentiometers
- POLGON polarized light goniometers

-- Recorder:

- SE 7000 16-track FM tape recorder

2) Playback and A/D conversion:

-- Datel ST-PDP twelve-bit A/D converter and PDP 11/45 with the PSP system

-- Sampling rate = 200 Hz

3) Low-pass filter:

-- 7-point Bartlett (triangular) window

-- sample:	weight:
t-3	0
t-2	1/9
t-1	2/9
t	1/3
t+1	2/9
t+2	1/9
t+3	0

-- response:
- cutoff frequency (1/2 amplitude, -6 dB) = 31 Hz
- phase shift = constant = 0

4) Differentiation routine:

-- Two-point central-difference algorithm:

$$x'(t) = (x(t+1) - x(t-1))/2h$$

(h = sample interval = 5 ms)

-- sample:	weight:
t-1	-1/2h
t+1	+1/2h

-- response:
- within 10% of ideal at 25 Hz
- 3 dB attenuation at 44 Hz
- 6 dB attenuation at 60 Hz
- phase shift = constant = 90 degrees

of resolution corresponds to a maximum signal-to-noise ratio of 4096:1 or 72 dB, greatly exceeding that of the FM signal. However, an additional source of noise enters here, since the resolution of each sample is limited to twelve bits: The difference between the real voltage (theoretically of infinite precision) and its finite-precision digital representation is known as quantization noise, a factor that poses a problem for deriving velocity and acceleration data. Second, the rate at which the signal is sampled must be carefully chosen, since the highest frequency component one can reliably represent in a sampled signal is exactly one-half the sampling rate, the Nyquist frequency. The spectral components of any time-series are unknown a priori, but, according to Woltring (1984), a general guideline for reasonable sampling rates is twice the frequency at which noise becomes the dominant characteristic. Figures 1a and 1b show the amplitude spectra[2] of a typical lip movement sequence (the utterance /əpæpæpə/) and a calibration trial (no movement), respectively, the latter representing the baseline noise present in the system. No divergence from the noise floor can be seen beyond 40 Hz, confirming that our sampling rate of 200 Hz amply satisfies the above criterion.

IV. Digital Filtering

As shown in Figure 1, there is a certain amount of noise in the digitized movement signal, due to quantization and other noise sources in the signal path from transducer to computer. The root mean square (RMS) amplitude of the calibration trial is twelve machine units, indicating that the maximum signal-to-noise ratio is 341:1 (4096:12), or 51 dB, for full range data. This noise affects the accuracy of any measurements made on the displacement data. Furthermore, much of the noise is in the high-frequency range, above any frequencies of interest, and since differentiation acts as a high-pass filter (spectral components are amplified in direct proportion to their frequencies), the dominant characteristic of a doubly-differentiated displacement waveform is usually the noise itself. Velocity data derived from a single differentiation step are also seriously affected by the amplified high frequency noise. It is thus extremely important to minimize the effects of noise while at the same time leaving the "signal" minimally affected. The severity of the high-frequency amplification introduced by differentiation requires that we low-pass filter the raw displacement data when deriving velocity. We use the same filter on the resultant velocity data when deriving acceleration. Since there are still significant high-frequency noise components in the observed acceleration waveforms, we again use the same filter on the derived acceleration data. Figure 2 shows parallel examples of unfiltered (left) and filtered (right) displacement, velocity, and acceleration data, from lower lip movements in speech production. As can be seen, the filter removes only the high-frequency components of the signals.

The filter we use is a seven-point Bartlett (triangular) window (Oppenheim & Schafer, 1975), a smoothing algorithm that is very simple in the time domain (as opposed to filters which use many more points). In addition to simplicity, it has two further desirable properties. First, high-frequency components are greatly reduced while the lower frequencies are relatively unaffected. Second, this window is symmetrical in shape and so introduces no phase shift. The computed frequency response of the filter is shown in Figure 3; see table 1 for further details.

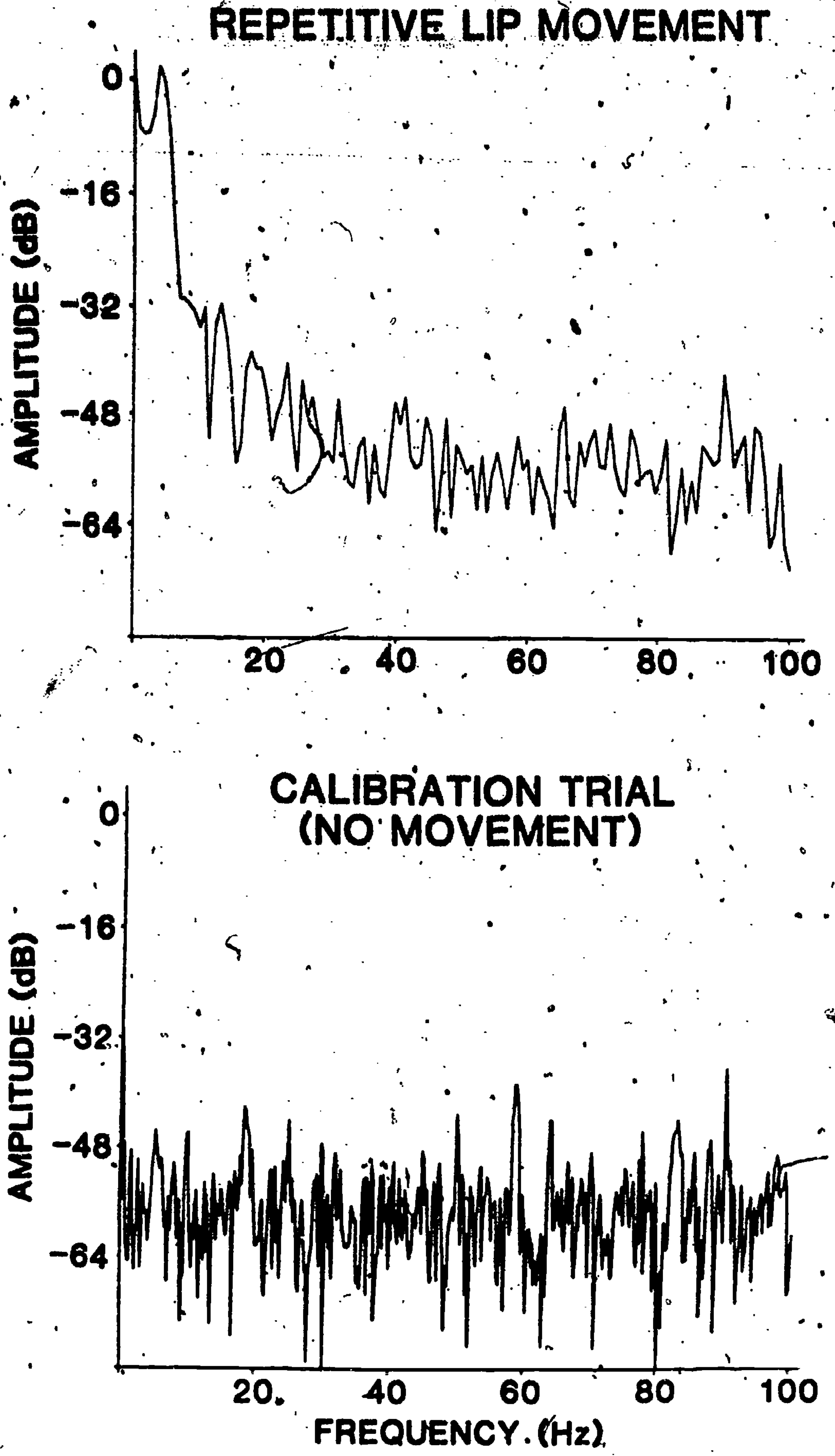


Figure 1. Amplitude spectra (logarithmic; see footnote 2) of a typical repetitive lip movement sequence (/əpǽpæpə/) (top; 256 points in the FFT) and a calibration trial (bottom).

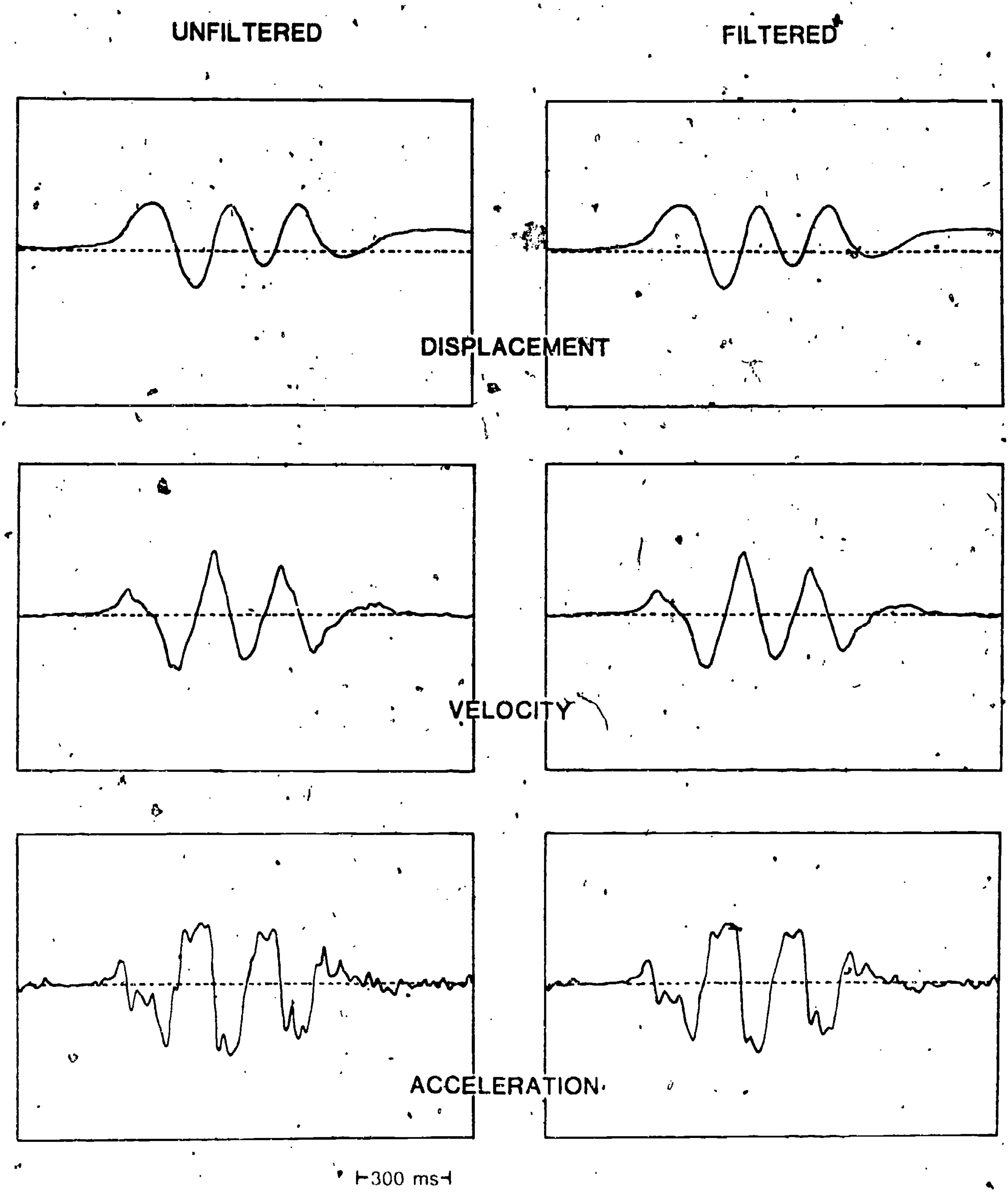


Figure 2. Examples of unfiltered (left) and filtered (right) displacement, velocity, and acceleration waveforms, reiterant speech.

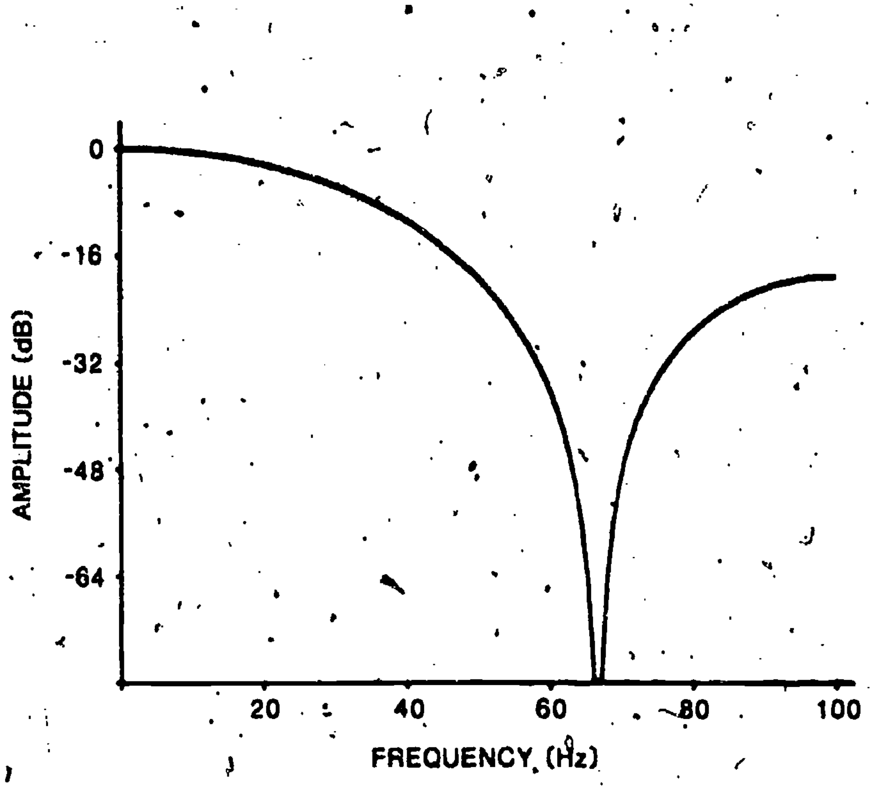


Figure 3. Computed frequency response of the seven-point Bartlett window.

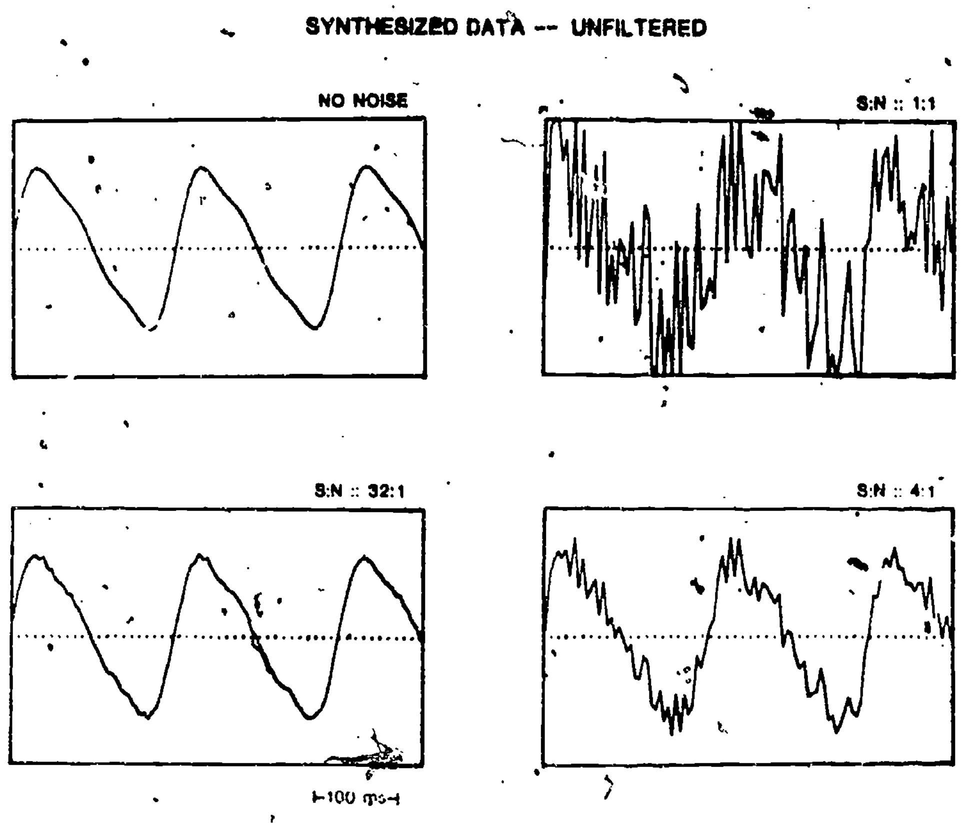


Figure 4. Waveforms of synthesized data, unfiltered. Clockwise from upper left: no added broadband noise, signal-to-noise ratio = 1:1, 4:1, and 32:1.

To observe the effects of the Bartlett filter on a known signal similar to collected data, we filtered simulated data created [illegible] our digital sinewave synthesis package (Rubin, 1982), using a fundamental of 5 Hz and two harmonics at 10 and 15 Hz. To this signal we added three levels of broadband noise, with resultant signal-to-noise ratios of 1:1, 4:1, and 32:1. Note that the latter ratio is still less than that observed during the above calibration trial. The waveforms are shown in Figure 4, and their respective amplitude spectra in Figure 5. All of these signals were then filtered using the seven-point Bartlett window: the resultant waveforms and their amplitude spectra are shown in Figures 6 and 7. Comparing the two spectral plots for the noise-free signal, unfiltered and filtered, it can be seen that the filter leaves the peaks of the three sinewave components largely unaffected--the 15 Hz peak is lowered only 1 dB. The same relative lack of attenuation of the three main peaks is observed for the noisy waveforms, while the noise content at the higher frequencies is lowered considerably. Visual inspection of the smoothed waveforms shows that even in the 1:1 signal-to-noise condition the overall signal pattern is retained after filtering. The filter performs as intended across the entire frequency range, reducing high frequency noise while leaving low frequency variations largely unaffected.

V. Digital Differentiation

With filtered displacement data, reasonable estimates of time derivatives can be computed. Many algorithms exist for the digital computation of derivatives, some more complicated than others. As in the case of filtering, we use an algorithm which is very simple in the time domain, the two-point central difference algorithm (CDA):

$$x'(t) = (x(t+1) - x(t-1))/2h,$$

where $x'(t)$ is the computed derivative at time sample t, $x(t+1)$ the sampled value of the original waveform at t+1, $x(t-1)$ the sampled value at t-1, and h the sample time interval, in this case, 5 ms. As well as simplicity, this algorithm possesses desirable accuracy, frequency response, and phase shift properties. In an experimental comparison of five derivative algorithms (the two-point forward-difference algorithm, the two- and four-point CDA, and five- and seven-point second-order fits) Marble, McIntyre, Hastings-James, and Hor (1981) found this algorithm to be the most accurate for twelve-bit data. In the frequency domain, an ideal differentiator amplifies spectral components in direct proportion to their frequencies, but since the two-point CDA is an averaging process, i.e., computed over more than one sample value, it acts as an ideal differentiator in series with a low-pass filter (Bahill, Kallman, & Lieberman, 1982). The frequency response of the central-difference algorithm is shown in Figure 8 along with that of an ideal differentiator (note the linear scale). The response is within 10% (-.9 dB) of ideal at 25 Hz and 29% (-3 dB) at 44 Hz. Thus, the spectral components of movement, most of which are below 44 Hz (see Rozendal, 1984, for a review), are affected very little. The algorithm's attenuation of the higher frequencies is desirable, since noise components dominate in this range. That is, the low-pass filtering action of the CDA offsets the high frequency amplification introduced by differentiation to a certain extent. Finally, unlike forward- or backward-difference algorithms, the CDA is a purely anti-symmetric function and as such introduces no frequency-dependent phase shifts. As a consequence, exact time correspondence between sampled displacement, computed velocity, and computed acceleration data is obtained.

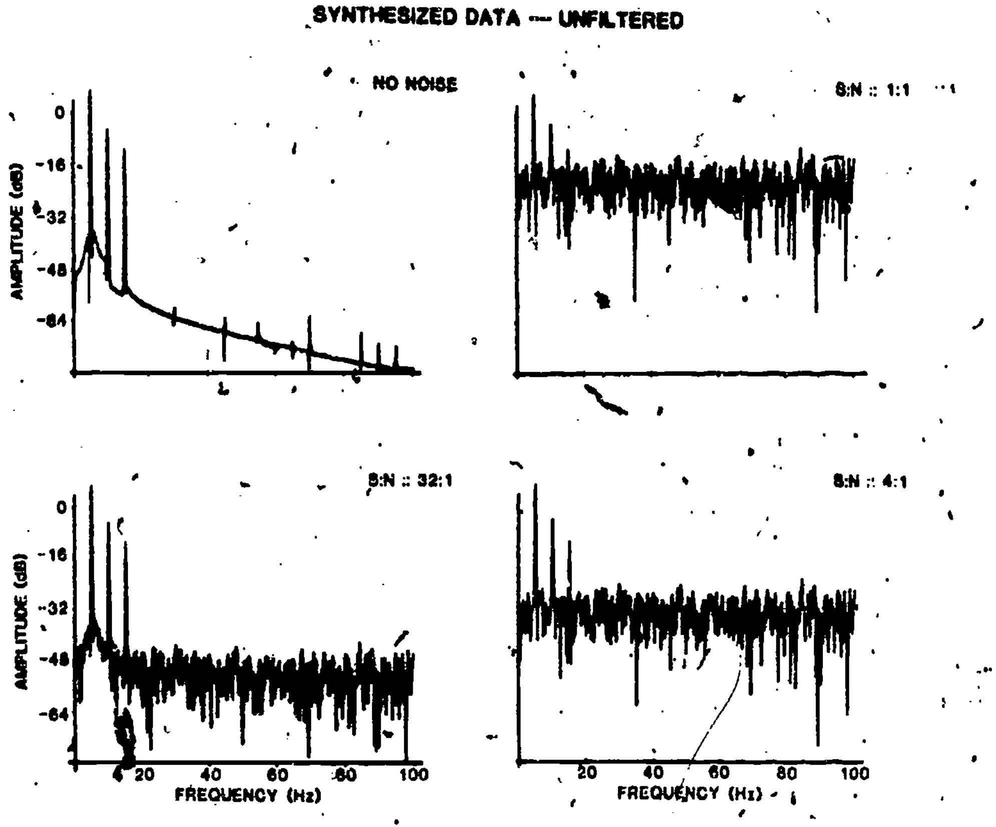


Figure 5. Amplitude spectra of the waveforms in Figure 4.

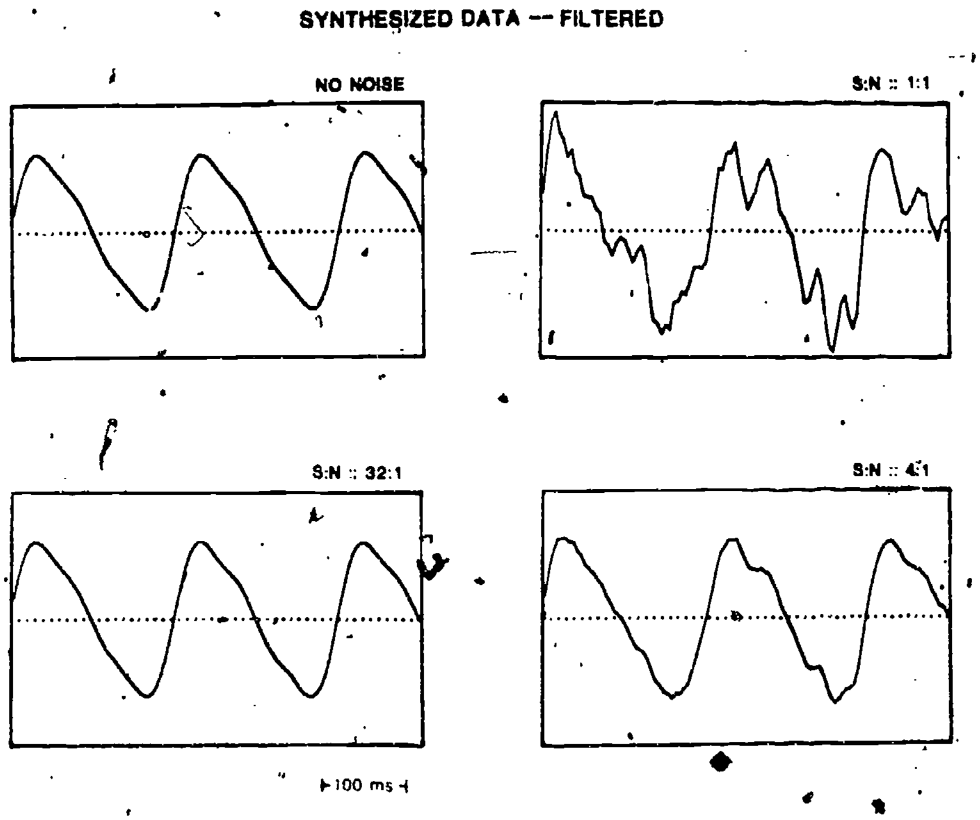


Figure 6. Waveforms of synthesized data, filtered with the seven-point Bartlett window. Clockwise from upper left: no added broadband noise, signal-to-noise ratio = 1:1, 4:1, 32:1.

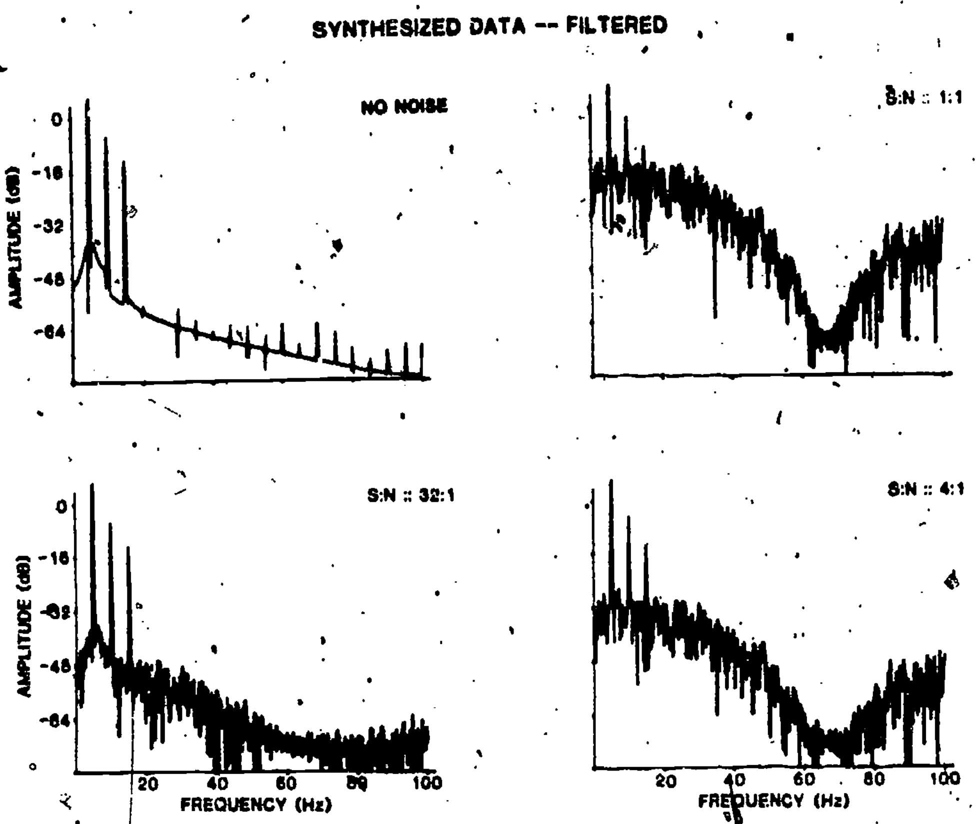


Figure 7. Amplitude spectra of the waveforms in Figure 6.

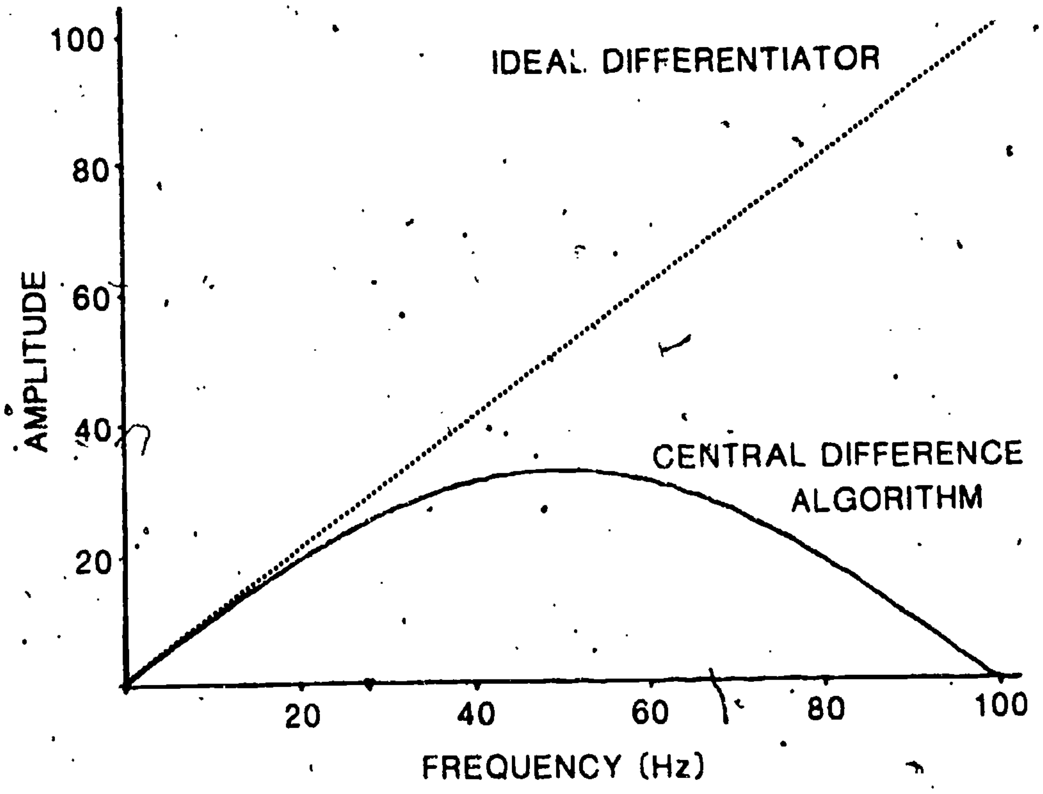


Figure 8. Frequency response (<u>linear</u> amplitude scale) of the two-point central difference algorithm (solid line) and the ideal differentiator (dotted line).

VI. Evaluation of the Digital Processing Sequence

As an overall evaluation of our processing sequence, we examined the acceleration characteristics of a set of repetitive movements whose fundamental frequency averaged 4.65 Hz, a rate similar to that observed in reiterant speech tasks (Kelso et al., 1985). Both angular displacement and acceleration were directly transduced (the former by a linear potentiometer, the latter by an Entran EGC-240-10D accelerometer), recorded on FM tape, and sampled at 200 Hz, as discussed above. The accelerometer's frequency response was found to be flat within .5 dB from 20 to 100 Hz. Derived acceleration data were obtained from the displacement data via the steps outlined in Section II, using the Bartlett filter and the central-difference algorithm. The amplitude spectra of the directly transduced and derived acceleration waveforms (see Figure 9) display strong similarities up to the frequency at which noise predominates in the directly transduced signal (approximately 35 Hz). There is some slight amplification in the derived signal's spectral components relative to those of the directly transduced signal from about 20 to 30 Hz; in this range, spectral components are proportionally amplified by the central-difference algorithm, and the filtering effects of both the CDA and the Bartlett filter are small. However, the main frequency characteristics of the directly transduced signal are closely reflected in the derived signal, and high-frequency noise is drastically reduced. Moreover, the waveforms themselves are quite similar, and a sample-by-sample cross-correlation of the two waveforms revealed a Pearson *r* of .98. These results indicate that the sequence of steps we use to derive acceleration from displacement produces data closely approximating that obtained from direct acceleration transduction.

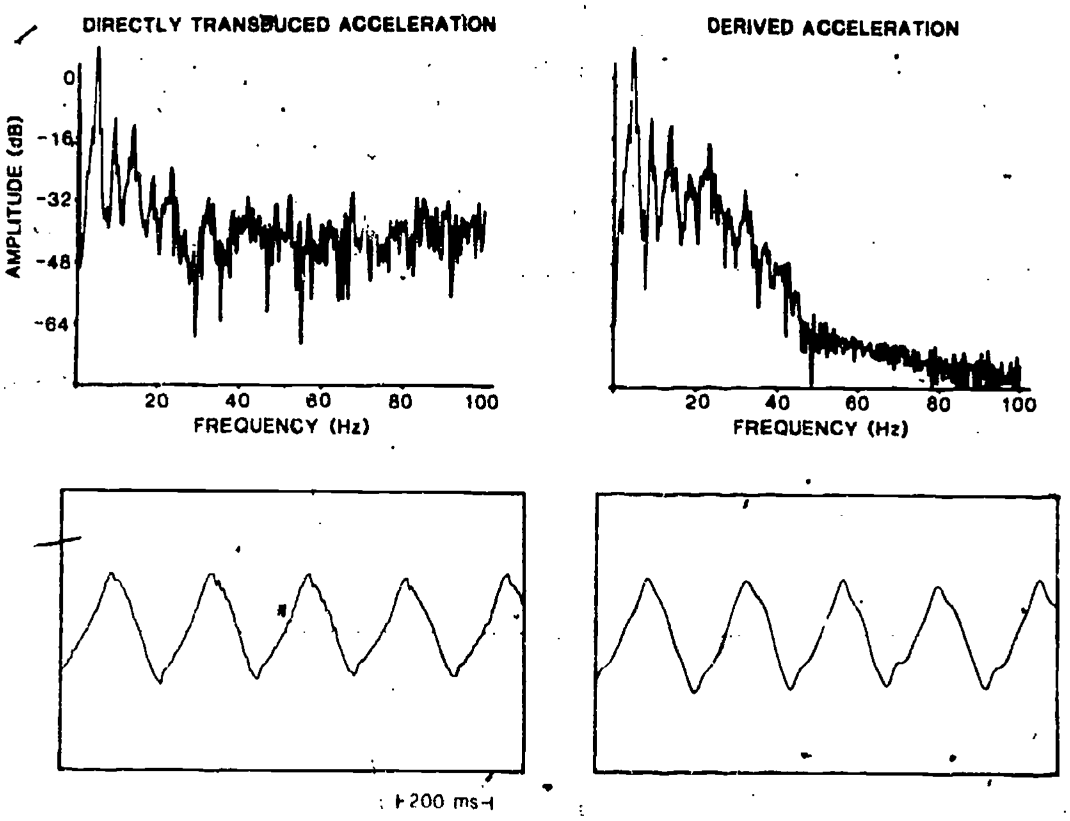


Figure 9. Acceleration waveforms and amplitude spectra (logarithmic) obtained from an accelerometer (left) and doubly differentiated displacement data (right).

VII. Summary

The processing of kinematic data raises a number of problems for the investigator, such as the presence of noise in the analog signals, quantization noise, and the characteristics of differentiation itself. Careful selection of filtering and differentiation procedures plays a crucial role in determining the precision of experimental results. These issues received attention in the present note in which we document and evaluate the various signal processing steps for treating movement data. To some extent, an acid test of the effectiveness of our system rests in the comparison between acceleration measurements derived from displacement data (an inherently noisy process due to double differentiation) and the raw signals directly provided by an accelerometer. As shown in part VI of the present note, the match between these independent measures of the same motion is quite impressive, both in terms of their waveforms and spectral properties. Few direct comparisons of this kind exist in the literature, and not all of them are well-matched (e.g., Pezzack, Norman, & Winter, 1977). This outcome is assuring, since the accuracy of one's measures is a major factor limiting their interpretation.

References

Bahill, A. T., Kallman, J. S., & Lieberman, J. E. (1982). Frequency limitations of the two-point central difference differentiation algorithm. *Biological Cybernetics*, *45*, 1-4.

Gulisano, V. (1982). *Physiological signal processing system: User's manual*. New Haven, CT: Haskins Laboratories.

Hogan, N. (1984). An organizing principle for a class of voluntary movements. *Journal of Neuroscience*, *4*, 2745-2754.

Jeannerod, M. (1984). The timing of natural prehension movements. *Journal of Motor Behavior*, *16*, 235-254.

Kaiser, J. F., & Reed, W. A. (1981). Data smoothing using low-pass digital filters. *Review of Scientific Instrumentation*, *48*, 1447-1457.

Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979). On the nature of human interlimb coordination. *Science*, *203*, 1029-1031.

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

Marble, A. E., McIntyre, C. M., Hastings-James, R., & Hor, C. W. (1981). A comparison of digital algorithms used in computing the derivative of left ventricular pressure. *IEEE Transactions in Biomedical Engineering*, *28*, 524-529.

Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics*, *46*, 135-147.

Oppenheim, A. V., & Schafer, R. W. (1975). *Digital signal processing*. Englewood Cliffs, NJ: Prentice-Hall.

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

Pezzack, J. C., Norman, R. W., & Winter, D. A. (1977). An assessment of derivative determining techniques used for motion analysis. *Journal of Biomechanics*, *10*, 377-382.

Rozendal, R. H. (1984). Biodynamics of locomotion, forty years later. In H. T. A. Whiting (Ed.), *Human motor actions--Bernstein reassessed*. Amsterdam: North Holland.

Rubin, P. (1982). Sinewave synthesis instruction manual. New Haven, CT: Haskins Laboratories.

Soechting, J. F., & Lacquaniti, F. (1981). Invariant characteristics of a pointing movement in man. Journal of Neuroscience, 1, 710-720.

Szubowicz, L. S. (1977). WENDY Version 1.6 Reference Manual: The Haskins Waveform Editing and Display System. New Haven, CT: Haskins Laboratories.

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

Winter, D. A. (1979). Biomechanics of human movement. New York: Wiley.

Woltring, H. J. (1984). On methodology in the study of human movement. In H. T. A. Whiting (Ed.), Human motor actions--Bernstein reassessed. Amsterdam: North-Holland.

Footnotes

[1] The basic movement analysis software package, which has been extended in many ways, was developed by P. Rubin, whose help is much appreciated.

[2] All amplitude spectra were computed via the Fast Fourier Transform (FFT) and plotted using the logarithmic dB scale, 20log(amplitude/2047.5) (except Figure 8, which is on a linear scale). A machine-unit value of 2047.5 corresponds to one-half of the full-scale twelve-bit data range (0 to 4095). Unless otherwise indicated, 1024 sample points were used in each of the FFT computations.

SPONTANEOUS KICKING IN VERY YOUNG INFANTS: EVIDENCE FOR A DYNAMIC BILATERAL SYSTEM

Esther Thelen,† Karl D. Skala,† and J. A. Scott Kelso††

Abstract. The effects of added mass to one leg on the spontaneous kicking movements of normal six-week-old infants were studied while infants kicked both in and out of water. Weighting decreased the kick rate of the weighted leg but increased the rate of the unweighted leg so that an overall rate was preserved. In the dry condition, infants maintained the baseline amplitude and velocity of the weighted leg, but showed significant increases in these kinematic variables in the unweighted leg. These effects were ameliorated by submersion. Involuntary infant kicks exhibit a similiar kinematic structure to adult voluntary movements, and like adult movements, appear to be organized bilaterally as a single functional unit.

When people perform skilled actions, their movements are smooth and efficient and show a high degree of coordination among the limbs and body segments. Analyses of such varied tasks as walking, reaching, posture control, and speech suggest that coordination is achieved by recruiting the muscle groups involved as a single functional synergy rather than as independent muscles (Bernstein, 1967; Craik, Herman, & Finley, 1976; Fowler, 1977; Grillner, 1975; Kelso, Southard, & Goodman, 1979; Nashner, 1977; Saltzman, 1979; Shik & Orlovskii, 1976; Turvey, 1977). Such functional linkages may span several joints or limbs, and even muscle groups quite distant from the moving segment may participate (Belen'kii, Gurfinkel, & Pal'tsev, 1967; Marsden, Merton, & Morton, 1983). Recent studies of speech movements, for example, show that when one element in the ensemble is perturbed, adjustments are made in the entire functional unit (Abbs & Gracco, 1984; Kelso, Tuller, V.-Bateson, & Fowler, 1984).

What are the developmental origins of coordination and when do such functional units develop? In stark contrast to adult performance, the movements of very young infants appear at first glance to be jerky and uncoordinated. Nonetheless, we report here that a unitary structuring and dynamic organization of movements is present at a very early age. First, we demonstrate that in six-week-old infants the motions of the legs are coupled: perturbations of the dynamic parameters of one leg are reflected in the space-time behavior of the opposite leg. Additionally, we show that although infants of this age have little or no voluntary control of their movements (Conel, 1941), the

---

†Department of Psychology, University of Missouri.

††Also University of Connecticut.

Acknowledgment. Support for this study was provided by NSF Grant BNS 8200434 and NICHD RCDA HD-00492 to E. Thelen, and ONR contract N00014-83-0083 and NINCDS Grant 13617 to Haskins Laboratories. We thank John Scholz and Glenn Hausfater for their comments.

kinematic structure of infant leg kicks is similar to that of adult voluntary movements.

When awake infants are placed supine, they perform rapid, cyclic leg kicks. This kicking in young infants is a spontaneous behavior; it is seemingly involuntary and requires no specific eliciting stimuli. As infants become more behaviorally aroused, they move and kick more (Thelen, Bradshaw, & Ward, 1981). Kinematic analyses have shown that within each limb, movements are indeed not random, but highly coordinated. The three joints of an individual leg--hip, knee, and ankle--are tightly phase-locked and move in temporal and spatial synchrony (Thelen & Fisher, 1983). Such phase coupling between the limbs was not evident, however. Although at two weeks of age, leg kicks largely alternate between right and left legs, infants become more asymmetrical between one and four months of age and often kick with only one leg (Thelen, Ridley-Johnson, & Fisher, 1983).

Contemporary theories of motor control emphasize the importance of the biodynamic characteristics of the moving system--its mass, stiffness, damping, and the equilibrium position of the muscles, in determining the kinematics of the resulting movement (Bizzi, Polit, & Morasso, 1976; Cooke, 1980; Fel'dman, 1966; Kelso, 1977; Kugler, Kelso, & Turvey, 1980). The early months of infancy are a period of especially rapid growth and a natural consequence of growth is an increase in leg mass. Thus, we asked whether infant movements would be sensitive to experimentally-induced changes in mass, and especially whether these dynamic alterations would have consequences for bilateral coordination. To do this, we added weight to one leg of six-week-old infants as they spontaneously kicked in the supine position both in and out of water. The resulting rate of kicking in each leg (loaded or unloaded) as well as the amplitude, velocity, and duration of kicks were examined. Systematic effects of this unilateral weighting on both legs would indicate that the bilateral system was functioning as a unitary ensemble. Submersion in water, on the other hand, should ameliorate the effects of weighting.

Twelve 6-week-old, normal infants (6 boys and 6 girls) served as subjects. Each infant was seen for two sessions separated by two or three days; one session was conducted with the infants' legs submerged in water. At each session, the infants' clothes were removed and the joints of their right legs marked with tape to facilitate movement analysis. They were placed on a lightly padded cot that inclined their torsos at a 25 degree angle and that allowed free movement of their limbs. During the submersion session the cot was placed in a 92 x 32 x 45 cm (32 gal) aquarium filled with sufficient pleasantly warm water to reach the infants' nipple line. Each session (wet and dry) commenced with a 3 minute baseline condition where infants kicked normally. This was followed by 3 minutes each of weighting either the right or left leg. The order of the wet and dry sessions and the laterality of the leg weighting was counterbalanced in the sample. Leg weighting consisted of a total of 185 grams of lead shot sewn into two fabric strips fastened with Velcro to the thigh and calf.[1] Sessions were videotaped from both a lateral and overhead view. Infants' arousal levels were monitored every 16 seconds during the taping session using a 6 point scale ranging from (1) asleep to (6) highly distressed.

In an earlier study, the rate of upright stepping in both legs was reduced in one-month-old infants by bilateral weighting of the legs (Thelen, Fisher, & Ridley-Johnson, 1984). We now asked whether the unilateral mass manipulations would affect the relative rate of right and left kicks. To do this two independent observers viewed the overhead videotapes and scored the number of right and left kicks.[2] In Table 1, we show that weighting one leg decreased the kick rate in the weighted leg and increased the rate in the unweighted leg in both wet and dry conditions. The overall kick rate, the sum of both right and left kicks over the time interval, did not change, however. Generalized arousal was not a factor; there was no significant difference in arousal scores as a function of weighting or water conditions. The effect of adding weights was to decrease the kick rate in the weighted leg and produce a concomitant increase in the rate of the unweighted leg. Submersion did not lessen the effect of weighting on kick rate. Thus, although the kick rate of individual legs changed systematically with weighting, the overall kick rate was preserved, strongly suggesting that bilateral kicking was produced as a single unitary activity.

We next asked if the kinematic variables of amplitude, velocity, and movement duration (period) were responsive to the unilateral weighting manipulation. From the 12 subjects observed, we chose the six with the highest overall kick rate for detailed movement analysis. For each of these six infants a naive observer selected a 5 sec representative sample of continuous kicking movements for each condition in each session. This procedure yielded an average of 7.8 flexion and extension movements per infant in each condiiton in the dry session (range, 4-12 movements) and 8.1 in the wet session (range, 2-18 movements). Movements were analyzed by digitizing the position of the hip, knee, and ankle joints of the right leg every 16 ms, i.e., the videoframe sampling rate was 60 Hz. Resulting displacements were filtered for high frequency noise using a numerical filter (Winter, 1979) based on a Fourier analysis for dominant frequencies (Thelen & Fisher, 1983). Movement amplitudes and durations were determined from the maxima and minima of the excursions of the knee joint along the x-axis, parallel to the torso of the infant. For each discrete movement excursion the peak velocity was identified as the maximum instantaneous velocity associated with that movement.

Figure 1 illustrates the typical right leg x-axis displacements in the unweighted, left-weighted, and right-weighted dry conditions, and in Table 2 we present the means for the sample. In the dry session, weighting the right leg resulted in infants' maintaining the baseline amplitude and velocity of the weighted leg. When the left leg was weighted, however, the amplitude and velocity of the right, unweighted leg was dramatically increased. The duration of the movement excursions remained unaffected.[3] In the water, there was no significant effect of weighting on amplitude, and a trend to a decreased velocity and increased duration in the right leg when it was weighted.

These experiments revealed remarkable self-organizing properties in the motor systems of six-week-old infants. When faced with an increased load perturbation to a single leg, infants acted to maintain a consistent overall kick rate within the bilateral system and the amplitude and velocity of the weighted leg. To do this, they must have sensed the load perturbation and increased the neural activation levels delivered to the system as a whole. This system-wide increase was, in turn, manifested as higher amplitude and velocity in the opposite, unweighted leg. Placing the infants in water substantially ameliorated the effects of the added mass; amplitude and velocity were not reliably increased in the unweighted leg and there was some evidence that water reduced these variables in the weighted leg, perhaps due to increased drag. Thus, the kinematics of the unperturbed leg reflected the dynamic

Table 1
Kick Rate and Arousal Scores as a Function of Submersion and Unilateral Leg Weighting

		Leg weight condition		
		No weight	Left weighted	Right weighted
Kick rate (mean kicks/min)[1]				
Dry				
	Right leg	6.25	10.57	5.41
	Left leg	7.77	6.32	10.54
	Total	14.02	16.89	15.95
Wet				
	Right leg	8.17	13.89	5.37
	Left leg	7.58	5.96	10.67
	Total	15.75	19.85	16.04
Arousal score [2]				
Dry		3.77	4.05	4.18
Wet		3.83	4.33	4.29

[1] Repeated measures ANOVA using kick rate as the dependent variable and water condition, weight condition, and kick laterality as independent variables showed no main effects for dry/wet condition, leg weight condition, or kick laterality. The interaction between leg weight condition and laterality was highly significant, $F(2,22) = 21.88$, $p < .001$. No other interactions were significant.

[2] Arousal scale: 1 (asleep), 2 (drowsy), 3 (awake, alert, few movements), 4 (awake, alert, active), 5 (fussy), 6 (highly distressed).

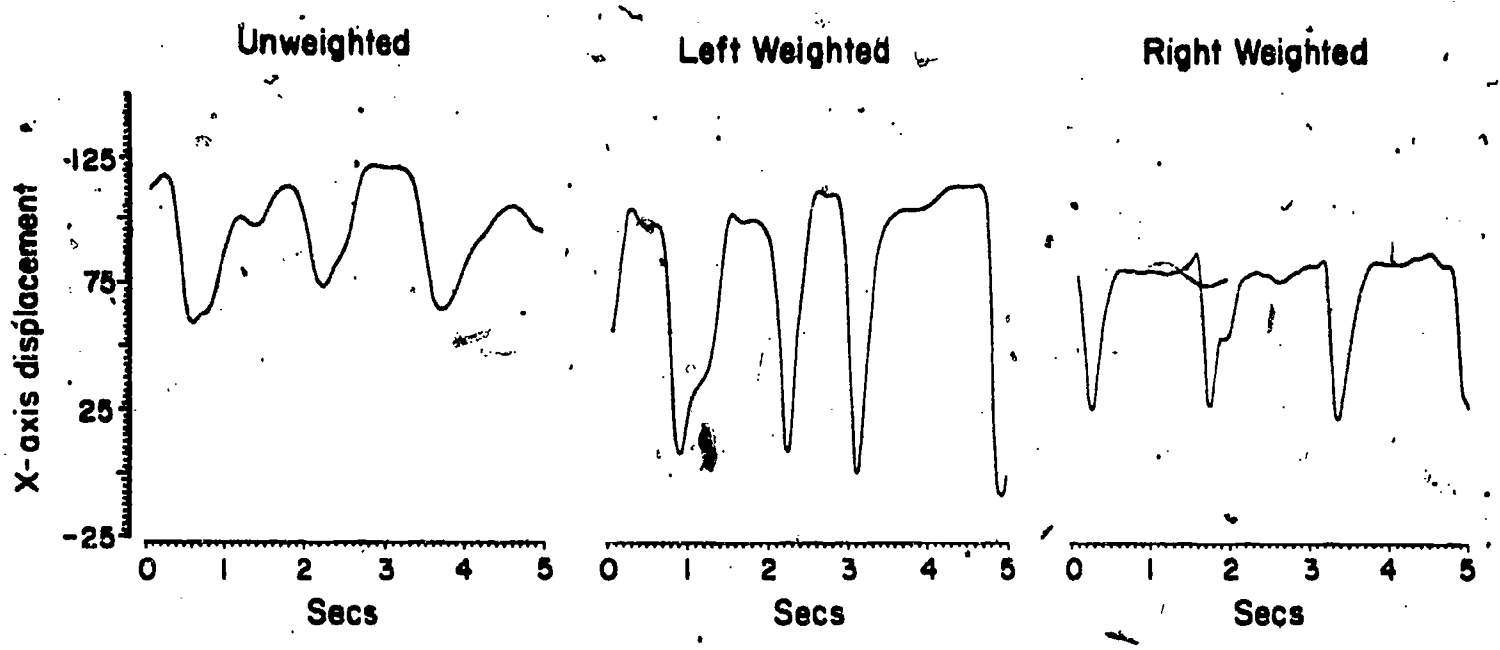


Figure 1. Horizontal (x-axis) excursions of the right leg of a six-week-old infant (dry condition) when the legs were unweighted, and when the left and right legs were weighted.

manipulations of the other limb. These results are consistent with a fixed power source being rationed out, as it were, over the whole kicking system, an effect accomplished presumably without cognitive intervention.[4]

Recent theories of motor control have shown that voluntary movements can be usefully modeled as a dynamical mass spring system whose parameters, stiffness and equilibrium position, can be set by the nervous system (Bizzi et al., 1976; Cooke, 1980; Fel'dman, 1966; Kelso, 1977; Kugler et al., 1980). In several ways, these spontaneous and involuntary infant movements meet predictions of such a model.[5] For example, as in many adult voluntary movements such as speaking and reaching (Cooke, 1980; Jeannerod, 1984; Ostry, Keller, & Parush, 1983), we found a strong linear relationship between movement amplitude and peak velocity in infant kicks (Table 2). Fifteen of eighteen correlation coefficients of the individual amplitude/peak velocity regression lines were significant at $p < .05$ or less in the dry session (all weight conditions) and 14 of the 18 possible were significant in the wet sessions (mean Pearson's $r = .818$ in dry and .848 in wet). According to the mass-spring model, the slope of the peak velocity-amplitude regression equation is an estimate of system stiffness, which can be adjusted by the length-tension relation between agonist and antagonist muscles (Bizzi et al., 1976; Cooke, 1980; Fel'dman, 1966; Kelso, 1977; Kugler et al., 1980). In Table 2, we show that in the dry condition, estimated stiffness increased in both weighted conditions over

Table 2
Kinematics of Right Leg Movements as a Function of Submersion and Unilateral Leg Weighting

	Amplitude (Arb. units)	Peak Vel. (Arb. units)	Amp-vel. corr.	Amp-vel. slope	Movement duration (s)
Dry Condition					
Right leg- no weights	50.95	118.97	.840	1.58	.387
Right leg- left weighted	85.77	220.41	.846	3.11	.397
Right leg- right weighted	50.07	140.15	.845	2.79	.360
Wet Condition					
Right leg- no weights	72.36	170.24	.881	2.37	.438
Right leg- left weighted	75.90	186.96	.740	1.79	.410
Right leg- right weighted	65.37	121.08	.765	1.32	.615

Repeated measures ANOVAs using amplitude, velocity, and movement duration as dependent measures showed a significant main effect of weighting on amplitude, $F(2,10) = 4.23$, $p < .05$, and velocity, $F(2,10) = 6.72$, $p < .02$, but no effect of dry or wet sessions and no interactions. Post-hoc comparisons showed significant ($p < .05$) differences in both amplitude and velocity in the dry condition: Left weighted differed significantly from the no weight and right weighted conditions, which did not differ significantly from each other. In the wet condition, there was a significant difference in velocity between left and right weighted conditions, as well as between the no weight and right weighted conditions. Each of the six infants showed this pattern of results in the dry condition: maintenance of the amplitude and velocity in weighted leg and increases in these variables in unweighted leg. There were no main effects on movement duration, but an interaction between dry/wet sessions and weight conditions, $F(2,10) = 4.61$, $p < .04$.

the baseline, but this effect was not observed in the water, presumably due in part to the combined influence of gravity and the fluid medium.

Although six-week-old infants do not show the tight in-phase or alternating coupling of mature limb movements, each limb does not act independently. In particular, the entire bilateral system appears dynamically responsive to scalar changes in activation levels. One implication is that self-correcting mechanisms are inherent in the neuromuscular system without higher level control. This may be one way to preserve movement topographies in growing organisms without the need for conscious recalibration. In addition, these results suggest we look at the motor asymmetries of early development in light of biodynamic changes. If the entire interlimb system is indeed sensitive to the biomechanical properties of one element, transient or asymmetric changes of mass or muscle tone may shift the lateral distribution of movements. Perhaps most important, however, is the demonstration that even at this early age, the kinematic structure of movement (the peak velocity-amplitude relation) shares a similarity with adult voluntary movements that involve quite different anatomical structures. This result hints at a basic law of biological motion common to developing and developed organisms.

References

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of the lip during speech. Journal of Neurophysiology, 51, 705-723.

Belen'kii, V. Y., Gurfinkel, V. S., & Pal'tsev, Y. I. (1967). Elements of control of voluntary movements. Biophysics, 12, 135-141.

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

Conel, J. L. (1941). The postnatal development of human cerebral cortex: Vol. 11. The cortex of the one-month old infant. Cambridge, MA: Harvard University Press.

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

Craik, R., Herman, R., & Finley, F. R. (1976). Human solutions for locomotion: Interlimb coordination. In R. M. Herman, S. Grillner, P. S. G. Stein, & D. G. Stuart (Eds.), Neural control of locomotion. New York: Plenum.

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

Fowler, C. A. (1977). Timing control in speech production. Bloomington, IN: Indiana University Linguistics Club.

Grillner, S. (1975). Locomotion in vertebrates. Physiological Reviews, 55, 247-304.

Jeannerod, M. (1984). The timing of natural prehension movements. Journal of Motor Behavior, 16, 235-254.

Kelso, J. A. S. (1977). Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 3, 529-543.

Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979). On the nature of human interlimb coordination. Science, 203, 1029-1031.

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

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

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.

Lampl, M., & Emde, R. N. (1983). Episodic growth in infancy: A preliminary report on length, head circumferences and behavior. In K. W. Fischer (Ed.), Levels and transitions in children's development (pp. 21-36). San Francisco: Jossey-Bass.

Levy, J., & Levy, J. M. (1978). Human lateralization from head-to-foot: Sex-related factors. Science, 200, 1291-1292.

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

Nashner, L. M. (1977). Fixed patterns of rapid postural responses among leg muscles during stance. Experimental Brain Research, 30, 13-24.

Ostry, D. J., Keller, E., & Parush, A. (1983). Similarities in the control of speech articulators and the limbs: Kinematics of tongue dorsum movement in speech. Journal of Experimental Psychology: Human Perception and Performance, 9, 622-636.

Pande, B. S., & Singh, I. (1971). One-sided dominance in the upper limits of human fetuses as evidence of asymmetry in muscle and bone weight. Journal of Anatomy, 109, 457-459.

Provine, R. R. (1982). Preflight development of bilateral wing coordination in the chick (Gallus domesticus): Effects of induced bilateral wing asymmetry. Developmental Psychobiology, 15, 245-255.

Saltzman, E. (1979). Levels of sensorimotor representation. Journal of Mathematical Psychology, 20, 91-163.

Schultz, A. H. (1926). Fetal growth of man and other primates. Quarterly Review of Biology, 1, 465-521.

Shik, M. L., & Orlovskii, G. N. (1976). Neurophysiology of locomotor automatism. Physiological Reviews, 56, 465-501.

Thelen, E., Bradshaw, G., & Ward, J. A. (1981). Spontaneous kicking in month-old infants: Manifestation of a human central locomotor program. Behavioral and Neural Biology, 32, 45-53.

Thelen, E., & Fisher, D. M. (1983a). The organization of spontaneous leg movements in newborn infants. Journal of Motor Behavior, 15, 353-377.

Thelen, E., & Fisher, D. M. (1983a). From spontaneous to instrumental behavior: Kinematic analysis of movement changes during very early learning. Child Development, 54, 129-140.

Thelen, E., Fisher, D. M., & Ridley-Johnson, R. (1984). The relationship between physical growth and a newborn reflex. Infant Behavior and Development, 7, 479-493.

Thelen, E., Ridley-Johnson, R., & Fisher, D. M. (1983). Shifting patterns of bilateral coordination and lateral dominance in the leg movements of young infants. Developmental Psychobiology, 16 29-46.

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

Winter, D. A. (1979). Biomechanics of human movement (pp. 32-39). New York: Wiley.

Footnotes

[1] The amount of mass added to the legs was determined to simulate actual growth changes in leg dynamics, specifically, the amount of weight gained by infants in their legs between 1 and 2 months of age. Calculations were based on longitudinal anthropometric measurements of leg circumference from an earlier study (Thelen et al., in press), and converted to volume and mass using the formulas in Winter (1979).

[2] Interobserver reliability on kick rate averaged over all infants, sessions, and conditions: $r = .97$.

[3] Presumably, infants maintained constant movement duration within a variable overall rate of kicking by adjusting the pauses between movements. Although pause lengths were not measured directly in this study, we found consistent movement durations (app. 300 ms) and pause durations inversely proportional to kick rate in several earlier studies (Thelen et al., 1981; Thelen & Fisher, 1983a, 1983b).

[4] Provine (1982) weighted one wing of 3-5 day old chicks and found that weighting reduced the frequency of both wings in bilaterally synchronous drop-evoked flapping. In the supine position, human infants rarely perform bilaterally synchronous leg movements until about 5 months of age (Thelen et al., 1983). Further research should ask whether unilateral weighting in human simultaneous kicks would also produce such a "matching" type of interlimb coordination.

[5] Movements according to the underlying dynamics of an undamped mass-spring system may be characterized by the following equation of motion:

$m\ddot{x} + k\Delta x = 0$

where m = mass, k = stiffness, $\Delta x = (x - x_0)$ with x_0 = rest position; and x and $\ddot{x}$ represent position and acceleration, respectively. Such systems display cyclic motions with a) period $(T) = 2\pi/\omega_0$, where $\omega_0 = (k/m)^{\frac{1}{2}}$; and b) amplitude (A) which is determined solely by initial conditions (i.e., initial Δx and $\ddot{x}$) and is independent of k or m (and hence period); and c) peak velocity ($V_p = \omega_0 A$). Since $\omega_0 A$ is the peak velocity of simple harmonic motion then the slope of $\omega_0 A$ versus A is ω_0. Hence, assuming constant mass this slope is proportional to $(K)^{\frac{1}{2}}$. See Kelso, V-Bateson, Saltzman, and Kay (1985). In infant kicking movements, there were highly significant correlations ($r = .801$ in dry; $r = .940$ in wet) between mean T (the inverse of ω_0) predicted from mean Vp-A relation for each infant in each condition and the actual measured duration of motion (T).

[6] There is scattered evidence for both episodic (Lampl & Emde, 1983) and asymmetric growth patterns (Levy & Levy, 1978; Pande & Singh, 1971; Schultz, 1926).

ON THE COORDINATION AND REGULATION OF COMPLEX SENSORIMOTOR SYSTEMS*

E. L. Saltzman

In his foreword to the English edition of N. A. Bernstein's The Coordination and Regulation of Movements (1967), A. R. Luria predicted that readers "will find in this book the brilliant essays of a great man of science and, no doubt, the ideas contained in these essays will have a marked influence on the future work" (ibid., p. viii) in biology, physiology, biomechanics, cybernetics, mathematics, psychology, and philosophy. It is unfortunate that this seminal collection of articles (translated from Russian and sampling Bernstein's scientific output from 1934-1962) has been out of print for the last five years. Now, however, thanks to H. T. A. Whiting and North-Holland Publishers these papers have been reprinted in an updated format by including with each original chapter a set of two companion pieces by current researchers in the area of coordinated sensorimotor activity. Judging from the conceptual ferment and interdisciplinary breadth evident in these new chapters (representing, additionally, the more "recent" disciplines of artificial intelligence and robotics), it is clear that Luria's prediction has been amply validated. Bernstein's ideas are as alive and vital today as they were when originally published, and his chapters alone should make this book required reading for anyone seriously interested in how actions of multidegree-of-freedom systems are coordinated and controlled.

However, this volume is clearly more than the original Bernstein since, with only a few exceptions, the companion pieces admirably serve the editor's purpose of "reassessing" Bernstein's work. The new chapters allow the reader to see how the field has developed since Bernstein's day and show not only the extent to which his themes have been explored, amplified, and extended, but also the extent to which new developments have superceded the old. Some aspects of Bernstein's work welcome updating. Most notably, his pioneering cyclogrammetric techniques for analyzing repetitive cyclic motions such as locomotion or hammering have been greatly improved upon in the intervening years. Both Woltring's comprehensive chapter on methodology and Rosendal's methodology section in his chapter on locomotion provide the reader with an admirable state-of-the-art "synopsis of developments in kinematic data acquisition and processing, mass distribution parameter estimation, and kinetic modelling" (p. 38). In a more theoretical vein, Bernstein realized that one could not assume a simplistic one-to-one "keyboard" mapping between patterns of efferent neural impulses and resultant articulator movements. Such

---

*A review of H. T. A. Whiting (Ed.), Human motor actions: Bernstein reassessed. New York: North-Holland, 1984. Review to appear in Contemporary Psychology, in press.

Acknowledgment. This research was supported in part by NIH Grant NS-13617 and Contract No. N0014-83-K-0083 from the U. S. Office of Naval Research.

"functional non-univocality" (p. 214) is due to the fact that this mapping is sensitive to the evolving physical context of the movement, e.g., the ongoing positions and velocities of the articulators, and the spatial relationship between the articulators and the gravitational field. As pointed out in the chapters by Boylls and Greene, by Rozendal, and by Hinton, this non-univocality no longer poses an insurmountable problem for theories of coordinated movement, and recent biomechanical and robotic control schemes incorporate and even exploit such indeterminacies.

Additionally, several of the companion chapters provide scholarly biographical and historical information regarding Bernstein's scientific career (Boylls and Greene; Requin, Somjen, and Bonnet; Pickenhain; Arbib) as well as translated excerpts from work previously unavailable in English (Pickenhain). Thus, for example, we learn that Bernstein won a Soviet State prize in 1947 for his monograph On the construction of movements (1947) and are disappointed to find that there is no English translation.

The companion chapters also provide intriguing glimpses into the subcurrents and crosscurrents that are shaping the field of movement science today. A healthy discipline is not without its controversies and movement science is no exception. In these chapters we see that movement science's health is abundantly apparent, as in the reactions of Requin et al. to the relatively recent "ecological" approach to action (particularly as outlined in Reed [1982]) that is itself represented in the chapters by Turvey and Kugler and by Reed. One sympathizes with Requin et al.'s objections, since the ecological approach is by no means an easy one to live with intellectually. But who ever said intellectual life was supposed to be easy, especially if it involves questioning some of our basic assumptions? And that is exactly what the ecological scientists would have us do concerning the organization of sensorimotor behavior. For example, synthesizing Bernstein's theoretical views with J. J. Gibson's theory of visual perception (e.g., Gibson, 1966, 1979; Michaels & Carello, 1981), Turvey and Kugler argue vigorously against taking out the "loan of intelligence" (p. 381) that is necessarily entailed when one invokes processes of cognitive mediation to account for the "construction" of percepts from sensations, or actions from commands. Viewed in the context of such challenges to the status quo in the field, these three chapters represent, in my opinion, a rather lively subunit of the new volume. However, before leaving the topic of this controversy, I feel compelled to inject a polite rejoinder to one aspect of Requin et al.'s chapter. They allude to a group of movement scientists at Haskins Laboratories as proselytizing "exegetes" (p. 471) or "disciples of a prophet" (p. 470) who "played a decisive role in the circulation of Bernstein's view" (p. 469). In fact, in the introductory chapter, Boylls and Greene mention a "Bernstein 'cult'" (p. xix) with reference to Requin et al.'s chapter. Now, I have been working at Haskins Laboratories for the last three years on problems of speech and limb coordination, and from my experience these nonsecular metaphors seem a bit baroque. It seems more appropriate to describe the Haskins group as critical admirers of Bernstein, and these do not a cult make. A fan club, maybe, but a cult? Heaven forbid!

Finally, several of the companion chapters highlight developments that, in this reviewer's opinion, lie at the cutting edge of the field and that both echo and significantly expand upon two of Bernstein's central themes: his emphasis on the task or "motor problem" (p. 344) in identifying functional units of action, and his insistence that coordinated biological actions be "regarded as *morphological objects*...(that)...do not exist as homogenous wholes at every moment but *develop* in time, that in their essence they incorporate time coordinates" (p. 178-179, italics added). Regarding the former issue, Boylls and Greene highlight the importance of studying movements at the "task level of description" (p. xix) and of conceptualizing coordination in terms of task-dynamical functions that define "a family of dynamical 'flows' of the sensorimotor state during the elaboration of a task in time" (p. xxii; see also Saltzman & Kelso, 1983, for related discussions of task-dynamics). Similarly, Arbib argues for the importance of qualitatively distinct controller structures in the performance of correspondingly distinct tasks like ball-catching and waltzing. But what is most significant in Boylls and Greene's discussion, perhaps, is the incorporation of current concepts from the nonlinear dynamical systems literature (e.g., "potential functions...bifurcations and catastrophes" [p. xxiv]) into their theoretical framework. This approach, in contrast with a traditional linear control system account, provides a set of mathematical and theoretical tools for dealing with issues of self-organization and pattern formation in physical and biological systems (e.g., Gierer, 1981; Haken, 1983; Prigogine & Stengers, 1984). As discussed in the chapters by Turvey and Kugler and by Arbib, these approaches should allow one to rigorously formulate problems of sensorimotor coordination, as Bernstein might have wished, in terms of morphogenetic or embryological changes in the structure of the "'biodynamic tissue' of live movements" (p. 178). For example, Bernstein's notion of time coordinates that are incorporated into the essence of an action (see earlier quote) has been developed into the concept of an action's *intrinsic timing* (e.g., Fowler, 1977; Kelso & Tuller, in press). According to this view, a movement's temporal structure unfolds as an implicit consequence of its intrinsic task-dynamics rather than as the explicit result of durational commands issued by an extrinsic executive time-keeper. Significantly, the notion of time as intrinsic to a system's dynamics appears well grounded in the related physical concept of *internal time* derived in the field of irreversible thermodynamics (e.g., Prigogine, 1984; see also Richardson, 1980, and Richardson & Rosen, 1979).

In conclusion, the present volume lives up to its title and offers a successful reassessment and update of Bernstein's classic 1967 publication. Taken together, the original chapters and the companion pieces confirm the belief expressed by Boylls and Greene in their introduction that Bernstein's ideas remain a "renewable resource" (p. xxix) for our attempts to understand the coordination, regulation, and growth of complex sensorimotor systems.

References

Bernstein, N. A. (1947). *On the construction of movements*. Moscow: Medgiz. (Monograph: In Russian)

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

Fowler, C. (1977). *Timing control in speech production*. Bloomington, IN: Indiana University Linguistics Club.

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.

Gierer, A. (1981). Generation of biological patterns and form: Some physical, mathematical, and logical aspects. *Progress in Biophysics and Molecular Biology*, *37*, 1-47.

Haken, H. (1983). Synergetics: An introduction (3rd ed.). Heidelberg: Springer Verlag.

Kelso, J. A. S., & Tuller, B. (in press). Intrinsic time in speech production: Theory, methodology, and preliminary observations. In E. Keller (Ed.), Sensory and motor processes in language. Hillsdale, NJ: Erlbaum.

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

Prigogine, I. (1984). Irreversibility and space-time structure. In W. Horsthemke & D. K. Kondepudi (Eds.), Fluctuations and sensitivity in nonequilibrium systems. Berlin: Springer-Verlag.

Prigogine, I., & Stengers, I. (1984). Order out of chaos: Man's new dialogue with nature. New York: Bantam Books.

Reed, E. S. (1982). An outline of a theory of action systems. Journal of Motor Behavior, 14, 98-134.

Richardson, I. W. (1980). The metrical structure of aging (dissipative) systems. Journal of Theoretical Biology, 85, 745-756.

Richardson, I. W., & Rosen, R. (1979). Aging and the metrics of time. Journal of Theoretical Biology, 79, 415-423.

Saltzman, E., & Kelso, J. A. S. (1983). Skilled actions: A task dynamic approach. Haskins Laboratories Status Report on Speech Research, SR-76, 3-50.

PUBLICATIONS
APPENDIX

PUBLICATIONS

Borden, G. J., Baer, T. B., & Kenney, M. K. (in press). Onset of voicing in stuttered and fluent utterances. *Journal of Speech and Hearing Research.*

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

Gurjanov, M., Lukatela, G., Moskovljević, J., Savić, M., & Turvey, M. T. (1985). Grammatical priming of inflected nouns by inflected adjectives. *Cognition*, 55-71.

Haken, H., Kelso, J. A. S., & Bunz, H. (in press). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*.

Hanson, V. L. (in press). Cognitive processes in reading: Where deaf readers succeed and where they have difficulty. In D. S. Martin (Ed.), *Cognition, education, and deafness: Trends in research and instruction.* Washington, DC: Gallaudet College Press.

Katz, R. B., & Shankweiler, D. (in press). Repetitive naming and the detection of word retrieval deficits in the beginning reader. *Cortex*.

Kelso, J. A. S., & Tuller, B. (in press). Intrinsic time in speech production: Theory, methodology, and preliminary observations. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language*. Hillsdale, NJ: Erlbaum.

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

Levitt, A. G., & Healy, A. F. (in press). The roles of phoneme frequency, similarity, and availability in the experimental elicitation of speech errors. *Journal of Memory and Language*.

Lisker, L. (1985). The pursuit of invariance in speech signals. *Journal of the Acoustical Society of America*, *77*, 1199-1202.

Mann, V. A. (1984). Longitudinal prediction and prevention of early reading difficulty. *Annals of Dyslexia*, *34*, 117-136.

Mann, V. A. (in press). A cross-language perspective on the relation between temporary memory skills and early reading ability. *Remedial and Special Education*.

Mann, V. A. (in press). Temporary memory skills and the acquisition of reading ability in Japan. To appear in H. K. Wong & O. Tzeng (Eds.). *Psychological aspects of the Chinese language*.

Mann, V. A. (in press). Why some children encounter reading problems: The contribution of difficulties with language processing and linguistic sophistication to early reading difficulty. In J. K. Torgessen & B. Y. Wang (Eds.), *Learning disabilities: Some new perspectives*. New York: Academic Press. Also published under the same title as *MIT Center for Cognitive Science Occasional Paper*, *29*, 1-44 (1984).

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory: Can they presage early reading problems? *Journal of Learning Disabilities*, *17*, 592-599.

Mann, V. A., Shankweiler, D., & Smith, S. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. *Journal of Child Language*, *11*, 627-643.

Mann, V. A., Sharlin, H. M., & Dorman, M. (1985). Children's perception of sibilants: The relation between articulation and perceptual development. *Journal of Experimental Child Psychology*, *39*, 252-264.

McGarr, N. S., Head, J., Friedman, M., Behrman, A., & Youdelman, K. (in press). Use of visual and tactile sensory aids in speech production training: A preliminary report. Journal of Rehabilitative Research and Development.

Munhall, K. G., Ostry, D. J., & Parush, A. (in press). Characteristics of velocity profiles of speech movements. Journal of Experimental Psychology: Human Perception and Performance.

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

Rakerd, B., & Verbrugge, R. R. (1985). Linguistic and acoustic correlates of the perceptual structure found in an individual differences scaling study of vowels. Journal of the Acoustical Society of America, 77, 296-301.

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

Recasens, D. (1984). Timing constraints and coarticulation: Alveolo-palatals and sequences of alveolar + [j] in Catalan. Phonetica, 41, 125-139.

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

Repp, B. H., & Bentin, S. (1984). Parameters of spectral/temporal fusion in speech perception. Perception & Psychophysics, 36, 523-530.

Repp, B. H., & Williams, D. R. (in press). Influence of following context on perception of the voiced-voiceless distinction in syllable-final stop consonants. Journal of the Acoustical Society of America.

Saltzman, E. L. (in press). On the coordination and regulation of complex sensorimotor systems. Review (H. T. A. Whiting (Ed.), Human motor actions: Bernstein reassessed. New York: North-Holland, 1984.). Contemporary Psychology.

Shankweiler, D., Smith, S. T., & Mann, V. A. (1984). Repetition and comprehension of spoken sentences by reading-disabled children. Brain and Language, 241-257.

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 197[illegible]	AD A026196	ED 123-678
SR-47	July - September [illegible]6	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		ED 252-907
SR-81	January - March 1985	**	**

Information on ordering any of these issues may be found on the following page.

**DTIC and/or ERIC order numbers not yet assigned.

SR-81 (1985)
(January-March)

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 <u>Language and Language Behavior Abstracts</u>, 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 06510

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

N/A

3. REPORT TITLE

Haskins Laboratories Status Report on Speech Research,

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

May, 1985

7a. TOTAL NO. OF PAGES

336

7b. NO. OF REFS

801

8a. CONTRACT OR GRANT NO.

HD-01994	NS13870
HD-16591	NS13617
N01-HD-5-2910	NS18010
RR-05596	N00014-83-K-0083
BNS-8111470	

9a. ORIGINATOR'S REPORT NUMBER(S)

SR-81

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

None

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited*

11. SUPPLEMENTARY NOTES

N/A

12. SPONSORING MILITARY ACTIVITY

See No. 8

13. ABSTRACT

This report (1 January-31 March) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

- Segmentation of coarticulated speech in perception
- Intrinsic time in speech production: Theory, methodology, and preliminary observations
- A theoretical model of phase transitions in human hand movements
- Repetitive naming and the detection of word retrieval deficits in the beginning reader
- Linguistic abilities and spelling proficiency in kindergarteners and adult poor spellers
- Errors in short-term memory for good and poor readers
- Longitudinal prediction and prevention of early reading difficulty
- Temporary memory for linguistic and nonlinguistic material in relation to the acquisition of Japanese Kana and Kanji
- Cognitive processes in reading: Where deaf readers succeed and where they have difficulty
- Discovering messages in the medium: Speech perception and the prelinguistic infant
- Vowel-to-vowel coarticulation in Catalan VCV sequences
- Categorical trends in vowel imitation: Preliminary observations from a replication experiment
- Influence of following context on perception of the voiced-voiceless distinction in syllable-final stop consonants
- Differential processing of phonographic and logographic single-digit numbers by the two hemispheres
- A note on processing kinematic data: Sampling, filtering, and differentiation
- Spontaneous kicking in very young infants: Evidence for a dynamic bilateral system
- On the coordination and regulation of complex sensorimotor systems

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

UNCLASSIFIED

Security Classification

*This document contains no information not freely available to the general public. It is distributed primarily for library use.

A-31408

UNCLASSIFIED
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Perception: segmentation, coarticulation, vowels, imitation, categories, following context, voiced-voiceless distinction, stop consonants--syllable-final, infants, prelinguistic						
Speech Articulation: timing, theory, methodology, vowels, coarticulation, Catalan, VCV sequences						
Motor Control: phase transitions, hand movement, data processing, sampling, filtering, differentiation, kicking, infants, coordination, regulation, sensorimotor systems, review						
Reading: short-term memory, good readers, poor readers, beginning readers, deaf readers, cognitive processes, differential processes, memory logographic numbers, phonographic numbers, hemispheric effects, repetitive naming, word retrieval, linguistic ability, spelling proficiency, poor spellers, prediction of skills, prevention of difficulty, Kana, Kanji						

DD FORM 1 NOV 65 1473 (BACK)
S/N 0101-807-6821

UNCLASSIFIED
Security Classification

A-31409