AD-A099 958   HASKINS LABS INC NEW HAVEN CT   F/G 5/10
                SPEECH RESEARCH: A REPORT ON THE STATUS AND PROGRESS OF STUDIES--ETC(U)
   MAR 81   A M LIBERMAN   PHS-HD-01994
UNCLASSIFIED   SR-65(1981)   NL

1 of 3
AD A
A099958

SR-65 (1981)

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 1981,

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

ACKNOWLEDGMENTS

iii

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President and Research Director
Franklin S. Cooper,* Associate Research Director
Patrick W. Nye, Associate Research Director
Raymond C. Huey, Treasurer
Alice Dadourian, Secretary

Investigators	Technical and Support Staff	Students*
Arthur S. Abramson*	Eric L. Andreasson	Claudia Carello
Peter Alfonso*	Elizabeth P. Clark	Tova Clayman
Thomas Baer	Vincent Gulisano	David Dechovitz
Fredericka Bell-Berti*	Donald Hailey	Steven Eady
Catherine Best*	Terry Halwes	Jo Estill
Gloria J. Borden*	Sabina D. Koroluk	Carole E. Gelfer
Susan Brady*	Agnes M. McKeon	David Goodman
Robert Crowder*	Nancy O'Brien	Janette Henderson
William Ewan*	Marilyn K. Parnell	Charles Hoequist
Carol A. Fowler*	William P. Scully	Robert Katz
Louis Goldstein*	Richard S. Sharkany	Aleksandar Kostić
Vicki Hanson	Leonard Szubowicz	Peter Kugler
Katherine S. Harris*	Edward R. Wiley	Anthony Levas
Alice Healy*	David Zeichner	Harriet Magen
Kiyoshi Honda[1]		Suzi Pollock
Daisy Hung		Patti Jo Price
Leonard Katz*		Sandra Prindle
Scott Kelso		Brad Rakerd
Andrea G. Levitt*		Daniel Recasens
Isabelle Y. Liberman*		Rosemarie Rotunno
Leigh Lisker*		Arnold Shapiro
Virginia Mann*		Susan Smith
Charles Marshall		Louis G. Tassinary
Ignatius G. Mattingly*		Janet Titchener
Nancy S. McGarr*		Emily Tobey-Cullen
Lawrence J. Raphael*		Douglas Whalen
Bruno H. Repp		Deborah Wilkenfeld
Philip E. Rubin		
Donald P. Shankweiler*		
Michael Studdert-Kennedy*		
Betty Tuller*		
Michael T. Turvey*		
Ovid Tzeng[2]		
Mario Vayra[3]		
Robert Verbrugge*		

---

*Part-time
[1]Visiting from University of Tokyo, Japan
[2]Visiting from University of California, Riverside
[3]Visiting from Scuola Normale Superiore, Pisa, Italy

CONTENTS

I. Manuscripts and Extended Reports

I. <u>MANUSCRIPTS</u> <u>AND</u> <u>EXTENDED</u> <u>REPORTS</u>

DISTINGUISHING TEMPORAL INFORMATION FOR SPEAKING RATE FROM
TEMPORAL INFORMATION FOR INTERVOCALIC STOP CONSONANT VOICING

Hollis L. Fitch+

Abstract. Temporal differences between voiced and voiceless conso-
nants include the duration of the acoustic segment corresponding to
vocal tract closure, and the duration of the vocalic section
preceding the closure (the "vowel"). Yet temporal differences such
as these are also produced by changes in speaking rate. A potential
problem for perceivers of speech would thus seem to be the confound-
ing of temporal information for voicing and temporal information for
rate. Experiment I confirms that both closure duration and vowel
duration can cue the intervocalic phonemic voicing difference
between /dabi/ and /dapi/, and it confirms that both closure
duration and vowel duration can cue the rate difference between fast
and slow speech. A consideration of the more temporally extensive
patterns over vowel and closure, however, establishes a distinction
between the two kinds of articulatory changes. A voicing change
from /b/ to /p/ lengthens the closure and shortens the vowel; a rate
change from fast to slow lengthens both. The inverse relationship
between closure duration and vowel duration (as in a voicing
change), expressed as a difference in the ratio of the two, was
found in Experiment I to affect judgments of voicing more than
judgments of rate. The direct relationship between closure duration
and vowel duration (as in a rate change), expressed as a difference
in the sum of the two, was found to significantly affect judgments
of rate but not judgments of voicing.

The adequacy of the "duration of a single acoustic segment" as
a descriptor was further tested in Experiment II, where vowel
duration was varied orthogonally to produced rate. It was found
that adjusting the vowel in its steady-state, vowel nucleus section
to equate its duration to that produced by a rate change did not
have the same perceptual effect as a rate change. The same ratio of
closure-to-vowel bounded /b/ and /p/ at the three naturally produced
rates, but different ratios bounded /b/ and /p/ when the vowel
involved in the ratio was not the result of a natural rate change.

Temporal patterns within the vowel were assumed to be the cause
of the rate-change and duration-change difference. Experiment III,

---

with the use of synthetic speech, demonstrated that the relationship between the duration of the initial consonant transitions and the duration of the steady-state vowel nucleus is perceptually salient. For vowels of equal duration, the longer the transitions and the shorter the vowel nucleus, the shorter was the closure needed to change /b/ to /p/. It was suggested that such fine-grained temporal relationships, like the coarser-grained temporal relationships explored in Experiment I, might serve to distinguish information for voicing from information for rate.

INTRODUCTION

Many phonetic distinctions can be perceptually cued by a change in the duration of an acoustic segment. One phoneme will be heard when an appropriately chosen segment of speech is short, and a different phoneme will be heard when that same segment is extended. Yet temporal aspects of the speech signal, such as the durations of acoustic segments, are altered by changes in speaking rate. In general, if two phonemes are distinguished by differences in segment duration, the perceptual boundary between them will come at a shorter segment duration if the phonemes are produced at a faster speaking rate. Contrasts between voiced and voiceless consonants (Summerfield & Haggard, 1972; Summerfield, 1974, 1975a, 1975b, Note 1; Port, 1976, 1979), single and double consonants (Pickett & Decker, 1960; Fujisaki, Nakamura, & Imoto, 1975), consonants and semi-vowels (Ainsworth, 1973; Minifie, Kuhl, & Stecher, 1977; Miller & Liberman, 1979), and short and long vowels (Ainsworth, 1974; Verbrugge, Strange, Shankweiler, & Edman, 1976; Verbrugge & Shankweiler, 1977) are known to be so affected (see Miller, in press, for a review). And in fact, speaking rate itself is assumed to be cued primarily by duration.

A potential problem thus arises for perceivers of speech in that temporal aspects of phonetic information would seem to be confounded with temporal aspects of rate information. The general problem is this: a given segment duration is not invariantly related to a given percept. Its ambiguity arises from the fact that a single duration reflects both the particular phoneme being spoken and the particular rate at which it is being spoken. Although the duration is informative about both phoneme identity and rate, it does not independently specify either; a given duration may be a "short" phoneme spoken slowly or a "long" phoneme spoken rapidly. How, then, can the phonetic message be isolated?

This research is an attempt to disentangle temporal information for rate from temporal information for one particular phonetic distinction: intervocalic /b/ versus /p/ (a distinction often referred to as one of phonemic "voicing").

In general, any consistent acoustic difference in the way two phonemes are produced is likely to provide a perceptual "cue" to that phonemic contrast when other differences are neutralized (see Bailey & Summerfield, 1980). In the case of an intervocalic voicing contrast in American English, two known acoustic differences are the duration of the silent or nearly silent portion

2

of the syllable corresponding to vocal tract closure, and the duration of the vocalic portion of the syllable preceding the closure.

When a voiceless stop (like /p/) is produced in the middle of a word, the vocal tract is held closed longer than when a voiced stop (like /b/) is so produced (Lisker, 1957; Port, 1976; also Slis & Cohen, 1969, for Dutch). There is already considerable evidence that, other things being equal, a long closure interval will perceptually cue a voiceless stop and a short closure interval will perceptually cue its voiced counterpart. Take, for example, the now classical minimal pair of rabid versus rapid. When a continuum of silent intervals is substituted for the acoustic segment corresponding to vocal tract closure, perceptual judgments systematically shift from rabid to rapid as the amount of silence, or "closure" duration, increases (Lisker, 1957; Port, 1976, 1978, 1979).

A voiceless stop is also produced with a shorter period of sound before the closure (House & Fairbanks, 1953; Denes, 1955; Peterson & Lehiste, 1960; House, 1961; Raphael, 1972; Klatt, 1975; Umeda, 1975; Port, 1976; also Delattre, cited by Belasco, 1953, for French; Zimmermann & Sapon, 1958, for Spanish; and Slis & Cohen, 1969, for Dutch). (This pre-closure vocalic section is often referred to as the "vowel," and will be so termed here for convenience.)[1] Although there is no direct evidence that vowel duration is a cue for intervocalic voicing, it is clear that when there is no second syllable after the stop, a short vowel cues a voiceless stop and a long vowel cues a voiced stop (Denes, 1955; Raphael, 1972). Lengthening the vowel in "gape," for example, can make it sound like "Gabe" (Raphael, 1972). It is reasonable, then, to expect that a change in either the closure duration or the vowel duration will cue voicing (see Lisker, 1978).

The problem as outlined above, however, is that the durations of the closure and the vowel segments do not change only with contrasts in phonemic voicing. Contrasts in speaking rate are also marked by changes in the durations of acoustic segments. As speaking rate slows, both the closure and the vowel parts of the word lengthen (Peterson & Lehiste, 1960; Gaitenby, 1965; Kozhevnikov & Chistovich, 1965; Port, 1976; Gay, 1978). It has been commonly assumed (but not, to my knowledge, verified) that the longer a segment the slower the rate of speech cued. (Although duration manipulations of various types have been interpreted as rate changes, experiments employing these duration manipulations have usually assumed the change in perceived rate and have measured the change in phonetic judgments (cf. Lindblom & Studdert-Kennedy, 1967; Ainsworth, 1973, 1974; Summerfield, 1974, 1975a, 1975b, Note 1; Fujisaki, Nakamura, & Imoto, 1975; Verbrugge & Isenberg, 1978; Miller & Grosjean, 1979; Miller & Liberman, 1979). In the few experiments in which rate judgments have been explicitly elicited, the only duration manipulations have been on the pauses between words (Grosjean & Lane, 1974, 1976). Thus, as Miller (in press) says in her recent and thorough review of rate effects, "the nature of the information that actually specifies tempo...has not been made explicit.")

A confound is thus established between voicing and rate as they relate to closure duration, and between voicing and rate as they relate to vowel duration.

3

The hypothesis put forward is that the confounding is only apparent, and due to the descriptors chosen. The choice of an individual acoustic segment as the object of attention and of duration as the variable used to characterize it are, after all, arbitrary choices. And a consideration of the way in which speech is produced--an analysis of the source event--suggests that those choices may not be best.

Take first the question of the appropriate unit of analysis. A single phonetically significant speech act usually produces a number of acoustic segments. The acoustic information for a phoneme, therefore, is rarely confined to just one acoustic segment, but is more often distributed over a wider temporal cross section of speech. Note, for example, that information about the nature of an intervocalic stop is carried in the vowel <u>and</u> the closure segments. While each part of the total acoustic consequence may partially specify the phoneme--each may be a perceptual cue for the phoneme-- no one part alone results in what is heard as "the phoneme." To restrict an acoustic analysis to one segment at a time may be to exclude from analysis the very aspects of the signal that are perceptually invariant. Therefore, the first strategy will be to consider a more temporally extensive description.

Next, take the question of the variable appropriate to describe a given stretch of speech. Again taking direction from speech production, it seems likely, at least according to some, that temporal duration is a measurable result of a movement or act but not a variable regulated by an actor (Fowler, 1977, 1980; Fitch, 1980; see also Fitch & Turvey, 1978; and Kugler, Kelso, & Turvey, 1980, for a discussion of this point in relation to motor coordination in general, and see Bernstein, 1967; Greene, 1972; Turvey, 1977a; and Turvey, Shaw, & Mace, 1978, for concepts of motor coordination that are the basis for this view). If it is not duration, per se, that is regulated, it may also be the case that it is not duration, per se, to which a perceiver of that act is actually sensitive. "Duration," being one acoustic consequence, may again be a cue for a phoneme but not a specification of it. Therefore, the description will not be limited to that one, single-dimensional consequence. Instead, variables will be used that take relationships among segments into account. Such higher-order variables may carry much of the character of the event, in that they may be the signature of the regulatory variables in effect.

Temporally extensive higher order variables allow the characteristics of more than one consequence of a single speech act to be incorporated into a single descriptor of that act. This type of description is more nearly compatible with the unitary nature of the phoneme heard. It is hoped that this will more closely approximate an invariant phonetic description--one unconfounded with rate.

This search for a different, more nearly invariant, description of the acoustic signal is motivated by the hypothesis that information for both speaking rate and phoneme identity are, for a speech perceiver, unambiguously present in that acoustic signal, and that an appropriate description will allow us as theorists to understand how we as hearers distinguish the two.

4

When faced with the apparent confounding, it is tempting to say that we are able to interpret the phonetic message by virtue of a knowledge of rate ("Since this is fast, it must be /p/"). The cue is disambiguated by its context. But the question as to how that "context" is specified remains. That it is "context" and not the "cue" itself is due simply to the focus, or definition, of the problem. One could as easily say that we are able to perceive speaking rate by virtue of a knowledge of the phonemes ("Since this is /p/, it must be fast"). Certainly a metric such as "number of phonemes per second" would seem a reasonable basis for rate perception, were the presupposed phonemic knowledge not a concern.

Is there information to specify both rate and phonetic identity? Inspiration is taken here from James J. Gibson's conviction that there is information that specifies important aspects of the source event to an appropriately attuned perceiver (Gibson, 1950, 1966, 1979; Turvey 1977b; see Turvey & Shaw, 1979, for a philosophical extension of this point that constitutes a reformulation of perception). The "depth perception" problem in vision might serve as a helpful analogy to the problem at hand. The problem is that a given size retinal image relates ambiguously to the object that produces it. That object could be small and close by, or large and far away. The one physical variable (retinal image size) corresponds to two perceptual dimensions (object size and object distance). Now, in accordance with the cue-normalized-by-context scheme of things, it could be (and has been) said that distance can be perceived by virtue of a knowledge of the normal sizes of objects ("Since this is a house [which is large], it must be far away"). Alternatively, it could be (and has been) said that size can be perceived by virtue of a knowledge of distance ("Since that is far away, it must be large"). (A knowledge of distance is usually invoked courtesy of prior experience gained through touch.) If retinal image size parallels closure duration, the difference between a close, small object and a distant, large object can be likened to the difference between a slow /b/ and a fast /p/.

A redefinition of the optic variable makes the vision problem tractable. Rather than confining the description to the temporally unextended, first-order variable of retinal size, a description compatible with the concern for "source event" may be used. Considering that the object and the eyeball will be moving relative to each other in a temporally extended event (either because the object is moving toward the person, or the person is moving toward the object), the rate of expansion of the retinal image can be defined. The rate of expansion of the retinal image of a close small object is not the same as the rate of expansion of the retinal image of a large distant object as they are approached at the same velocity by the perceiver. The closer object will have a greater rate of optical expansion than the farther object (Schiff, 1965; Lee, 1974). Thus, distinguishing the two becomes possible when a temporally extended source event (rather than a retinal snapshot; see Turvey, 1977b) is described in terms of a higher-order variable (rather than simply size). It is hoped that, likewise, a redefinition of the acoustic variables will make the speech problem tractable.

EXPERIMENT I

Introduction

Experiment I has two purposes. The first is to verify the perceptual salience of the two acoustic variables already described (1. closure duration, and 2. vowel duration) for both perceptual dimensions (a. voicing, and b. rate). Acoustic variable 1 is tested in condition 1 of this experiment. Does closure duration contribute to the perception of voicing? Does it contribute to the perception of rate? Acoustic variable 2 is tested in condition 2 of this experiment. Does vowel duration contribute to the perception of voicing? Does it contribute to the perception of rate? (See Figure 1.) Positive answers to these questions would establish the lack of a one-to-one correspondence between the acoustic signal (as defined by these variables) and the resulting percept.

Acoustic variable   Perceptual dimension
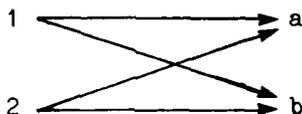
Figure 1

The second purpose is to see whether a more nearly one-to-one correspondence between signal and percept can be established by choosing different descriptors of the acoustic signal. To this end, two new variables will be defined. One is based on an inverse relationship between closure and vowel durations, and the other is based on a direct relationship between closure and vowel durations.

Recall that a long closure accompanies a _voiceless_ stop and a _slow_ rate of speech. This one _acoustic variable of closure duration_ correlates with both voicing and rate. Vowel duration, also, is a correlate of both voicing and rate; a long vowel accompanies a _voiced_ stop and a _slow_ rate of speech. But notice that the pattern of duration change that accompanies a voicing contrast is different from that which accompanies a rate contrast. A change from /b/ to /p/ lengthens the closure and shortens the vowel; a slowing of rate lengthens both. The inverse relationship between closure and vowel durations in a voicing contrast means that the ratio of these two durations will change, although their total duration may not. On the other hand, the direct relationship between closure and vowel durations in a rate contrast guarantees that total duration will change.

6

This difference provides a solution in principle to the problem of perceptually differentiating rate and voicing. The perceptual salience of this potential information is tested in the second two conditions of this experiment. The closure-to-vowel ratio (C/V) is tested in condition 3; total closure-plus-vowel duration (C+V) is tested in condition 4. Do both acoustic variables contribute to both perceptual dimensions (as above), or does one variable indicate voicing and the other indicate rate? (See Figure 2.) The hypothesis is that, since one type of temporal pattern (inverse closure/vowel relationship) corresponds to a voicing contrast, and a different temporal pattern (direct closure/vowel relationship) corresponds to a rate contrast, it should be possible to create one pair of stimuli that are easy to discriminate in terms of voicing but not rate, and another pair of stimuli that are easy to discriminate in terms of rate but not voicing.

Acoustic variable   Perceptual dimension

3 ─────────────▶ a

4 ─────────────▶ b

Figure 2

Method

Each condition was composed of one pair of stimuli, corresponding to one of the acoustic variables described above. Thus there were four pairs of stimuli in all. In one pair, closure duration was varied while vowel duration was held constant. In another, vowel duration was varied while closure duration was held constant. A third pair of stimuli was created by varying the closure-to-vowel ratio of the two members (thus embodying the inverse voicing relationship), while equating the closure-plus-vowel duration. The fourth pair of stimuli was created by varying the closure-plus-vowel duration of the two members (thus embodying the direct rate relationship), while equating the closure-to-vowel ratio.

These stimuli were made from recordings of a woman saying the nonsense words /dabi/ (pronounced "dah' bee") and /dapi/ (pronounced "dah' pee") in a sentence frame. These recordings were digitized, and out of each were electronically spliced the parts necessary for building the stimuli for the four conditions.

The following considerations determined the construction of the stimuli. First, cues other than duration to the voicing distinction must not overpower the potential effects of duration. Therefore, the first syllable of /dabi/ and the second syllable of /dapi/ were used. Thus, the formant transitions into the closure were more suggestive of /b/, but the formant transitions out of the closure were more suggestive of /p/ (no burst was included). In addition, the procedure of using a silent closure interval (as in previous experiments reported in the Introduction) was adopted. This prevented any voicing during the closure from overpowering the other cues (see Lisker & Price, 1979), and also made it easy to manipulate closure duration.

Second, the specific durations used must not preclude evidence of the potential effects of duration by falling totally into one or the other perceptual category. In other words, the ranges of durations chosen must span the /b/-/p/ perceptual boundary, at least for most subjects.

The third consideration was that the durations chosen had simultaneously to satisfy the various constraints imposed by each of the four conditions. Thus, for example, while it would have been possible to use one extremely short and one extremely long closure duration if only closure duration was being tested, the choice of those values was here guided by the requirement that, in another condition, the difference between "short" and "long" closure had to match the difference between "short" and "long" vowel in order to equate total stimulus duration. In other words, [long vowel plus short closure] had to equal [short vowel plus long closure].

The duration of the vowel was varied by having the recorded words spoken at two different rates: conversational, and slow. The duration of the closure was varied by computer manipulation, using a program that allows insertion of the desired amount of silence into a file (Szubowicz, Note 2). The duration of the second syllable was not varied. It was taken from the sentence recorded at the conversational rate, and was 174.1 ms.

Pilot testing was done to determine appropriate durations, and the following four pairs of stimuli (one for each condition) were created (see Figure 3).

Condition 1. Closure duration. The first pair of stimuli was created to test the perceptual salience of closure duration. One member of the pair had a "short" closure of 70 ms; the other member of the pair had a "long" closure of 112 ms. The vowel was the same for each: the 254 ms slow /dab/.

To the extent that closure duration is a cue to voicing, the member with the short closure should sound more like /dabi/ and the member with the long closure should sound more like /dapi/.

To the extent that closure duration is a cue to rate, the member with the short closure should sound faster and the member with the long closure should sound slower.

Condition 2. Vowel duration. A second pair of stimuli was created to test the perceptual salience of vowel duration. One member of the pair had a "short" vowel of 212 ms, which was the conversational rate /dab/; the other

8

Figure 3. Schematic of stimuli for the four conditions of Experiment I.

had a "long" vowel of 254 ms, which was the slow /dab/. Both had a 112 ms closure.

To the extent that vowel duration is a cue to voicing, the member with the short vowel should sound more like /dapi/ and the member with the long vowel should sound more like /dabi/.

To the extent that vowel duration is a cue to rate, the member with the short vowel should sound faster and the member with the long vowel should sound slower.

Condition 3. Closure-to-vowel ratio. A third pair of stimuli was created to test the perceptual salience of C/V. One member of the pair had the short (212 ms) vowel and the long (112 ms) closure; the other member of the pair had the long (254 ms) vowel and the short (70 ms) closure. The [short vowel, long closure] stimulus had a closure-to-vowel ratio of about .5. The [long vowel, short closure] stimulus had a closure-to-vowel ratio of about .3. Both had a closure-plus-vowel duration of 324 ms.

To the extent that C/V is a cue to voicing, the stimulus with the large C/V ratio should sound more like /dapi/, and the stimulus with the small C/V ratio should sound more like /dabi/.

Condition 4. Closure-plus-vowel duration. The fourth pair of stimuli was created to test the perceptual salience of C+V. One member of the pair had the long (254 ms) vowel and the long (112 ms) closure, making the total closure-plus-vowel duration 366 ms. The other member of the pair had the short (212 ms) vowel and a 96 ms closure, making the total closure-plus-vowel duration 308 ms. The "short" closure in this case was somewhat longer than the "short" closure in the other conditions so that the closure-to-vowel ratios of the stimuli would both be about .4.[2]

To the extent that C+V is a cue for rate, the long stimulus should sound slower and the short stimulus should sound faster.

Twenty tokens of each of the four pairs (10 tokens in each order) were randomized and recorded on audio tape. Each of the resulting 80 pairs of stimuli constituted one trial of a listening test. There was a 1 sec pause between the members of each pair, and a 3 sec pause between trials. There was a longer pause after every 20 trials, separating the test into 4 lists.

This test tape was played twice. The first time subjects were asked to judge which member of each pair sounded faster. After each trial, they were to check the first column on an answer sheet if the first word sounded faster, or to check the second column on the answer sheet if the second word sounded faster. The second time the tape was played, subjects were asked to judge which member of each pair sounded more as if it contained /p/ (rather than /b/), and to mark the answer sheet appropriately after each trial.

Subjects were volunteers from an introductory psychology class, paid for their participation. All were native speakers of American English and had no known hearing loss. Fourteen subjects participated.

10

Results and Discussion

The difference between the proportion of responses accorded one member of a pair and chance (50%) was assessed by t-test. Two t-tests were performed on each pair of stimuli; one tested whether there was a significant effect on the proportion of "P" responses, and the other tested whether there was a significant effect on the proportion of "faster" responses.

First consider the results of conditions 1 and 2. As expected, both closure duration and vowel duration contributed to the perception of both voicing and rate. That is, subjects were able to make both reliable voicing judgments and reliable rate judgments when either duration alone was varied. In condition 1, closure duration significantly affected voicing judgments, $t(13)=6.70$, SE=.97, p<.001, with the long-closure stimulus sounding more like /dapi/, and it significantly affected rate judgments, $t(13)=5.69$, SE=.51, p<.001, with the short-closure stimulus sounding faster. In condition 2, vowel duration significantly affected voicing judgments, $t(13)=5.13$, SE=.78, p<.001, with the short-vowel stimulus sounding more like /dapi/, and it significantly affected rate judgments, $t(13)=11.57$, SE=.70, p<.001, with the short-vowel stimulus sounding faster.

Thus, previous results showing that intervocalic voicing is cued by closure duration were corroborated. The inference that intervocalic voicing is cued by the duration of the previous vowel was justified. The two assumptions about the perception of rate were verified; closure duration contributes to the perception of rate, and vowel duration contributes to the perception of rate. In summary, all four relationships between acoustic variables and perceptual dimensions, as diagrammed in Figure 1, were highly significant. The potential confounding on which the puzzle addressed in this thesis rests is thereby established.

These results, if examined more closely, however, also offer the first hint that voicing and rate are not supported similarly in these temporal aspects of the acoustic signal. While it is true that both dimensions are significantly affected by both acoustic variables, it is interesting that not all four relationships are equally strong. Closure duration affected voicing more than rate, but vowel duration affected rate more than voicing. This can be seen by examining the left half of Figure 4. In condition 1, the closure duration difference led to a 64% difference (82% versus 18%) in how often the two stimuli were heard to be more p-like; it led to only a 30% difference (65% versus 35%) in how often the two stimuli were heard to be faster. Conversely, in condition 2, the vowel duration difference led to an 82% difference (92% versus 9%) in how often the two stimuli were heard to be faster, but it led to only a 40% difference (70% versus 30%) in how often the two stimuli were heard to be more p-like.

Turn next to the results of the third and fourth conditions. If defining the acoustic signal in terms of these two variables had been totally successful in distinguishing temporal information for rate and temporal information for voicing, condition 3 would have resulted in completely consistent judgments of which member of its pair sounded more p-like, and in no significant difference between the members in terms of which was judged faster; condition 4 would have resulted in completely consistent judgments of

11
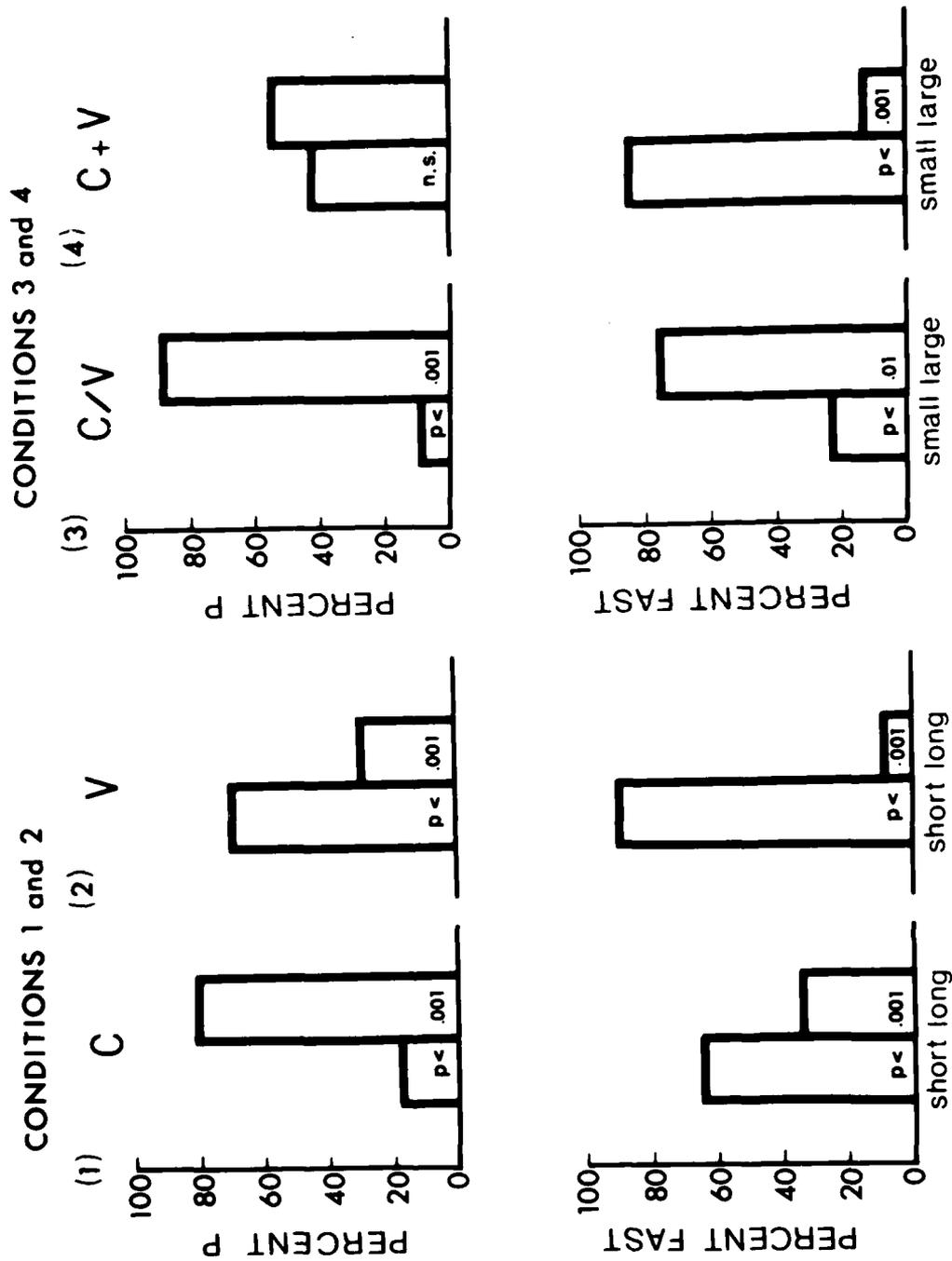
Figure 4. Results of Experiment I.

12

which member of the pair sounded faster, and in no significant difference in which was judged more p-like.

As expected, C/V (variable 3) significantly affected voicing judgments, $t(13)=10.00$, SE=.78, p<.001. It may be noted that this variable led to better discrimination of voicing than did either closure duration (variable 1) or vowel duration (variable 2). While all three of these acoustic variables produced highly significant voicing results (p<.001), the ratio condition produced a larger t-value (10.00) and more of a difference between pair members (78%) than did the other two (6.70 and 64% for closure duration; 5.13 and 40% for vowel duration). C/V also (but to a lesser extent) significantly affected rate judgments, $t(13)=4.02$, SE=1.27, p<.01, producing a 42% difference (76% versus 24%) between pair members. This was not anticipated, but is understandable given the results of the first two conditions. Since closure and vowel durations contribute unequally to the perception of rate, changing the ratio of these two segments upsets the perceptual balance.

The effect of C+V (variable 4) on rate judgments was highly significant, $t(13)=9.36$, SE=.78, p<.001. It led to a 72% difference (86% versus 14%) between pair members. As hypothesized, however, the change in total duration (with C/V held constant) did not significantly affect voicing judgments, $t(13)=1.43$, SE=.84, p>.1.

Thus, C/V was more effective in allowing the discrimination of voicing than of rate, and C+V was effective in allowing the discrimination of rate but was not effective in allowing the discrimination of voicing. Although some asymmetry was also noted in the percentage scores of conditions 1 and 2, there is a suggestion from the patterns of significance that variables 3 and 4 produce more differentiation of the rate and voicing results than do variables 1 and 2.

In summary, this experiment demonstrates that the temporal aspects of phonetic information may be distinct from the temporal aspects of rate information.

EXPERIMENT II

Introduction

Experiment II pursues the distinction between temporal information for voicing and temporal information for rate.

While there was a suggestion in Expermiment I that the perceptual dimensions of voicing and rate were better differentiated by the acoustic variables of closure duration-to-vowel duration ratio (C/V) and closure duration-plus-vowel duration sum (C+V) than by the acoustic variables of closure duration and vowel duration, most of that improvement was due to the contribution of the variable C/V. Varying C/V led to more consistent voicing judgments than varying closure duration or vowel duration, and equating C/V led to rate judgments not significantly different from chance. The complement was not the case for the variable C+V. In fact, simply varying vowel duration led to slightly better rate discrimination than did varying both closure and

13

vowel duration. Yet we know that vowel duration as information for rate is confounded with voicing: the vowel could be shorter because the rate of speaking is faster, but it also could be shorter because the syllable-final consonant is devoiced.

Doubt about segment duration as an appropriate and adequate variable led, in Experiment I, to a consideration of temporal patterns defined over a grain coarser than the individual segment. It now prompts an investigation of temporal patterns at a grain finer than the individual segment.

If "duration" per se really is the variable to which a perceiver is sensitive, then how a segment (in this case, the vowel) gets its duration should not matter. If it is "duration" per se that is regulated, the same result would obtain whether a decrease in duration was due to a decrease in rate or due to devoicing. A particular vowel duration, no matter what its linguistic origin, would always be produced in the same way, by a specification of a duration parameter.

On the other hand, if duration is merely a by-product of what is regulated, the same duration could arise in more than one way. This allows room for the possibility that a change in vowel duration due to rate is distinguishable from a change in vowel duration due to voicing. That is, while vowel duration is one measurable consequence of a change in rate, and it is one measurable consequence of a change in voicing, presumably the articulatory dimension used to regulate rate is different from the articulatory dimension used to regulate voicing. While both articulatory dimensions may overlap in their vowel duration consequences (and so duration can cue both), their overall patterns of change within the vowel (as well as over more temporally extensive stretches) might be different.

An analogy might help. Think of two springs, each of which is moving a mass (like a block of wood) attached to its end. These two mass-spring systems are meant to represent the speech producing system at two different times. The spring sytems can vary on two dimensions. One is the stiffness of the spring itself, and the other is the resistance against which the mass is moving. These are meant to represent the articulatory dimensions used to control rate and voicing.

Now say that the stiffer spring is also the one that is moving its mass against less resistance. If the springs are pulled and then released, setting the masses in motion, they might return to their resting positions in the same amount of time. The duration of that motion would be one measurable consequence of each system. But there would be differences in the pattern of each motion. The duration of the flight back to equilibrium would be the same, but the flight pattern would be different. A stiff spring moving against a low resistance would be distinguishable from a loose spring moving against a high resistance, even if their movement durations were the same.

Is vowel duration, likewise, a variable of result rather than of regulation, and is a syllable-final /p/ said slowly distinguishable from a syllable-final /b/ said rapidly, even if the syllable (or "vowel"; see footnote 1) durations are the same? Perhaps. There is reason to believe, from activities other than talking, that a change in the rate of an activity

14

is brought about by the regulation of an underlying dynamic variable (like stiffness and resistance in the previous example), which in turn gives rise to a differentiated pattern of kinematic results (like duration) (see Runeson, 1977; and Fitch & Turvey, 1978, for a discussion of dynamic versus kinematic variables). Locomotion is an extensively studied activity in which this is demonstrated. The step cycle of an individual limb is one complete cycle of stepping--up and forward, down and back. Two components of that cycle are easily discernible: the stance phase, when the foot is planted on the ground and the body is moving over it, and the swing phase, when the foot is off the ground (Philippson, 1905). As the rate of locomotion increases, the duration of the total step cycle decreases (there are more steps per minute), and the distance covered during that cycle increases. An analysis of the stance and swing phases shows that there is differentiation within those overall changes. The duration of the stance phase decreases, but the duration of the swing phase stays almost the same. Conversely, the distance covered during the stance phase stays about the same, but the distance covered during the swing phase increases (Grillner, 1975; Shik & Orlovsky, 1976). Now, it turns out that both these results--the change in the duration of the stance phase and the change in the distance covered during the swing phase--can be rationalized by a change in just one underlying variable. That is the amount of force applied at the beginning of the stance phase (Orlovsky, Severin, & Shik, 1966; Shik & Orlovsky, 1976). When the force of the leg against the ground is increased at the beginning of the stance phase (at which time the foot is on the ground in front of the body), the body is propelled over the foot in a shorter amount of time (the duration of the stance phase decreases). Of course, since the foot is planted, the distance that the body can travel over the foot during that phase is limited. Therefore, since the same distance is covered in a shorter amount of time, more thrust is developed, and the body automatically travels a further distance once the foot leaves the ground for the swing phase (the distance covered during the swing phase increases). Thus, a change in just one variable (force) can create a differentiated pattern of results within the overall cycle that corresponds to a change in the rate of locomotion.

So, to return to the question of vowel duration, perhaps a change in the rate of speaking, like a change in the rate of locomotion, is caused by a variable that gives rise to a differentiated pattern of temporal results within the total vowel duration. If a rate change is not a duration change per se--if duration is instead but the result of a rate-producing mechanism, and if that mechanism produces temporal patterns different from other articulatory mechanisms--then equating duration would not necessarily equate perceived rate, and a duration change due to rate would be potentially differentiable from other duration changes. On the other hand, if a change in rate is, in fact, a change in duration, then equating duration would equate perceived rate.

To test the hypothesis that the effect of a change in speaking rate is not duplicated by an equivalent, but differently implemented, change in duration, vowel duration was varied orthogonally from produced speaking rate. The non-rate duration change was effected on the vowel nucleus only, by computer.

Now let us return to the question of C/V as information for voicing. Remember that the objective is not to find simply another cue (albeit a more effective one) for intervocalic voicing. It is to find temporal information for voicing that is not confounded with temporal information for rate.

There is both theoretical and empirical support for the notion that C/V is a rate-invariant signature of the state of the voicing mechanism. First, since it is a relational variable, it carries the possibility of preserving the essence of an act while allowing its details to change. The details--in this case, the absolute durations of closure and vowel--would be free to vary as necessary (within the prescribed constraint of the relationship) to accommodate changes in rate. There is encouraging evidence, again from walking rather than talking, that relational variables like ratios can characterize an activity such that they do remain invariant as rate is changed. As the velocity of locomotion changes, there are, as one might expect, changes in the amplitudes of the electromyographic (EMG) records of muscle activity in the leg. The EMG amplitudes, in other words, are rate-dependent. However, the ratios of the EMG amplitudes of the leg extensor muscles do not change as the running speed of the animal increases or decreases (Engberg & Lundberg, 1969; Grillner, 1975). Those ratios do, of course, change when the animal switches to an activity other than locomotion, thus moving its legs in a characteristically different style. The EMG ratios, therefore, are rate-invariant characteristics of locomotion. In another example, the ratios of EMG activity in the muscles of the hip, knee, and ankle joints show a similar invariance during the act of regaining balance after different perturbations (Nashner, 1977). These examples illustrate how a pattern or relationship, such as that expressed by a ratio, can be a signature of the regulatory constraints in effect. (For instances in speech production of other kinds of relationships that are preserved by such "coordinative structures," see Fowler, 1980.)

While any expression of a closure-vowel relationship would be a potential candidate for a rate-invariant voicing signature, there is empirical evidence to favor C/V in particular. This evidence comes from work by Port (1978) on the rabid-rapid contrast cited earlier. Port recorded two sentences containing the word rabid, one at a slow speaking rate and the other at a fast speaking rate. Both rabids were excised, made into test continua by substituting a range of silent intervals for the naturally produced closure, and reinserted into the sentences. When the slow rabid was substituted for the fast rabid, a longer closure was needed to make it sound like rapid. However, when the voicing judgments were considered not in terms of closure duration but in terms of the ratio of closure duration to vowel (/rab/) duration, the perceptual boundary between /b/ and /p/ for the word spoken slowly and the word spoken rapidly were close.

To see whether C/V is rate-invariant information for voicing, the duration of the closure that is needed to change /dabi/ to /dapi/ at different speaking rates will be examined. To the extent that C/V is rate-invariant information for voicing, the perceptual boundary between /b/ and /p/ should fall at the same ratio for all (naturally produced) rates.

Notice that a further hypothesis may be drawn at this point. It relates to the fact that the rate-invariance of C/V is being tested, and the

16

information for rate itself is being questioned here. If C/V is rate-invariant, but the artificial duration change is not perceptually equivalent to the naturally produced rate change, the perceptual boundary between /b/ and /p/ might not fall at the same ratio in the computer manipulated conditions. Perhaps only a rate change causes that kind of duration change that preserves voicing information such that the ratio bounding /b/ and /p/ is unaffected. Another kind of duration change might, in fact, happen to resemble something of a voicing change, which would certainly interact with other voicing information. Therefore, while it might very well turn out to be the case that a constant closure-to-vowel ratio perceptually bounds /b/ and /p/ at different speaking rates, it might also be true that that relationship is disturbed (voicing judgments are shifted) when the vowel duration involved in the ratio does not arise from saying the same word at a different rate.

To the extent, then, that this duration change is not the equivalent of a rate change, the voicing boundary should fall at different amounts of closure for the same-duration vowels.

Method

Three speaking rates were used in this experiment. They were determined by making preliminary recordings of the test sentences at what the talker considered a comfortable conversational rate, at what she considered the fastest rate she could produce without deleting phonemes, and at what she considered the slowest rate she could produce without sounding very unnatural or inserting pauses. The average duration of the slow sentences was 26% longer than the average duration of the conversational rate sentences, and the average duration of the fast sentences was 15% shorter than the average duration of the conversational rate sentences. This is roughly in line with data obtained by Port (1976) using a similar procedure. He found slow sentences to be about 20% longer and fast sentences to be about 30% shorter than conversational rate sentences. These rates were then matched to metronome settings by adjusting the metronome beats to coincide with the stressed syllables. The sentences were constructed to have a regular rhythm. They were: "I think' that it sounds' like a dah'bee," and "I think' that it sounds' like a dah'pee." The metronome rates were then used to control the final recording. The slow rate was 92 beats per minute, the conversational rate was 120 beats per minute, and the fast rate was 160 beats per minute. Both sentences were recorded at each of the three rates (making six sentence types). Three tokens of each sentence type were produced and the recordings digitized.

As in Experiment I, stimuli were built from the first vocalic section (or "vowel"; see footnote 1) of /dabi/, an interval of silence, and the second vocalic section of /dapi/. One /dab/ from each speaking rate and the /pi/ from the conversational rate were spliced out of the sentences for this purpose. The median duration tokens of the three recordings in each category were used.

To carry out the orthogonal rate x duration design of the experiment, pitch pulses were either duplicated or deleted from each /dab/ so as to match the durations of the /dab/'s from the other two rates. For example, a sufficient number of pitch pulses was deleted from the slow /dab/ to match the

duration of the conversational rate /dab/, and then more pitch pulses were deleted to match the duration of the fast rate /dab/. Thus, there were altogether nine /dab/'s: three speaking rates (slow, conversational, fast) x three durations (long, medium, short). The editing was performed on the steady-state, vowel nucleus part of the syllable. This region was determined on spectrograms by drawing a line parallel to the time axis through the first formant. (A spectrogram of the slow /dabi/ is shown in Figure 5.) The amplitude envelope of the syllable and the waveform shape of the pitch pulses also aided in identifying the regions of least change.

A female talker was used, and pitch pulses averaged 5 ms. It was thus possible to match durations to within less than 3 ms. The actual durations of all nine /dab/'s are shown in Table 1.

Table 1

Duration of /dab/ in ms for the nine conditions of Experiment II.

Final Duration	Original Speaking Rate		
	Slow	Conversational	Fast
Long	254.4	256.6	254.0
Medium	210.2	211.9	213.0
Short	175.6	177.1	177.9

To each of these nine /dab/'s were appended from 50 ms to 100 ms silence (in increments of 10 ms), and the (constant) second syllable, creating nine /dabi/ to /dapi/ continua. There were 54 stimuli in all: 9 continua x 6 intervals of silence in each. Ten tokens of each stimulus, or 540 stimuli in all, were randomized and recorded on audio tape. There was a 3 sec pause between stimuli during which time subjects were to mark "B" if the word on that trial sounded more like /dabi/, or "P" if the word on that trial sounded more like /dapi/. The test was broken into 20 lists, with a longer pause between lists.

Eleven volunteers from introductory psychology courses participated in the experiment for course credit. All were native speakers of American English and had no known hearing loss.
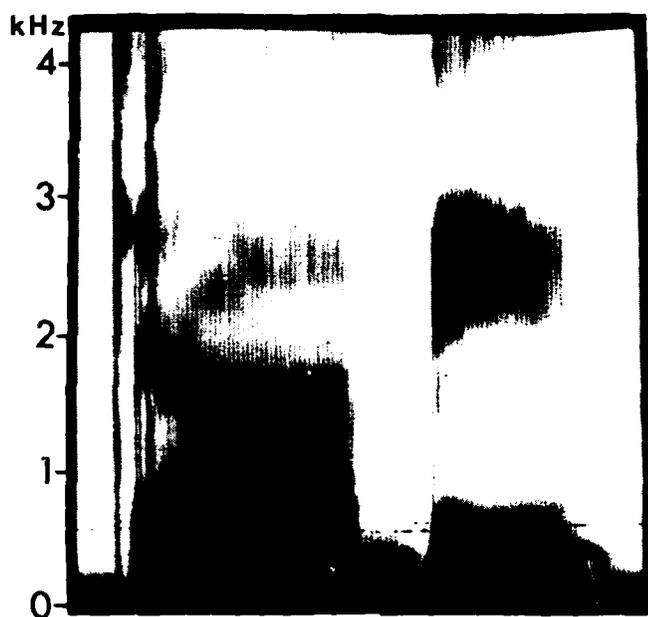
dαbi

Spectrogram of slow /dαbi/

Figure 5. Spectrogram of slow /dabi/.

19

Results and Discussion

To assess whether both the duration and the rate factor were significant, a three by three repeated measures analysis of variance was performed on the number of "P" responses in each condition. The duration factor was highly significant, $F(2,20)=144.93$, $MSe=106.6$, $p<.001$, accounting for most of the variance. The rate factor was also highly significant, however, $F(2,20)=18.00$, $MSe=95.82$, $p<.001$. In other words, the original rate at which the word was spoken influenced the amount of silence needed to hear /p/ above and beyond the contribution to that judgment due to duration. Equating duration does not fully account for the effect of rate. Thus, the major hypothesis is supported. A change in vowel duration is not the perceptual equivalent of a change in speaking rate.

The rate x duration interaction also reached significance, $F(4,40)=7.14$, $MSe=18.30$, $p<.001$, but was small compared to the main effects, and did not appear to counteract their interpretation.

These results become clearer when displayed graphically. The total number of "P" responses in each condition can be plotted as a function of closure duration. This allows a picture of the responses to each stimulus rather than simply a summary of the responses to each condition. These nine identification functions, averaged over the 11 subjects, are shown in Figure 6.

It can be seen that each function increases regularly with closure duration. When that duration is short, there are few "P" responses. The stimuli sound like /dabi/. As closure duration increases, the stimuli sound less like /dabi/ and more like /dapi/. When closure duration is longest, "P" responses predominate. This is a parametric confirmation of the voicing results of the closure duration condition in Experiment I.

It can also be seen that the functions depicting the three short /dab/ conditions (dotted lines) are displaced to the left of the functions depicting the three medium duration /dab/ conditions (dashed lines), which are in turn displaced to the left of the three functions depicting the long /dab/ conditions (solid lines). This indicates that the shorter the vocalic section preceding closure (or "vowel"; see footnote 1), the less silence is necessary to hear /p/. This is a parametric confirmation of the voicing results of the vowel duration condition in Experiment I.

At each level of duration, the curves from the three original speaking rates (slow = wide, conversational = medium, fast = thin) are spread out. Their staggering is an indication of the effect due to original speaking rate, with duration held constant.

That the ordering of rates within a duration is not the same for all three levels of duration is an indication of the rate x duration interaction.

From each identification function it is possible to determine the perceptual (in this case, voicing) boundary for that condition, defined as that point along the continuum where "B" and "P" judgments are equally likely (50% "P" judgments). With less silence, /b/ is more likely to be heard; with

20
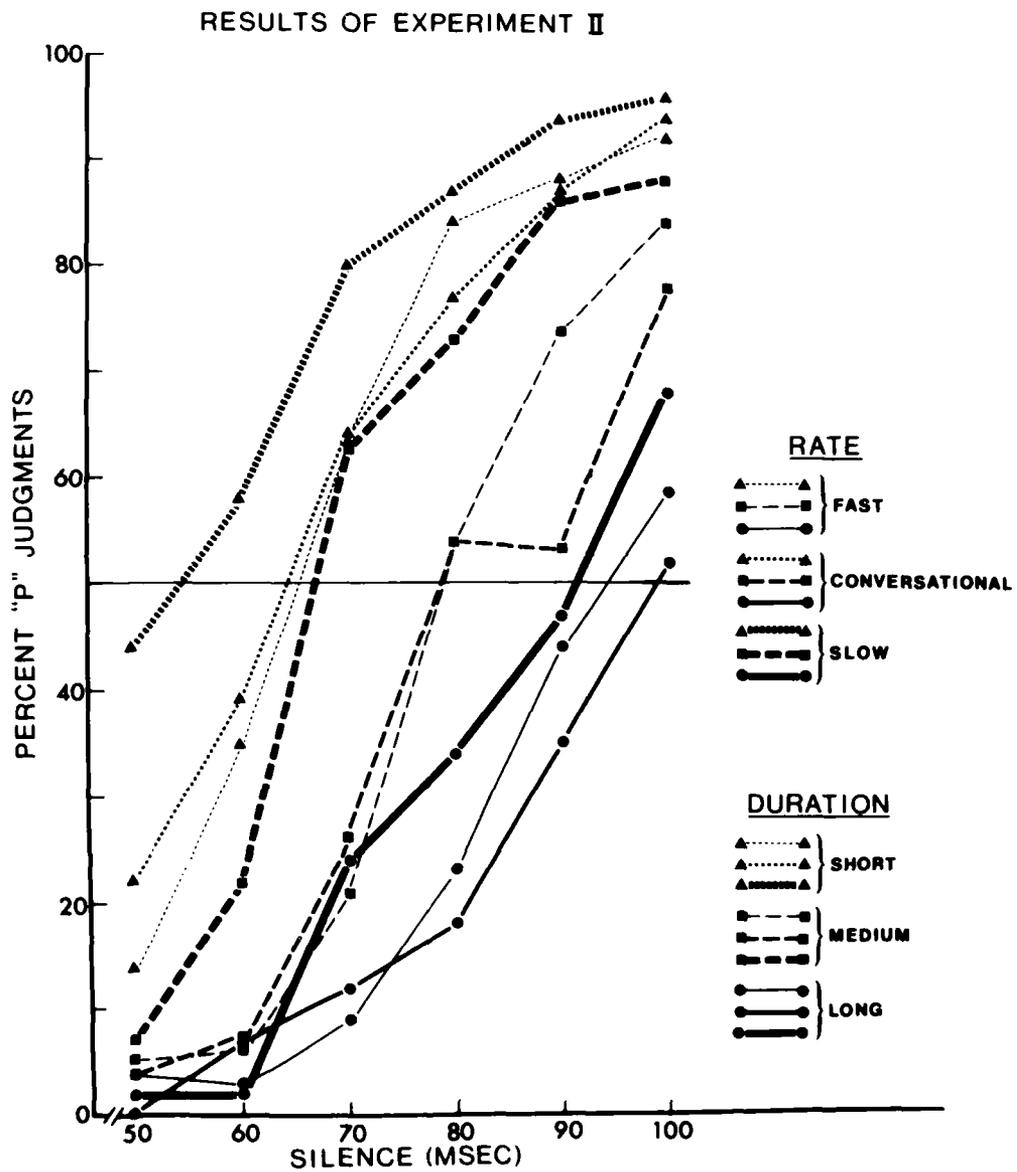
Figure 6. Results of Experiment II. Each curve shows the results of one
condition.

more silence, /p/ is more likely to be heard. The value of closure duration at this point is shown for each condition in Table 2. It can be seen, by looking at the three unaltered /dab/ conditions displayed along the diagonal, that the faster the speaking rate, the shorter the closure needed to hear /p/. The voicing boundary was at 92 ms in the slow condition, 78 ms in the conversational rate condition, and 65 in the fast condition. This, of course, is an illustration of how rate affects phonetic perception.

To answer the question of the rate-invariance of C/V, the voicing boundary was recalculated in terms of this acoustic variable. Table 3 shows the results of dividing the closure duration at the voicing boundary by the vowel duration for each condition. It can be seen that for the original speaking rate conditions, the voicing boundary was nearly rate invariant. It was at .36 in the slow condition, at .37 in the conversational rate condition, and at .36 in the fast condition. Thus, the rate-invariant nature of C/V is supported.

It can also be seen in Table 3 that the voicing boundary did not stay the same in the conditions where the vowel duration was the result of computer manipulation. The perceptual boundary ranged from .31 for some of the shortened syllables to .38 for some of the lengthened syllables. One might say that C/V is not invariant under a non-rate change.

These results lend support to the idea that different causes of changes in vowel duration are perceptually differentiable, due to differences in resulting temporal patterns within the vowel. The importance of this temporal differentiation at the finer grain can now be seen, but the nature of the difference between one temporal pattern and the other can only be inferred from the fact that the artificial duration change, presumably unlike the real rate change, was wrought on the vowel nucleus only. Experiment III explores this difference.

EXPERIMENT III

Introduction

The kind of fine-grained temporal patterning difference that may have been effective in Experiment II can be illustrated by contrasting two vowels of the same duration. Consider the originally slow but shortened /dab/ and the originally fast, unaltered /dab/. Remember that the originally slow but shortened syllable was shortened only in the steady-state region of the vowel nucleus. If an increase in speaking rate shortens the whole syllable to some extent, and not just the vowel nucleus, then the slow shortened syllable would have a disproportionately short vowel nucleus. The conjecture is that the relative durations of initial d-transitions and steady-state vowel nucleus are critical. This can be tested using synthetic speech. Rather than editing the waveform of real speech (as in Experiment II) and indirectly effecting formant changes, the formant structure will be directly manipulated with a formant synthesizer. The (/dab/) vowel duration will be held constant and the relative durations of the initial transitions and the steady-state vowel nucleus varied.

22

Table 2

Voicing Boundary for the nine conditions of Experiment II
in terms of ms closure

Final Duration	Original Speaking Rate		
	Slow	Conversational	Fast
Long	92	99	94
Medium	67	78	79
Short	54	64	65

Table 3

Voicing boundary for the nine conditions of Experiment II
in terms of C/V

Final Duration	Original Speaking Rate		
	Slow	Conversational	Fast
Long	.36	.38	.37
Medium	.32	.37	.37
Short	.31	.36	.36

23

Method

The synthetic stimuli used in this experiment were made by taking formant and amplitude measurements of the slow /dabi/ from Experiment II, and using these to control the parameters of the OVE III synthesizer at Haskins Laboratories. As in the previous experiment, the conditions differed in terms of the variety of /dab/ used. All were 170 ms long. In condition one, the /dab/ had long transitions and a short vowel nucleus. In condition two, the /dab/ had medium duration transitions and a medium duration vowel nucleus. In the third condition, the /dab/ had short transitions and a long vowel nucleus.

The three transition durations and three vowel nucleus durations were constructed as follows. The syllable as copied from real speech provided the longest version of both. The shorter transitions were formed by shifting the rising amplitude contour at the onset of the syllable farther into the syllable, thus, in effect, starting the syllable later into the formant transitions. A 10 ms (1 data frame) shift formed the medium duration transitions. A 20 ms (2 data frame) shift formed the short duration transitions. The shorter vowel nuclei were formed by deleting frames in the central, steady-state portion of the syllable. Ten ms (1 frame) were deleted to form the medium duration vowel nucleus. Twenty ms (2 frames) were deleted to form the short vowel nucleus.

Again, as in Experiment II, a /dabi/ to /dapi/ continuum was formed in each condition by appending a range of silent intervals, and the (constant) second syllable. In all conditions, the silent interval ranged from 10 ms to 90 ms in 20 ms increments. There were 15 stimuli in all: 3 continua x 5 intervals of silence in each. Ten tokens of each stimulus, or 150 stimuli in all, were randomized and recorded on audio tape. There was a 3 sec pause between stimuli during which time subjects were to mark "P" if the word on that trial sounded more like /dapi/, or "B" if the word on that trial sounded more like /dabi/. There was a longer pause between every 25 trials.

Eleven volunteers from introductory psychology courses participated in the experiment for course credit. All were native speakers of American English and had no known hearing loss.

Results and Discussion

The effect of the temporal patterning within the vowel was significant, as tested by a one-way repeated measures analysis of variance performed on the number of "P" responses in each condition, $F(2,20)=9.08$, $MSe=9.63$, $p<.005$. Thus, the amount of closure needed to hear /p/ differs even though the duration of the vowel preceding the closure is the same. One 170 ms token of the vowel is not perceptually equivalent to another 170 ms token of the vowel.

To show the direction of the difference, the identification functions are plotted in Figure 7. It can be seen that the shorter the d-transitions and the longer the vowel nucleus, the longer is the closure needed to hear /p/. The voicing boundary was at 41 ms in condition 1 (long transition, short vowel nucleus); it was at 43 ms in condition 2 (medium duration transitions, medium duration vowel nucleus); and it was at 49 ms in condition 3 (short transitions, long vowel nucleus). (These results are given in terms of closure

24

RESULTS OF EXPERIMENT III
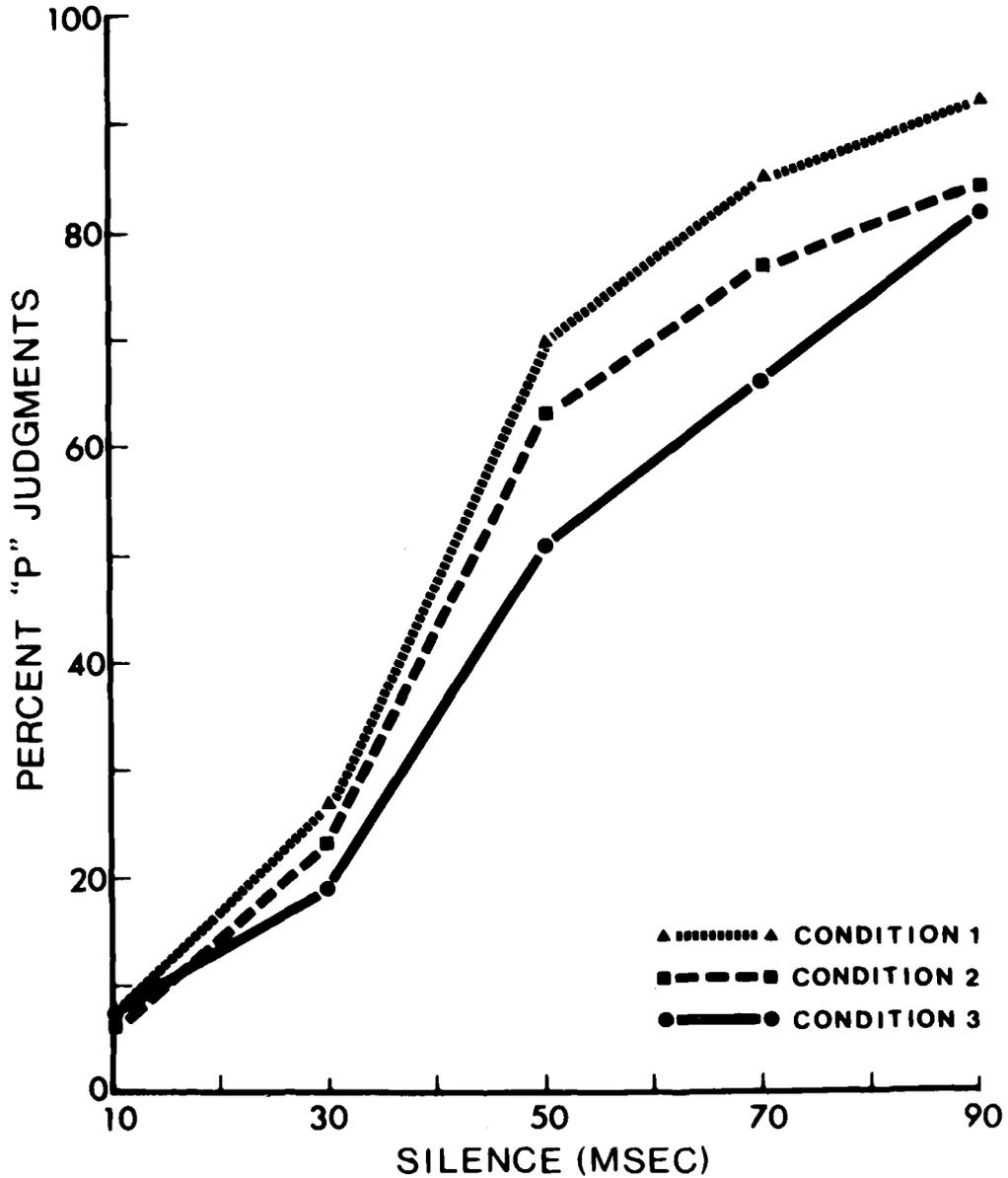
Figure 7.   Results of Experiment III.   Condition 1 = long transitions, short
           vowel nucleus.   Condition 2 = medium transitions, medium vowel
           nucleus.   Condition 3 = short transitions, long vowel nucleus.   25

duration, but of course since all vowel durations are the same, it is also true that the voicing boundaries are at different ratios of closure to vowel in all three conditions.)

The direction of these results is consistent with the results of Experiment II. They support the hypothesis that the difference in the perceptual effect of the artificial duration change and the rate change was due to the different resulting proportions of initial transitions and vowel nucleus. The slow, shortened syllable is represented by condition 1, and the fast syllable is represented by condition 3. Evidently longer transitions and a shorter vowel nucleus in the artificially shortened stimulus allowed /dapi/ to be heard at a shorter closure duration.

Ideally, one would like to be able to link these different patterns to different kinds of articulatory changes, as in the analogy of the mass-spring system (see Introduction to Experiment II). Unfortunately, such an attempt would be premature, since the lesser amount of silence needed to hear /p/ in condition 1 could be rationalized in at least two ways. On one account, the shorter vowel nucleus might be an indication of devoicing, thus requiring less silence to reach the voicing boundary. Alternatively, the shorter vowel nucleus might be an indication of a faster rate of speech, thus requiring less silence to reach the voicing boundary. Further work will be necessary to show how the relationship between transitions and vowel nucleus might distinguish rate and voicing. This should be guided by a consideration of production patterns, as was the investigation of the coarser-grained relational variables in Experiment I.

These results do confirm the inadequacy of simple vowel duration as a descriptor, and they confirm the importance of temporal relations within that duration.

GENERAL DISCUSSION

Let us now reconsider the effort with which we began: a redefinition of the acoustic signal. The purpose of this redefinition was to approach a description of the speech signal that makes clear the acoustic basis for the perception of rate and the acoustic basis for the perception of intervocalic stop consonant voicing. It was to come closer to a specification of the information for both rate and voicing, rather than to proliferate cues for each. The intent in that regard was limited; only temporal aspects of the information have been considered. A full specification is the ultimate but not the immediate goal.

The need for redefinition arises because the search for the acoustic basis of perceived linguistic units has proved so unyielding. The phoneme is an elusive creature; the conclusion that there does not exist a one-to-one correspondence between signal and percept has seemed inescapable (cf. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). This lack of correspondence is sometimes expressed as a one-to-many problem, wherein one acoustic cue relates to more than one perceptual dimension (e.g., closure duration relates to voicing and rate); and it is sometimes expressed as a many-to-one problem, wherein more than one acoustic cue relates to one

26

perceptual dimension (e.g., closure duration and vowel duration relate to voicing) (Liberman & Pisoni, 1977; Liberman & Studdert-Kennedy, 1978). It was felt that the best strategy by which to avoid this conundrum was to consider simultaneously a matched number of acoustic variables and perceptual dimensions, since unique solutions are only possible in a properly dimensioned problem. (This thought is expressed more formally in Shaw and Cutting, 1980, where the relationship between physical variables and information space is discussed.)

Here, two perceptual dimensions were considered simultaneously. The potential "solutions" to the information for voicing were constrained by the requirement that that information be invariant under a rate transformation, and the potential solutions to the information for rate were constrained by the requirement that that information be invariant under a voicing transformation (see Mark, Todd, & Shaw, in press, for a discussion of group properties in relation to visual perception). Information that would distinguish each was sought.

We began with a situation in which two acoustic variables and two perceptual dimensions were confounded. It was verified that both closure duration and vowel duration cued both voicing and rate. The acoustic variables, therefore, had to be redefined.

Two aspects of that redefinition were addressed. One concerned the temporal extent of the unit to which a descriptor is to be applied; the other concerned the nature of the descriptive variable. Alternatives to the traditional "duration of a single acoustic segment" were sought. The alternatives were prompted by a consideration of the production of speech and other coordinated actions, in the belief that an understanding of the source event can best guide a search for the information to which a perceiver of that event is sensitive. In regard to the first aspect, this consideration makes one wary of violating the natural boundaries of the event by chopping the signal into segments along the time line using a criterion oblivious to the source event (such as the smallest unit that stands out in a visual display). In regard to the second aspect, knowing that complex acoustic results may arise from a single source of control makes one wary of using too simple an acoustic variable, which may confine one to the realm of "cues." Taken together, these considerations are consonant with other recent efforts to define the essential nature of linguistic units in accordance with a certain understanding of the production of coordinated movements. This understanding does not preclude overlapping, but distinct information for phonemes in the acoustic stream, and has been used to argue that invariant phonetic descriptions need not be ruled out by the fact of context-produced variability due to coarticulation (Fowler, Rubin, Remez, & Turvey, 1979; Fowler, 1980).

Such a description is of necessity abstract. The move toward this more abstract type of specification is also advocated by Bailey and Summerfield (1980) who, while noting that any consistent acoustic difference between phonemes can serve as a "cue," also note that "the perception of events in general, including articulatory events, may involve the direct apprehension of patterns of change over time and may not, therefore, require the perceptual integration of a succession of discrete cues" (p. 562).

In Experiment I, relational variables at a temporal grain coarser than the individual segment were defined by taking different closure-vowel production patterns into account. This allowed a good differentiation of the rate perception results and the voicing perception results, and provided a basis for distinguishing temporal information for rate from temporal information for voicing.

This distinction was pursued in Experiment II by considering temporal patterns within an individual segment (the vowel). A change in vowel duration due to a rate change was contrasted with a change in vowel duration due to a computer manipulation of only the vowel nucleus. The two kinds of duration changes were not equivalent. They affected the voicing judgments differently.

Experiment III confirmed that the temporal relation between initial consonant transitions and vowel nucleus is perceptually salient, and supported the hypothesis that it was this difference in Experiment II that made the artificial duration change different from the rate change. It was concluded that this relational variable may further distinguish rate and voicing.

In accordance with the overall goal, one would eventually like to see relations within the vowel and relations between vowel and closure integrated into a single variable. In fact, even richer variables will undoubtedly be necessary to reach the level of description that qualifies as "specification." A strategy for proceeding toward that enrichment is to look for information that progressively distinguishes a greater number of perceptual dimensions.

REFERENCE NOTES

1. Summerfield, Q. On articulatory rate and perceptual constancy in phonetic perception. Upublished manuscript, 1978.
2. Szubowicz, L. L. WENDY, Reference Manual, Version 1.5, Haskins Laboratories, New Haven, Ct., 1979

REFERENCES

Ainsworth, W. A. Durational cues in the perception of certain consonants. Proceedings of the British Acoustical Society, 1973, 2, 1-4.

Ainsworth, W. A. The influence of precursive sequences on the perception of synthesized vowels. Language and Speech, 1974, 17, 103-109.

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

Belasco, S. The influence of force of articulation of consonants on vowel duration. Journal of the Acoustical Society of America, 1953, 25, 1015-1016.

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

Denes, P. Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, 1955, 27, 761-764.

Engberg, I., & Lundberg, A. An electromyographic analysis of muscular activity in the hindlimb of the cat during unrestrained locomotion. *Acta Physiologica Scandia*, 1969, 75, 614-630.

Fitch, H. L. The influence of preceding syllable structure on intervocalic voicing boundary. *Journal of the Acoustical Society of America*, 1980, 67, S50. (Abstract)

Fitch, H. L., & Turvey, M. T. On the control of activity: Some remarks from an ecological point of view. In D. M. Landers & R. W. Christina (Eds.), *Psychology of motor behavior and sport*. Champaign, Ill.: Human Kinetics, 1978.

Fowler, C. A. *Timing control in speech production*. Unpublished doctoral dissertation, University of Connecticut, 1977.

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

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production*. New York: Academic Press, 1979.

Fujisaki, H., Nakamura, K., & Imoto, T. Auditory perception of duration of speech and non-speech stimuli. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech*. London: Academic Press, 1975.

Gaitenby, J. H. The elastic word. *Haskins Laboratories Status Report on Speech Research*, 1965, SR-2, 3.1-3.12.

Gay, T. Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, 1978, 63, 223-230.

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

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

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

Grillner, S. Locomotion in vertebrates. *Physiological Reviews*, 1975, 55, 247-304.

Grosjean, F., & Lane, H. Effects of two temporal variables on the listener's perception of reading rate. *Journal of Experimental Psychology*, 1974, 102, 893-896.

Grosjean, F., & Lane, H. How the listener integrates the components of speaking rate. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 538-543.

Greene, P. H. Problems of organization of motor systems. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (Vol 2). New York: Academic Press, 1972.

House, A. On vowel duration in English. *Journal of the Acoustical Society of America*, 1961, 33, 1174-1178.

House, A. S., & Fairbanks, G. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, 1953, 25, 105-113.

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

Kozhevnikov, V. A., & Chistovich, L. A. Rech 'artikuliatsiia i vospriiatiie, Moscow-Leningrad. Translated as: *Speech Articulation and Perception*.

29

Washington, D.C.: Clearinghouse for Federal Scientific and Technical Information, 1965, JPRS 30-543.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On the concept of coordinative structure as dissipative structure. I. Background and theory. In G. E. Stelmach (Ed.), Tutorials in motor behavior. Amsterdam: North-Holland, 1980.

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

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

Liberman, A. M., & Pisoni, D. B. Evidence for a special speech-perceiving subsystem in the human. In T. H. Bullock (Ed.), Recognition of complex acoustic signals. Berlin: Dahlem Konferenzen, 1977, 59-76.

Liberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), Handbook of sensory physiology, Vol. VIII: Perception. New York: Springer-Verlag, 1978, 143-187.

Lindblom, B. E. F., & Studdert-Kennedy, M. On the role of formant-transitions in vowel recognition. Journal of the Acoustical Society of America, 1967, 42, 830-843.

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

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

Lisker, L., & Price, P. J. Context-determined effects of varying closure duration. In J. Wolf, & D. Klatt (Eds.), Speech communication papers presented at the 97th meeting of the Acoustical Society of America. New York: Acoustical Society of America, 1979, 45-48.

Mark, L. S., Todd, J. T., & Shaw, R. E. The perception of growth: A geometric analysis of how different styles of change are distinguished. Journal of Experimental Psychology: Human Perception and Performance, in press.

Miller, J. L. The effect of speaking rate on segmental distinctions: Acoustic variation and perceptual compensation. In P. D. Eimas & J. L. Miller (Eds.), Perspectives on the study of speech. Hillsdale, N.J.: Lawrence Erlbaum Associates, in press.

Miller, J. L., & Grosjean, F. Further studies of the influence of sentence rate on the closure duration cue for voicing. In J. Wolf & D. Klatt (Eds.), Speech communication papers presented at the 97th meeting of the Acoustical Society of America. New York: Acoustical Society of America, 1979, 41-44.

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

Minifie, F., Kuhl, P., & Stecher, B. Categorical perception of [b] and [w] during changes in rate of utterance. Journal of the Acoustical Society of America, 1977, 62, S79. (Abstract)

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

Orlovsky, G. N., Severin, F. V., & Shik, M. L. Effect of speed and load on coordination of movements during running of the dog. Biophysics, 1966, 11, 414-417.

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

Philippson, M.   L'Autonomie et la centralisation dan le systeme des animaux. *Trav. Lab. Physiol. Inst. Solvay (Bruxelles)*, 1905, 7, 1-208.

Pickett, J. M., & Decker, L. R.   Time factors in perception of a double consonant. *Language and Speech*, 1960, 3, 11-17.

Port, R. F.   *The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words*. Unpublished doctoral dissertation, University of Connecticut, 1976.

Port, R. F.   Effects of word-internal versus word-external tempo on the voicing boundary for medial stop closure. *Journal of the Acoustical Society of America*, 1978, 63, S20.   (Abstract)

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

Raphael, L. J.   Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. *Journal of the Acoustical Society of America*, 1972, 51, 1296-1303.

Runeson, S.   *On visual perception of dynamic events*. Doctoral dissertation, University of Uppsala, Sweden, 1977.

Schiff, W.   Perception of impending collision: A study of visually directed avoidant behavior. *Psychological Monographs*, 1965, 79, Whole No. 604.

Shaw, R. E., & Cutting, J. E.   The structuring of language: Clues from an ecological theory of event perception. In U. Bellugi & M. Studdert-Kennedy (Eds.), *Signed language and spoken language: Biological constraints on linguistic form* (Dahlem Konferenzen). Weinheim/Deerfield Beach, Fl./Basel: Verlag Chemie, 1980.

Shik, M. L., & Orlovsky, G. N.   Neurophysiology of locomotor automatism. *Physiological Reviews*, 1976, 56, 465-501.

Slis, I. H., & Cohen, A.   On the complex regulation of the voiced-voiceless distinction I. *Language & Speech*, 1969, 12, 80-104.

Summerfield, Q.   Towards a detailed model for the perception of voicing contrasts.   In *Speech Perception* (3), Department of Psychology, Queen's University of Belfast, 1974.

Summerfield, Q.   Cues, contexts and complications in the perception of voicing contrasts.   In *Speech Perception* (3), Department of Psychology, Queen's University of Belfast, 1975.   (a)

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

Summerfield, Q., & Haggard, P.   Speech rate effects in the perception of voicing.   In *Speech Synthesis and Perception* (6), *Psychology Laboratory, University of Cambridge*, 1972.

Turvey, M. T.   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, N.J.: Erlbaum, 1977.   (a)

Turvey, M. T.   Contrasting orientations to a theory of visual information processing. *Psychological Review*, 1977, 84, 67-88.   (b)

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

Turvey, M. T., Shaw, R. E., & Mace, W. Issues in the theory of action: Degrees of freedom, coordinative structures and coalitions. In J. Requin (Ed.), Attention and performance VII. Hillsdale, N.J.: Erlbaum, 1978.

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

Verbrugge, R. R., & Isenberg, D. Syllable timing and vowel perception. Journal of the Acoustical Society of America, 1978, 63, S4. (Abstract)

Verbrugge, R. R., & Shankweiler, D. P. Prosodic information for vowel identity. Journal of the Acoustical Society of America, 1977, 61, S39. (Abstract)

Verbrugge, R. R., Strange, W., Shankweiler, D. P., & Edman, T. R. What information enables a listener to map a talker's vowel space? Journal of the Acoustical Society of America, 1976, 60, 198-212.

Zimmermann, S. A., & Sapon, S. M. Note on vowel duration seen cross-linguistically. Journal of the Acoustical Society of America, 1958, 30, 152.

FOOTNOTES

[1]Since information for successive phonemes overlaps in the acoustic signal, it is not possible to temporally segment a syllable into discrete vowel and consonant components. The vowel is co-produced with the surrounding consonants, and vowel information seems to be spread throughout the vocalic region. Vowel duration is often defined, therefore, as the total extent of the vocalic region. To simplify exposition, the term "vowel" will be used here to refer to the pre-closure vocalic region.

[2] This change works in the direction of a more conservative test of the hypothesis. No difference in voicing is expected, and increasing the closure duration of the short vowel stimulus would make it even more like /p/, thus less like the other, evenly biased stimulus. A difference in rate is expected, and increasing the closure duration of the short vowel stimulus would make its total duration longer, thus more like the other, long, stimulus.

ARTICULATORY MOTOR EVENTS AS A FUNCTION OF SPEAKING RATE AND STRESS

Betty Tuller,+ Katherine S. Harris,++ and J. A. Scott Kelso+++

Abstract. Two basic types of explanation have been proposed for the changes in segmental timing that occur when speakers change rate or stress of component syllables. One view is that the segmental "commands" for syllables spoken quickly and for unstressed syllables show more extensive temporal overlap than the same syllables spoken more slowly or with greater syllabic stress. An alternative view is that the temporal relations among articulations remain constant over changes in speaking rate and stress, but that the individual gestures themselves vary. Experiment 1 explored the temporal relations among electromyographic measures of articulatory events, and the pattern of changes in individual muscle actions, over suprasegmental variations in syllable stress and speaking rate. Large variations were found in the magnitude and duration of activity in each muscle; variations accompanying speaking rate change were not equivalent to the variations accompanying a change in stress. The electromyographic activity underlying lip movements for bilabial stop consonants (orbicularis oris) and tongue fronting for the vowels /i/ and /e/ (genioglossus) appeared to maintain a tight timing pattern. In a second experiment, X-ray microbeam data were collected for the same types of utterances used in tne first experiment. Kinematic patterns, like EMG patterns, showed that temporal relations between tongue and lip movements were preserved over changes in speaking rate and syllable stress.

Investigations of speech production have often focused on a search for invariant units that correlate with aspects of a speaker/hearer's linguistic competence. Many of these studies share an assumption about linguistic units: namely, that they are discrete, static, and context-invariant entities, selected and ordered prior to their execution by peripheral motor mechanisms. Most experiments have consisted of a search for discrete stretches in the acoustic or physiological output in the hope that they might correlate with linguistic units. However, such studies have met with little success, whether

---

looking for invariant units in the acoustic signal (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; but see Stevens, 1973), patterns of muscle activity (Harris, Lysaught, & Schvey, 1965; MacNeilage & DeClerk, 1969), articulatory movements (MacNeilage, 1970), or vocal tract area functions. The repeated failure to find invariant correlates of abstract linguistic units has promoted the claim that abstract representations used to describe linguistic competence are obscured when translated into linguistic performance, because the latter are subject to the physical constraints of human speech to which the former are indifferent (cf. Öhman, 1972).

The foregoing conception of linguistic units as abstract and discrete is inherent in those current models of speech production which assume that articulatory control of suprasegmental changes is independent of segmental articulation. Articulatory control over variations in speaking rate and syllable stress, for example, is considered as "...the consequence of a timing pattern imposed on a group of (invariant) phoneme commands" (Shaffer, 1976, p. 387; parentheses his). Similarly, Lindblom (1963) suggested that each phoneme has an invariant "program" that is unaffected by changes in lexical stress and speaking rate (tempo).[1] According to Lindblom, when successive programs are executed, their temporal overlap results in coarticulation between segments. Thus, when a vowel coarticulates with a following consonant, it is because the consonant program begins before the vowel program is finished (see also Stevens & House, 1963). When speaking rate increases or stress decreases, the command for a new segment arrives at the articulators before the preceding segment is fully realized. As a consequence, there is temporal shortening and articulatory undershoot, both of which characterize unstressed syllables and fast speaking rates (see also Kozhevnikov & Chistovich, 1965). In such models, therefore, increases in speaking rate and decreases in syllable stress are accomplished with comparable strategies and hence have similar acoustic consequences. They predict that the "commands" for some aspects of articulation of a given phoneme stand in a fixed relation to commands for other aspects of the same phoneme, but that the relative temporal alignment of control signals for successive segments, and their kinematic realizations, vary with stress and speaking rate.

The models discussed above suggest that changes in speaking rate and syllable stress are both characterized by invariant segments with variable temporal relations between them. One prediction of this view is that the relation between target formant frequency and duration is fixed; that is, when the duration of a vowel shortens, it will undershoot the articulatory "target," resulting in more centralized formant frequencies than occur with longer vowel durations. However, Harris (1978) performed a spectrographic analysis of a small set of nonsense utterances produced at two speaking rates and with two levels of stress, and found that changes in vowel formant frequencies were not fixed in relation to changes in vowel duration. Her results suggest that extant models for suprasegmental changes cannot be supported at an acoustic level.

A similar conclusion follows from a small body of electromyographic (EMG) data showing that segmental articulation varies considerably with speaking rate (Gay & Hirose, 1973; Gay & Ushijima, 1974; Gay, Ushijima, Hirose, & Cooper, 1974) and syllable stress (Harris, 1971, 1973; Harris, Gay, Sholes, & Lieberman, 1968; Sussman & MacNeilage, 1978). However, these studies have not

34

examined temporal relations among successive segments (i.e., intersegmental timing). Furthermore, no experiments exist in which speaking rate and syllable stress have been orthogonally varied in the same experiment. It is possible, for example, that the timing of articulation for successive segments remains fixed over suprasegmental changes, but that the segments themselves vary.

The present experiments explored the temporal relations among articulatory events as a function of syllable stress and speaking rate. Specifically, Experiment 1 sought to determine whether variations in stress and rate change the timing of EMG activity for successive phonetic segments while maintaining the segmental articulations constant, or whether such suprasegmental variations change the EMG activity for individual segments but maintain the timing relations between successive segments. As we shall see, fairly constant temporal relations were evident between selected articulatory muscles (orbicularis oris and genioglossus) in the face of metrical variations in rate and stress. In addition, the patterns of EMG activity in orbicularis oris and genioglossus were different when stress rather than speaking rate was varied. Thus, the results do not support the notion that acoustic shortening, which typically accompanies both decreases in syllable stress and increases in speaking rate, is the product of a single style of articulatory change. Experiment 2, although more restricted in scope, examined whether the EMG timing patterns observed were also evident in the kinematics of lip and tongue movements. Such data are important for two reasons: first, because of the possibility that peripheral biomechanical factors can cloud the relation between EMG and kinematics, and second, because both sources of data (along with relevant acoustic evidence) may provide a more comprehensive picture of intersegmental timing than either one alone. A pleasing aspect of the present experiments is that both the EMG and kinematic data allow us to converge on the same conclusions regarding stress and rate effects on articulatory patterns.

EXPERIMENT 1

Method

Subjects. The subjects were two female adults (KSH and FBB), both of whom were native speakers of American English.

Materials and procedures. The speech sample consisted of four-syllable nonsense utterances of the form /əpipipə/, /əpipibə/, /əpepepə/, and /əpepebə/, with stress placed on either the first or the second medial syllable. Subjects read quasi-random lists of these four utterances at two self-selected speaking rates, "slow" (conversational) and "fast." Although 25 repetitions were produced of each utterance, later processing failures reduced the lists to 20 repetitions for KSH and 21 for FBB.

Data recording. Electromyographic activity was recorded from the genioglossus and orbicularis oris muscles. Bipolar hooked-wire electrodes, prepared and inserted as described by Hirose (1971) were used to record EMG activity

35

from the anterior portion of the genioglossus muscle. Genioglossus bunches the main body of the tongue and brings it forward and is active in production of the vowel /i/ (e.g., Alfonso & Baer, 1981; Raphael & Bell-Berti, 1975; Smith, 1971).

Electromyographic activity was recorded from orbicularis oris (superior and inferior) using paint-on surface electrodes (Allen & Lubker, 1972) spaced at about one-half centimeter from the vermilion border of the lips. Orbicularis oris is known to participate in bilabial closure (Harris et al., 1965; Fromkin, 1966).

The EMG data were rectified, computer-sampled, integrated using a time constant of 35 msec, and averaged for each utterance type (Kewley-Port, 1974). In order to ensure at least one successful recording for each muscle for each subject, input of two or three electrodes was recorded from each muscle. Those electrodes whose recordings appeared on preliminary inspection to show the clearest onset and offset points were selected for further analysis.

Acoustic recordings were made simultaneously with the EMG recordings and both were analyzed on subsequent playback from multichannel FM tape. The EMG tokens were realigned and reaveraged three times, at the end of periodic vibration in ·the acoustic signal for the first, second, and third vowels, respectively. In this way, average muscle activity could be examined at specific points of interest without the time-smearing effects of averaging tokens that were aligned at a temporally distant point.

Figure 1 shows typical averaged interference patterns for orbicularis oris activity (the thin line) and genioglossus activity (the thick line). The patterns on the left- and right-hand sides of the figure represent the same utterance; a schematic acoustic signal appears above each pattern. The pattern on the left is the average of twenty tokens aligned at the end of the acoustic periodicity for the first vowel (the schwa); the point of alignment for tokens comprising the pattern on the right was the end of acoustic periodicity for the third vowel.

Onsets and offsets of EMG activity were determined from data averaged around the temporal line-up closest to the activity of interest. The averaging program provides a listing of the mean amplitude of each EMG signal in microvolts during successive 5-msec intervals. Baseline and peak values for each muscle were determined from this numerical listing; the time of onset (and offset) was defined as the point in time when the relevant muscle activity increased (or decreased) to 10% of its range of activity. Typically, 10% of the range was just slightly higher than the background level of activity in each muscle. In the present experiment, the genioglossus muscle is active for the second and third vowels of each utterance type. In this environment, the trough between the peaks of activity for successive vowels, evident in Figure 1, is the only measure of "onset" or "offset" in the relevant syllables. The duration of activity in genioglossus for syllable one, for example, was taken to be from onset to the lowest point in the trough (see Figure 1). Similarly, orbicularis oris is active for all three consonants and, particularly in fast and unstressed utterances, does not always return to its baseline value between successive consonant peaks.

36
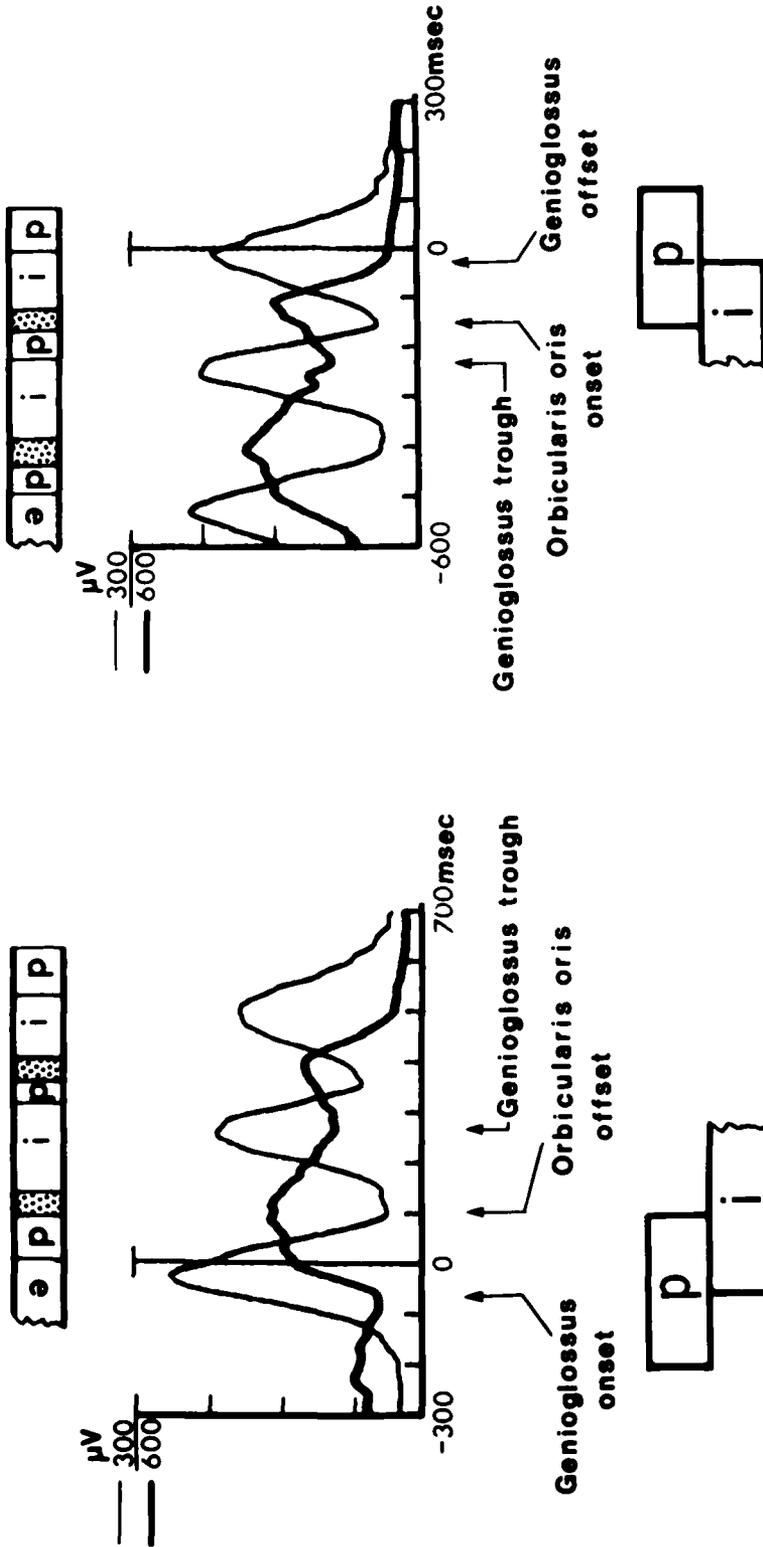
Figure 1. Typical averaged interference patterns for orbicularis oris activity (the thin line) and genioglossus activity (the thick line). The left- and right-hand sides of the figure represent the same utterance averaged at the end of acoustic periodicity for the first vowel (the schwa) and the third vowel, respectively. A schematic acoustic signal is above each pattern. The trough between peaks of activity is indicated, as are the onsets and offsets of activity in genioglossus and orbicularis oris.

37

The acoustic recordings were measured for their durational characteristics, using an interactive computer program that displays the acoustic waveform. The duration of voicing was measured for the first and second medial vowels, as well as devoicing durations for the /p/ and /b/ closures. Measures were made of the interval from the first acoustic evidence of closure (defined here as the point when the high frequency components of the periodic wave disappear) to the second acoustic evidence of closure. For ease of communication, this interval will be referred to below as the "acoustic duration of the first syllable." The measured interval from the second acoustic evidence of closure to the third will be referred to as the "acoustic duration of the second syllable." These measures were averaged, omitting tokens for which there were EMG processing failures.

RESULTS

In the analyses that follow, binomial tests and z-scores were used to determine the effects of speaking rate (fast vs. slow), syllable stress (stressed vs. unstressed), vowel identity (/i/ vs. /e/), and final consonant identity (/p/ vs. /b/) on the observed acoustic and electromyographic measures. Because of the small sample size used in this experiment, the binomial test and z-scores corrected for continuity (Siegel, 1956), both nonparametric statistics, were deemed more appropriate than parametric statistics. These analyses examine the direction of change, not the magnitude of change. Unless z-scores are explicitly given, the analysis used was a binomial test. Significance levels given are for two-tailed analyses.

I. Acoustic Analysis and Discussion

The acoustic duration of each syllable was examined to determine the effects of changing speaking rate (fast vs. slow), syllable stress (stressed vs. unstressed), vowel (/i/ vs. /e/), syllable position (first vs. second syllable), and final consonant (/p/ vs. /b/). Figure 2 presents the mean acoustic syllable durations for the two levels of each of these five variables. The analyses showed an effect of speaking rate on acoustic syllable duration ($z=-5.48$, $p < .001$). Not surprisingly, syllables spoken slowly were significantly longer than the same syllables spoken quickly. Acoustic syllable duration also shortened with decreases in syllable stress ($z=5.48$, $p < .001$). The magnitude of the changes in acoustic syllable duration was not equivalent for these variables; acoustic syllable duration was shortened more by an increase in speaking rate than by a decrease in syllable stress (70 vs. 30 msec).

These changes in acoustic syllable duration are in general agreement with the pattern of acoustic changes documented in the literature. Acoustic vowel durations have often been observed to shorten as speaking rate increases (e.g., Lindblom, 1963; Kozhevnikov & Chistovich, 1965; Lehiste, 1970; Port, 1976; Verbrugge & Shankweiler, 1977). Stressed syllables are usually measured to be longer than unstressed syllables (Fry, 1955, 1958; Gaitenby, 1965; Lieberman, 1960; Tiffany, 1959).

38

Mean Acoustic Syllable Durations

*p < .01
**p < .001
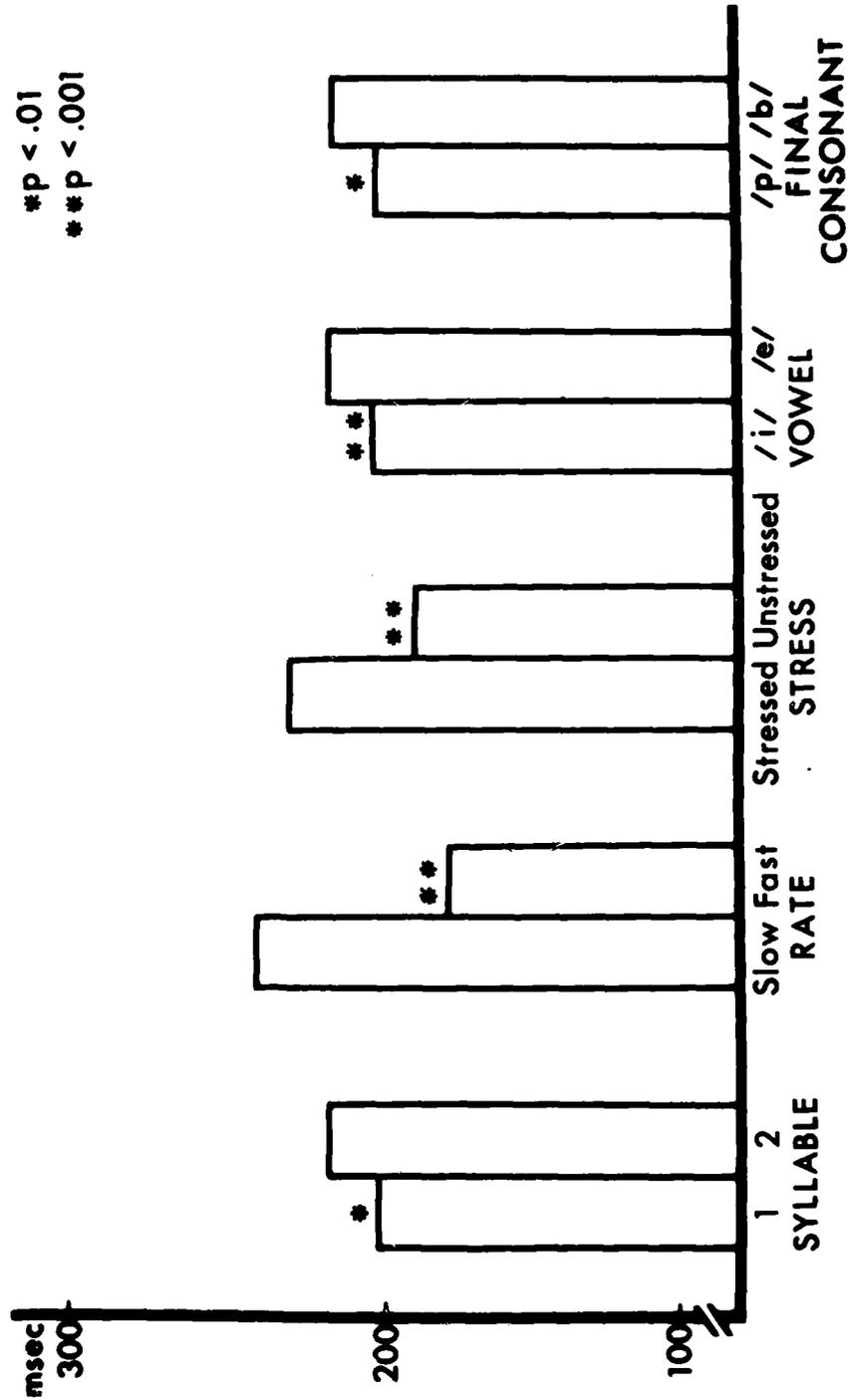
Figure 2. Mean acoustic syllable durations as a function of syllable (1 vs. 2), speaking rate (fast vs. slow), syllable stress (stressed vs. unstressed), vowel (/i/ vs. /e/), and final consonant (/p/ vs. /b/).

The acoustic duration of syllables also differed as a function of vowel identity. Syllables containing the vowel /e/ were significantly longer than syllables containing the vowel /i/ ($z=-5.13$, $p < .001$). The second syllable of these utterances was consistently longer than the first syllable ($z=-2.65$, $p < .01$), and those (second) syllables ending with /b/ closure were longer than those syllables ending with /p/ closure ($p < .01$).

Similar analyses were performed examining the effects of speaking rate, syllable stress, vowel identity, and syllable position on the measured closure durations of the bilabial stop consonant /p/. The closure duration of final /p/ or /b/ was not measured because, using the criterion of acoustic syllable duration defined here, this interval is part of the final stop consonant-schwa syllable. Closure durations shortened when speaking rate increased ($p < .01$) or stress decreased ($p < .01$). There appeared to be an interaction of syllable position and stress on closure duration for bilabial stops. In the first syllable, stressed syllables had initial bilabial stops with longer closure durations than did unstressed syllables ($p < .01$); in the second syllable, the initial bilabial closure in unstressed syllables was longer than in stressed syllables ($p < .001$). No other variable affected the duration of bilabial closure.

Although changes in closure duration are not well documented, Gay et al. (1974), Kent and Moll (1972), and Port (1976) have reported limited evidence that closure durations tend to decrease with increasing rates of speech. In contrast, Gay and Hirose (1973) found no change in closure duration over changes in stress or rate.

The general pattern of acoustic duration changes reported here concurs with the available literature. This observation suggests that the subjects were indeed following the instructions to speak faster or to vary stress. The next step in the analysis was to examine the duration of electromyographic activity in the genioglossus and orbicularis oris muscles, their peak values, and their temporal relations to determine whether these measures vary as a function of syllable position, speaking rate, syllable stress, final consonant, and vowel identity.

II. EMG Analysis: Variations in Individual Muscle Actions

a. Genioglossus. Z-scores, corrected for continuity, showed that the duration of genioglossus activity varied significantly with changes in speaking rate and syllable stress ($z=-4.42$, $p < .001$ and $z=-5.13$, $p < .001$, respectively), being longer for slow and stressed syllables than for syllables spoken quickly or without primary stress (see Table 1). Genioglossus activity was also found to be longer in the first syllable than in the second syllable ($p < .01$).

40

```
-----------------------------------------------------------------------
```

Table 1

Mean duration (in msec) and peak amplitude (in microvolts) of
genioglossus and orbicularis oris as a function of speaking rate
and syllable stress.

| | Duration | | Peak Amplitude | |
	Slow	Fast	Slow	Fast
Orbicularis Oris	169**	149	488	497
Genioglossus	229**	185	254	260
	Stressed	Unstressed	Stressed	Unstressed
Orbicularis Oris	165*	143	525**	459
Genioglossus	228**	186	293**	255

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


```
-----------------------------------------------------------------------
```


The peak amplitude of activity in genioglossus varied with changes in
syllable stress, being higher in stressed syllables than in unstressed
syllables ($z = -3.71$, $p < .001$). Genioglossus peak amplitude did not vary
significantly with changes in speaking rate ($z = .18$, $p > .2$), syllable position
($p > .2$), or vowel identity ($p > .2$).

Subjects' genioglossus recordings were also examined individually. For
both subjects, genioglossus duration was longer in slow than in fast syllables
($p < .05$) and longer in stressed than in unstressed syllables ($p < .01$). Peak
amplitude of activity in genioglossus for each subject did not change with
speaking rate ($p > .08$).

The two subjects showed different patterns of change in peak amplitude of
genioglossus activity as a function of vowel (/i/ vs. /e/). For KSH, the peak
amplitude of genioglossus activity was higher for /e/ than for /i/ ($p < .01$),
although genioglossus duration did not alter ($p > .2$). Figure 3A shows
genioglossus activity for /i/ and /e/ for this subject. However, genioglossus
activity for /e/ shows two clear peaks, indicating that the vowel was produced
as a diphthong. In contrast, for subject FBB (Figure 3B) peak amplitude of
genioglossus was higher ($p < .01$), and genioglossus duration shorter ($p < .01$),
for /i/ than for /e/. Genioglossus activity for /i/ and /e/ shows only one

41

Genioglossus

a) **KSH**

b) **FBB**

—— pipib
—— pepeb
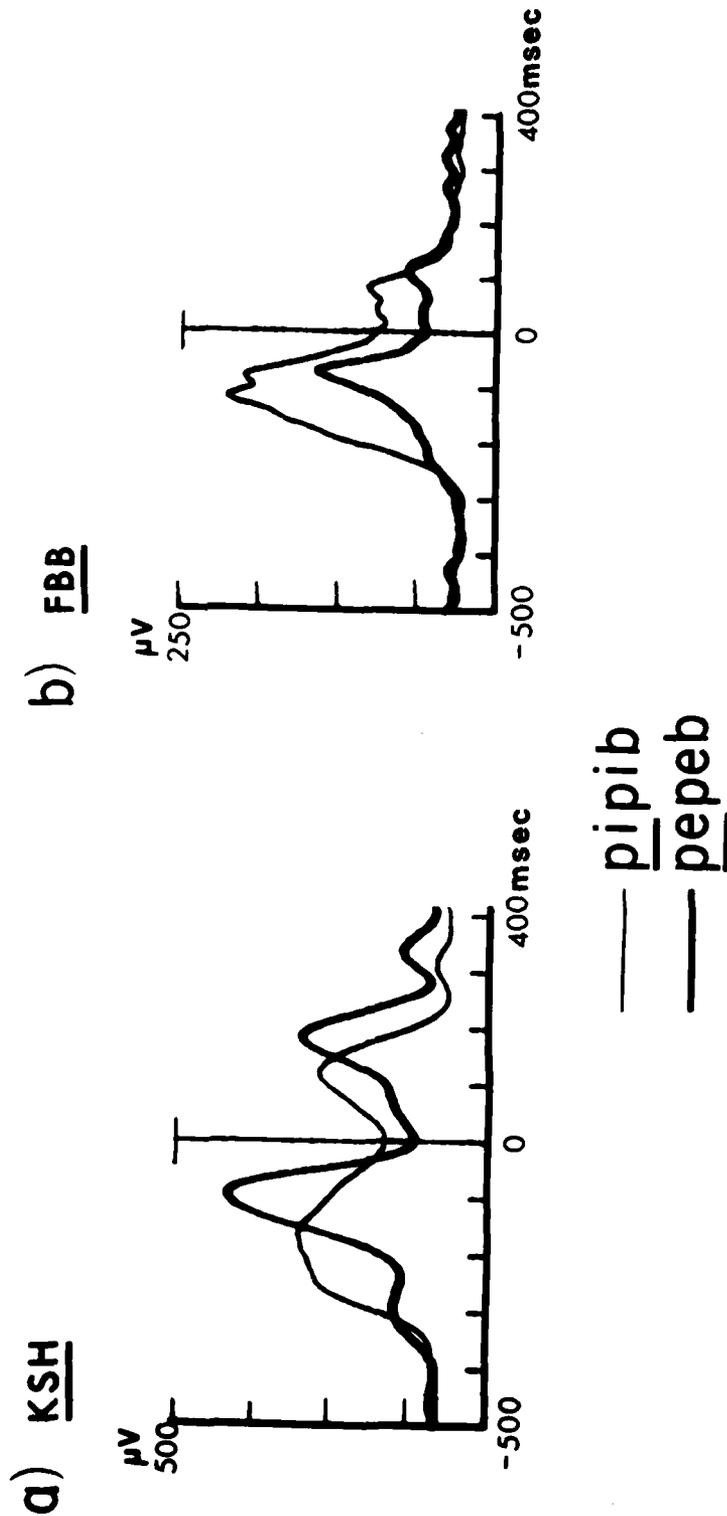
Figure 3. Genioglossus activity for production of /i/ (the thin line) and /e/ (the thick line) for a) KSH and b) FBB.

42

clear peak. Because genioglossus activity is higher for /i/ than for /e/, and only one peak is evident for each vowel, the production of /e/ by this subject was probably more open than the production of /i/, and was not produced as a true diphthong.

b. <u>Orbicularis oris</u>. An increase in speaking rate and a decrease in syllable stress decreased the duration of orbicularis oris activity (z=-4.42, p <.001, and z=-2.65, p <.01, respectively; see Table 1). Orbicularis oris duration was also longer when /p/ rather than /b/ was the final consonant (z=-2.65, p <.01). The durations of orbicularis oris activity were statistically equivalent for the first and second syllable (p >.2), and there was no effect of vowel identity (z=-.18, p >.2).

Orbicularis oris peak amplitude was higher for stressed than unstressed syllables (z=-4.42, p <.001). Peak amplitude was also higher when the bilabial stop occurred in the first syllable rather than the second (p <.01). Syllables spoken quickly tended to have larger amplitudes than syllables spoken slowly (z=-1.94, p <.052), but this comparison did not reach significance. There was no effect on orbicularis oris peak amplitude of vowel (z=-1.24, p >.2) or final consonant (z=-.28, p >.2).

In summary, when results for the two subjects are considered together, as speaking rate increased from "conversational" to "fast," the duration of each muscle's activity shortened; mean genioglossus activity shortened from 229 to 185 msec; mean orbicularis oris activity shortened from 169 to 149 msec (see Table 1). Thus, genioglossus duration varied proportionally more than orbicularis oris duration. With an increase in speaking rate, the peak amplitude of activity in genioglossus was unaffected; peak amplitude of activity in orbicularis oris increased somewhat, but this increase was not significant. When the syllable was stressed rather than unstressed, activity in both genioglossus and orbicularis oris was of longer duration and higher peak amplitude. With a shift from unstressed to stressed production, mean duration of genioglossus activity lengthened from 186 to 288 msec; mean orbicularis oris duration lengthened from 143 to 162 msec. Mean peak amplitude of genioglossus rose from 255 to 293 uV; mean orbicularis oris peak amplitude rose from 459 to 525 uV. There were no systematic effects of phonetic context on genioglossus activity or orbicularis oris peak amplitude but the duration of orbicularis oris was longer for production of /p/ than /b/.

The foregoing summary underscores the considerable variation observed in duration and peak amplitude of muscle activity. The range of variation in individual muscles and in the acoustic syllable duration is presented in Table 2 (A and B). For example, the value in the upper left-hand cell represents the difference between the longest and shortest measured acoustic duration of the syllable /pi/ produced by KSH. Obviously, the acoustic duration varied substantially (101 msec). An examination of parts A and B of Table 2 indicates that the acoustic syllable durations and the durations and peak amplitudes of muscle activity are generally quite variable over changes in syllable stress, speaking rate, and phonetic context (that is, the numbers in all cells are relatively large). In the next section, we examine whether temporal relations among muscles are as variable.

Table 2

The range of variation in measured acoustic syllable duration, duration
and peak amplitude of individual muscles, and temporal relations
between muscles, over changes in speaking rate and syllable stress.

	KSH				FBB			
	pi	pe	ip,ib	ep,eb	pi	pe	ip,ib	ep,eb
A. Durations (msec)								
Acoustic Syllable	101	108	123	147	120	139	131	139
Orbicularis Oris	65	55	95	100	45	50	30	20
Genioglossus	115	110	110	60	135	120	90	100
B. Peak Amplitude (μV)								
Orbicularis Oris	41	34	63	52	185	283	196	211
Genioglossus	88	122	70	172	39	60	163	107
C. Timing Relations (msec; see text)								
Onset-to-onset time	40	60	110	65	35	30	90	110
Offset-to-offset time	110	85	70	35	140	125	10	35
Peak-to-peak time	85	125	60	40	100	125	50	55
Time of simultaneous activity	30	30	30	20	20	20	25	25

44

III. EMG Analysis: Temporal Relations Among Muscle Actions

The onsets and offsets of EMG activity were determined in order to examine temporal relations between orbicularis oris and genioglossus activity as a function of speaking rate, syllable stress, and phonetic context. Onset-to-onset times, peak-to-peak times, offset-to-offset times, and durations of simultaneous activity (overlap) were determined for both muscles. Part C of Table 2 presents the range of variation measured for each of these four interval types. Each value represents the difference between the smallest and the largest measure of the relevant temporal interval.

Table 2C indicates that certain aspects of the timing of lip and tongue fronting activity (orbicularis oris and genioglossus activity) in relation to each other, vary widely with changes in speaking rate and syllable stress. Large variations occurred in onset-to-onset times, offset-to-offset times, and peak-to-peak times. These temporal relations between muscles, like the duration and magnitude of activity in individual muscles, appear free to vary with suprasegmental change.

In contrast, one aspect of the timing of lip and tongue fronting activity in relation to each other, remained fairly stable over variations in speaking rate and syllable stress. Specifically, variations in the duration of simultaneous activity in genioglossus and orbicularis oris were small compared with the large variations observed in the other measured temporal relations.

The actual durations of the measured intervals are presented in Table 3. Each pair of values represents the smallest and the largest measure of the relevant temporal interval. For example, values in the upper left-hand cell indicate that for production of the syllable /pi/ by KSH, the temporal interval from the onset of orbicularis oris activity to the onset of activity in genioglossus ranged from 55 to 95 msec over changes in stress and rate. The individual measures comprising Table 3 were converted to scores indicating their difference from the cell mean. For each subject, the variance of the difference scores was calculated for each temporal measure (including all four syllable types) and the differences between variances were tested for significance using t-tests for correlated variances. The variance of the overlap interval was significantly smaller than the variance of the onset-to-onset interval ($t(30)=4.58$, $p <.01$, and $t(30)=7.21$, $p <.01$, for KSH and FBB, respectively), the offset-to-offset interval ($t(30)=6.43$, $p <.01$, and $t(30)=9.3$, $p <.01$), and the peak-to-peak interval ($t(30)=7.18$, $p <.01$, and $t(30)=9.01$, $p <.01$). Thus, the variance of the temporal overlap of genioglossus and orbicularis oris activity was smaller than the variance of any other measured interval. This temporal stability was evident over substantial individual changes in durations and peak amplitudes of genioglossus and orbicularis oris, and changes in acoustic syllable duration, described above.

Two systematic variations in the temporal relation between orbicularis oris and genioglossus were observed. For subject KSH, a change in syllable stress affected the mean duration of overlap of genioglossus and orbicularis oris activity for the syllables /pi/ and /pe/ ($p < .05$), such that stressed syllables showed more overlap than unstressed syllables (136 vs. 125 msec). An increase in speaking rate also affected overlap duration for the syllables /pi, pe/ ($p < .05$); the mean duration of overlap was greater at slow than fast

45

Table 3

Measured temporal relationships between activity of genioglossus
(GG) and orbicularis oris (OO) for each subject and each syllable type.
Pairs of values represent the shortest and longest measure (in msec) of
the indicated temporal interval.

<u>KSH</u>

	OO onset to GG onset	OO offset to GG offset	OO peak to GG peak	GG onset to OO offset
/pi/	55- 95	80-190	45-130	125-155
/pe/	35- 95	145-230	110-235	100-130

	GG onset to OO onset	GG offset to OO offset	GG peak to OO peak	OO onset to GG offset
/ip,ib/	90-200	60-130	80-140	70-100
/ep,eb/	20- 85	85-120	45- 85	85-105

<u>FBB</u>

	OO onset to GG onset	OO offset to GG offset	OO peak to GG peak	GG onset to OO offset
/pi/	10- 45	55-195	40-140	65- 85
/pe/	30- 60	70-195	55-180	55- 75

	GG onset to OO onset	GG offset to OO offset	GG peak to OO peak	OO onset to GG offset
/ip,ib/	70-160	45- 55	65-115	45- 70
/ep,eb/	50-160	20- 55	35- 90	45- 70

rates (137 vs. 124 msec). Both of these changes are in the direction opposite to that predicted by the models discussed earlier.

Figure 4 illustrates the general preservation of timing relations over changes in syllable stress, speaking rate, and phonetic context for one speaker. The temporal overlap of genioglossus and orbicularis oris (the y-axis) is plotted against acoustic syllable duration (the x-axis) for the syllables /pi/, /pe/, /ip,ib/, and /ep,eb/. Points are labeled as to the stress and rate characteristics of the syllable. Note that the dispersion along the y-axis is quite small (25 msec) although the values on the x-axis vary substantially, illustrating that the timing relation between genioglossus and orbicularis oris activity is fairly stable relative to the large variations in acoustic syllable duration.

The best-fitting straight line was computed for the data from each of the four plots, and the slopes were tested for significant differences from zero. No value reached significance. Notice that if the temporal overlap of successive segments increased as acoustic syllable duration decreased (with an increase in rate or a decrease in stress), one would predict the regression lines of Figure 4 to show a negative slope. For each subject's productions of each syllable type, we computed the linear regression of the relevant temporal interval on acoustic syllable duration, orbicularis oris and genioglossus duration and peak amplitude. The majority of regression lines (31 out of 40) showed a slope of zero. Although there were nine best-fitting straight lines whose slopes differed significantly from zero, all nine were of positive slope and thus in the direction opposite to that predicted by the speech production models discussed above (e.g., Lindblom, 1963). (For a complete set of figures comparing the temporal overlap of activity with changes in individual muscles' activity, see Tuller, 1980.)

The analyses with nonzero slopes may be understood as a consequence of limitations in the experimental design. In those cases where activity in orbicularis oris does not return to its baseline value between successive bilabial stops, the measure of "genioglossus onset to orbicularis oris offset" is underestimated by the measure "genioglossus onset to orbicularis oris trough." This may happen in unstressed or quickly spoken utterances, which are also of short duration, thus "tilting" the regression line in the positive direction. In fact, trough amplitude shows an inverse linear relationship to the two muscles' temporal overlap (r =-.80 for /pi/ and r =-.77 for /pe/). As the "offset" amplitude of orbicularis oris increases, the measured duration of overlap of activity in the two muscles decreases, resulting in a regression line of positive slope.

EXPERIMENT 2

Experiment 2 was performed to supplement the results of Experiment 1, using a kinematic analysis of the movements of lip and tongue in a single speaker. Since the two experiments were not performed simultaneously, and the exact relationship between EMG activity in selected muscles and articulatory movement is as yet unclear, measures could not be defined in parallel. However, we believe this experiment provides additional information on suprasegmental effects on articulatory patterns.
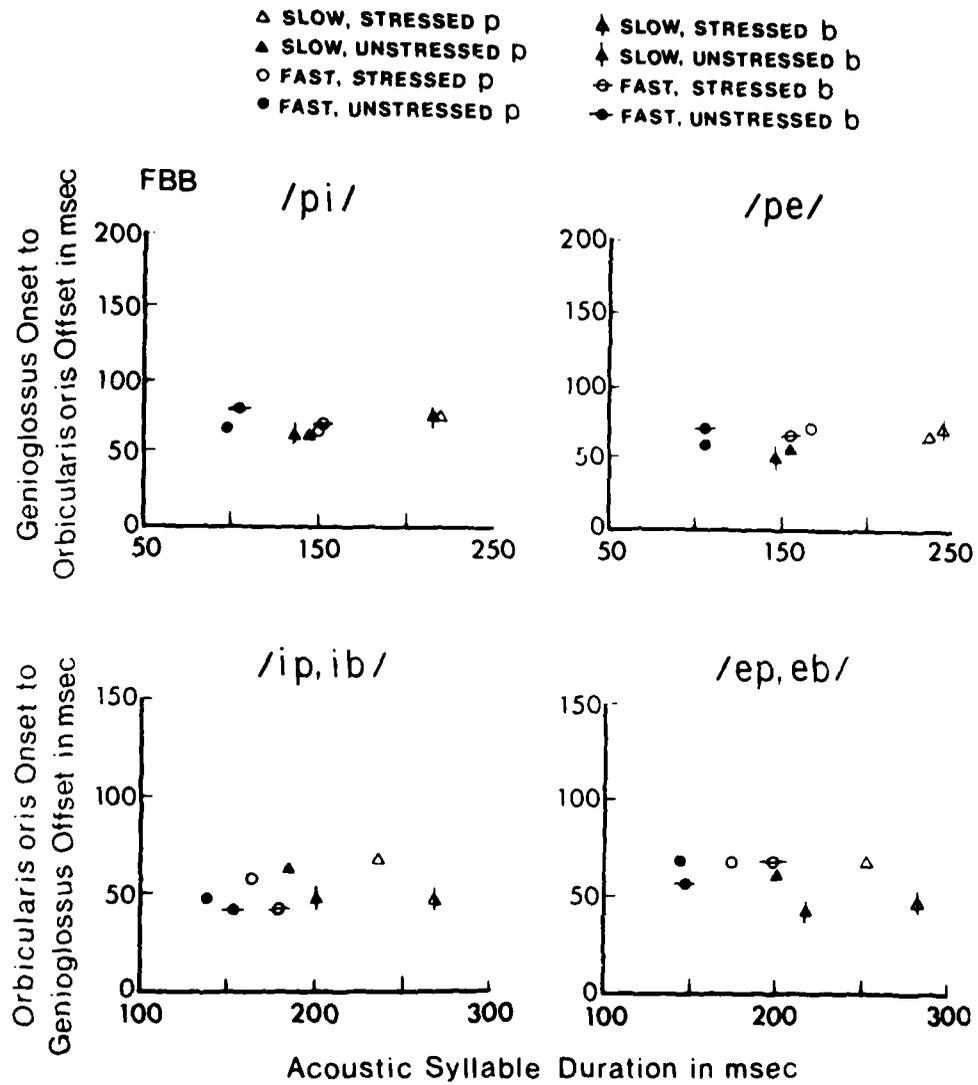
Figure 4. Acoustic syllable duration plotted against the temporal overlap of genioglossus and orbicularis oris activity for production of the four syllable types by FBB.

Method

Subjects. The subject was a single male adult (TB), a native speaker of American English.

Materials and procedures. The speech sample consisted of sixteen four-syllable nonsense utterances of the form /əpipipə/, /əpikipə/, /əpihipə/ and /əpiʔipə/, produced in sets of four utterances with stress on either the second or third syllable, uttered at either of two self-selected speaking rates. In the original set, /əpitipə/ was produced as well, but instrumental failures reduced the set of intervocalic consonants to four.

Data recording. Articulatory movements were recorded with a new method, the X-ray microbeam, which is a variant on cinefluorographic techniques as they are used in conventional modern speech research (Kent & Moll, 1972). In such techniques, films are taken of a subject with radiopaque markers placed on significant articulators. In subsequent analysis, the films are projected frame-by-frame, and the rectilinear coordinates of the pellets identified; coordinates are then stored under computer control (Zimmerman, Kelso, & Lander, 1980). Subsequently, x and y trajectories can be plotted. In the X-ray microbeam system (Kiritani, Itoh, & Fujimura, 1975; Kiritani, 1977), radiopaque markers are tracked by an X-ray microbeam under on-line computer control of the beam deflection. Thus, the only information preserved in the initial data recording is the x and y coordinate positions of the pellets as a function of time. This has the desirable result both of reducing human interactive analysis time and substantially reducing radiation dosage to the subject. Conceptually, however, the system provides data that are equivalent to a conventional analysis.

Figure 5 shows pellet positions used in the experiment. The pellets labeled R_1 and R_2 provide references for the coordinate system, and, using routines in the data analysis package, eliminate the effects of head movement on pellet position. The pellets measured were LL (lower lip), TB (tongue blade) and TM (tongue "middle" or dorsum). Pellets labeled TR and MN were not analyzed. Acoustic recordings were made with a close-talking microphone, and were synchronized with the X-ray microbeam system output. Frame rate was 126 f.p.s.

Figure 6 shows a plot of the output of the system for the y-axis displacement of tongue and lip movements for the utterance /əpipipə/, spoken at a fast rate. Each dot represents one frame. Computer analysis included a smoothing algorithm (Fujimura, Miller, & Nelson, Note 1).

In this experiment, we wished to make measurements that would be congruent with the measures of EMG activity of Experiment 1. Since the fronting and raising activity of the tongue is well correlated with genioglossus activity (Alfonso & Bae   ,, as is the relationship of pursing and closure with orbicularis oris activity (Gay, et al., 1974; Abbs & Kennedy, 1980), we measured the onset and peak displacements for tongue and lip. Onset of movement was defined as that time when a pellet reached 15% of maximum displacement. Transition time was defined as the period over which the pellet showed continuous increase. Maximum displacement was defined as the difference between displacement at onset and displacement at the end of transition

49
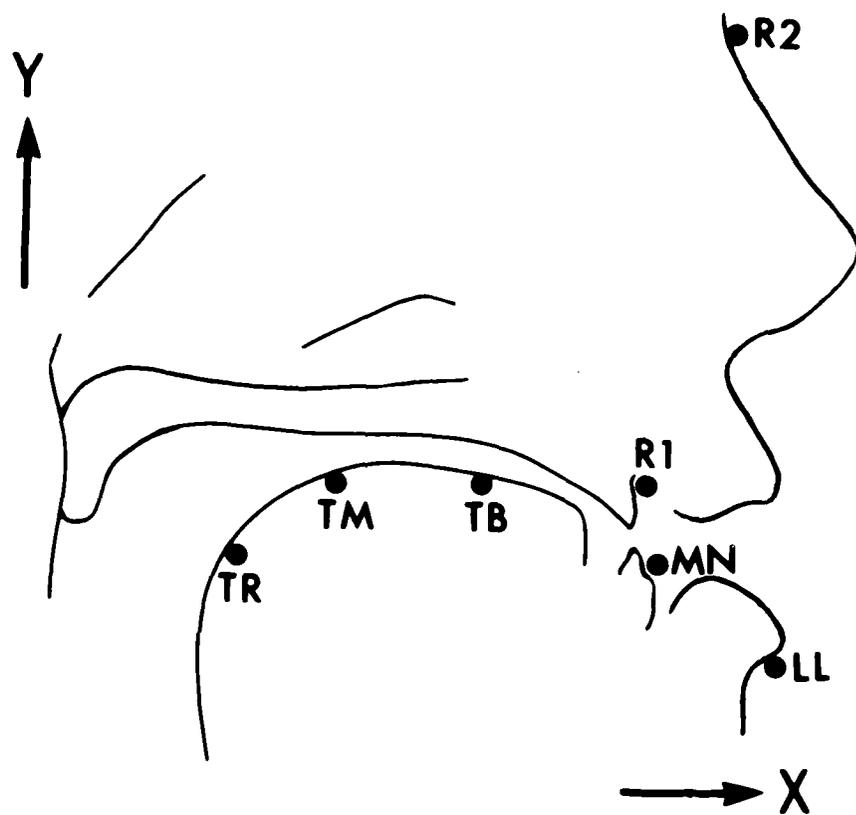
Figure 5. Pellet positions for tongue body (TB), tongue "middle" or dorsum (TM), and lower lip (LL). R_1 and R_2 are reference pellets for the coordinate system. (TR and MN were not analyzed.)
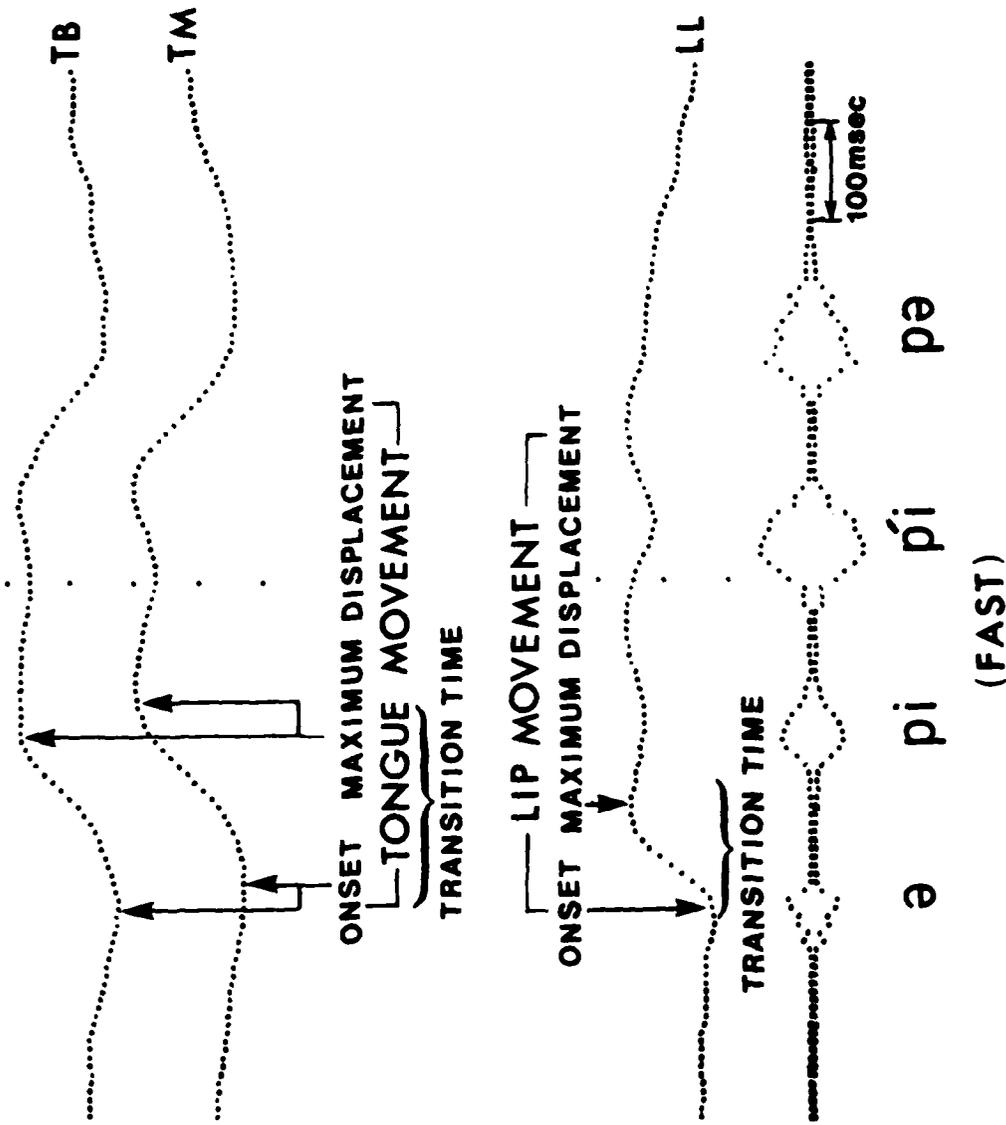
Figure 6. Tongue and lip movements (y-axis displacement) for the utterance /əpípipə/ spoken at a fast rate. Onset of movement, transition time, and maximum displacement ₂e indicated (see text). The acoustic waveform appears underneath the movement tracings.

TB

TM

LL

100msec

ONSET MAXIMUM DISPLACEMENT
TONGUE MOVEMENT
TRANSITION TIME

LIP MOVEMENT
ONSET MAXIMUM DISPLACEMENT
TRANSITION TIME

e   pi   pí   pi   ed

(FAST)

AMPLITUDE (ARBITRARY)

time. These measures are indicated in Figure 6. Measures were made of the second syllable, but not the third syllable, because the consonant between them varied. For the same reason, two definitions were used of acoustic syllable duration. For syllables ending in /ʔ/, /p/, and /k/, acoustic evidence of closure was used as the right-most syllable boundary, as in Experiment 1. For /h/, time of friction offset was used.

RESULTS

As in the analysis of the preceding experiment, we used the binomial test, a nonparametric statistic, to examine the effects of speaking rate (fast vs. slow) and the effects of stress (stressed vs. unstressed) on the various acoustic and kinematic parameters. The size of the sample was too small to assess the effects of the intervocalic consonants. However, inspection revealed no obvious effects of the consonant that closed the syllable of interest on events occurring at syllable onset.

I. Acoustic Analysis

The mean acoustic syllable durations are shown in Figure 7. Not surprisingly, and in accord with the previous results, there are significant effects of both speaking rate ($p < .01$) and stress ($p < .01$). Interestingly, the average syllable durations adopted by the speaker in this experiment were not very different from those observed in the previous experiment. Mean values for the different intervocalic consonant conditions are included in the figure, although the significance of differences cannot be tested. Again, the results are as we would expect from the existing literature.

II. Kinematic Analysis: Variations in Articulator Movement

Values for transition time and maximum displacement are shown in Table 4. There are no significant differences in maximum displacement for either stress or speaking rate. Indeed, average values do not show a systematic pattern. This result is somewhat surprising, in view of the literature indicating systematic effects of stress, although not speaking rate, on formant values (Gay, 1977; Harris, 1978; Verbrugge & Shankweiler, 1977). The only obvious explanation is that the pellet placements used here may not have been maximally sensitive to position of the tongue front. For example, the TB pellet is quite far back on the tongue body. Transition time, however, shows significant effects of stress for four out of six cases, and of speaking rate for two out of six cases. Furthermore, mean differences are, with one exception [LL (x coordinate)], always in the expected direction--that is, the duration of articulator movement is always shorter for movements in "fast" syllables, and for unstressed syllables. Thus, acoustic duration, duration of EMG activity, and lip and tongue transition times all show the same general effects of stress and speaking rate.

52

Mean Acoustic Syllable Durations

* p < .01
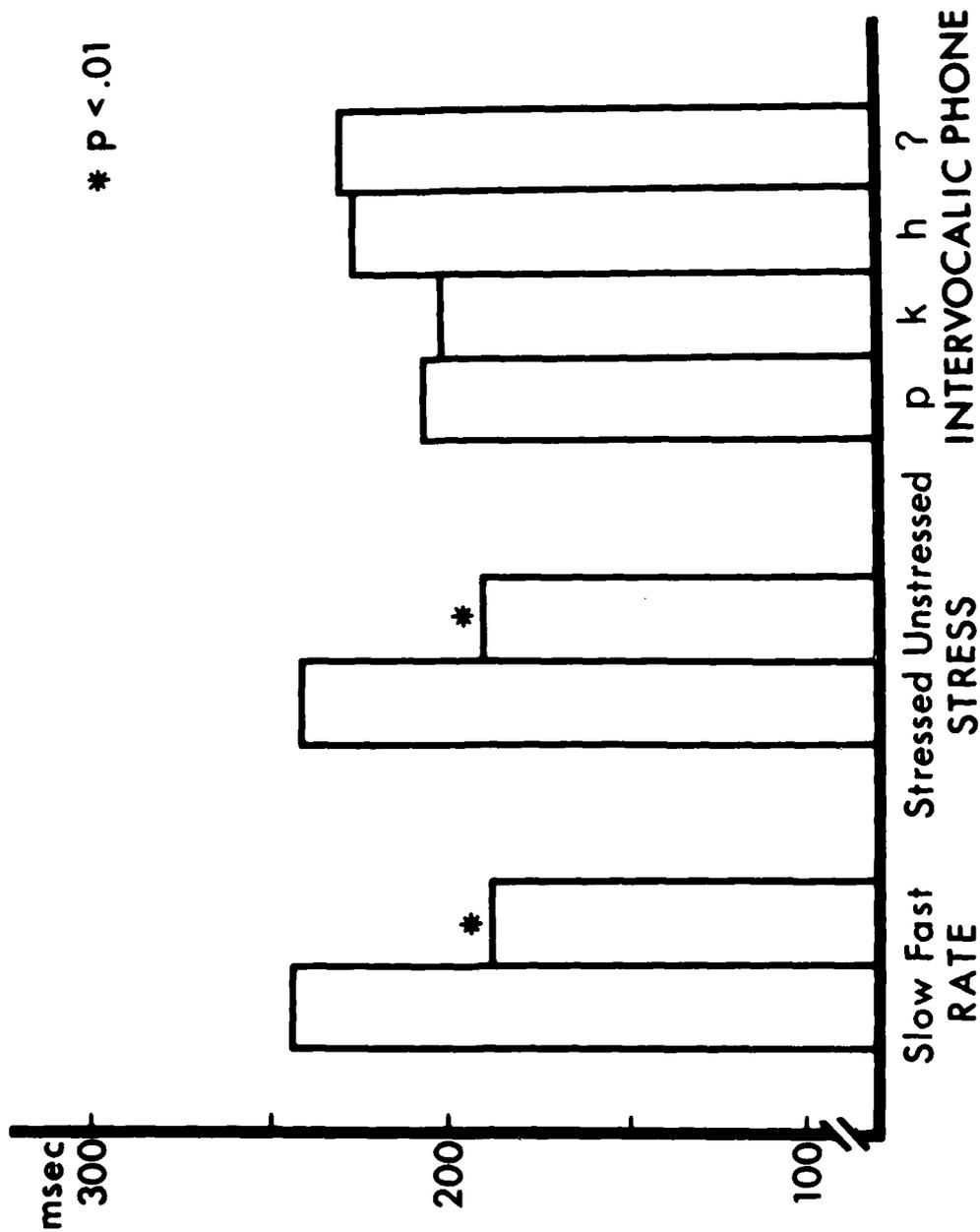
Figure 7. Mean acoustic syllable durations as a function of speaking rate, syllable stress, and intervocalic phone.

53

Table 4

Means and standard deviations (sd) for maximum displacements (in
arbitrary units) and movement transition times (in frames) as a function
of stress and rate. x and y coordinates are indicated for the lower
lip (LL), tongue middle or dorsum (TM), and tongue blade (TB) pellets.

	Maximum Displacement				Transition Time			
	Slow		Fast		Slow		Fast	
	Mean	(sd)	Mean	(sd)	Mean	(sd)	Mean	(sd)
LL(x)	19.3	(1.2)	19.7	(3.8)	12.0	(2.0)	12.3	(2.7)
LL(y)	32.3	(2.9)	34.3	(3.1)	12.9	(2.9)*	10.1	(1.6)
TM(x)	45.4	(2.8)	45.8	(2.8)	25.0	(2.7)*	22.1	(3.6)
TM(y)	49.9	(3.3)	45.6	(4.5)	20.6	(3.2)	19.2	(3.5)
TB(x)	49.9	(2.4)	48.0	(2.3)	25.9	(2.5)	22.0	(3.8)
TB(y)	40.3	(3.2)	38.5	(3.6)	22.4	(2.9)	19.0	(3.6)

	Stressed		Unstressed		Stressed		Unstressed	
	Mean	(sd)	Mean	(sd)	Mean	(sd)	Mean	(sd)
LL(x)	20.0	(1.5)	19.0	(3.6)	13.7	(2.1)*	10.5	(1.1)
LL(y)	33.5	(3.3)	33.0	(3.0)	13.2	(2.6)**	9.7	(1.2)
TM(x)	45.4	(2.8)	45.8	(2.8)	24.9	(3.7)	22.2	(2.7)
TM(y)	48.2	(4.1)	47.3	(4.9)	21.9	(2.8)*	18.0	(2.8)
TB(x)	49.1	(2.3)	48.7	(2.7)	25.5	(3.6)*	22.4	(3.2)
TB(y)	37.8	(2.4)	40.9	(3.6)	22.0	(3.0)	19.6	(3.7)

N=8

*p < .05
**p < .01

54

III. Kinematic Analysis: Temporal Relations Among Articulator Movements

A finding of Experiment 1 was that temporal relations among some aspects of EMG activity remained stable relative to large changes in the duration of other variables. The same type of relationship can be seen in the present experiment. Table 5 shows the ranges of acoustic syllable duration, transition time, and the relationship of peak lip displacement (approximately, greatest closure) to the onset of tongue movement. An examination of these values shows that the range of acoustic duration is large and the range of overlap is small. Both sets of values are comparable with those of the preceding experiment. Transition time shows an intermediate range of variation over suprasegmental change. The variances of the acoustic duration and transition time measures were tested for significance against the variance of one overlap measure (lip peak to tongue onset, LL and TM pellets, y-coordinates). With one exception (LL, x-coordinate), the variability of acoustic duration and transition time was greater than that of articulatory overlap (ps <.05).

Table 5

The range of variation in acoustic syllable duration, transition time
and the time of peak lip displacement to the onset of
tongue activity, over changes in speaking rate and syllable stress, in msec.

Acoustic Duration	Transition Times		Lip peak to tongue onset (overlap)	
114.3	TM(x)	111.1	TM(x)	39.7
	TM(y)	103.2	TM(y)	39.7
	TB(x)	95.2	TB(x)	39.7
	TB(y)	103.2	TB(y)	59.5
	LL(x)	63.5		
	LL(y)	63.5		

Binomial tests were performed on the measured temporal overlaps, separately for x and y values, for lip and the two tongue pellets. No effect of speaking rate (p >.3) or syllable stress (p >.3) was significant. In this experiment, the lack of significant effect of speaking rate or stress is less dramatic than in the previous one, because the range of transition times is relatively small, compared to the range of muscle activity times. However, the results are substantively similar, although it might be remarked that the data corpus for this experiment is much smaller than in the previous one.

55

Figure 8 illustrates the stability of timing relations over changes in speaking rate and syllable stress for y coordinates for the LL and TM pellets. Plots for x coordinates, and for the TB pellet, look similar. Again, the slope of the best-fitting straight line is not significantly different from zero, so that the overlap changes little relative to large variations in acoustic syllable duration. Thus, while this experiment is of smaller scope than the previous one, and the measures used are not precisely the same, the results are quite similar.

DISCUSSION

When investigators have examined acoustic, electromyographic, and kine-matic patterns over several speaking rates and levels of stress, the results have often been very variable, both among subjects and among experiments. One measure that is extremely consistent, however, is the acoustic duration of syllables; unstressed syllables and syllables spoken quickly are typically shorter than their stressed or slowly spoken counterparts. Similarly, meas-ures of acoustic syllable durations in Experiments 1 and 2 showed shorter durations for fast and unstressed syllables relative to syllables spoken slowly or with primary stress, suggesting that subjects consistently changed rate and stress of their speech when instructed to do so.

The effects of changes in speaking rate and syllable stress on EMG activity are not as clearly understood. In the present experiment, the observed patterns of muscle activity that occurred over variations in speaking rate and syllable stress were less consistent than the measures of acoustic duration. First consider the effects of changing speaking rate. In Experi-ment 1, the observed decrease in duration of genioglossus activity with an increase in speaking rate is in agreement with that reported by Gay and Ushijima (1974) and Gay et al. (1974). In Experiment 1, peak amplitude of activity in genioglossus did not vary as a function of speaking rate; Gay and his colleagues report decreases in genioglossus activity as speaking rate increases. The pattern of changes in orbicularis oris activity did not confirm the pattern of changes reported by Gay and his colleagues for two speakers (Gay & Hirose, 1973; Gay & Ushijima, 1974; Gay et al., 1974). In their experiment, peak amplitude of orbicularis oris activity increased with increases in speaking rate; in Experiment 1, no changes in peak amplitude as a function of speaking rate were observed. The duration of activity in orbicularis oris, which here decreased with an increase in speaking rate, was not reported by Gay et al.

The EMG patterns resulting from changes in syllable stress are compatible with the small body of data available on this subject. The peak amplitude of activity in genioglossus was higher, and its duration of activity longer, when the vowel was stressed rather than unstressed. Identical observations have been reported by Harris (1971, 1973) for genioglossus activity during produc-tion of /i/. The peak amplitude of EMG activity in orbicularis oris during the production of bilabial stops was also observed to increase with increased stress, in agreement with a finding by Harris et al. (1968). The duration of orbicularis oris activity increased with an increase in syllable stress, an observation that has not, to our knowledge, been previously reported.

56
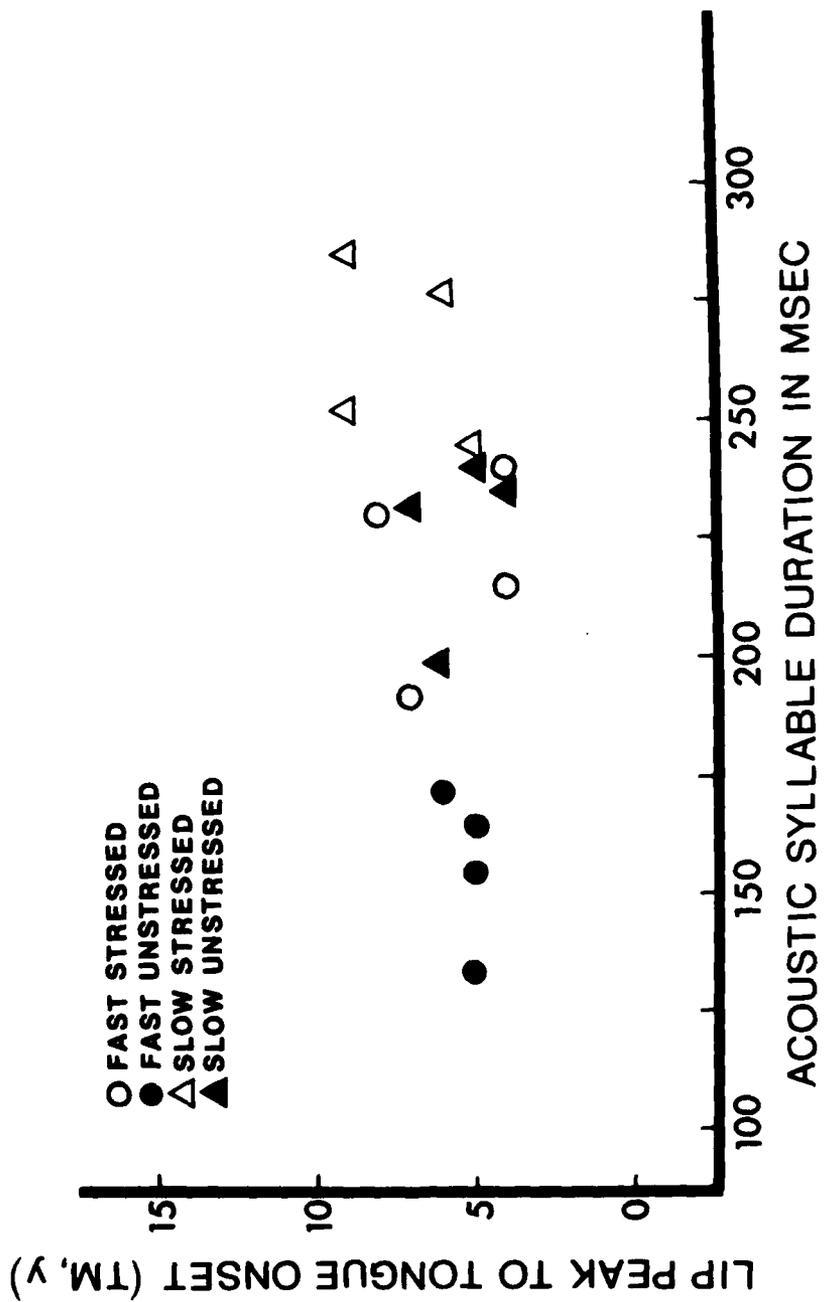
Figure 8. Acoustic syllable duration plotted against the interval (in frames) from peak lip movement to the onset of movement of the tongue dorsum (pellets LL and TM, respectively; y-coordinates).

The pattern of duration changes in genioglossus and orbicularis oris activity indicates that the vowel portion of CV and VC syllables is more "elastic" than the consonant portion (Gaitenby, 1965; Gay, 1978; Kozhevnikov & Chistovich, 1965; Lehiste, 1970; Port, 1976). Specifically, with an increase in speaking rate or a decrease in syllable stress the duration of genioglossus activity shortened more than did the duration of orbicularis oris activity (in both absolute and relative time).

The hypothesis (Lindblom, 1963) that changes in acoustic duration that result from either changes in speaking rate or level of stress are the product of a change in a single production rule was not supported by the results. Although variations in both stress and rate affected acoustic syllable duration, they apparently produced these durational changes by distinct effects on muscle behavior. For both orbicularis oris and genioglossus, decreases in speaking rate lengthened the duration of EMG activity, but had no effect on peak amplitude of activity. However, increases in syllable stress not only lengthened the duration, but also increased the peak amplitude of activity, of both muscles.

The results of Experiment 2 did not give very clear evidence for production differences between rate and stress changes. There was no evidence for significant differences in maximum displacement as a consequence of stress or rate change. Both stress and rate affected the measured duration of articulator movement (transition time). However, the pattern of results supports the notion of somewhat larger effects of stress than of speaking rate.

It should be apparent that the effect of rate or stress changes on motor events cannot be simply to speed up or slow down the execution of putative invariant motor commands (phonemic or otherwise; Kozhevnikov & Chistovich, 1965; Lindblom, 1963; Shaffer, 1976). If it is argued that articulatory events are the consequences of motor commands, rules must be established governing how the motor activity underlying commands for any given segment alters as a function of variations in speaking rate or syllable stress (see also Harris, Gay, Sholes, & Lieberman, 1968; Harris, 1971, 1973, 1978; Gay, 1978). A single rule (as proposed by Lindblom, 1963) will not suffice if one considers that the systematic alterations in patterns of EMG activity may themselves be specific to the type of linguistic transformation. It should be underscored that a talker has two very different aims when changing speaking rate and when changing stress; for the former the talker must move the articulators slower (or faster), whereas for the latter the talker must make certain syllables more (or less) prominent. Intuition also suggests that changing stress and rate are not equivalent motor transformations. It is very difficult for a speaker to alternate fast and slow speaking rates syllable-by-syllable, but very easy (and common) for a speaker to alternate stressed and unstressed syllables.

In the literature, decreases in syllable stress and increases in speaking rate have often been described as having similar acoustic consequences. Vowels in unstressed syllables and syllables spoken quickly are usually characterized as shorter and more centralized in the F1/F2 vowel space than their stressed or more slowly spoken counterparts (e.g., Lindblom, 1963; Stevens & House, 1963). In contrast, spectrographic measures of the speech

58

signal have indicated different effects of stress and rate on vowel acoustics. Verbrugge and Shankweiler (1977), for example, reported the usual changes in syllable duration when speaking rate or syllable stress was varied. However, formant frequency measures of the vowel spectra revealed no centralization in fast relative to slow speech, but large vowel formant shifts in unstressed relative to stressed syllables. Similar findings were reported by Harris (1978) and Gay (1977). Gay (1977) also reported that unstressed syllables show reduced F_0 and amplitude contours relative to quickly spoken stressed syllables, even when they are of equal duration.

Compared to the considerable individual variations in measures of orbicularis oris and genioglossus, temporal relations between genioglossus and orbicularis oris remained relatively fixed over changes in speaking rate and syllable stress. Similarly, peak lip closure and tongue onset relations, in Experiment 2, varied very little over suprasegmental change. Thus, aspects of the motor activity underlying lip movements for the bilabial stop and tongue fronting for the vowel, and their kinematic consequences, remained within relatively tight temporal boundaries.

It should be noted that the importance of temporal relations in speech production has been emphasized elsewhere. For example, Lisker and Abramson (1964, 1971) argue that the diverse acoustic consequences of a voicing contrast in stop consonants result primarily from a coordinated timing relation between glottal and supraglottal events. That is, the timing of the release of oral occlusion relative to the onset of glottal pulsing has acoustic consequences that distinguish voiced from voiceless stops in syllable-initial position. Raphael (1975), in an investigation of the effects of final consonant voicing on vowel duration, observed that the vowel gesture lengthens before a voiced consonant but the onset of muscle activity for the following consonant occurs at approximately the same time relative to the offset of muscle activity for the preceding vowel--exactly what we found in Experiment 1.

In the experiments described here, we presented evidence that the relative timing of EMG activity in two articulatory muscles, and the relative timing of lip and tongue movements, remained fairly stable compared with the large variations observed in individual variables. Although the relationship between muscle activity and movement patterns (or, for that matter, between EMG, movement, and acoustics), is as yet unclear, we find it encouraging that both the electromyographic and kinematic data converge on the same general finding concerning stress and rate effects on speech motor control.

This finding, that temporal relations among aspects of motor activity or kinematic events remain relatively stable over large changes in magnitude or duration of individual variables, is not unique to speech production but is common to diverse problems of motor control and coordination (see Kelso, Tuller, & Harris, 1981, for a review). The temporal patterns observed here, however, involved a very restricted set of articulatory muscles and linguistic elements. In order to explore whether the results are indicative of a general constraint on articulatory timing, we performed an extension of these experiments in which we examined intersegmental timing relations within a larger group of muscles over more varied utterances. The results are presented in the following paper (Tuller, Kelso, & Harris, 1981) and suggest that the

relative timing of activity in various muscles is in fact preserved over metrical variations in speaking rate and syllable stress.

REFERENCE NOTE

1. Fujimura, O., Miller, J., & Nelson, W. A speech research center with a computer controlled X-ray microbeam system. Bell Telephone Laboratories, unpublished manuscript, 1980.

REFERENCES

Abbs, J. H., & Kennedy, J. G. III. Neurophysiological processes of speech movement control. SMCL Preprints (University of Wisconsin, Madison), 1980, (Spring-Summer), 62-99.

Alfonso, P., & Baer, T. An electromyographic-cinefluorographic-acoustic study of dynamic vowel production. Haskins Laboratories Status Report on Speech Research, 1981, SR-65 this volume.

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

Fant, G., Stalhammar, U., & Karlsson, I. Swedish vowels in speech material of various complexity. In Speech Communication Seminar, Stockholm, 1974. Uppsala: Almqvist and Wiksell, 1974.

Fromkin, V. A. Neuromuscular specifications of linguistic units. Language and Speech, 1966, 9, 170-199.

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

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

Gaitenby, J. H. The elastic word. Haskins Laboratories Status Report on Speech Research, 1965, SR-2, 3.1-3.12.

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

Gay, T. Effect of speaking rate on vowel formant movements. Journal of the Acoustical Society of America, 1978, 63, 223-230.

Gay, T., & Hirose, H. Effect of speaking rate on labial consonant production: A combined electromyographic/high speed motion picture study. Phonetica, 1973, 27, 44-56.

Gay, T., & Ushijima, T. Effect of speaking rate on stop consonant-vowel articulation. In Proceedings of the Speech Communication Seminar, Stockholm, 1974. Uppsala: Almqvist and Wiksell, 1974, 205-209.

Gay, T., Ushijima, T., Hirose, H., & Cooper, F. S. Effect of speaking rate on labial consonant-vowel articulation. Journal of Phonetics, 1974, 2, 47-63.

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

Harris, K. S. Stress and syllable duration change. Haskins Laboratories

Status Report on Speech Research, 1973, SR-35/36, 31-38.

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

Harris, K. S., Gay, T., Sholes, G. N., & Lieberman, P. Some stress effects on electromyographic measures of consonant articulations. Haskins Laboratories Status Report on Speech Research, 1968, SR-13/14, 137-151.

Harris, K. S., Lysaught, G. F., & Schvey, M. M. Some aspects of the production of oral and nasal labial stops. Language and Speech, 1965, 8, 135-147.

Hirose, H. Electromyography of the articulatory muscles: Current instrumentation and techniques. Haskins Laboratories Status Report on Speech Research, 1971, SR-25/26, 73-86.

Kelso, J. A. S., Tuller, B., & Harris, K. S. A 'dynamic pattern' perspective on the control and coordination of movement. To appear in P. MacNeilage (Ed.), Speech production, Also Haskins Laboratories Status Report on Speech Research, 1981, SR-65, this volume.

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

Kewley-Port, D. An experimental evaluation of the EMG data processing system: Time constant choice for digital integration. Haskins Laboratories Status Report on Speech Research, 1974, SR-37/38, 65-72.

Kiritani, S. Articulatory studies by the X-ray microbeam system. In M. Sawashima & F. S. Cooper (Eds.), Dynamic aspects of speech production. Tokyo: University of Tokyo Press, 1977.

Kiritani, S., Itoh, K., & Fujimura, O. Tongue-pellet tracking by a computer-controlled X-ray microbeam system. Journal of the Acoustical Society of America, 1975, 57, 1516-1520.

Kozhevnikov, V., & Chistovich, L. Speech: Articulation and perception. Moscow-Leningrad, 1965 (English translation: J.P.R.S., Washington, D.C. No. JPRS 30543).

Lehiste, I. Suprasegmentals. Cambridge, Mass.: M.I.T. Press, 1970.

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

Lieberman, P. Some acoustic correlates of word stress in American English. Journal of the Acoustical Society of America, 1960, 32, 451-454.

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

Lindblom, B. E. F. Motor control mechanisms. Papers from the Institute of Linguistics, University of Stockholm, 1974, 26, 1-19.

Lisker, L., & Abramson, A. A cross-linguistic study of voicing in initial stops. Word, 1964, 20, 384-422.

Lisker, L., & Abramson, A. S. Distinctive features and laryngeal control. Language, 1971, 47, 767-785.

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

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

Öhman, S. E. G. Discussion paper on acoustics of speech linguistic theory and speech research. In J. H. Gilbert (Ed.), Speech and cortical functioning. New York: Academic, 1972.

Port, R. F. The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words. Unpublished doctoral dissertation, University of Connecticut, 1976.

Raphael, L. J. The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English. Journal of Phonetics, 1975, 3(1), 25-33.

Raphael, L. J., & Bell-Berti, F. Tongue musculature and the feature of tension in English vowels. Phonetica, 1975, 32(1), 61-73.

Siegal, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Shaffer, L. H. Intention and performance. Psychological Review, 1976, 83, 375-393.

Shik, M. L., & Orlovskii, G. N. Neurophysiology of locomotor automatism. Physiological Reviews, 1976, 56, 465-501.

Smith, T. S. A phonetic study of the extrinsic tongue muscles. UCLA Working Papers in Phonetics, 1971, 18.

Stevens, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), Human communication: A unified view. New York: McGraw Hill, 1973.

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

Sussman, H. M., & MacNeilage, P. F. Motor unit correlates of stress: Preliminary observations. Journal of the Acoustical Society of America, 1978, 64, 338-340.

Tiffany, W. R. Nonrandom sources of variation in vowel quality. Journal of the Acoustical Society of America, 1959, 58, 434-445.

Tuller, B. Coordination among articulators in speech. Unpublished doctoral dissertation, University of Connecticut, 1980.

Tuller, B., Kelso, J. A. S., & Harris, K. S. Phase relationships among articulator muscles as a function of speaking rate and syllable stress. Haskins Laboratories Status Report on Speech Research, 1981, SR-65, this volume.

Verbrugge, R. R., & Shankweiler, D. P. Prosodic information for vowel identity. Journal of the Acoustical Society of America, 1977, 61, S39. (Abstract)

Zimmerman, G., Kelso, J. A. S., & Lander, L. Articulatory behavior pre- and post-full-mouth tooth extraction and alveoloplasty. Journal of Speech and Hearing Research, 1980, 23, 630-645.

FOOTNOTE

[1]Although Lindblom's later work does not adhere to the originally described model (e.g., Lindblom, 1968, cited in 1974), it has strongly influenced recent experimental work (e.g., Fant, Stalhammer, & Karlsson, 1974; Gay, 1978; Gay et al. 1974; Harris, 1978) and is, we believe, representative of a class of theories of speech motor control.

PHASE RELATIONSHIPS AMONG ARTICULATOR MUSCLES AS A FUNCTION OF SPEAKING RATE AND STRESS

Betty Tuller,+ J. A. Scott Kelso,++ and Katherine S. Harris+++

Abstract. The present experiment--continuous with our earlier work--examined temporal aspects of muscle activity over suprasegmental changes in speaking rate and syllable stress. Five muscles known to be associated with lip, tongue, and jaw movements were sampled. Large variations were observed in magnitude and duration of activity in individual muscles. However, analysis of the phase relationships among muscles suggested that the timing of consonant-related muscle activity remained fixed relative to activity for the flanking vowels. This style of control, in which the relative timing of activity among muscles is preserved across metrical changes, is a characteristic of many nonspeech motor activities and may rationalize certain findings in speech production and perception.

Two basic types of explanation have been proposed for the changes in segmental timing that occur with variations in speaking rate and syllable stress. One view is that the segmental "commands" for syllables spoken quickly and for unstressed syllables show more extensive temporal overlap than the same syllables spoken more slowly or with greater syllabic stress (e.g., Kozhevnikov & Chistovich, 1965; Lindblom, 1963; Shaffer, 1976). An alternative view is that the temporal relationships among articulations remain constant over changes in stress and speaking rate, but the individual gestures themselves change (e.g., Kent & Moll, 1975; Kent & Netsell, 1971; Löfqvist & Yoshioka, 1980, 1981). In earlier papers (Kelso, Tuller, & Harris, 1981; Tuller & Harris, 1980; Tuller, Harris, & Kelso, 1981), we provided evidence for the latter hypothesis. Compared with the large variations that were observed in the magnitude and duration of electromyographic (EMG) activity in individual muscles, the temporal relationship between consonant-and vowel-related activity in a given consonant-vowel (CV) or vowel-consonant (VC) pair (and the resulting kinematics) remained comparatively stable over suprasegmental change. However, no broader conclusions could be drawn concerning the preservation of temporal aspects of articulation because the phonetic structure of the utterances used did not allow investigation of intersegmental

timing over more than two phonetic segments. It may be that individual articulatory events are temporally constrained relative to some longer period of articulation than examined in previous experiments. The longer period of activity may vary as a function of changes in speaking rate and syllable stress and, possibly, may be a factor in the perceptual specification of these changes.

The present experiment was designed to explore the possibility that relative timing of articulatory events is preserved over suprasegmental change. There are some a priori grounds from two quite disparate sources that might motivate a relative timing hypothesis. The first comes from the speech perception literature. For example, long and short vowel pairs are distinguished perceptually (at least in part) by vowel duration in relation to perceived rate of speech and not by absolute vowel duration (Rakerd, Verbrugge, & Shankweiler, 1980). The second comes from emerging work on other motor activities that suggests that relative timing (phasing) among muscles and kinematic events is preserved over metrical changes in force or rate. For example, MacMillan (1975) observed that in a freely locomoting lobster, activity in the limb muscles occurs at a constant phase position relative to the step cycle, even when a load is attached to the limb. As yet, however, no experiment in the speech production literature has been sufficiently expanded to evaluate relative timing among segmental articulations. In the present experiment, electromyographic recordings from lip, tongue, and jaw muscles were obtained during production of utterances whose phonetic structure allowed intersegmental timing relationships to be examined over more than two phonetic segments. The results suggest that the preservation of relative timing of muscle activity over metrical change is characteristic of the temporal organization of speech.

METHOD

Subjects

The subjects were five adult females: four were native speakers of American English, and one was an English-speaking native of New Zealand. Four of the five subjects were naive as to the purpose of the experiment. It may be remarked at the outset that neither dialect nor experimental sophistication had any conspicuous effects.

Materials and Procedures

The speech sample consisted of eight two-syllable nonsense utterances of the form $/pV_1CV_2p/$, where C was either /p/ or /k/ and V_n was either /i/ or /a/. Each utterance was spoken with stress placed on either the first or second syllable. The subjects read quasi-random lists of these utterances at two self-selected speaking rates, "slow" (conversational) and "fast." Two of the five subjects were not able to produce the utterances at a consistently faster rate than the "slow" rate they had chosen; these two subjects did not complete the utterance list at the "fast" rate. Each utterance was embedded in the carrier sentence "It's a _____ again," thus minimizing the effects of initial and final lengthening and prosodic variations. Twelve repetitions were produced of each utterance.

64

Data Recording

Electromyographic activity was recorded from orbicularis oris (OO) using paint-on surface electrodes (Allen & Lubker, 1972) spaced at about one-half centimeter from the vermilion border of the lips. Orbicularis oris is known to participate in bilabial closure (Harris, Lysaught, & Schvey, 1965; Fromkin, 1966).

Electromyographic activity was also recorded from the anterior portion of genioglossus (GG), anterior belly of the digastric (ABD), medial (internal) pterygoid (MP), and the inferior head of lateral (external) pterygoid (LPI), using bipolar hooked-wire electrodes (Hirose, 1971). Genioglossus bunches the main body of the tongue and brings it forward, and is active in production of the vowel /i/ (e.g., Alphonso & Baer, 1981; Raphael & Bell-Berti, 1975; Smith, 1971). The functional properties of the additional muscles have been described in detail elsewhere (Tuller, Harris, & Gross, in press). The anterior belly of digastric and the inferior head of lateral pterygoid are active in association with jaw lowering during speech (e.g., for the production of /a/). Medial pterygoid acts to raise the jaw during speech.

During insertion of the hooked-wire electrodes, the subject was in a slightly reclined position and breathed nitrous oxide to reduce discomfort. Detailed descriptions of electrode placement and insertion techniques may be found in Ahlgren (1966) and Gross and Lipke (Note 1). Verification of electrode placements used maneuvers for which the role of each muscle is well established (Ahlgren, 1966; Carlsöo, 1952, 1956; Harris et al., 1965; Møller, 1974; Moyers, 1950; Smith, 1971).

The EMG potentials from the various muscles were recorded on multichannel FM tape, rectified, computer-sampled, software integrated with a time constant of 35 msec, and averaged using the Haskins Laboratories EMG system described by Kewley-Port (1974). Acoustic recordings were made simultaneously with the EMG recordings and both were analyzed on subsequent playback.

The EMG tokens were realigned and reaveraged three times, at the onset of the acoustic release burst for the first, second, and third stop consonants, respectively. In this way, average muscle activity could be examined at specific points of interest without the time-smearing effects of averaging tokens that were aligned at a temporally distant point.

Onsets and offsets of activity were determined from data averaged around the acoustic line-up point closest to the activity of interest. The averaging program provides a numerical listing of the mean amplitude of each EMG signal in microvolts during successive 5-msec intervals. Baseline and peak values for each muscle were determined from this numerical listing; the time of onset (and offset) was defined as the time when the relevant muscle activity increased (or decreased) to 10% of its range of activity. Typically, 10% of the range was just slightly higher than the background level of activity in each muscle. Some of the electrodes were displaced during the course of the experiment or recorded EMG activity from a neighboring muscle as well as the muscle of interest; data from these electrodes were not used in the analyses that follow. Table 1 shows the electrode placements for each subject that had stable, uncontaminated EMG activity.

The acoustic recordings were measured for their durational characteristics, using an interactive computer program that displays the acoustic waveform. Measures were made of the interval from the first acoustic evidence of closure for the initial /p/ (defined here as the point when the high frequency components of the periodic wave disappear) to the second acoustic evidence of closure (for the medial stop consonant). For ease of communication, this interval will be referred to below as the "acoustic duration of the first syllable." The measured interval from the second acoustic evidence of closure to the third (for the final /p/) will be referred to as the "acoustic duration of the second syllable." These measures were averaged, omitting tokens for which there were EMG processing failures.

Table 1

Adequate electrode placements and EMG recordings for each subject.

| | Subject | | | | |
	PS*	BT*	JT*	GC	VR
Orbicularis Oris	X	X	X	X	X
Genioglossus	X	X	X	X	
Medial Pterygoid	X	X		X	X
Lateral Pterygoid- Inferior Head	X	X	X	X	X
Anterior Belly of Digastric	X	X		X	X

Asterisks (*) denote those subjects who produced the utterances at two different speaking rates.

RESULTS AND DISCUSSION

In this experiment, the sample size was sufficiently small to warrant the use of nonparametric statistics, specifically binomial tests and z-scores corrected for continuity (Siegal, 1956). Unless z-scores are explicitly given, the analysis used was a binomial test, and all analyses were two-tailed. We should emphasize that this analysis examines the direction of change, not the magnitude of change.

I. Acoustic Analysis

The acoustic durations of syllables were examined to determine the effects of syllable stress (stressed vs. unstressed), speaking rate (fast vs. slow), vowel (/i/ vs. /a/), consonant (/p/ vs. /k/), and syllable (first vs. second). Mean durations for each syllable type are given in Figure 1. Stressed syllables, and syllables spoken slowly, were significantly longer than the same syllables destressed or spoken quickly (z = -6.50, p <.001 and z = -5.79, p <.001, respectively). Vowel identity also affected syllable duration: syllables containing /a/ were significantly longer than syllables containing /i/ (z = -7.63, p <.001; see Peterson & Lehiste, 1960). Mean acoustic duration for the first syllable was not different from mean acoustic duration for the second syllable (z = .15, p >.2), and the effect of consonant identity was not significant (z = -1.14, p >.2).

The effects of changes in speaking rate and stress on the acoustic durations of syllables are by now well established in the speech production literature. Unstressed syllables and syllables spoken quickly are generally found to be shorter than stressed syllables and syllables spoken slowly (e.g., Fry, 1955, 1958; Gaitenby, 1965; Kozhevnikov & Chistovich, 1965; Lehiste, 1970; Lindblom, 1963; Tiffany, 1959). Measures of acoustic syllable durations in this experiment support these general findings, suggesting that subjects consistently changed speech rate and stress when instructed to do so.

II. EMG Analysis: Variations in Individual Muscle Actions

Binomial tests examining the effects of speaking rate on the duration and peak amplitude of activity in each muscle were performed on the data from the three speakers who were able to produce the utterances at two different rates (PS, JT, BT). Analyses examining the effects of syllable stress and syllable position were performed on all five speakers. Separate analyses were performed for each muscle. Utterances containing /k/ will not be discussed because no muscle showed clear activity for that segment alone. The basic results are presented in Table 2.

a. Lip muscle activity

Orbicularis oris. Orbicularis oris duration was longer when the /p/ occurred in syllables spoken slowly rather than quickly (p <.01) and when the /p/ occurred in the second rather than the first syllable (z = -2.65, p <.01). It should be noted that the initial /p/ in the first syllable is preceded by a schwa (from the carrier phrase "It's a..."), whereas the initial /p/ in the second syllable is preceded by a point vowel. Thus, the lips may have to travel farther to accomplish the bilabial closure for the second syllable than the first. Variations in syllable stress and vowel identity did not affect the duration of orbicularis oris activity (z = -.88, p >.2 and z = -.53, p >.2, respectively).
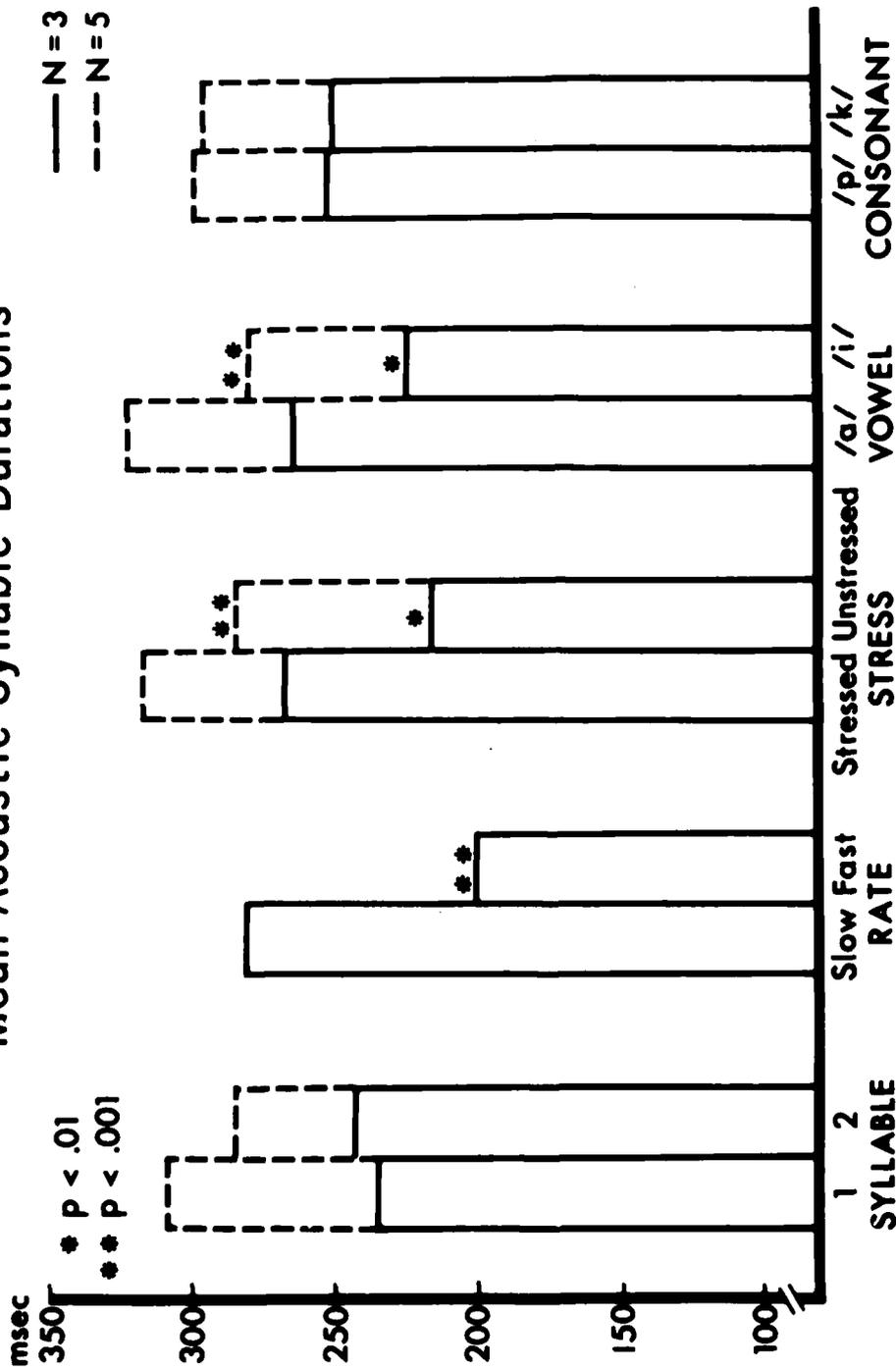
Figure 1. Mean acoustic syllable durations for syllable (1 vs. 2), speaking rate (fast vs. slow), syllable stress (stressed vs. unstressed), vowel (/i/ vs. /a/), and consonant (/p/ vs. /k/). Solid lines represent data from the three subjects who produced the utterances at two speaking rates. Broken lines represent data from all five subjects.

Table 2

Mean duration (in msec) and peak amplitude (in microvolts) in five
muscles as a function of speaking rate (for those subjects who produced
the utterances at two speaking rates) and syllable stress (for all
subjects with good recordings from the indicated muscles).

	Slow	Fast	Stressed	Unstressed
Orbicularis Oris				
Duration	185**	160	197	189
Peak amplitude	283	274	288***	265
Genioglossus				
Duration	280**	207	326**	278
Peak amplitude	133	124	154**	129
Lateral Pterygoid-Inferior head				
Duration	177	160	211**	156
Peak amplitude	173**	203	184**	150
Anterior Belly of Digastric				
Duration	232	217	237*	170
Peak amplitude	168**	253	174*	123
Medial Pterygoid				
Duration	131	112	148	152
Peak amplitude	73*	104	96	89

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

The lack of duration change in orbicularis oris with changes in syllable stress is not consistent with the results of our earlier work (Tuller et al., 1981); in that experiment, orbicularis oris duration increased with stress. To our knowledge, durational changes in oris activity have not been reported elsewhere.

The peak amplitude of orbicularis oris activity increased as a function of increases in syllable stress (z = -3.36, p <.001) and was higher when the /p/ occurred in the second syllable than in the first syllable (z = -2.65, p <.01). The latter effect may be due to the different vowels preceding the bilabial consonant in each syllable. No effects of speaking rate (p >.2) or vowel (p >.2) were observed.

The increase in orbicularis oris peak amplitude of activity with an increase in syllable stress agrees with data reported by Harris, Gay, Sholes, and Lieberman (1968). The lack of variation in orbicularis oris peak amplitude as a function of speaking rate agrees with the results of our previous experiment but differs from reports by Gay and his colleagues (Gay & Hirose, 1973; Gay & Ushijima, 1974; Gay, Ushijima, Hirose, & Cooper, 1974). In those experiments, peak amplitude of activity in orbicularis oris increased with increases in speaking rate for two speakers.

b. <u>Tongue muscle activity</u>

<u>Genioglossus</u>. Variations in speaking rate and stress resulted in different changes in the activity of genioglossus for the production of /i/. An increase in speaking rate was accompanied by a shortened duration of genioglossus activity (p <.01), but peak amplitude was unchanged (p >.2). Increases in syllable stress were associated with increases in both the duration (p <.01) and peak amplitude (p <.01) of genioglossus activity. Syllable position had no effect on either genioglossus duration (p >.05) or peak amplitude (p >.2). This pattern of results is identical to that observed in our earlier work and agrees with data reported by Gay and Ushijima (1974), Gay et al. (1974), and Harris (1971, 1973).

c. <u>Jaw muscle activity:</u> Depressors

<u>Lateral pterygoid (inferior head)</u>. As reported in Tuller, Harris, and Gross (in press), the inferior head of lateral pterygoid was consistently active for production of the vowel /a/. Activity in this muscle was longer and of higher amplitude for stressed syllables containing the vowel /a/ than for the same syllables spoken without primary stress (ps <.01). In contrast, increased speaking rates were associated with increases in peak amplitude of inferior head of lateral pterygoid (p <.01), although the duration of its activity remained unchanged (p >.2). Syllable position had no effect on lateral pterygoid duration or peak amplitude (ps >.2).

<u>Anterior belly of the digastric</u>. The changes in duration and peak amplitude of anterior belly of digastric were similar to the changes observed in inferior head of lateral pterygoid. (Both muscles act to lower the jaw for the open vowel /a/.) Increases in syllable stress were associated with

70

significantly increased anterior belly of digastric duration (p <.05) and peak amplitude (p <.05). In contrast, increases in speaking rate were associated with increases in peak amplitude of activity in anterior belly of digastric (p <.01), but the duration of activity was unaffected (p >.2). Duration and peak amplitude were both unaffected by syllable position (p >.2 and p >.1, respectively).

d. Jaw muscle activity: A jaw elevator

Medial pterygoid. Medial pterygoid activity could only be examined following the vowel /a/; this muscle often showed low levels of activity during /i/ so that an accurate measure of onset of activity in association with jaw raising could not be obtained.

The duration of medial pterygoid activity was not significantly affected by changes in speaking rate, syllable stress, or syllable position. The peak amplitude of medial pterygoid activity was similarly unaffected by variations in syllable stress. However, speaking rate did affect the peak amplitude of activity in this muscle, which was higher during fast speech than during slow speech (p <.05). In addition, peak amplitude of activity was higher in the second syllable than the first (p <.05). This effect is probably the result of different vowels preceding the consonant in each of the two syllables. The initial consonant in the first syllable is preceded by a schwa, whereas the initial consonant in the second syllable is preceded by the open vowel /a/. Thus, the jaw may travel farther for the consonant closure in the second syllable than the first.

To summarize, the changes in each muscle's activity was different for variations in speaking rate than for variations in syllable stress (see Table 2). With increases in rate of speech, the duration of activity in the single tongue muscle observed (genioglossus) and in the lip muscle (orbicularis oris) shortened significantly, but the peak amplitude of activity was unaffected. In contrast, as speaking rate increased, activity in the two jaw depressors (inferior head of lateral pterygoid and anterior belly of digastric) and the jaw raiser (medial pterygoid) increased in peak amplitude of activity but did not change in duration. With a shift from stressed to unstressed syllable production, orbicularis oris decreased in duration of activity but showed no change in peak amplitude; genioglossus, inferior head of lateral pterygoid, and anterior belly of the digastric decreased both duration of activity and peak amplitude.

But is there any consistency as to how different muscles act with changes in speaking rate and syllable stress? One possibility is that muscles active for vowel gestures show one pattern of change with variations in rate and stress, whereas muscles active for consonant gestures show a different pattern of change. This is probably not the case: The effects of variations in speaking rate on genioglossus are very different from the effects on lateral pterygoid (inferior head) and anterior belly of digastric. In fact, genioglossus and orbicularis oris show similar patterns of electromyographic change with variations in rate of speech.

71

Another possibility is that the variations in muscle activity that occur with changes in rate and stress are determined by the articulator involved. For example, the two muscles examined that lower the jaw show identical patterns of change as a function of speech rate and stress. Similarly, Gay et al. (1974) observed the same pattern of change in orbicularis oris amplitude as a function of speaking rate whether the muscle was active for /p/ or for /u/. In the present experiment no lip or tongue muscle was examined other than orbicularis oris and genioglossus, so this hypothesis could not really be tested.

It is important to ask whether the differences observed among muscles in their response to speaking rate and stress variations are consistent with other reports. An increase in speaking rate of utterances containing the vowel /i/ resulted in a decrease in the duration of genioglossus activity with no change in its peak amplitude. An increase in speaking rate of utterances containing the vowel /a/ resulted in lateral pterygoid and anterior belly of digastric maintaining the same duration of activity as during slow productions of /a/, but the peak amplitude of activity in each muscle increased. For both utterance types, the measured acoustic duration was shorter at the fast than the slow speaking rate. This suggests that fast productions of /i/ "undershoot" (relative to /i/ spoken slowly) to a greater degree than do fast productions of /a/ (relative to /a/ spoken slowly).

There is both acoustic and kinematic support for this hypothesis. For example, fast productions of /i/ and /a/ have higher first formants than when the same vowel is produced slowly (Gay, 1974), which suggests articulatory undershoot for /i/ and overshoot for /a/ as speaking rate increases. X-ray tracings also indicate more articulatory undershoot for /i/ than for /a/ when speaking rate increases (Gay et al., 1974). Kent and Moll (1972), using cinefluorography, found the mandible to be relatively lower for fast /a/ than for slow. Both of these kinematic observations support the acoustic results (Gay, 1974) and the pattern of EMG changes observed in the present experiment.

With regard to stress changes, however, when /i/ and /a/ are spoken in an unstressed manner, they both show acoustic changes consistent with articulatory undershoot (e.g., Delattre, 1969; Verbrugge & Shankweiler, 1977, among many others), and the change in formant frequency tends to be greater than that occurring with variations in speaking rate (Verbrugge & Shankweiler, 1977). As measured electromyographically, genioglossus, lateral pterygoid (inferior head), and anterior belly of digastric all decrease in duration and peak amplitude with a reduction in syllable stress, a finding that supports one aspect of Öhman's (1967) "extra energy" hypothesis: An increase in peak amplitude and duration of EMG activity can be considered as "more energetic" articulation (Harris, 1973). However, the increased energy does not appear to be distributed equally over components of the production system.

In summary, this experiment demonstrated different effects of speaking rate and syllable stress on the pattern of activity in the muscles examined. However, the data could not elucidate whether the pattern of change across muscles as a function of suprasegmental changes was constrained by phonetic or anatomic considerations. It is suggested that the different patterns across muscles are genuine since they are supported by available acoustic and kinematic data.

III. Temporal Constraints on Muscle Actions

A. Intrasegmental timing

The utterance and muscle set used in this experiment allowed an examination of temporal aspects of muscle activity between members of a muscle pair that act synergistically for a given gesture. The muscle pairs examined were orbicularis oris and medial pterygoid, a lip and a jaw muscle both active for the vowel-to-consonant gesture in /əp/ and /ap/, and anterior belly of digastric and lateral pterygoid (inferior head), both jaw muscles active for the consonant-to-vowel jaw lowering in /pa/. The intervals examined included the time from the onset of the first active muscle of the pair to the onset of the second muscle of the pair (onset-to-onset time), the time from the first muscle of the pair to reach peak amplitude to the time of peak amplitude in the second muscle (peak-to-peak time), and the time from the first muscle's offset to the second muscle's offset (offset-to-offset time).

a. Orbicularis oris and medial pterygoid for production of /əp/ and /ap/

The onset of orbicularis oris activity preceded the onset of medial pterygoid activity and medial pterygoid offset preceded orbicularis oris offset (see Fig. 2a). No measured interval was found to vary systematically with changes in speaking rate (onset-to-onset time, peak-to-peak time, and offset-to-offset time, $ps > .2$). However, the onset-to-onset time of orbicularis oris and medial pterygoid varied as a function of syllable stress, this interval being shorter when the vowel in VC syllables was stressed rather than unstressed ($p < .05$). Variations in syllable stress did not affect peak-to-peak time, or offset-to-offset time ($ps > .2$).

Vowel identity was also found to affect the onset-to-onset time of orbicularis oris and medial pterygoid ($p < .01$); the interval from orbicularis oris onset to medial pterygoid onset was shorter for the VC gesture in /ap/ than in /əp/. That is, medial pterygoid activity began earlier relative to orbicularis oris onset when the necessary excursion of jaw movement increased (cf. Öhman, 1965). Peak-to-peak and offset-to-offset times were unaffected ($ps > .2$).

b. Anterior belly of digastric and lateral pterygoid (inferior head) for production of /pa/.

The onset of activity in anterior belly of digastric usually preceded the onset of activity in lateral pterygoid (inferior head); peaks and offsets of activity in anterior belly of digastric and lateral pterygoid usually occurred at approximately the same time. The temporal relationships between these muscles were not systematically affected by changes in speaking rate (onset-to-onset time, peak-to-peak time, and offset-to-offset time; $ps > .2$). Similarly, syllable stress did not systematically affect the measure of peak-to-peak time ($p > .2$) or offset-to-offset time ($p > .2$). However, the measure of onset-to-onset time was significantly affected by changes in syllable stress, being shorter for stressed than unstressed syllables ($p < .01$).

These results indicate that aspects of the EMG patterns of different muscles acting on a single articulator during production of a single phonetic
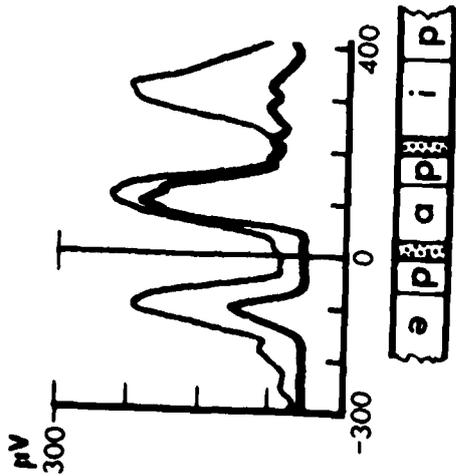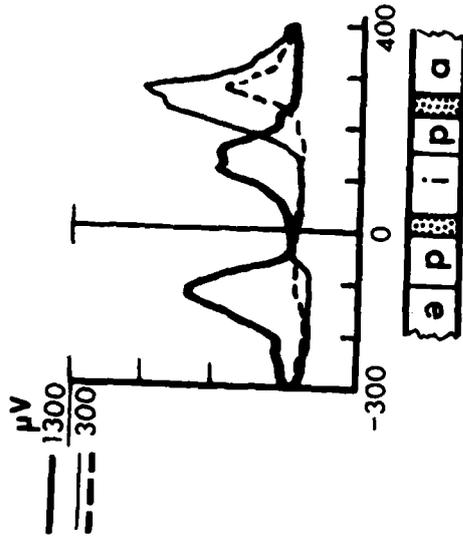
73

Figure 2. Two examples of muscles active for the same phonetic segment. a) orbicularis oris (the thin line) and medial pterygoid (the thick line) for production of /p/; b) anterior belly of digastric (the thin line) and inferior head of lateral pterygoid (the dotted line) for production of /a/. Orbicularis oris (the thick line) is also shown. Schematic acoustics appear below each figure.

segment may change in relation to each other with changes in rate or stress. There is kinematic evidence, however, that movements of <u>different</u> articulators for a single phonetic segment maintain a fixed temporal relationship across rate and stress changes. For example, Kent and Netsell (1971) used cinefluorography to examine tongue body and lip articulations during the production of the syllable /wi/. The relationship between onsets of tongue body and lip movements remained invariant over changes in lexical stress, although the magnitude and velocity of the movements were not preserved (see also Kent & Moll, 1975; Löfqvist, 1981; Löfqvist & Yoshioka, 1981; Lubker, McAllister, & Lindblom, 1977; McAllister, Lubker, & Carlsson, 1974).

B. <u>Intersegmental</u> <u>timing</u> <u>over</u> <u>two</u> <u>phonetic</u> <u>segments</u>

In the following analyses, the action of one muscle is related only to the consonant gesture and the action of a second muscle is related only to the vowel gesture in a CV or VC pair. The temporal overlap of activity in the two muscles was examined to determine whether this measure, earlier observed to be relatively stable (Tuller & Harris, 1980; Tuller et al., 1981), varied as a function of speaking rate or syllable stress.

a. Orbicularis oris and genioglossus for production of /pi/ and /ip/.

In the articulation of /pi/ and /ip/, orbicularis oris moves the lips for the consonant and genioglossus moves the tongue body for production of the vowel. The temporal overlap of activity in these two muscles for the production of /pi/ (the interval from the onset of activity in genioglossus to the offset of activity in orbicularis oris) was unaffected by variations in speaking rate ($p > .2$), or syllable stress ($p > .2$). Similarly, the temporal overlap of activity in these two muscles for the production of /ip/ (the interval from the onset of activity in orbicularis oris to the offset of genioglossus activity) was unaffected by changes in speaking rate and syllable stress ($p > .2$).

b. Inferior head of lateral pterygoid and orbicularis oris, and inferior head of lateral pterygoid and medial pterygoid, for production of /pa/ and /ap/.

For production of the syllable /pa/, there was no significant effect of speaking rate ($p > .2$) on the interval from lateral pterygoid (inferior head) onset of activity to orbicularis oris offset. However, syllable stress did affect the overlap of activity in orbicularis oris and lateral pterygoid inferior ($p < .05$) such that stressed syllables showed longer durations of overlap than unstressed syllables.

The duration of the interval from onset of activity in lateral pterygoid (inferior head) to the offset of activity in medial pterygoid was examined for production of the syllable /pa/. The duration of this interval was not affected by speaking rate ($p > .2$) or syllable stress ($p > .2$).

For production of /ap/, the temporal overlap of orbicularis oris and lateral pterygoid (inferior head) and the temporal overlap of medial and lateral pterygoid were unaffected by changes in speaking rate ($p > .2$) or syllable stress ($ps > .2$).

75

c. Orbicularis oris and anterior belly of digastric, and medial pterygoid and
   anterior belly of digastric, for production of /pa/ and /ap/.

The temporal overlap of activity in orbicularis oris and anterior belly
of digastric for production of /pa/ was unaffected by changes in speaking rate
(p >.2) or syllable stress (p >.2). For production of the syllable /ap/,
however, the temporal overlap of orbicularis oris and digastric was affected
by speaking rate (p <.05) such that the duration of overlap was longer for
fast than slow syllables. Changes in syllable stress had no systematic effect
on the temporal overlap of activity in these muscles (p >.2).

The duration of the interval from the onset of activity in anterior belly
of digastric to the offset of activity in medial pterygoid for production of
the syllable /pa/ was not affected by changes in syllable stress (p >.2), but
did change with variations in speaking rate (p <.01); this interval was longer
for syllables spoken slowly than for syllables spoken quickly. For production
of the syllable /ap/, the interval from onset of activity in medial pterygoid
to offset of activity in anterior belly of digastric was not significantly
affected by changes in speaking rate or syllable stress (ps >.2).

Most of the above comparisons gave the same results as reported by Tuller
and Harris (1980) and Tuller et al. (1981). The temporal overlap of activity
in muscles specific to only the vowel or only the consonant of CV and VC
syllables remained relatively stable over changes in speaking rate or syllable
stress. However, two comparisons resulted in variations in the duration of
overlapping activity as a function of changes in speaking rate. The first
showed a longer interval from orbicularis oris onset to anterior belly of
digastric offset in /ap/ spoken quickly than in /ap/ spoken slowly. The
second comparison showed the opposite direction of change in the temporal
overlap of two muscles' activity; the interval from the onset of activity in
anterior belly of digastric to the offset of activity in medial pterygoid was
longer for /pa/ spoken slowly than for /pa/ spoken quickly. One comparison
showed changes in duration of temporal overlap with changes in syllable
stress. The interval from lateral pterygoid (inferior head) onset to orbicu-
laris oris offset was longer in stressed /pa/ than unstressed /pa/. These
last two effects are in the direction opposite to that predicted by models of
speech production that posit invariant segmental articulations that show
increasing temporal overlap with decreasing syllable stress or increasing
speaking te (Kozhevnikov & Chistovich, 1965; Lindblom, 1963; Shaffer, 1976).

The durations of overlapping muscle activity that could be determined for
each subject and for each syllable type, pooled across rate and stress
conditions, are presented in Table 3. Each pair of values represents the
smallest and the largest measure of the relevant temporal interval.
Examination of Table 3 reveals that the range of values determined for each
subject generally did not exceed the integration time constant of 35 msec.
However, the range of temporal overlap of medial pterygoid and anterior belly
of digastric was 70 msec for BT, PS, and VR. For PS, the range of temporal
overlap of orbicularis oris and anterior belly of digastric was 60 msec.
Thus, although the variability in timing of muscle activity in CV or VC pairs
is relatively small compared with the changes in duration and magnitude of
activity in individual muscles, it may not be small enough to conclude that
the temporal overlap of activity remains fixed over metrical variations in
speaking rate and syllable stress.

76

Table 3

Measured temporal overlaps of activity in the muscles indicated, for
each subject and for each syllable type. Pairs of values represent
the shortest and the longest measure (in msec) of the indicated interval.

<div align="center">Subject</div>

		BT*	JT*	PS*	VR	GC
OO & GG	/pi/	125-145	120-130	100-115	-------	120-140
	/ip/	95-130	45- 65	70- 85	-------	135-140
OO & ABD	/pa/	65- 85	-------	20- 65	65- 75	60- 65
	/ap/	60- 95	-------	5- 65	30- 40	50- 60
OO & LPI	/pa/	35- 60	35- 50	20- 30	25- 45	40- 50
	/ap/	35- 45	30- 45	15- 40	25- 35	35- 45
MP & ABD	/pa/	5- 75	-------	10- 80	45-110	10- 70
	/ap/	15- 45	-------	25- 60	25- 30	50- 60
MP & LPI	/pa/	15- 45	-------	10- 40	10- 30	30- 45
	/ap/	10- 40	-------	25- 45	15- 30	25- 55

*Asterisks denote those subjects who produced the utterances at two rates of
speech. Empty cells denote no adequate recording of activity from the
subject and muscle indicated.

C. Intersegmental timing over three phonetic segments

In this section we examine whether the timing of intersegmental events
remains constant relative to the changing duration of some longer period of
articulatory activity. The relative timing of articulator activity is ana-
lyzed in terms of the phase relationships among muscle actions. This analysis

requires demarcation of some period of articulatory activity and the latency of occurrence of activity for an articulatory event within the defined period (cf. von Holst, 1973; Stein, 1971, 1976, among others). To this end, several periods of articulatory activity were demarcated, defined as the time between two successive occurrences of some electromyographic event in one segment type. One such period was the time between onsets of muscle activity (the EMG event) underlying the production of two occurrences of a consonant (one segment type). For example, for the first CVC in the utterance /pipap/, this period could be the time between the onset of orbicularis oris activity for the initial /p/ to the onset of medial pterygoid activity for the medial /p/.

Within each defined period, the latency of the same sort of electromyographic event was determined for a different segment type. The latency was defined as the time between the occurrence of the EMG event in one segment type (the onset of the articulatory period) and the next occurrence of the same sort of EMG event in a different segment type. For example, in the first CVC of the utterance /pipap/, the time from the onset of activity in orbicularis oris for the initial /p/ to the onset of activity in genioglossus for production of the vowel /i/ was defined as the latency of the event within the articulatory period.

Nine "periods" and corresponding events within each period were defined in this way and are described below for utterances of the form $C_1V_1C_2V_2C_3$. Each of the nine describes the timing of some articulatory event relative to a defined period.

1. The period from the onset of muscle activity for C_1 to the onset of muscle activity for C_2; the latency from the onset of activity for C_1 to the onset of activity for V_1. This examines whether the onset of V_1 activity occurs at a constant time relative to the onsets of the flanking consonants.

2. The period from the time of peak amplitude of muscle activity for C_1 to the peak amplitude of muscle activity for C_2; the latency from the peak amplitude of activity for C_1 to the peak amplitude of activity for V_1. This examines whether the peak amplitude of V_1 activity occurs at a constant time relative to the time of peak amplitude of the flanking consonants.

3. The period from the offset of muscle activity for C_1 to the offset of muscle activity for C_2; the latency from the offset of activity for C_1 to the offset of activity for V_1. This examines whether the offset of activity for V_1 occurs at a constant time relative to the offsets of the flanking consonants.

4. The period from the onset of muscle activity for V_1 to the onset of muscle activity for V_2; the latency from the onset of activity for V_1 to the onset of activity for C_2. This measure examines whether the onset of C_2 activity occurs at some constant time relative to the onset of activity for the flanking vowels.

78

5. The period from the peak amplitude of muscle activity for V_1 to the peak amplitude of muscle activity for V_2; the latency from the peak amplitude of activity for V_1 to the peak amplitude of activity for C_2. This examines whether the peak amplitude of activity for C_2 occurs at a constant time relative to the time of peak amplitude in the flanking vowels.

6. The period from the offset of muscle activity for V_1 to the offset of muscle activity for V_2; the latency from the offset of activity for V_1 to the offset of activity for C_2. This measure examines whether the offset of activity for C_2 occurs at a constant time relative to the offset of activity for the flanking vowels.

7. The period from the onset of muscle activity for C_2 to the onset of muscle activity for C_3; the latency from the onset of activity for C_2 to the onset of activity for V_2. This examines whether the onset of muscle activity for V_2 occurs at a constant time relative to the onsets of C_2 and C_3.

8. The period from the peak amplitude of muscle activity for C_2 to the peak amplitude of muscle activity for C_3; the latency from the peak amplitude of activity for C_2 to the peak amplitude of activity for V_2. This measures whether the peak amplitude of activity in V_2 occurs at a constant time relative to the time of peak amplitude of activity in the flanking consonants.

9. The period from the offset of muscle activity for C_2 to the offset of muscle activity for C_3; the latency from the offset of activity for C_2 to the offset of activity for V_2. This examines whether the offset of activity in V_2 occurs at a constant time relative to the offsets of activity in C_2 and C_3.

These nine pairs of what, for ease of communication, will be called "periods" and "latencies" were obtained for all possible muscle combinations and for all utterances within each of the four speaking conditions (i.e., slow rate with the first syllable stressed, slow rate with the second syllable stressed, fast rate with the first syllable stressed, and fast rate with the second syllable stressed).[1] One analysis, then, would consist of four coordinate pairs for a given speaker and muscle combination, each pair corresponding to the period and latency measures for an utterance under one speaking condition. Pearson's product-moment correlations were calculated on each set of four coordinate pairs. A high linear correlation would indicate that the latency of the measured event relative to the measured period remained fairly constant over variations in speaking rate and syllable stress.

Figures 3, 4, and 5 show the distributions of correlations for the different measures. Figures 3a, 3b, and 3c correspond to the definitions of period and latency described above as 1, 2, and 3, respectively. Figures 4a, 4b, and 4c correspond to definitions 4, 5, and 6, respectively, and Figures 5a, 5b, and 5c correspond to definitions 7, 8, and 9, respectively. All muscle combinations and utterances are displayed together. One measure shows a higher correlation, and less variability, than all other measures.
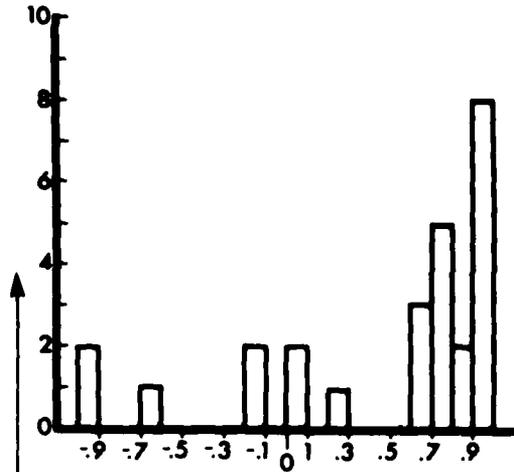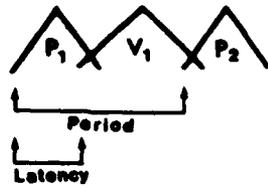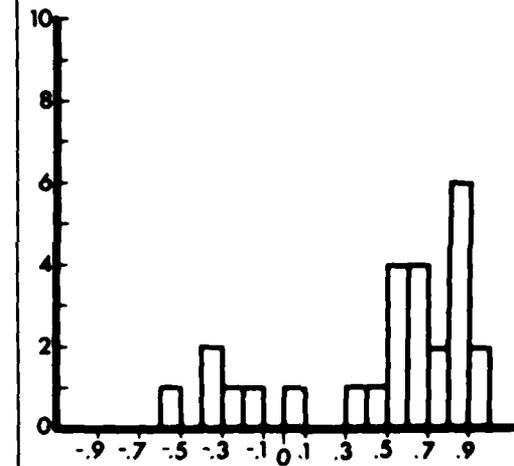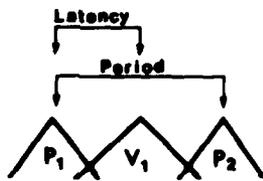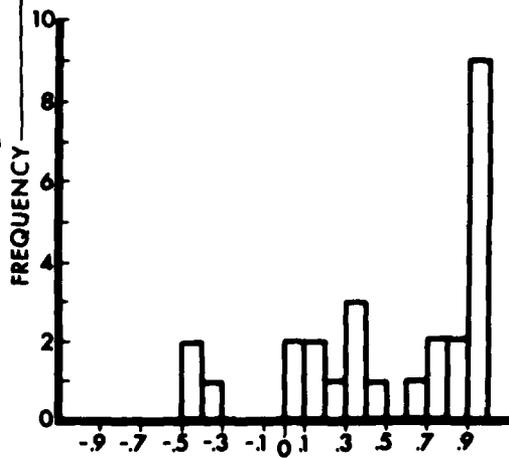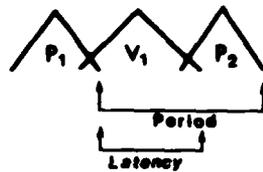
Figure 3. Distribution of correlations for periods and latencies as indicated, for $P_1V_1P_2$ utterances.
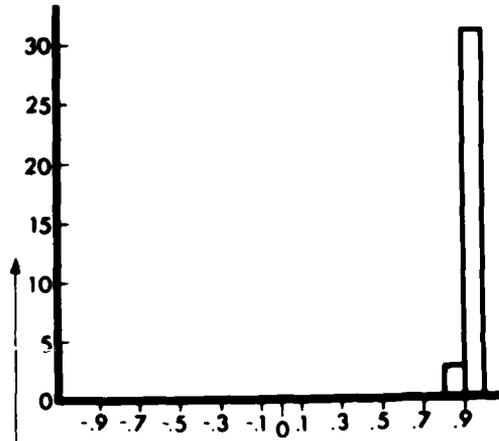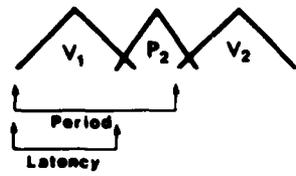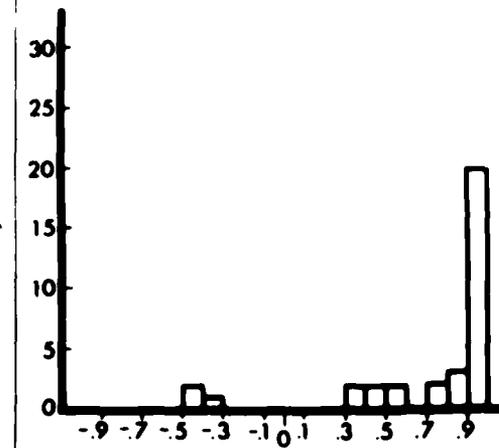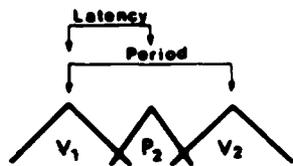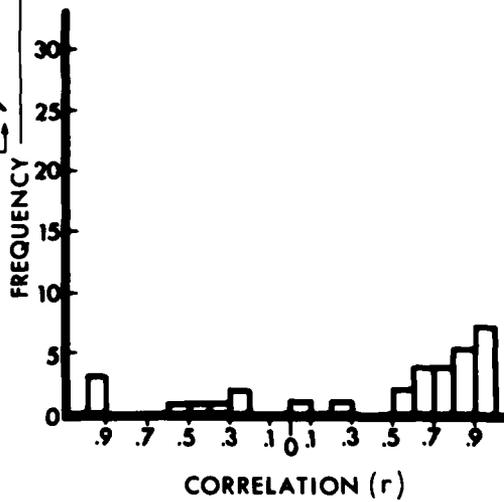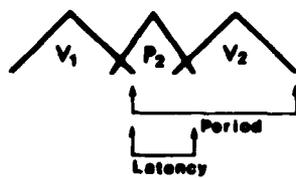
Figure 4. Distribution of correlations for periods and latencies as indicated, for $V_1P_2V_2$ utterances.
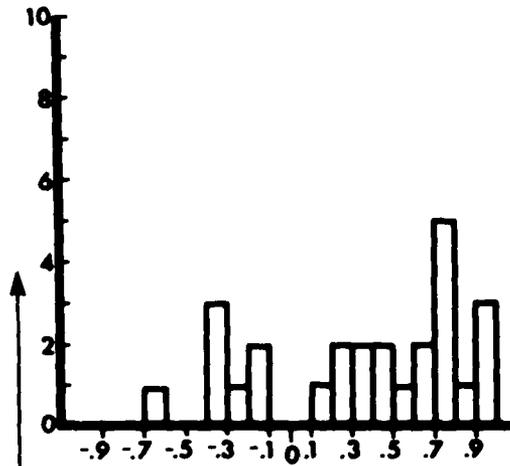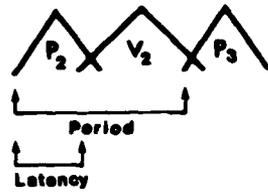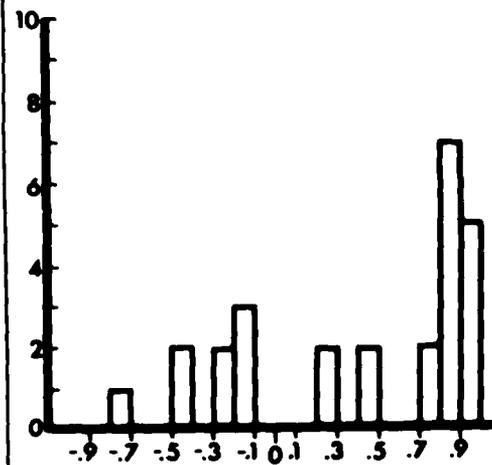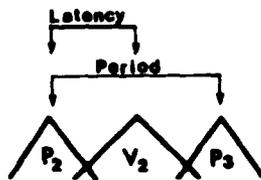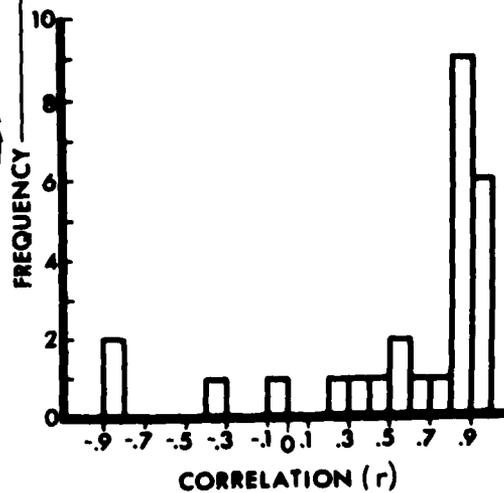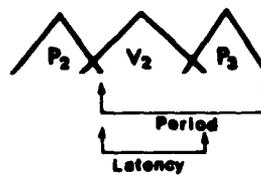
Figure 5. Distribution of correlations for periods and latencies as indicated, for $P_2V_2P_3$ utterances.

82

Specifically, a high linear correlation (ranging from $r=.87$ to $r=.99$) obtains between the period from the onset of muscle activity for V_1 to the onset of muscle activity for V_2, and the latency from the onset of activity for V_1 to the onset of activity for C_2 (Fig. 4a). All other definitions of period and latency produced wider distributions whose shapes differed significantly from the curve in Figure 4a, for correlations greater than .8 (Kolmogorov-Smirnov, $ps < .01$, one-tailed).

It should be underscored that although the high correlation is obtained over the various possible muscle combinations and utterances, this is not to say that the actual ratio of the latency divided by the period remains constant regardless of the specific muscles or utterances involved. The same combination of muscles specific to production of two different utterances will likely show two different ratios of latency to period for the two utterance types. Similarly, two different combinations of muscles will often show different ratios of latency to period for production of the same utterance type. Consider the period and latency measures with consistently high linear correlations (period defined as V_1 onset to V_2 onset, latency defined as V_1 onset to C_2 onset). For PS, for example, when the appropriate intervals of genioglossus, orbicularis oris, and lateral pterygoid (inferior head) activity were determined for the VCV /ipa/, the mean ratio of latency divided by period for the four stress-rate conditions (stressed slow, unstressed slow, stressed fast, and unstressed fast) was .55 (sd=.05). For production of /api/, these same three muscles showed a mean ratio of latency divided by period of .77 (sd=.04). When a different muscle trio (genioglossus, orbicularis oris, and anterior belly of digastric) was examined in relation to the same VCV utterances /ipa/ and /api/, spoken by the same subject, the mean ratios of latency to period were .59 (sd=.05) and .87 (sd=.04), respectively.

To summarize, in a VCV utterance the timing of onsets of activity for successive vowel and consonant segments appeared to be temporally constrained in relation to a longer period of articulation than previously examined, namely, the period between onsets of activity for successive vowels. Thus, relative timing of muscle activity remained fixed over changes in speaking rate and syllable stress and over concomitant changes in duration and peak amplitude of activity in the individual muscles (see also Kelso et al., 1981, Figure 1).

GENERAL DISCUSSION

The results of the present experiment suggest that an appropriate description of temporal aspects of articulation is relative to a longer articulatory period than previously examined in the speech production literature. For eight of the nine experimentally-defined articulatory periods and latencies, linear correlations of period and latency produced a very wide distribution of correlations. In contrast, for $p_1V_1p_2V_2p_3$ utterances, when the articulatory period was defined as the interval from V_1 onset to V_2 onset, and the latency defined as the interval from V_1 onset to p_2 onset, the correlations of latency and period produced a distribution that was extremely narrow. Moreover, the correlations themselves were extremely high. In other words, the timing of consonant articulation remained fixed relative to the

surrounding vowel articulations. This suggests that the preservation of temporal relationships over metrical change may characterize speech motor activity and that the appropriate definitions of period and latency are understandable within a traditional linguistic framework.

Let us consider these two implications in a little more detail. It is important to note that when speaking rate or syllable stress vary, the major durational changes occur in vocalic portions of the utterance (e.g., Gaitenby, 1965; Kozhevnikov & Chistovich, 1965; Lehiste, 1970). Consider the articulatory period as the interval between successive consonant onsets and the latency within the period as the interval from the onset of the first consonant of the period to the vowel onset (Figs. 3a and 5a). When syllable stress or speaking rate vary, the bulk of the durational change will affect only the measured period, not the latency. In other words, in the measure "latency divided by period," changes in stress and rate will affect mainly the denominator, so that the ratio cannot be maintained. Similarly, in CVC utterances, when the time between the peak of activity for successive consonants is defined as the period, and the interval from the first consonant peak to the vowel peak is defined as the latency (Figs. 3b and 5b), changes in stress and rate affect the measure of period proportionally more than the measure of latency.

But consider when the period is defined as the interval from the onset of muscle activity for V_1 to the onset of activity for V_2, and the latency as the interval from V_1 onset to the onset of muscle activity for the medial consonant (Fig. 4a). The major durational changes that occur as stress and speaking rate vary will affect both measures, leaving at least the possibility that their ratio remains fixed. Thus, the common formulation of "phase position" is appropriate for speech production when the period and latency are demarcated within the muscle events by reference to linguistic segments.

One strong indication that an appropriate description of speech motor control is in terms of relative timing constraints is the congruence of the present data with recent descriptions of speech perception. For example, Summerfield (1975a, 1975b) found that the temporal boundary of voice-onset-time (VOT) between perception of voiced and voiceless stop consonants is dependent on the speaking rate of the carrier phrase. Similarly, Port (1978, 1979) examined the influence of speech rate on the perception of the voicing distinction in medial stop consonants, cued in part by the duration of silence preceding the consonant release. The duration of silence necessary to specify that the medial stop consonant was voiceless, and not voiced, decreased as speaking rate increased. These examples suggest that the relative timing of acoustic events may characterize the perception of voicing distinctions. That is, the category distinctions are perceived relative to total speech time (interpreted as speaking rate).

Other evidence for the importance of relative timing in speech perception is available in Miller and Grosjean's (in press) replication of Port's results, Miller and Liberman's (1979) demonstration showing evidence of a rate-dependent phonetic boundary between stops and semivowels, and Pickett and Decker's (1960) result showing similar rate effects on the perception of geminate consonants. In addition, long and short vowel pairs are distinguished, at least in part, by vowel duration in relation to perceived rate of

84

speech and not by absolute vowel duration (Rakerd et al., 1980). These results suggest that the timing of some event contributing to a phonetic distinction is not constrained within absolute temporal boundaries, but is perceived in relation to some longer period specifying speech rate.

Just as relative timing has significance for speech perception so is it important in the control and coordination of nonspeech skills. The following discussion is an attempt to highlight the similarities between speech motor control and control of some of these activities with an eye to how changes in rate or magnitude of movement are accomplished. It will become obvious that the data are analogous in many ways to the data presented here. It is suggested that types of analyses common to investigations of these other motor skills might profitably guide studies of speech production (see also Fowler, Rubin, Remez, & Turvey, 1980; Kelso, in press; Kelso et al., 1981; Moll, Zimmerman, & Smith, 1977).

The sort of timing relationships evident in the present experiment are illustrated in investigations of freely locomoting animals, such as humans (Herman, Wirta, Bampton, & Finley, 1976), cats (Engberg & Lundberg, 1969), cockroaches (Delcomyn, 1971; Pearson & Iles, 1973), lobsters (MacMillan, 1975) and turtles (Stein, 1978). When these animals increase the speed of their locomotion, the duration of the "step cycle" in each limb may decrease markedly. However, the phase relationships among the limbs (whether measured electromyographically or kinematically) are constant over a wide range of stepping frequencies (see Grillner, 1975; Shik & Orlovskii, 1976).

Timing relationships within a limb may also be preserved over speed. For example, MacMillan (1975) reported that in the lobster, both agonists and antagonists maintained a constant phase position relative to the limbs' cycle duration over a wide range of stepping frequencies. When a load was attached to the limb, the depressor and elevator muscles (the primary determinants of the power and return strokes, respectively) preserved their phase positions within the step cycle, although the duration of the elevator activity shortened considerably. This very brief discussion should suffice to convey the more general implication--that constant phase relationships among variables characterize locomotion in many different species.

One possible objection to drawing parallels between the control of locomotion and the control of speech is that locomotion is an activity easily described as a fundamental pattern of events that recurs over time. The observed pattern is not strictly stereotypic, however, because it is modifiable in response to environmental changes, such as bumps in the terrain. The question remains whether a style of coordination in which temporal relationships are preserved over changes in individual components holds for nonspeech activities that are less obviously rhythmic and whose fundamental pattern is not immediately apparent. Examinations of kinematic aspects of one such activity, handwriting, reveals this style of coordination.

When individuals were asked to vary their writing speed without varying movement amplitude (Viviani & Terzuolo, 1980), the relative timing of certain movements did not change with speed. Specifically, the tangential velocity records resulting from different writing speeds revealed that overall duration changed markedly across speeds. But when the individual velocity records were

85

adjusted to approximate the average duration, the resulting pattern was highly
invariant. In other words, major features of writing a given word occurred at
a fixed time relative to the total duration taken to write the word (see also
Terzuolo & Viviani, 1979, for a similar analysis of typewriting). The same
timing relationships are preserved over changes in magnitude of movements,
over different muscle groups, and over different environmental
(e.g., frictional) conditions (cf. Denier van der Gon & Thuring, 1965; Holler-
bach, 1980; Wing, 1978).

Thus, for some animals, the timing of activity in individual limb muscles
during locomotion remains fixed relative to the step cycle, and in handwrit-
ing, the timing of individual strokes remains fixed relative to the period for
writing the entire word. The experiment described here suggests that speech
production is organized in a manner similar to these other motor activities,
at least at the electromyographic level. A temporal patterning of components,
in this case muscle activities, was preserved independent of changes in the
duration and absolute magnitude of activity in the individual muscles.

It should be emphasized that this result does not entail the notion that
speech production is organized as continuous vowel-to-vowel production with
consonants superimposed on this basic organization (see Fowler, 1977; Öhman,
1966; Perkell, 1969). In locomotion, the timing of extensor activity may
remain fixed relative to the time between successive flexions (see Engberg &
Lundberg, 1969), yet the organization of locomotion is not described as
continuous flexion-to-flexion with extension superimposed on this basic cycle.

In summary, the results presented here suggest a view of interarticulator
relationships that is compatible with the style of temporal organization in
other motor activities. The temporal organization proposed is one that
maintains relative timing for the preservation of correct articulation.
Although not highlighted previously in theories of speech timing, the exis-
tence of relative timing constraints in speech production should not be
surprising, given their salience in speech perception. Rather, the observed
temporal constraints are compatible with and, as suggested earlier, may
rationalize several findings in perception and linguistics.

REFERENCE NOTE

1. Gross, B. D., & Lipke, D. A technique for percutaneous lateral pterygoid
   electromyography. Manuscript submitted for publication, 1979.

REFERENCES

Ahlgren, J. Mechanism of mastication. A quantitative cinematographic and
    electromyographic study of masticatory movements in children, with spe-
    cial reference to occlusion of the teeth. Acta Odontologica
    Scandinavica, 1966, 24, Suppl. 44.

Alfonso, P., & Baer, T., An electromyographic-cinefluorographic-acoustic study of dynamic vowel production. _Haskins Laboratories Status Report on Speech Research_, 1981, _SR-65_, this volume.

Allen, G. D., & Lubker, J. F. New paint-on electrodes for surface electromyography. _Journal of the Acoustical Society of America_, 1972, _52_, 124. (Abstract)

Carlsöo, S. Nervous coordination and mechanical function of the mandibular elevators. _Acta Odontologica Scandinavica_, 1952, _10_, Suppl. 11.

Carlsöo, S. An electromyographic study of the activity and an anatomical analysis of the mechanism of the lateral pterygoid muscle. _Acta Anatomica_ (Basel), 1956, _26_, 339-351.

Delattre, P. An acoustic and articulatory study of vowel reduction in four languages. _International Review of Applied Linguistics_, 1969, _7_, 295-325.

Delcomyn, F. The locomotion of the cockroach _Periplaneta Americana_. _Journal of Experimental Biology_, 1971, _54_, 443-452.

Denier van der Gon, J. J., & Thuring, J. Ph. The guiding of human writing movements. _Kybernetik_, 1965, _4_, 145-147.

Engberg, I., & Lundberg, A. An electromyographic analysis of muscular activity in the hindlimb of the cat during unrestrained locomotion. _Acta Physiologica Scandia_, 1969, _75_, 614-630.

Fowler, C. _Timing control in speech production._ Bloomington, Ind.: Indiana University Linguistics Club, 1977.

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

Fromkin, V. A. Neuromuscular specifications of linguistic units. _Language and Speech_, 1966, _9_, 170-199.

Fry, D. B. Duration and intensity as physical correlates of linguistic stress. _Journal of the Acoustical Society of America_, 1955, _27_, 765-768.

Fry, D. B. Experiments in the perception of stress. _Language and Speech_, 1958, _1_, 126-152.

Gaitenby, J. H. The elastic word. _Haskins Laboratories Status Report on Speech Research_, 1965, _SR-2_, 3.1-3.12.

Gay, T. Jaw movements during speech: A cinefluorographic investigation. _Haskins Laboratories Status Report on Speech Research_, 1974, _SR 39/40_, 219-229.

Gay, T., & Hirose, H. Effect of speaking rate on labial consonant production: A combined electromyographic/high speed motion picture study. _Phonetica_, 1973, _27_, 44-56.

Gay, T., & Ushijima, T. Effect of speaking rate on stop consonant-vowel articulation. In _Proceedings of the Speech Communication Seminar, Stockholm, 1974._ Uppsala: Almqvist and Wiksell, 1974.

Gay, T., Ushijima, T., Hirose, H., & Cooper, F. S. Effect of speaking rate on labial consonant-vowel articulation. _Journal of Phonetics_, 1974, _2_, 47-63.

Grillner, S. Locomotion in vertebrates. _Physiological Reviews_, 1975, _55_, 247-304.

Harris, K. S. Vowel stress and articulatory reorganization. _Haskins Laboratories Status Report on Speech Research_, 1971, _SR-28_, 167-178.

Harris, K. S. Stress and syllable duration change. _Haskins Laboratories Status Report on Speech Research_, 1973, _SR-35/36_, 31-38.

Harris, K. S., Gay, T., Sholes, G. N., & Lieberman, P. Some stress effects on electromyographic measures of consonant articulations. Haskins Laboratories Status Report on Speech Research, 1968, SR-13/14, 137-151.

Harris, K. S., Lysaught, G. F., & Schvey, M. M. Some aspects of the production of oral and nasal labial stops. Language and Speech, 1965, 8, 135-147.

Herman, R., Wirta, R., Bampton, S., & Finley, R. Human solutions for locomotion: Single limb analysis. In R. M. Herman, S. Grillner, P. S. G. Stein, & D. Stuart (Eds.), Neural control of locomotion. New York: Plenum Press, 1976.

Hirose, H. Electromyography of the articulatory muscles: Current instrumentation and techniques. Haskins Laboratories Status Report on Speech Research, 1971, SR-25/26, 73-86.

Hollerbach, J. M. An oscillation theory of handwriting. Cambridge, Mass.: MIT Artificial Intelligence Laboratory, 1980.

Holst, E. von. The behavioral physiology of animal and man: The collected papers of Erich von Holst (Vol. 1) (R. Martin, trans.). London: Methuen and Co., Ltd, 1973.

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

Kelso, J. A. S., Tuller, B., & Harris, K. S. A 'dynamic pattern' perspective on the control and coordination of movement. To appear in P. MacNeilage (Ed.), Speech production. Also Haskins Laboratories Status Report on Speech Research, 1981, SR-65, this volume.

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

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

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

Kewley-Port, D. An experimental evaluation of the EMG data processing system: Time constant choice for digital integration. Haskins Laboratories Status Report on Speech Research, 1974, SR-37/38, 65-72.

Kozhevnikov, V., & Chistovich, L. Speech: Articulation and perception. Moscow-Leningrad, 1965 (English translation: J.P.R.S., Washington, D.C. No. JPRS 30543).

Lehiste, I. Suprasegmentals. Cambridge, Mass.: M.I.T. Press, 1970.

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

Löfqvist, A., & Yoshioka, H. Laryngeal activity in Icelandic obstruent production. Haskins Laboratories Status Report on Speech Research, 1980, SR-63/64, 275-292. (Also Nordic Journal of Linguistics, in press.)

Löfqvist, A., & Yoshioka, H. Interarticulator programming in obstruent production. Haskins Laboratories Status Report on Speech Research, 1981, SR-65, this volume.

Lubker, J., McAllister, R., & Lindblom, B. On the notion of inter-articulator programming. Journal of Phonetics, 1977, 5, 213-226.

MacMillan, D. L. A physiological analysis of walking in the American lobster. Philosophical Transactions of the Royal Society of London, 1975, 270B, 1-59.

McAllister, R., Lubker, J., & Carlsson, J. An EMG study of some characteristics of the Swedish rounded vowels. Journal of Phonetics, 1974, 2, 267-278.

Miller, J. L., & Grosjean, F. How the components of speaking rate influence perception of phonetic segments. Journal of Experimental Psychology: Human Perception and Performance, in press.

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

Moll, K. L., Zimmerman, G. H., & Smith, A. The study of speech production as a human neuromotor system. In M. Sawashima & F. S. Cooper (Eds.), Dynamic aspects of speech production. Tokyo: University of Tokyo Press, 1977, 107-127.

Møller, E. Action of the muscles of mastication. In Y. Kawamura (Ed.), Physiology of mastication. Basel: S. Karger, 1974.

Moyers, R. E. An electromyographic analysis of certain muscles involved in temporomandibular movement. American Journal of Orthodontics, 1950, 37, 481-515.

Öhman, S. E. G. Durations of formant transitions. Royal Institute of Technology, QPRS, 1965, 1, 10-14.

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

Öhman, S. E. G. Numerical model of coarticulation. Journal of the Acoustical Society of America, 1967, 41, 310-320.

Pearson G. K., & Iles, J. F. Nervous mechanisms underlying intersegmental coordination of leg movements during walking in the cockroach. Journal of Experimental Biology, 1973, 58, 725-744.

Perkell, J. S. Physiology of speech production: Results and implications of a quantitative cineradiographic study. Cambridge, Mass.: MIT Press, 1969.

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

Pickett, J. M., & Decker, L. R. Time factors in perception of a double consonant. Language and Speech, 1960, 3, 11-17.

Port, R. F. Effects of word-internal versus word-external tempo on the voicing boundary for medial stop closure. Journal of the Acoustical Society of America, 1978, 63, S20. (Abstract)

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

Rakerd, B., Verbrugge, R. V., & Shankweiler, D. P. Speaking rate, syllable stress, and vowel identity. Haskins Laboratories Status Report on Speech Research, 1980, SR-62, 149-160.

Raphael, L. J., & Bell-Berti, F. Tongue musculature and the feature of tension in English vowels. Phonetica, 1975, 32(1), 61-73.

Shaffer, L. H. Intention and performance. Psychological Review, 1976, 83, 375-393.

Shik, M. L., & Orlovskii, G. N. Neurophysiology of locomotor automatism. Physiological Reviews, 1976, 56, 465-501.

Siegal, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Smith, T. S. A phonetic study of the extrinsic tongue muscles. UCLA Working Papers in Phonetics, 1971, 18.

2 OF 3
AD A
099958

Stein, P. S. G. Intersegmental coordination of swimmeret motorneuron activity in crayfish. Journal of Neurophysiology, 1971, 34, 310-318.

Stein, P. S. G. Mechanisms of interlimb phase control. In R. M. Herman, S. Grillner, P. S. G. Stein, & D. Stuart (Eds.), Neural control of locomotion. New York: Plenum Press, 1976.

Stein, P. S. G. Motor systems, with special reference to the control of locomotion. Annual Review of Neuroscience, 1978, 1, 61-81.

Summerfield, A. Q. Cues, contexts and complications in the perception of voicing contrasts. In Speech Perception, 1975, 4, Department of Psychology, Queen's University, Belfast. (a)

Summerfield, A. Q. How a full account of segmental perception depends on prosody and vice versa. In A. Cohen & S. G. Nooteboom (Eds.), Structure and process in speech perception. New York: Springer-Verlag, 1975. (b)

Terzuolo, C. A., & Viviani, P. The central representation of learned motor patterns. In R. E. Talbott & D. R. Humphrey (Eds.), Posture and movement. New York: Raven Press, 1979.

Tiffany, W. R. Nonrandom sources of variation in vowel quality. Journal of the Acoustical Society of America, 1959, 58, 434-445.

Tuller, B. Coordination among articulators in speech. Unpublished doctoral dissertation. The University of Connecticut, 1980.

Tuller, B., & Harris, K. S. Temporal models of interarticulator programming. Journal of the Acoustical Society of America, 1980, 67(S), S64. (Abstract)

Tuller, B., Harris, K. S., & Gross, B. An electromyographic study of the jaw muscles during speech. Journal of Phonetics, in press.

Tuller, B., Harris, K. S., & Kelso, J. A. S. Articulatory motor events as a function of speaking rate and stress. Haskins Laboratories Status Report on Speech Research, 1981, SR-65, this volume.

Verbrugge, R. R., & Shankweiler, D. P. Prosodic information for vowel identity. Journal of the Acoustical Society of America, 1977, 61, S39. (Abstract)

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

Wing, A. M. Response timing in handwriting. In G. E. Stelmach (Ed.), Information processing in motor control and learning. New York: Academic Press, 1978.

FOOTNOTE

[1]Data from the two subjects who spoke the utterances at only one speaking rate were not considered; these subjects would have only two coordinate pairs per utterance, guaranteeing a linear correlation of 1.

INTERARTICULATOR PROGRAMMING IN OBSTRUENT PRODUCTION*

Anders Löfqvist+ and Hirohide Yoshioka++

Abstract. Most work on speech motor control has been devoted to the spatial and temporal coordination of articulatory movements for successive units, segments or syllables, in the speech chain. An intrasegmental temporal domain has generally been lacking in speech production models, but such a domain is necessary at least for certain classes of speech sounds, e.g., voiceless obstruents, clicks, ejectives. The present paper examines the nature of laryngeal-oral coordination in voiceless obstruent production in different languages using the combined techniques of electromyography, transillumination and fiberoptic filming of the larynx together with aerodynamic and palatographic records for information on supralaryngeal articulations. The results suggest that laryngeal articulatory movements are organized in one or more continuous opening and closing gestures that are precisely coordinated with supralaryngeal events according to the aerodynamic requirements of speech production.

INTRODUCTION

The problem of speech motor control has usually been seen as one of accommodating and coordinating in space and time the articulatory demands for successive segments in the speech chain, and studies of coarticulation have generally been directed towards this problem (Daniloff & Hammarberg, 1973; Kent & Minifie, 1977). Since the articulatory units have usually been taken to be more or less identical with the units of linguistic analysis, the temporal resolution necessary in most speech production models has been of the order of magnitude of the segment. A segmental approach has been further encouraged by the fact that the feature representation of segments at a systematic phonetic level, with few exceptions, contains no intrasegmental temporal domain, and such feature representations have often been taken as the input to the speech production apparatus. For some classes of speech sounds such as voiceless obstruents, clicks, ejectives, and implosives, it is, however, necessary to posit a temporal domain for articulatory movements within one and the same linguistic and/or articulatory unit (cf. Lisker, 1974).

[HASKINS LABORATORIES: Status Report on Speech Research SR-65 (1981)]

Voiceless obstruent production requires control and coordination of several articulatory systems. The tongue, the lips and the jaw are engaged in the formation of the constriction or occlusion; the soft palate is elevated in order to close the entrance to the nasal cavity and prevent air from escaping that way; the vocal folds are abducted in order to prevent glottal vibrations and, by reducing laryngeal resistance to air flow, contribute to the high air flow and/or buildup of oral air pressure.

Voiceless obstruent production thus involves simultaneous activity at both laryngeal and supralaryngeal levels, and the oral and laryngeal articulations have to be temporally coordinated. The aim of the present paper is to examine the nature of laryngeal-oral coordination in voiceless obstruent production.

METHOD

Laryngeal articulations were monitored simultaneously by fiberoptic filming and transillumination. Filming was made though a flexible fiberscope at a film speed of 60 frames/second. The film was analyzed frame by frame, and the distance between the vocal processes measured as an index of glottal opening. The light passing through the glottis was also sensed by a phototransistor placed on the neck just below the cricoid cartilage. Recordings from 10-15 repetitions of each test utterance were computer averaged. Unless stated otherwise, the average transillumination signals have been integrated over 5 milliseconds. In order to obtain the speed of glottal opening change, the first derivative of glottal displacement was calculated by successive subtractions at 5 millisecond increments in the average transillumination records. Neither transillumination nor fiberoptic films can be calibrated at present. The scales thus differ between experimental runs, and numerical comparisons of glottal opening and velocity should only be made for a given subject within one and the same recording session.

The movement records were supplemented by EMG recordings from the posterior cricoarytenoid and the interarytenoid muscles, in order to determine if observed laryngeal movements were caused by muscular and/or nonmuscular, e.g., aerodynamic forces.

Implosion and release of voiceless stops were determined from records of oral egressive air flow and oral air pressure. Such records are, however, not reliable indicators of beginning and end of oral constriction in voiceless fricatives. Therefore, additional recordings were made using a custom-made artificial palate with implanted electrodes (cf. Kiritani, Kakita, & Shibata, 1977). Six electrodes at the alveolar ridge were connected in parallel; a battery and a resistor were connected in series between the six electrodes and a reference electrode. Onset and offset of tongue-palate contact could then be identified as changes in voltage across the resistor. A more detailed description of the experimental procedure can be found in Yoshioka, Löfqvist, and Hirose (1979), Löfqvist and Yoshioka (1980), and Löfqvist (in press).

The fiberoptic filming was made to assess the validity of the transillumination technique. Temporal patterns of glottal opening variations obtained by fiberoptic filming and by transillumination showed a high correlation and

proved to be practically identical (Yoshioka et al., 1979; Löfqvist & Yoshioka, 1980, in press). We will therefore mainly discuss the information obtained by transillumination. The electromyographic records will not be dealt with apart from the general observation that laryngeal articulatory movements were accompanied by distinct activity patterns in the two muscles investigated, with the posterior cricoarytenoid activated for abduction and the interarytenoid activated for adduction, respectively.

RESULTS

In single voiceless obstruents, the laryngeal articulation usually has the form of a single "ballistic" opening and closing gesture, cf. Figure 1. The timing of this gesture in relation to supraglottal articulatory events is tightly controlled, and apparently varies for fricatives and aspirated stops, Figure 1. Peak glottal opening occurs earlier during the fricative than during the stop. The abduction also appears to occur at higher velocity, and peak opening seems larger for the fricative.

In clusters of voiceless obstruents, one or more continuous glottal opening and closing gestures occur, as shown in Figure 2. In general, separate opening gestures are associated with fricatives and with aspirated stops. In a cluster of fricative + unaspirated stop, only one glottal gesture is found with peak glottal opening during the fricative. When several glottal gestures are found in a cluster, their relationship to oral articulations is similar to that found in single obstruents.

Variations in the relative timing of laryngeal and oral articulations are used to produce contrasts of aspiration in stop consonants. This is illustrated in Figure 3 with material from Icelandic, which has a three-way contrast of preaspirated, unaspirated, and postaspirated voiceless stops. The three stops in Figure 3 differ in at least two respects. First, the relative timing of glottal abduction/adduction and oral closure/release is different. For the unaspirated stop, glottal abduction starts at the implosion, and peak glottal opening, i.e., onset of glottal adduction, occurs close to the implosion. The postaspirated category has glottal abduction beginning at implosion and peak glottal opening at oral release. For the preaspirated stop, both glottal abduction and peak glottal opening precede oral closure.

A second difference in Figure 3 is that of glottal opening size. The present material suggests that postaspirated stops have larger glottal opening than their preaspirated and unaspirated cognates. Glottal opening is smaller for the preaspirated type, and very small for the unaspirated one. For the latter, the fiberoptic films showed a small, spindle-shaped opening in the membraneous portion of the glottis.

A closer view of interarticulator timing in Swedish voiceless stop production is given in Figure 4. This figure is based on measurements from repetitions of simple CVCVC nonsense words where the number of segments and the placement of stress were systematically varied. For aspirated stops, peak glottal opening is systematically delayed in relation to stop implosion as the duration of stop closure increases. Unaspirated stops in Swedish have longer closure duration, and peak glottal opening generally occurs closer to stop implosion for unaspirated than for aspirated stops.

93
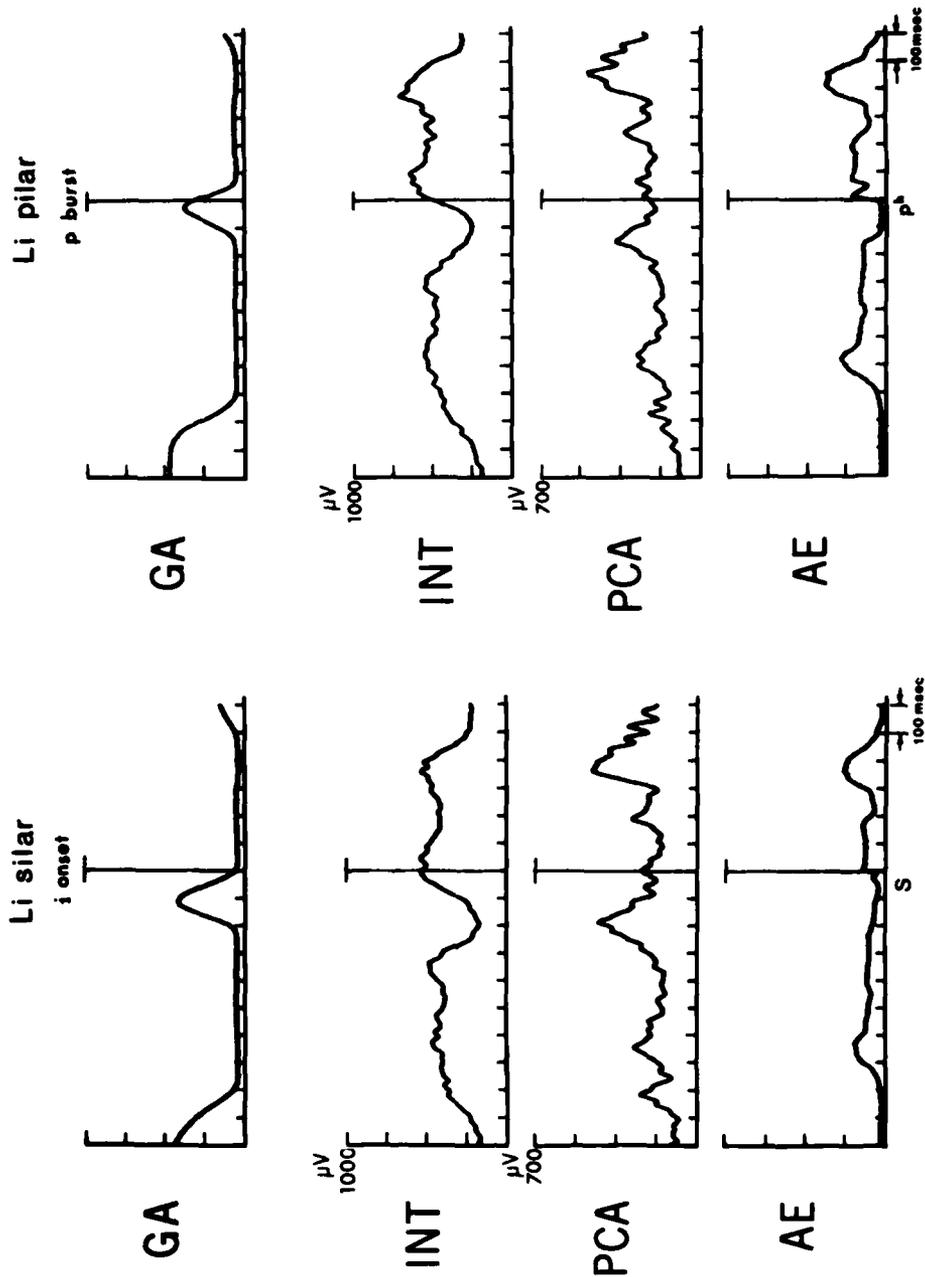
Figure 1. Average transillumination signal (GA), interarytenoid (INT) and posterior cricoarytenoid (PCA) EMG records, and audio envelope (AE) of Swedish utterances containing a voiceless fricative (left) and a voiceless postaspirated stop (right).

94
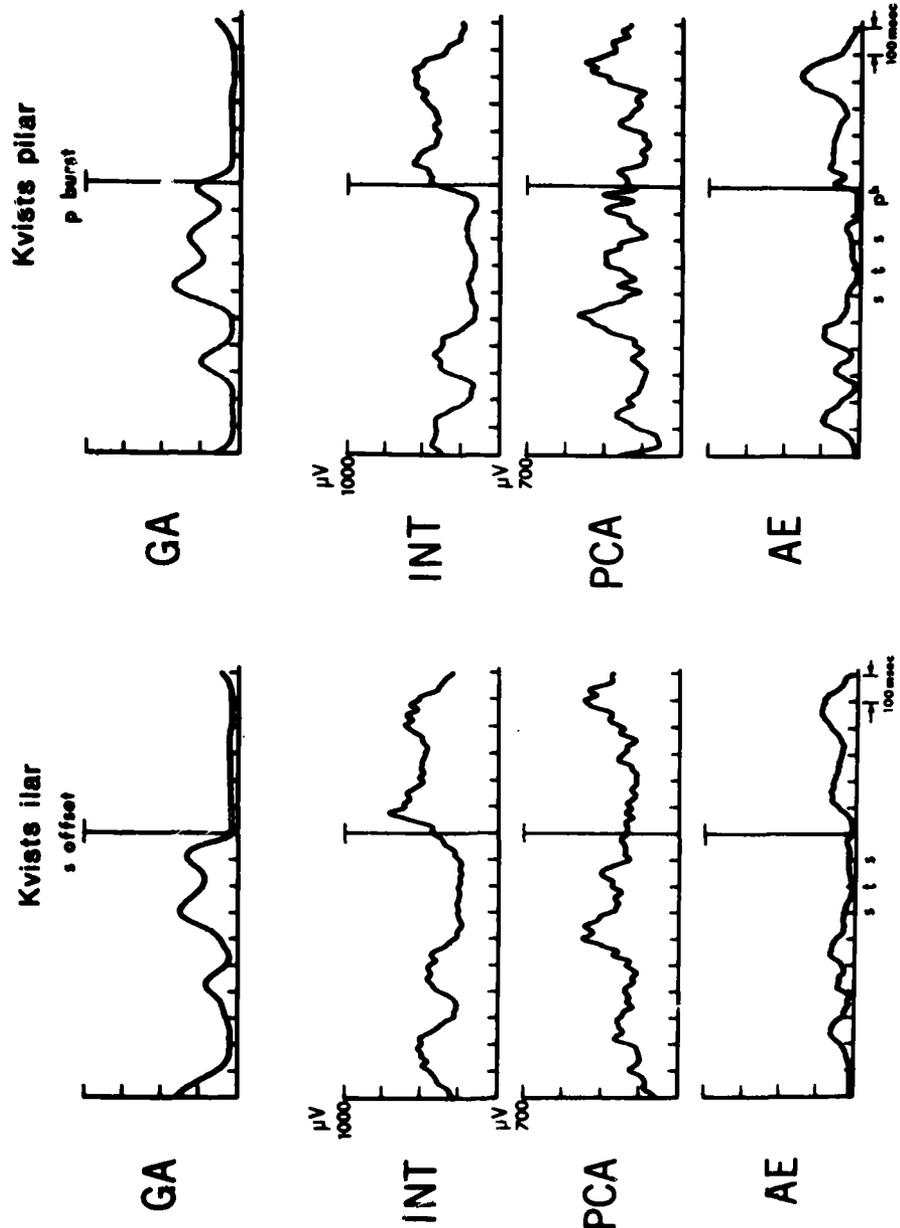
Figure 2. Glottal area, EMG and audio signals of two Swedish utterances containing different voiceless obstruent clusters.

95
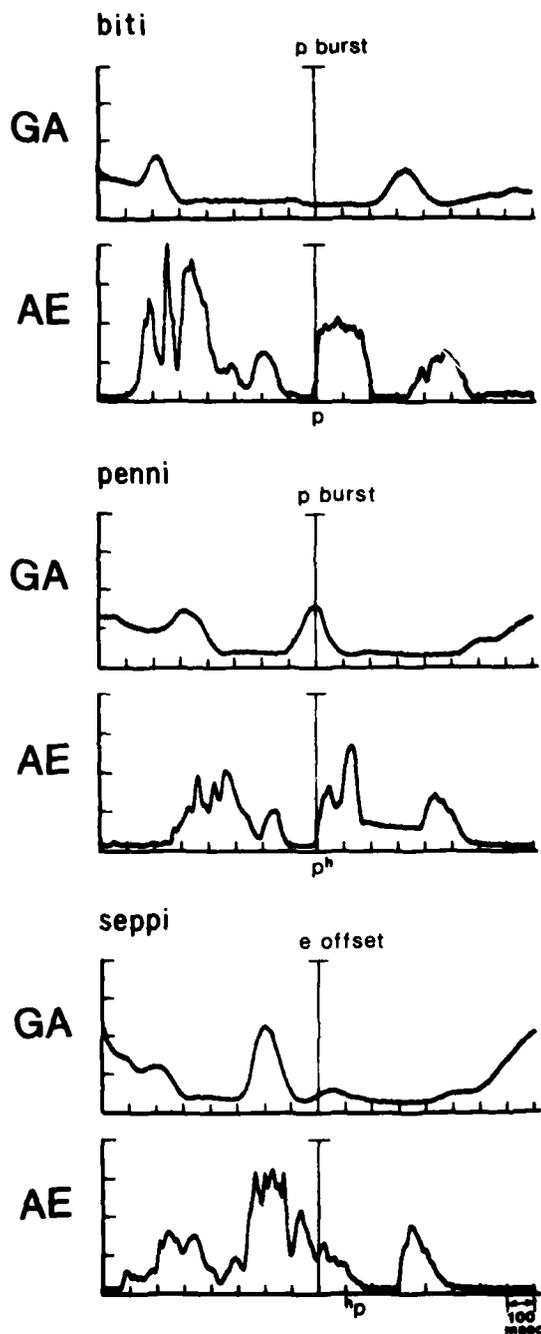
Figure 3. Glottal area and audio signals of Icelandic utterances containing unaspirated (top), postaspirated (middle), and preaspirated (bottom) voiceless stops.

96
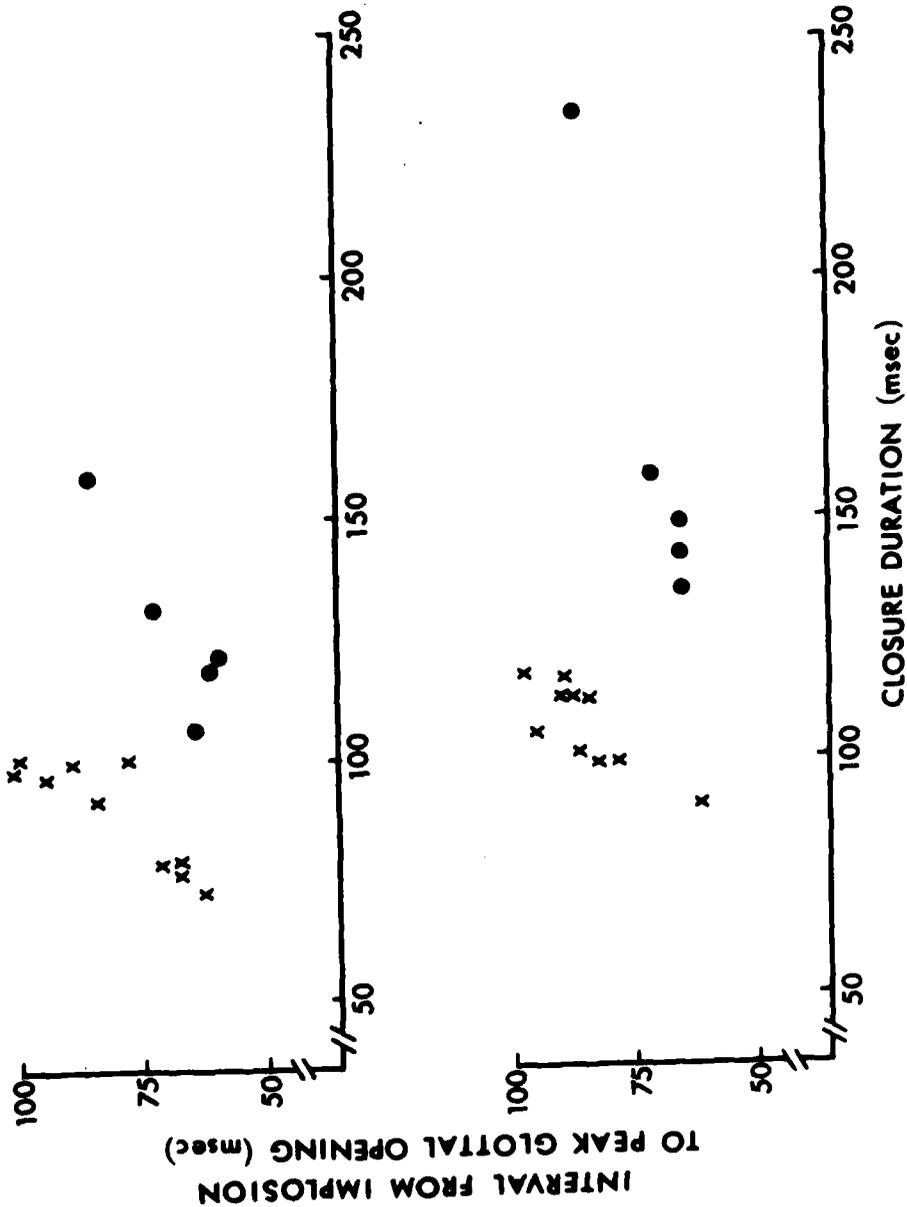
Figure 4. The interval from stop implosion to peak glottal opening plotted versus closure duration for Swedish voiceless stops in various positions and under different stress conditions. Each data point represents the mean of 20 tokens. Top and bottom graphs refer to two different speakers. Aspirated stops are denoted by X and unaspirated by O.

97

Figure 5 presents similar measurements for voiceless fricatives and aspirated stops in American English at two different speaking rates. As in Figures 1 and 3, peak glottal opening for aspirated stops occurs at oral release. In fricative production, peak opening occurs close to the middle of the oral constriction. Within the stops, the interval from implosion to peak glottal opening increases with increasing closure duration as in the Swedish material in Figure 4. A similar relationship is found for the fricatives, although the slope of the function is less steep. A comparison between the two speaking rates shows that the two sets of measurements form more or less a continuous function. These results thus indicate that the ratio between the interval from implosion to peak glottal opening and closure duration tends to remain constant.

A more detailed view of the laryngeal opening and closing gesture in voiceless obstruent production is presented in Figures 6, 7, and 8 for three different speakers and languages. The displacement averages were made with an integration of 15 milliseconds, and the velocity calculated by successive subtractions. All curves are aligned with reference to the offset of the preceding vowel. In the velocity plots, positive values indicate abduction and negative values indicate adduction. The linguistic material consisted of single voiceless stops and fricatives as well as clusters of stops and fricatives. The solid lines in the figures represent single fricatives or clusters beginning with a fricative, whereas the broken lines represent single stops or clusters beginning with a stop, irrespective of the nature of the following segments in the cluster. Japanese does not allow consonant clusters, and the Japanese material contains a devoiced vowel following the initial stop or fricative with a single or geminated stop or fricative occurring after the devoiced vowel.

In the displacement plots we observe again a difference in the timing of peak glottal opening with respect to the offset of the preceding vowel, i.e., peak opening occurs closer to the offset of the vowel when a fricative follows immediately after the vowel. From the velocity plots it is evident that peak abduction velocity is higher in the fricative case. The fricative abduction also has a narrow peak in the velocity plots, whereas the abduction gesture in the stop case is broader. For the Swedish subject in Figure 6, the stop abduction has an initial velocity peak followed by a second peak about 50 milliseconds later.

A striking similarity in the velocity plots for the different speakers is that peak velocity of the abduction gesture tends to occur at a fixed distance from the offset of the preceding vowel. This holds true for all the fricative cases, irrespective of variations in speed, size, duration, and timing of the glottal gesture. For the Japanese material in Figure 7, peak velocity of the stop abduction coincides in time with that for the fricatives. In the Icelandic case, Figure 8, peak abduction velocity occurs at two different times for fricatives and stops, respectively, but within the two families of curves, peak velocity tends to occur at the same time.
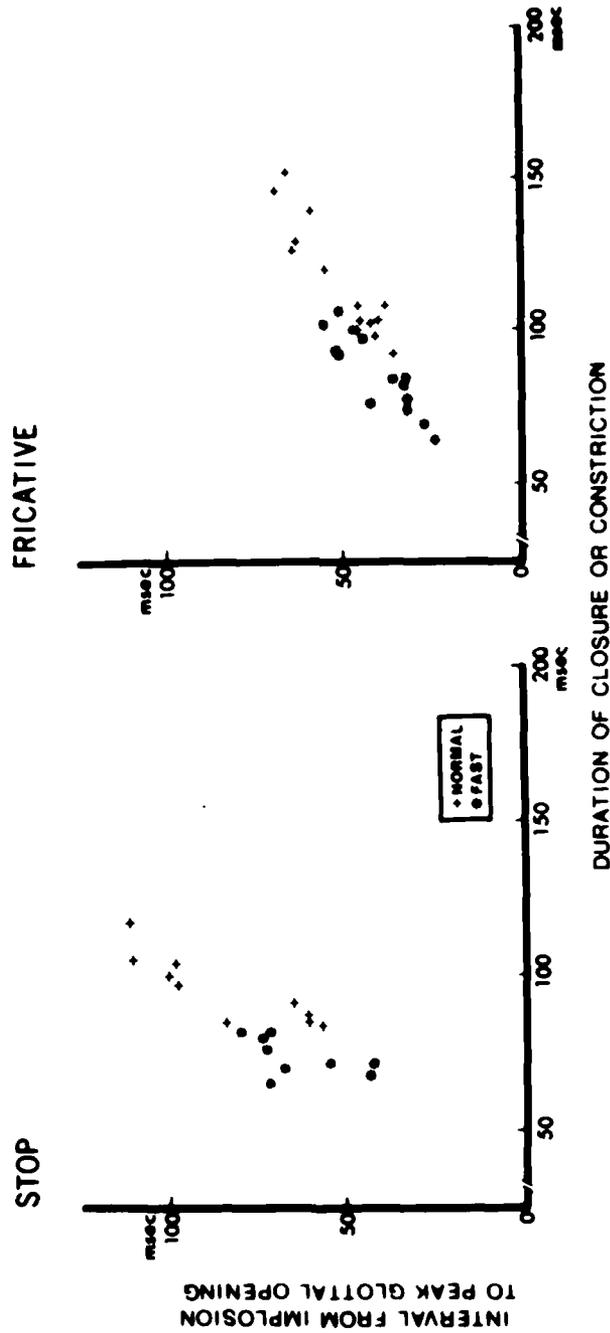
Figure 5. The interval from onset of tongue-palate contact to peak glottal opening plotted versus duration of oral closure or constriction for American English stops and fricatives in various positions and under different stress conditions at two speaking rates.

VELOCITY

DISPLACEMENT

DISTANCE FROM VOWEL OFFSET (MSEC)
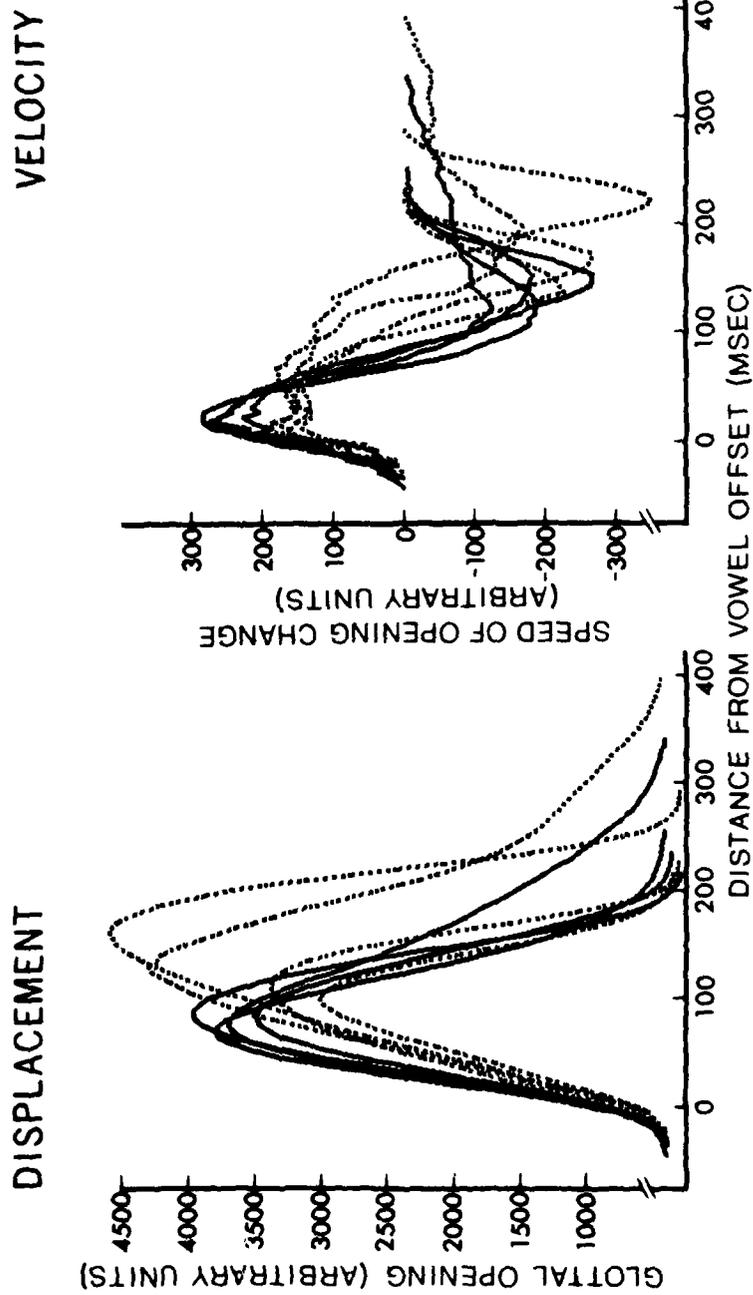
Figure 6. Plots of size and speed of the glottal abduction/adduction gesture for Swedish voiceless obstruents. Zero on x-axis represents offset of the vowel preceding the obstruents. Abduction velocity is shown with positive sign, adduction velocity with negative sign. See text for further explanation.

100
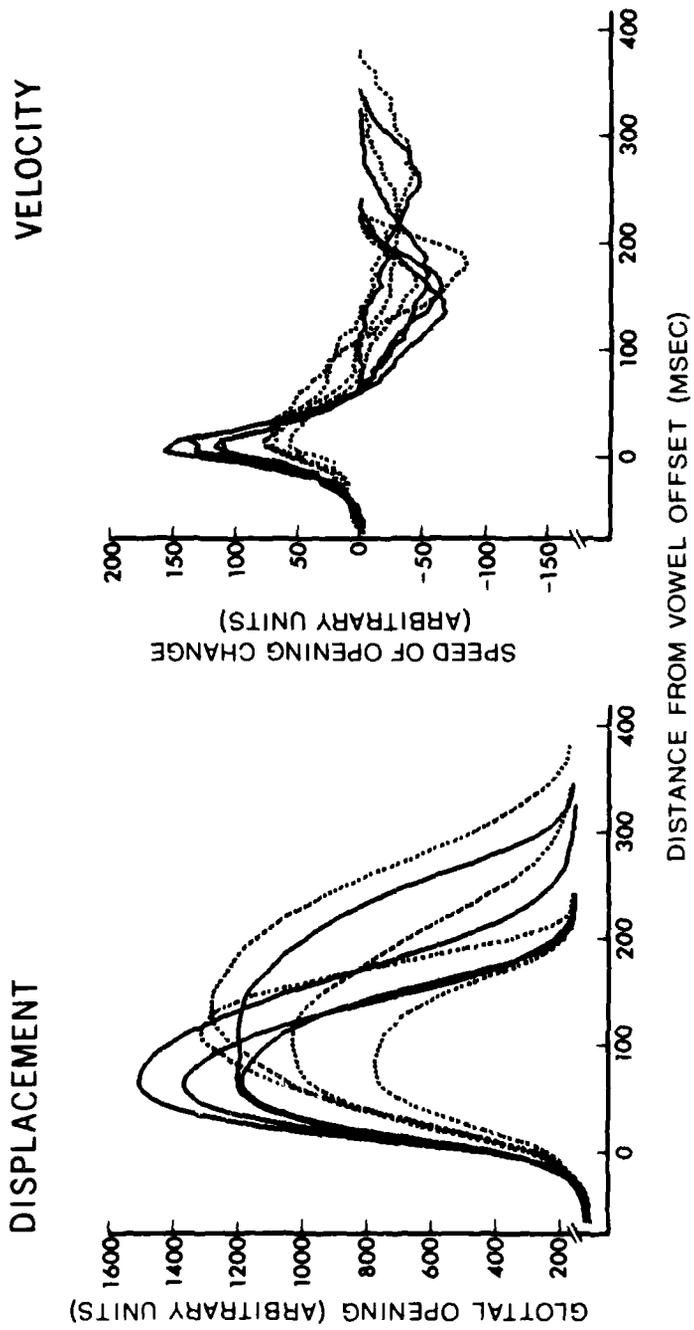
Figure 7. Plots of size and speed of the glottal abduction/adduction gesture for Japanese voiceless obstruents. Symbols as in Figure 6.

VELOCITY

DISPLACEMENT

SPEED OF OPENING CHANGE (ARBITRARY UNITS)

GLOTTAL OPENING (ARBITRARY UNITS)

DISTANCE FROM VOWEL OFFSET (MSEC)
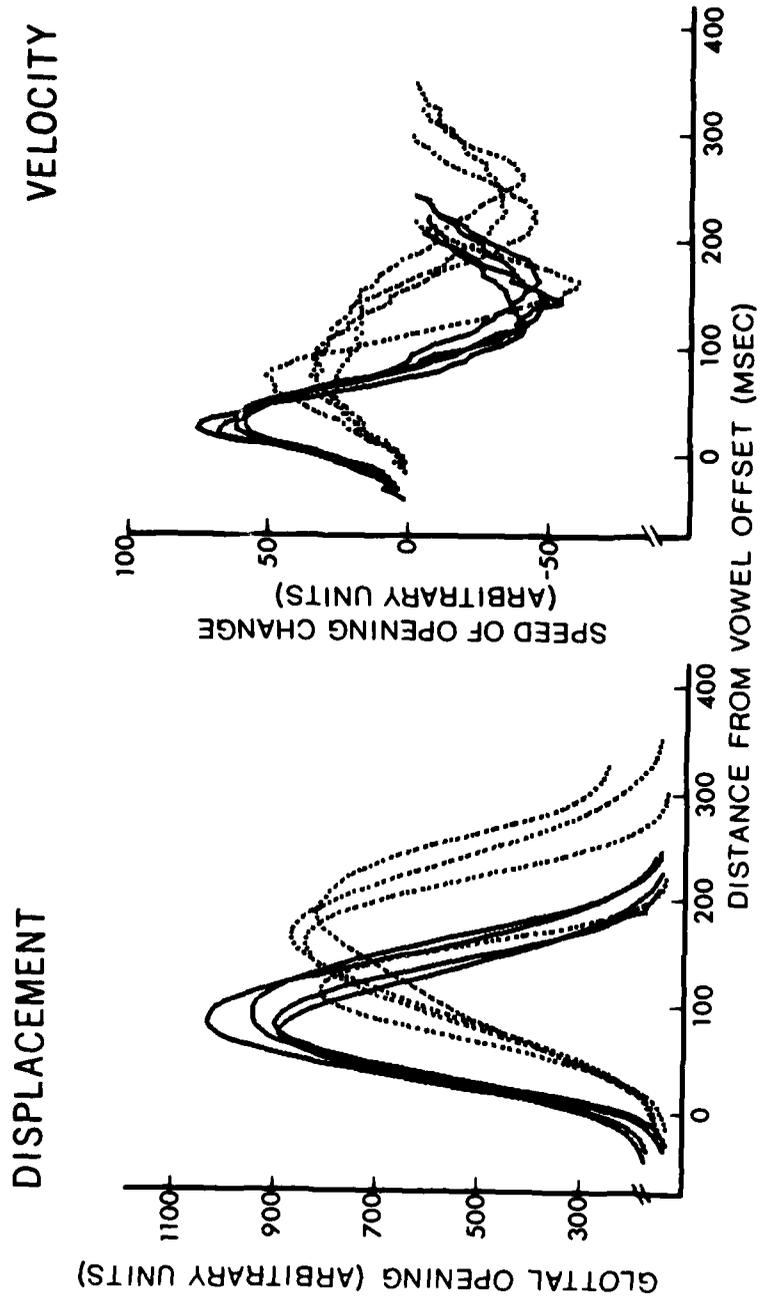
Figure 8. Plots of size and speed of the glottal abduction/adduction gesture for Icelandic voiceless obstruents. Symbols as in Figure 6.

102

DISCUSSION

The present results, as well as those of other studies reviewed in Löfqvist (in press), suggest that the glottis is continuously changing in voiceless obstruent production. Laryngeal articulations are thus organized in one or more opening and closing gestures. Static open glottal configurations rarely seem to occur in speech, and also appear difficult to maintain in some nonspeech conditions (cf. Löfqvist, Baer, & Yoshioka, 1980).

The laryngeal gestures are tightly coordinated with supralaryngeal events to meet the aerodynamic requirements for producing a signal with a specified acoustic structure. Variations in the relative timing of the laryngeal opening and closing gesture and the oral closing and opening gesture are used to produce contrasts of voicing and aspiration (cf. Abramson, 1977).

Initiation of glottal abduction before oral closure in voiceless stops produces preaspiration as shown in Figure 3. If glottal abduction starts after oral closure, prevoicing results, and if the abduction gesture starts at stop release, a voiced (or murmured) aspirated stop is produced. Similarly, a glottal gesture beginning at stop implosion and with peak glottal opening close to the implosion is used for producing voiceless unaspirated stops, whereas a gesture starting at implosion and with peak opening at stop release results in a postaspirated stop. These different obstruent categories are thus basically produced by differences in interarticulator timing.

Differences in the size of the laryngeal gesture seem to co-occur with the timing differences. Variations in size and timing of the laryngeal gesture are best regarded as interacting strategies for achieving a specific acoustic output. An early timing of peak glottal opening together with a small opening can thus be used in producing unaspirated voiceless stops, since they will both contribute to a glottal configuration suitable for voicing at stop release, cf. Figure 3. A comparatively small glottal opening for preaspirated stops could be related to the production of glottal frication noise during the period of preaspiration. Similarly, the size of the glottal gesture for a voiced (or murmured) aspirated stop would be adjusted to produce both glottal vibrations and frication noise. A large glottal opening at the release of voiceless postaspirated stops would not only contribute to the delay in voice onset but also create suitable aerodynamic conditions for noise generation at the oral place of articulation as the articulators are being separated immediately after the release.

The differences in glottal displacement and velocity between stops and fricatives in Figures 6, 7, and 8 are also most likely related to different aerodynamic requirements for stop and fricative production. A rapid increase in glottal area would allow for the high air flow necessary to generate the turbulent noise source during voiceless fricatives (Stevens, 1971). In stops, a slower increase in glottal opening together with the concomitant oral closure could be sufficient to stop glottal vibrations in combination with the buildup of oral air pressure (cf. Yoshioka, 1979). In the Icelandic material in Figure 8, glottal abduction starts considerably later relative to offset of the preceding vowel for stops than for fricatives. Although it is tempting to view this difference as a deliberate action by the speaker to avoid unwanted preaspiration, it is best regarded as a speaker-specific variation, since we

have found similar differences between stops and fricatives for speakers of American English, where preaspiration does not occur. The present results thus indicate that differences exist between stops and fricatives in the initial glottal abduction phase. The magnitude and form of these differences may, however, show some interspeaker variability.

The acoustic consequences of variations in interarticulator timing in obstruent production are complex and spread out over a period of time, involving differences in the sound source and the spectral composition of the signal.

The preaspirated stop in Figure 3 is thus associated with the following sequence of source changes: periodic voicing during the preceding vowel, aperiodic noise, silence, transient noise, periodic voicing during the following vowel. For the postaspirated stop in the same figure, the sequence would be voicing, silence, transient, noise, voicing. At the same time the spectral qualities of the signal would differ according to the nature of the preceding and following vowels and the place of articulation of the obstruent. This complex of acoustic cues, produced by a unified articulatory act, is integrated in speech perception to form a single percept (cf. Liberman & Studdert-Kennedy, 1978; Repp, Liberman, Eccardt, & Pesetsky, 1978).

As interarticulator timing appears to be an essential feature of voiceless obstruent production, one may question the descriptive adequacy and usefulness of feature systems with timeless representations for modeling speech production, whatever their merits may be for abstract phonological analysis. Specifying glottal states along dimensions of spread/constricted glottis and stiff/slack vocal cords (Halle & Stevens, 1971) would thus not only seem to be at variance with the phonetic facts but also to introduce unnecessary complications. The difference between postaspirated and unaspirated voiceless stops is rather one of interarticulator timing than of spread versus constricted glottis. Similarly, the difference between voiceless and voiced postaspirated stops is also one of timing rather than of stiff versus slack vocal cords. Preaspirated stops are naturally accounted for within a timing framework but cannot be readily differentiated from postaspirated ones in a timeless feature representation. It is, of course, possible to translate a timeless representation into differences in interarticulator timing, but if timing is of importance, it seems counterintuitive to derive it rather than represent it directly, especially if feature representations are to have a phonetic basis and describe parameters that the speaker can control independently. The importance of interarticulator timing in obstruent production is not a new idea, e.g., Rothenberg (1968), Lisker and Abramson (1971), Ladefoged (1973), Abramson (1977). It has, moreover, been noted by phonologists who, for reasons not entirely clear, still favor timeless phonological descriptions (e.g., Anderson, 1974).

The tight temporal coordination of laryngeal and oral articulations in voiceless obstruent production exemplified in the present material constitutes an important problem for any theory of speech production.

Models of speech production based on feature spreading (Daniloff & Hammarberg, 1973; Hammarberg, 1976; Bladon, 1979; see also Fowler, 1980) would seem incapable of handling this kind of interarticulator programming, at least

in their current form. One reason for this is that their temporal resolution is limited to quanta of phone or syllable size, whereas laryngeal-oral coordination in obstruents requires a finer grain of analysis. An additional limitation is that such models do not specifically address the general problem of interarticulator coordination in space and time. These limitations of current feature spreading models stem partly from the fact that they take as input the timeless units of abstract phonological theory.

Given the dynamic character of speech production and the need to coordinate different articulators in space and time, it seems rational to view speech production as an instance of control of coordinated movements in general. A powerful theory of motor control has been proposed by Bernstein (1967), and elaborated by Greene (1971, 1972; see also Boylls, 1975; Turvey, 1977; Kugler, Kelso, & Turvey, 1980; Kelso, Holt, Kugler, & Turvey, 1980; Fowler, Rubin, Remez, & Turvey, 1980). Designed to cope with the number of degrees of freedom to be directly controlled, this theory views motor coordination in terms of constraints between muscles, or groups of muscles that have been set up for the execution of specified movements. Areas of motor control where this theory has proved to be productive include locomotion (Grillner, 1975), posture control (Nashner, 1977), and hand coordination (Kelso, Southard, & Goodman, 1979). One merit of this view is that it predicts and rationalizes tight temporal relationships between articulators. In particular, it predicts that some such relationships should remain invariant across changes in stress and speaking rate, and material on oral articulations presented by Tuller and Harris (1980) is in agreement with this prediction. Some aspects of the present results can be rationalized within this theoretical framework.

Peak velocity of the glottal abduction gesture was found to occur almost at the same point in time relative to the offset of the preceding vowel, irrespective of variations in speed, size, duration, and timing of the gesture.

Another aspect is the relationship between laryngeal and oral articulations presented in Figures 4 and 5. Here, peak glottal opening was found to be delayed in relation to the formation of the oral constriction or occlusion as the latter increased. For aspirated stop consonants, this results in a constant temporal relation between peak glottal opening and oral release, ensuring an open glottis at the release to produce aspiration. The ratio between the interval from implosion to peak glottal opening and closure/constriction duration tends to remain constant across changes in overall obstruent duration.

We can regard such constant relationships as structural prescriptions for the articulators, specifying relations that have to be maintained in obstruent production across changes in stress and speaking rate. On the other hand, a metrical prescription specifies the activity levels of articulatory muscles. As suggested by Boylls (1975), the metrical prescription can be regarded as a scalar quantity multiplying the activities of the oral and laryngeal muscles in obstruent production while preserving the structural prescription.

REFERENCES

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

Anderson, S. R. The organization of phonology. New York: Academic Press, 1974.

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

Bladon, R. A. Motor control of coarticulation: Linguistic considerations. In Proceedings of the Ninth International Congress of Phonetic Sciences, Vol. 2. Copenhagen: Institute of Phonetics, 1979, 325-331.

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

Daniloff, R., & Hammarberg, R. On defining coarticulation. Journal of Phonetics, 1973, 1, 239-248.

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

Fowler, C. A., Rubin, P., Remez, R., & Turvey, M. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), Language production. London: Academic Press, 1980, 373-420.

Greene, P. H. Introduction. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge, Mass.: MIT Press, 1971, xi-xxxi.

Greene, P. H. Problems of organization of motor systems. In R. Rosen & F. Snell (Eds.), Progress in theoretical biology, Vol. 2. New York: Academic Press, 1972, 303-338.

Grillner, S. Locomotion in vertebrates: Central mechanisms and reflex interaction. Physiological Reviews, 1975, 55, 247-304.

Halle, M., & Stevens, K. A note on laryngeal features. Quarterly Progress Report (Research Laboratory of Electronics, Massachusetts Institute of Technology), 1971, 101, 198-213.

Hammarberg, R. The metaphysics of coarticulation. Journal of Phonetics, 1976, 4, 353-363.

Kelso, J. A. S., Holt, K., Kugler, P. N., & Turvey, M. T. On the concept of coordinative structures as dissipative structures: II. Empirical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland Publishing Company, 1980, 49-70.

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

Kent, R. D., & Minifie, F. D. Coarticulation in recent speech production models. Journal of Phonetics, 1977, 5, 115-134.

Kiritani, S., Kakita, K., & Shibata, S. Dynamic palatography. In M. Sawashima & F. S. Cooper (Eds.), Dynamic aspects of speech production. Tokyo: University of Tokyo Press, 1977, 159-168.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. 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. Amsterdam: North-Holland Publishing Company, 1980, 3-47.

106

Ladefoged, P. The features of the larynx. *Journal of Phonetics*, 1973, 1, 73-83.

Liberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology*, Vol. 8. Berlin: Springer Verlag, 1978, 143-178.

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

Lisker, L., & Abramson, A. S. Distinctive features and laryngeal control. *Language*, 1971, 47, 767-785.

Löfqvist, A. Interarticulator programming in stop production. To appear in *Journal of Phonetics*, in press.

Löfqvist, A., Baer, T., & Yoshioka, H. Scaling of glottal opening. *Journal of the Acoustical Society of America*, 1980, 67, S53. (Abstract)

Löfqvist, A., & Yoshioka, H. Laryngeal activity in Swedish obstruent clusters. *Journal of the Acoustical Society of America*, 1980, 68, 792-801.

Löfqvist, A., & Yoshioka, H. Laryngeal activity in Icelandic obstruent production. To appear in *Nordic Journal of Linguistics*, in press.

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

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

Rothenberg, M. *The breath-stream dynamics of simple-released-plosive production* (Bibliotheca Phonetica, Vol. 6). Basel: Karger, 1968.

Stevens, K. N. Airflow and turbulence noise for fricative and stop consonants: Static considerations. *Journal of the Acoustical Society of America*, 1971, 50, 1180-1192.

Tuller, B., & Harris, K. S. Temporal models of interarticulator programming. *Journal of the Acoustical Society of America*, 1980, 67, S64. (Abstract)

Turvey, M. T. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing*. Hillsdale, N.J.: Lawrence Erlbaum, 1977, 211-265

Yoshioka, H. Laryngeal adjustments during Japanese fricative and devoiced vowel production. *Haskins Laboratories Status Report on Speech Research*, 1979, SR-58, 147-160.

Yoshioka, H., Löfqvist, A., & Hirose, H. Laryngeal adjustments in the production of consonant clusters and geminates in American English. *Haskins Laboratories Status Report on Speech Research*, 1979, SR-59/60, 127-151.

107

AN ELECTROMYOGRAPHIC-CINEFLUOROGRAPHIC-ACOUSTIC STUDY OF DYNAMIC VOWEL PRODUCTION*

Peter J. Alfonso+ and Thomas Baer

There are many studies in the phonetics literature, based on various combinations of electromyographic (EMG), cinefluorographic, and acoustic data, that describe the positioning of various articulators, most notably the tongue, during the production of vowels. However, with the exception of a few experiments carried out at Haskins Laboratories and at the Research Institute of Logopedics and Phoniatrics at the University of Tokyo (e.g., Gay, Ushijima, Hirose, & Cooper, 1974; Borden & Gay, 1978; and Kiritani, Sekimoto, Imagawa, Itoh, Ushijima, & Hirose, 1976), none of these studies have incorporated simultaneous recording of all three types of measurement. The paucity of studies incorporating simultaneous measurements is most likely due to the inherent technical difficulties of the methodology, since the information gained from simultaneous monitoring of the different levels of speech production, namely neuromuscular, articulator movement, and acoustic, would contribute significantly to our understanding of dynamic speech production.

With respect to vowel articulation, it would be worthwhile to establish the agreement among muscle activity underlying tongue movement, positioning of the tongue, and the resultant acoustic output during the production of various vowels for the same speaker. For instance, Wood (1979) has pointed out that the controversy that still exists over the more appropriate level of vowel description, acoustic or articulatory, is related to the inconsistencies among different X-ray studies, and to the poor agreement between these studies and other acoustic studies. This seems to be the source of a recurring problem; often EMG, movement, and acoustic data collected from different experiments that usually use different talkers are used to make comparisons and assumptions about each measurement level. Certainly, the testing and formulation of models of vowel articulation would seem to depend upon a complete description provided only by simultaneous measures.

Other instances where simultaneous analysis of the three levels is more useful than any combination of the two are related to dynamic measurements of vowel production. That is, simultaneous measures not only allow for inter-articulator timing measurements, such as tongue and jaw relationships, but

---

also allow for intra-articulator timing measurements, for example, genioglossus muscle activity and tongue fronting. Furthermore, high correlations between patterns of EMG activity and movement lend support to the notion that the relationship between EMG activity and movement of the muscle-articulator system under study is causal.

The purpose of this study was to investigate the dynamics of vowel articulation by simultaneously monitoring muscle activity (using electromyography), articulatory movements (using lateral cinefluorography), and acoustic output. A single speaker of American English produced isolated syllables of the form /əpVp/, using ten different vowels. We will consider here only the dynamics associated with tongue movements for these syllables. More specifically, we will show that the timing of vertical tongue movements for both front and back vowels was time-locked to some component of the initial consonant, while the timing of horizontal movements began much earlier for back vowels than for front vowels. For back vowels, horizontal tongue movement began before voice onset for the schwa, whereas for front vowels horizontal tongue movement began at about the same time as their vertical movements. In addition, we will show that the differentiation in horizontal tongue movements during schwa production was perceptually significant.

PROCEDURE

Cinefluorographic films were made at a rate of 60 frames per second. For these films, pellets were glued to the tongue tip, blade, and dorsum and to the upper and lower incisors, as indicated in Figure 1. In addition, a gold chain was laid on the floor of the nasal tract for monitoring velar movements. However, we will consider here only movements of the tongue dorsum.

EMG signals were recorded from the orbicularis oris muscle and from two muscles of the tongue, the genioglossus and superior longitudinal. The paths of insertion of the hooked wire electrodes for these muscles are also indicated in Figure 1. Good quality acoustic recordings were made by using a close-talking directional microphone.

During the X-ray filming, the subject read a randomized 20-word list, producing two tokens each of the 10 vowels. He then continued without X-ray filming, producing an additional 20 tokens of each vowel to extend the base of the acoustic and electromyographic data. The subject's utterances from the experiment were later presented to a panel of listeners in an identification task, and all utterances were unambiguously perceived as intended by the talker.

Measurements of pellet movements with respect to the reference pellet (upper incisor) were made on a frame-by-frame basis with the aid of a digitizing tablet. Electromyographic and acoustic data were processed using standard methods at Haskins Laboratories.

110
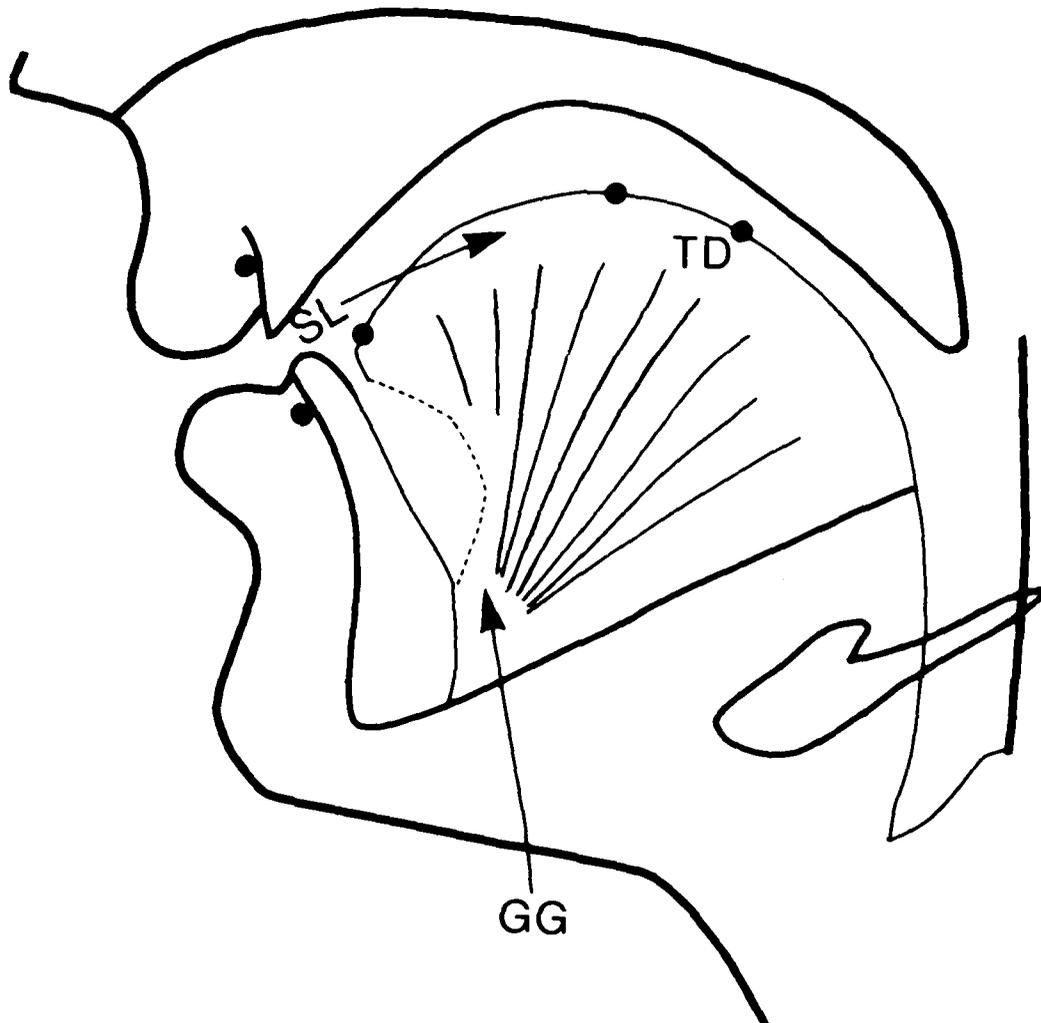
Figure 1. Schematic representation of lead pellets attached to the tongue tip, blade, and dorsum, and to the upper and lower incisors. Also shown is a gold chain laid on the floor of the nasal tract for monitoring velar movements. The arrows indicate the paths of insertion of the hooked wire electrodes for the genioglossus and superior longitudinal muscles.

111

RESULTS

The next three figures demonstrate the good agreement among the three types of measures made in this study.

Figure 2 shows results of acoustic measurements on vowels produced during the X-ray run. The back vowels, with the exception of /a/, were all relatively high and were tightly grouped. However, the front vowels were spread out approximately along a diagonal, with the vowels /i/ and /e/ higher and more forward than /I/ and /ε/.

Figure 3 shows the movement trajectories of the tongue dorsum pellet for each vowel during the interval from its voice onset until lip closure for the final consonant (that is, the vocalic period). Movements along all of these trajectories, except the one for /ɔ/, are in an ascending direction and away from the center. The pattern of locations of these trajectories grossly resembles the vowel pattern in the acoustic domain just shown.

Figure 4 shows the pattern of peak EMG activity for the genioglossus muscle for each of the ten vowels. Greatest activity is noted for /i/ and /e/ and somewhat less for /u/ and /o/. These vowels, traditionally termed tense, are also observed to be highest in the acoustic and articulatory domains. Among the remaining vowels, there is somewhat more activity for the front than for the back.

Next we turn our attention to articulator timing measurements. Simultaneous monitoring of different levels of speech production, namely muscle activity, articulator movement, and acoustic, allow for both intra- and inter-articulator timing measurements. As an example of intra-articulator measures, Figure 5 demonstrates the relationship between genioglossus EMG activity and tongue movements. This figure shows that correlation functions between patterns of genioglossus EMG activity with tongue horizontal and tongue vertical movements for the vowel /i/ nearly reach unity at latencies of about 110 msec. This latency seems to be a reasonable value for the mechanical response time of this muscle-articulator system. High correlations of this type, genioglossus EMG with tongue fronting and bunching movements in this example, lend support to the notion that the relationship between EMG activity and movement of the muscle-articulator system under study was causal.

Similar patterns of genioglossus activity were reported by Raphael and Bell-Berti (1974) for the same talker producing six of these vowels in a similar frame. The Raphael and Bell-Berti study, in addition, reports data from additional lingual muscles. Their data, as well as our own, demonstrate that the onset of genioglossus activity never preceded the onset of voicing for the vowel by more than 250 msec. For back vowels, however, styloglossus muscle activity begins at least 500 msec before the onset of voicing. This muscle is thought to participate in tongue backing. Thus, the EMG data suggest a timing difference for backing and fronting maneuvers.

With these comments in mind, we turn our attention to interarticulator timing measurements. Figure 6 shows sagittal plane trajectories for the tongue dorsum pellet for four vowels. The time interval for these plots begins at the voice onset of the schwa and ends at lip contact for the final consonant. The number of vowels has been limited here to simplify the figure.

112

F₁ vs F₂
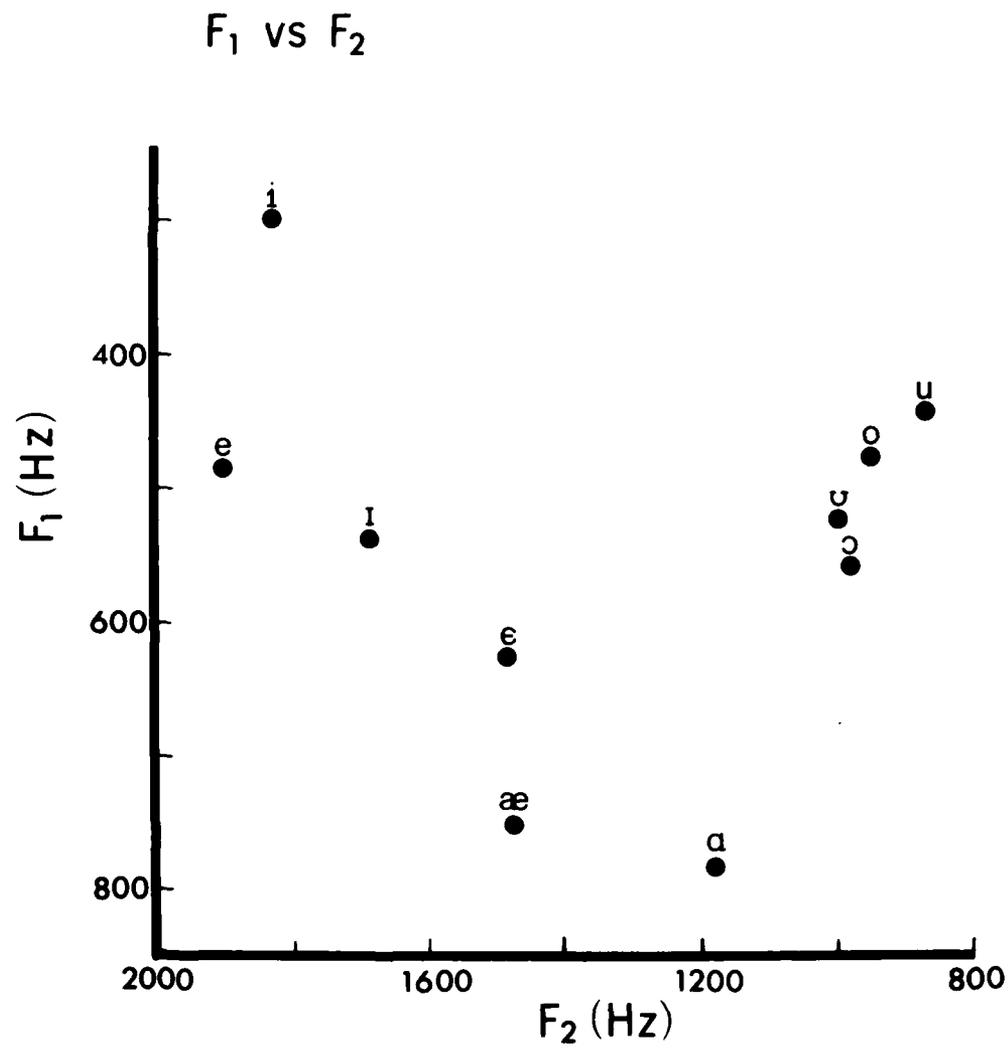
Figure 2. Peak center frequency values in Hz for the ten vowels used in this study. Each data point represents the average of the two tokens produced during the X-ray run.

TONGUE DORSUM TRAJECTORIES
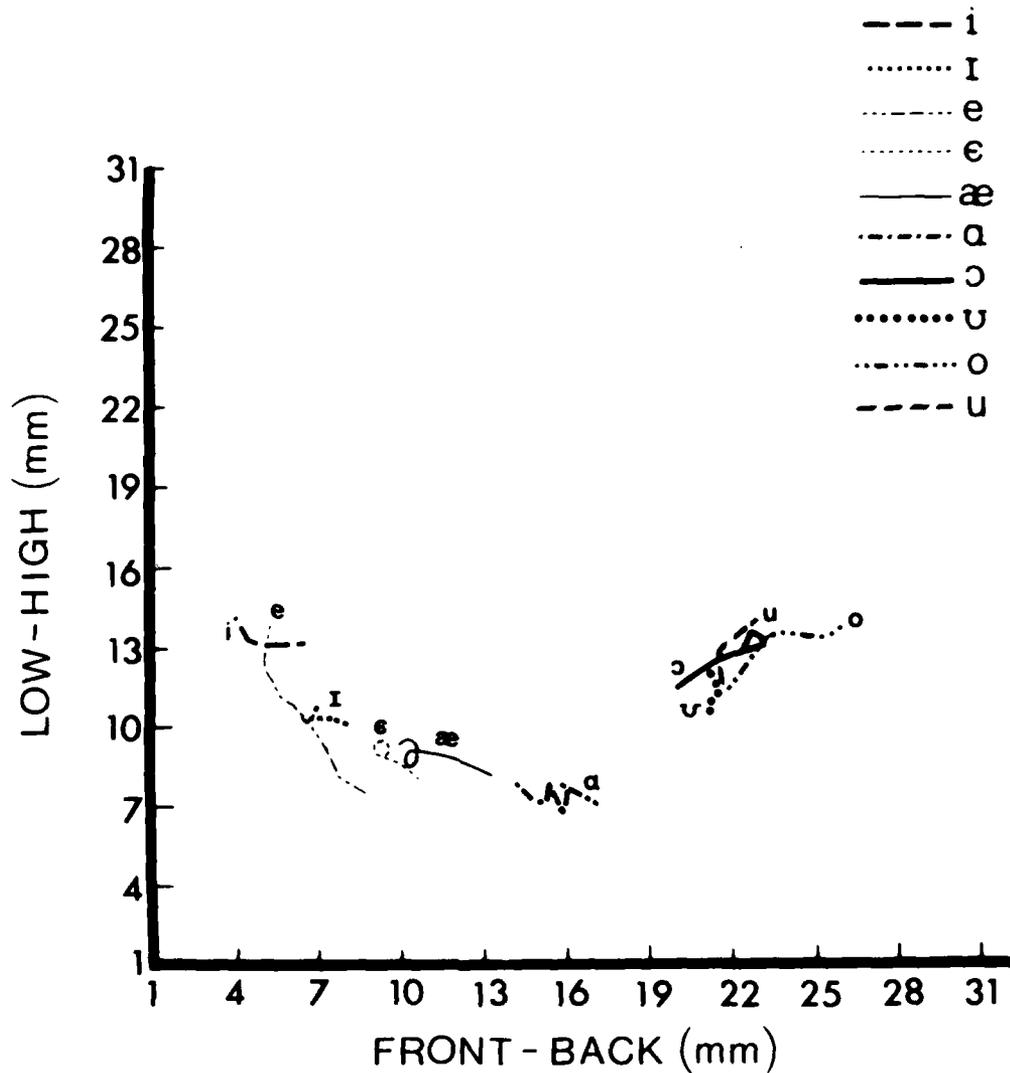
Figure 3. Movement trajectories of the tongue dorsum pellet during the interval from the voice onset for the vowel to the lip closure for the final consonant. With the exception of /ɔ/, movements along the trajectories are in an ascending direction and away from the center. Each trajectory represents the average movement of two tokens.

114

Figure 4. Peak genioglossus EMG activity for each of the ten vowels. Each data point represents the average of two tokens produced during the X-ray run.

115
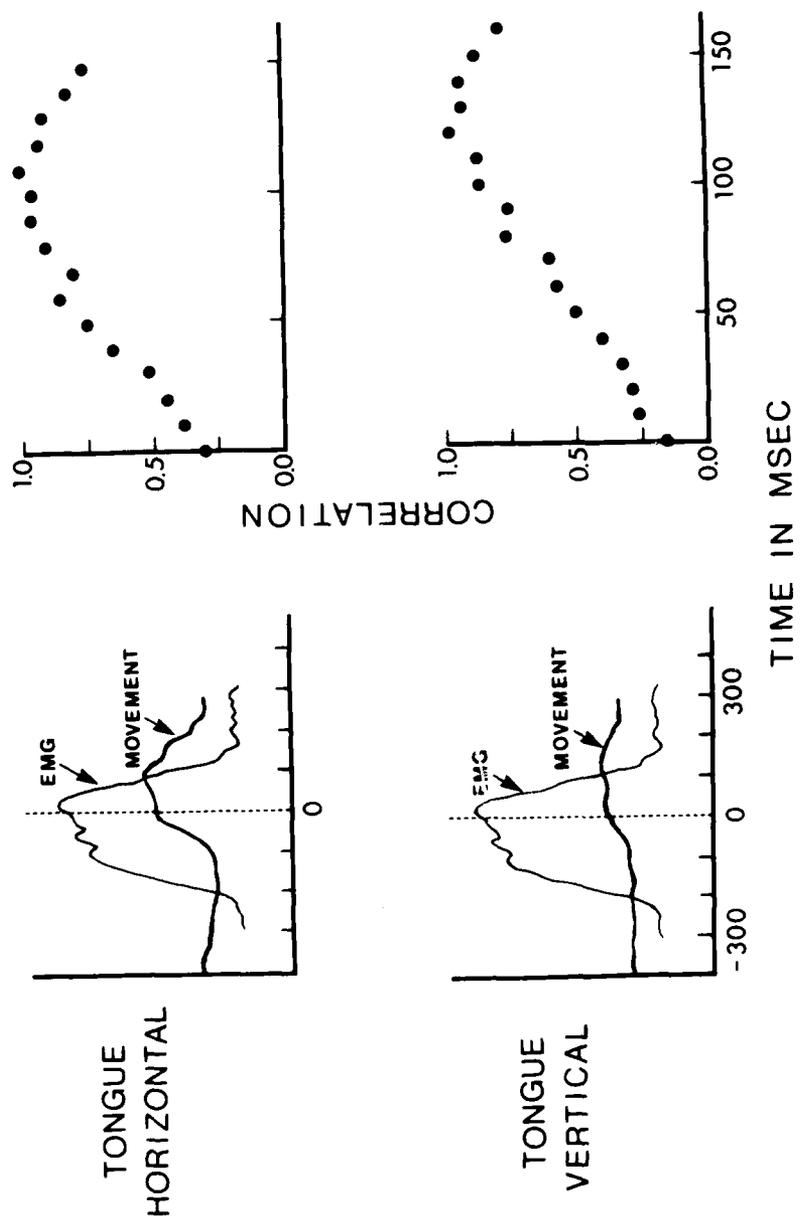
Figure 5. Genioglossus EMG activity with tongue dorsum horizontal movement (top left) and with tongue dorsum vertical movement (bottom left) during /i/. Correlation functions between the EMG curve and the respective movement curves are shown on the right.

TONGUE DORSUM TRAJECTORIES
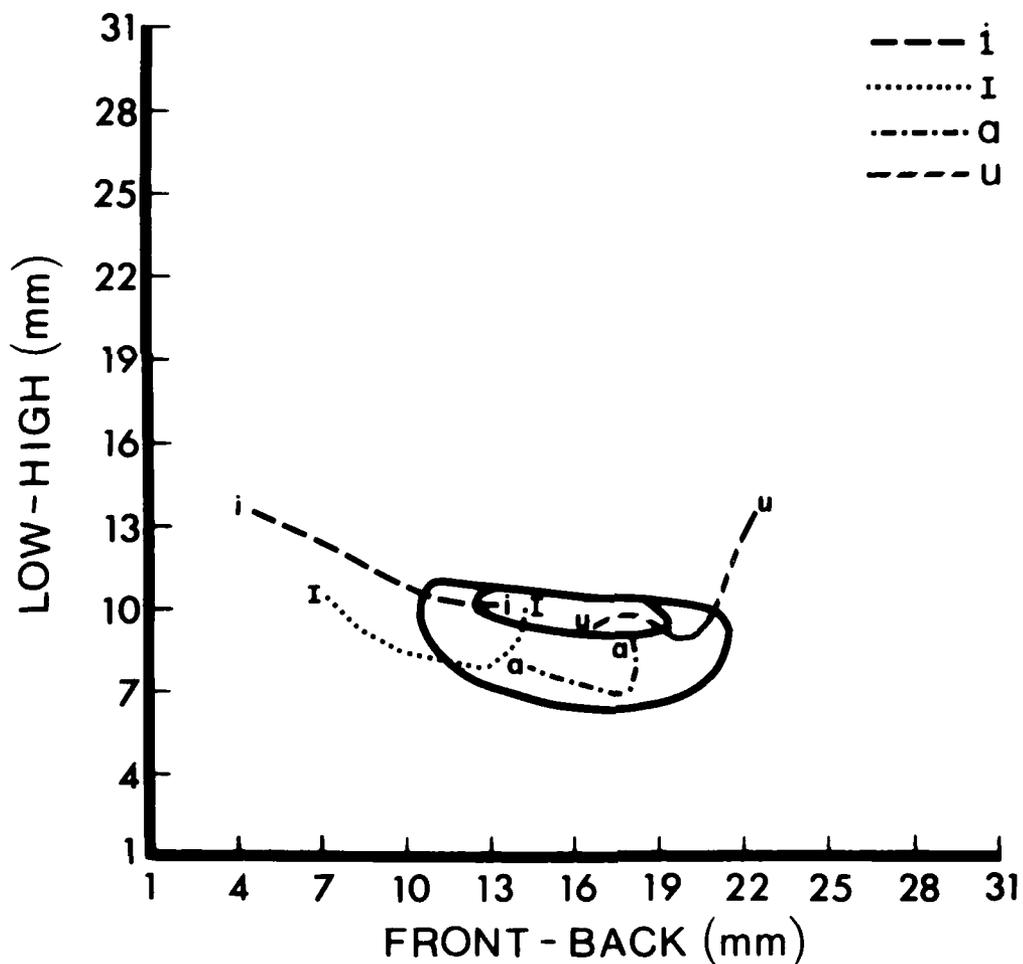
Figure 6. Movement trajectories of the tongue dorsum pellet during the interval beginning with the voice onset of the schwa, including the initial consonant and the vowel, and ending with the lip contact for the final consonant. Trajectories during the production of the schwa are enclosed by the inner black line, during the production of the initial bilabial closure are enclosed by the outer black line, and during the interval from the release of the initial consonant to the lip closure for the final consonant appear outside the black lines.

117

Lines have been superimposed on the trajectories in Figure 6 to indicate three different time intervals. The trajectories during the production of the schwa are enclosed by the inner line. The trajectories during the production of the bilabial closure are enclosed by the outer line. With the exception of /a/, trajectories after the consonant release appear outside the region enclosed by the lines.

Considering tongue positioning during the schwa, we can see that the region is long and flat; that is, anticipatory movements for the vowel occur primarily in the horizontal direction but very little in the vertical direction. Moving into the /p/ closure region, the trajectories continue to spread horizontally and also lower. Lowering movements during bilabial stops have been noted previously (Houde, 1967). It is unclear whether this movement is active or passive. In either case, there is a movement apparently related to the consonant that makes it difficult to determine the onset of vowel-related movements. Finally, the trajectories, moving upward and out toward the extremes of the space, demonstrate vowel-related movements.

The next two figures show the time course of tongue dorsum movements for all ten vowels. First, we consider the vertical dimension, shown in Figure 7. In this plot, the lineup point--zero time--was the onset of voicing for the vowel. Implosion for the consonant occurred at different times depending on vowel type, and ranges from about 120 to 160 msec. Vertical tongue position is the same for all vowels during the interval preceding implosion. The curves begin to diverge from each other at this point. Therefore, the onset of vertical vowel-related movements appears to be time-locked to some component of the consonant, so that they appear in these utterances at about the time of implosion.

Horizontal movements shown in Figure 8 are different. These curves are separate even at the earliest time measured, 350 msec before voice onset for the vowel. More significantly, the curves for back vowels and high front vowels begin to diverge from each other almost immediately. Notice that while backing movements for back vowels begin much earlier than their vertical movements, the fronting movements for front vowels begin only at about the same time as their vertical movements--that is, at about the moments of implosion.

We can perhaps explain the difference between fronting and backing on physiological grounds. At least for the high front vowels, a single muscle-- namely the genioglossus--may be responsible for moving the tongue both forward and upward. On the other hand, tongue backing is achieved by muscles other than the genioglossus--for example, the styloglossus. Thus, backing movements could occur independently from vertical movements in high back vowels.

Why they should be controlled independently, however, cannot be determined from the above data alone. Several explanations are possible. It may be that backing movements are intrinsically slower than raising and fronting movements and therefore must begin earlier. Other explanations might rest on acoustic or aerodynamic grounds. However, the results show, for this speaker, that front-back information about the vowel is available before high-low information, and that the information is available at the beginning of the syllable.

118

TONGUE DORSUM VERTICAL POSITION
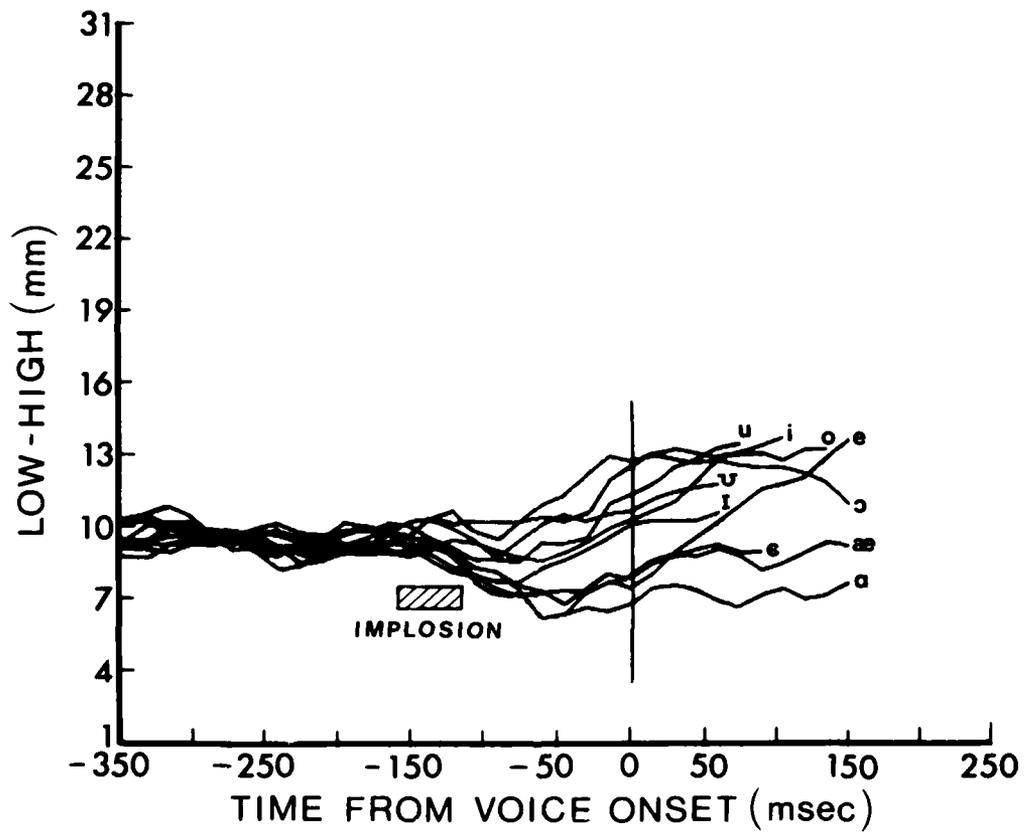
Figure 7.   Tongue dorsum vertical movements.   Zero time represents the onset
            of voicing for the vowel.   Implosion of the initial consonant
            ranged from -120 to -160 msec depending on vowel type, and is shown
            by the rectangle.

TONGUE DORSUM HORIZONTAL POSITION
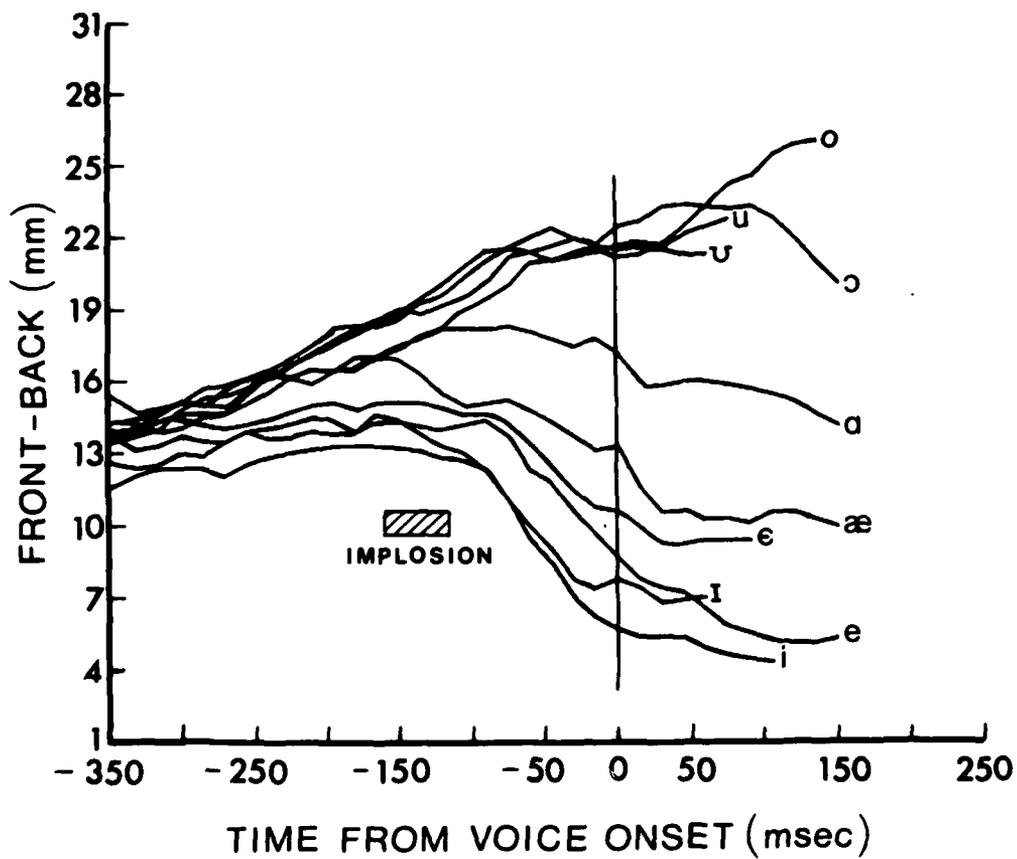
Figure 8. Tongue dorsum horizontal movements. Zero time represents the onset of voicing for the vowel. Implosion of the initial consonant ranged from -120 to -160 msec depending on vowel type, and is shown by the rectangle.

To test the notion that the anticipatory horizontal tongue movements during the production of the schwa were perceptually significant, AX discrimination and phoneme labeling tests were conducted. Specifically, we wanted to know if listeners could discriminate between schwas produced with front versus back tongue positions. Schwa segments from three productions of /əpip/, and from a single production of /əpIp/, /əpup/, and /əpap/ were excised by computer. Each of the six stimuli was about 25 msec in duration and consisted of about three pitch periods. Using the Haskins Pulse Code Modulation system, the six stimuli were digitized, and AX discrimination and labeling tests were prepared and presented to 12 subjects. The results of the discrimination test are shown in Figure 9. The ordinate represents the A stimulus and the abscissa represents the X stimulus of all possible AX discrimination pairs. The data are collapsed across the front group, which consisted of the three schwas taken from three different productions of /əpip/ (hereafter referred to as the /i/ schwas) and one schwa taken from /əpIp/ (hereafter the /I/ schwa), and a back group that consisted of one schwa each taken from a single production of /əpap/ and /əpup/ (the /a/ and /u/ schwas, respectively). For instance, the first row shows that when the first token of one of the three /i/ schwas, i1, was paired with front group schwas, i2, i3, and I schwas, discrimination performance was at chance level, 46 percent correct. However, when the i1 schwa was paired with back group schwas (the /a/ and /u/ schwas), discrimination performance improved to 82 percent correct. The summary data shown at the bottom of the figure demonstrate that discrimination performance across all front-back AX pairs was well above chance at 85 percent correct, whereas discrimination performance across front-front pairs was at a chance level of 46 percent correct. However, also note that discrimination performance across back-back pairs was also well above chance at 86 percent correct. Finally, note that overall discrimination performance, which included same as well as different AX pairs, was at 79 percent correct. These data led us to conclude that listeners were able to discriminate between the front and back group schwas produced by the same speaker. However, discrimination was probably based on the acoustic consequences of articulatory parameters other than fronting and backing alone, since discrimination performance between the back group schwas, as well as overall discrimination performance, was very high.

Based on the results of the discrimination test, we decided to test further the perceptual significance of the anticipatory horizontal movement and perhaps other differentiating articulatory gestures occurring during the production of the schwa by asking our subjects to label the stimuli as either /i/, /I/, /u/, or /a/. The same stimuli used in the discrimination test were used in the labeling tests, except that only one /i/ schwa was used. The results are shown in Figure 10. Here, each row represents the distribution of responses for 240 presentations of a stimulus. In each cell, the upper left score represents the frequency of that response, and the bottom right score represents percent occurrence. Overall correct performance, represented by scores of the main diagonal, is 42 percent correct, which is well above chance. Even though the schwa stimuli are only about 25 msec long, and represent reduced vocal tract shapes as plotted in both the movement and acoustic space, they appear to have a distinguishable vowel-like quality that results in the surprisingly accurate labeling.

'X' STIMULUS

	E$_1$	E$_2$	E$_3$	I	A	U
E$_1$			15/33 46		18/22 82	
E$_2$			11/33 33		22/22 100	
E$_3$			13/33 39		21/22 96	
I			21/33 64		17/22 77	
A			33/44 75		11/11 100	
U			38/44 86		8/11 73	

'A' STIMULUS — FRONT / BACK

ACROSS GROUPS = 149/176 = 85%

FRONT VS FRONT = 60/132 = 46%

BACK VS BACK = 19/22 = 86%

TOTAL CORRECT = 520/660 = 79%

Figure 9. Results of AX discrimination testing. The ordinate represents the A stimulus and the abscissa represents the X stimulus of all possible AX pairs. Data are collapsed across a front group consisting of three "/i/ schwas" and one "/I/ schwa," and across a back group consisting of a single "/a/ and /u/ schwa." The symbol "E" represents the vowel /i/.

122

RESPONSE

	E	I	U	A
E 240	72 / 30	71 / 30	57 / 24	40 / 17
I 240	54 / 23	93 / 39	33 / 14	59 / 25
U 240	43 / 18	27 / 11	126 / 53	48 / 20
A 240	2: / 11	63 / 26	35 / 15	113 / 47

STIMULUS

404/960   42

Figure 10. Results of the labeling tests. Each row represents the distribution of the responses for 240 presentations of a stimulus. In each cell, the upper left score represents the frequency of that response, and the bottom right score represents percent occurrence. The symbol "E" represents the vowel /i/.

123

Finally, notice that the subjects appeared to have more difficulty labeling the front schwas than the back. The /i/ stimulus, for example, was labeled as /i/ 72 times and as /I/ 71 times, whereas the /u/ and /a/ stimuli were labeled correctly 126 and 113 times, respectively. Although it is quite probable that other vocal tract parameters contributed to the increased accuracy in which the back schwas are labeled, we submit that the anticipatory backing gesture observed in the movement data during schwa production is at least one of the articulatory parameters contributing to this effect. That is, the anticipatory tongue backing during schwa production appears to be perceptually significant.

In conclusion, the major findings of this experiment indicate that studies of coarticulation must consider the different components of tongue movement since they appear to have different constraints, and that the consequences of the anticipatory tongue movements appear to be perceptually significant.

REFERENCES

Borden, G. B., & Gay, T. S. On the production of low tongue tip /s/: A case report. Journal of Communication Disorders, 1978, 11, 425-431.

Gay, T. J., Ushijima, T., Hirose, H., & Cooper, F. S. Effect of speaking rate on labial consonant-vowel articulation. Journal of Phonetics, 1974, 2, 47-63.

Houde, R. A. A study of tongue body motion during selected speech sounds. Unpublished doctoral dissertation, University of Michigan, 1967.

Kiritani, S., Sekimoto, S., Imagawa, H., Itoh, K., Ushijima, T., & Hirose, H. Simultaneous recording of EMG with pellet tracking by use of x-ray microbeam. Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 1976, No. 10, 13-14.

Raphael, L. J., & Bell-Berti, F. Tongue musculature and the feature of tension in English vowels. Phonetica, 1974, 32, 61-73.

Wood, S. A radiographic analysis of constriction location for vowels. Journal of Phonetics, 1979, 7, 25-43.

SHOULD READING INSTRUCTION AND REMEDIATION VARY WITH
THE SEX OF THE CHILD?

Isabelle Y. Liberman+ and Virginia A. Mann++

We have been asked to consider the possibility that methods of reading
instruction and remediation should vary with the sex of the child. However,
our research suggests that the critical problems underlying reading disability
may very well be the same for both boys and girls--the problems may simply be
more prevalent among boys. Therefore, we would prefer to begin a discussion
of this question not by a consideration of sex differences, but rather by
describing the characteristics that we have found among the reading disabled
which distinguish them from children who read well. We will then present some
recent evidence from our laboratory about how sex may or may not relate to
some of these characteristics, and finally will offer some thoughts about
instruction and remediation.

The research effort over the past several years or so by the Haskins
reading research group has led us to the conviction that the difficulty of
most, though perhaps not all, of the children who have problems in learning to
read is basically linguistic in nature--not visual, or auditory, or motor, or
whatever, but rather in the ineffective use of phonologic strategies. Thus
far, we have found this linguistic deficiency of poor readers in regard to two
major requirements of the reading process--lexical access and representation
in short-term memory.

LINGUISTIC STRATEGIES IN READING

Linguistic Awareness and Lexical Access

First, a few words about the requirements of lexical access--that is,
what the would-be reader needs if he is to get from the visual stimulus to the
word it represents. Here we have considered that one critical requirement is

---

a kind of linguistic awareness--the ability to stand back from one's language and analyze it into its component segments. Where the speaker-listener can usually make do with an understanding of linguistic structures that is only passive, the reader-writer is often required to deal with those structures in a more explicit way. To that extent, the would-be reader-writer must be a kind of linguist. At the very least, he must become aware of the segmental units represented by the orthography. In an alphabetic system, the basic segmental unit is, of course, the phoneme.

We have learned from speech research (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) that the phoneme should be particularly difficult to abstract from the speech stream. Because of the way we articulate and co-articulate, phonemes are merged in the sound in such a way that a word like dog, for example, has three phonological segments and three orthographic segments but only one isolable segment of sound. The information for the three phonological segments is there, but so thoroughly overlapped in the sound that the phonemes cannot be made to stand alone. This characteristic of speech is no problem for the speaker-hearer because he is apparently equipped with a neurophysiology that functions automatically below the level of awareness to extract the phonological structure for him. To understand a spoken utterance, therefore, the speaker-hearer need not be explicitly aware of its phonological structure any more than he need be aware of its syntax. But that explicit awareness of phonological structure of his language is precisely what we believe to be required if the beginning reader is to take full advantage of the alphabetic system. First, he must realize that spoken words consist of a series of separate phonemes. Second, he must understand how many phonemes the words in his lexicon contain and the order in which these phonemes occur. Without this awareness, he will find it hard to see what reading is all about (Liberman, 1971, 1973).

Consider the child who sees the printed word dog for the first time. If he has never been exposed to language analysis skills, he will see the printed word only as a visual pattern of risers and descenders and squiggles of one sort or another and will be at a loss to pronounce it at all. But suppose he has been taught to identify the letters and, as they say, "sound them out." No matter how skilled he is at reading the letters and approximating their sounds, he must still match the printed word dog to the real word /dɔg/ he already has in his lexicon. To do that, however, he must understand that the word /dɔg/ that he already knows consists of these three segments. Otherwise, given the impossibility of producing the phonemic segments in isolation, the best he can do in reading the word is to produce [də-ɔ-gə], a nonsense trisyllable that bears no certain relationship to the lexical item /dɔg/. Moreover, another consequence of the merging of the phonemes in the sound stream is that if he is to arrive at the correct phonological representation of the word, he had better not pronounce each letter separately. Instead, he will have to pronounce the syllable that is represented by two or three or more letters, the number varying with the nature of the word. In the case of the word /dɔg/, the number is three. We suspect that acquiring the ability to do this--that is, to know how to combine the letters of the orthography into the appropriate coding units and, moreover, to do that quickly and automatically (Laberge & Samuels, 1976)--is an aspect of reading skill that as much as any other separates the fluent reader from the beginner.

Given all these considerations, we can see why we might expect a reader to find it difficult to become aware of the phonemic segments and why this might be a problem for him as he begins to read. Let us now look very briefly at some of the evidence that the child does indeed have difficulty with phonemic analysis.

In our own research (Liberman, Shankweiler, Fischer, & Carter, 1974), we have found that in a sample of four-, five-, and six-year-olds, none of the nursery-age children could segment by phoneme, whereas half managed to do syllable segmentation. Only 17 percent of the kindergarteners could segment by phoneme, while again about half of them could segment by syllable. At six, whereas 90 percent of the children could do syllable segmentation, only 70 percent were successful with phoneme segmentation. It is certainly clear from this research and from the many other studies that followed that awareness of phoneme segments is harder to achieve than awareness of syllable segments and develops later, if at all.

Having suggested that the linguistic awareness necessary for a proper appreciation of an alphabetic orthography is, in fact, hard to achieve, we can turn again to its role in reading and summarize the empirical evidence available. To save space, we will touch only on the correlational evidence; there is considerable other corroborative evidence from the analysis of the errors of beginning readers (Shankweiler & Liberman, 1972; Fowler, Liberman, & Shankweiler, 1977; Fowler, Shankweiler, & Liberman, 1979), but we will have to omit that here.

In considering the correlational studies, we should begin by remarking on the spurt in awareness of phoneme segmentation at age six, from 17 percent correct at age five to 70 percent correct at age six. Six is, of course, the age at which the children in our schools begin to receive instruction in reading and writing. It goes without saying that age is important for both linguistic awareness and for reading, because, being cognitive achievements of sorts, both linguistic awareness and reading must require the attainment of a certain degree of intellectual maturity. But we also suspect that these two abilities are reciprocally related: While phonetic awareness may be important for the acquisition of reading, being taught to read may at the same time help to develop phonetic awareness (Liberman, Liberman, Mattingly, & Shankweiler, 1980; Alegria, Pignot, & Morais, in press; Morais, Cary, Alegria, & Bertelson, 1979).

Our own research speaks only to the first point--that linguistic awareness may be necessary for the acquisition of reading. What we have found in numerous experiments is that despite widely diverse subject populations, school systems, and measurement devices, there is a strong positive correlation between awareness of phoneme segmentation and later success in learning to read (Blachman, 1980; Helfgott, 1976; Treiman, Note 1; Zifcak, 1977).

A longitudinal study in preparation by our group (Mann, Liberman, & Shankweiler, Note 2) has just recently replicated an earlier finding of ours (Liberman & Shankweiler, 1979) that the ability to segment a word at all, even at the syllable level, is very highly correlated with reading ability. It was found that 85 percent of the good readers in the first-grade group were among the kindergarteners who had been able to segment by syllable the year before,

whereas only 24 percent of the poor readers had been able to do so. The segmenting ability of the average readers fell in between. We will return to this study later when we look at differences between the sexes.

Now as to the second point, the possibility that instruction in reading is important in the development of linguistic awareness (or the reciprocal nature of its relationship with reading), there is some work by a team of Belgian psychologists that is both relevant and interesting. One paper, from the Belgian laboratory (Alegria, Pignot, & Morais, in press), compares the syllable and phoneme segmentation performances of two groups of first graders-- one which had been taught by a largely whole-word method (the global group) and the other which had been taught by a largely phonics method (the synthetic group). The synthetic group did somewhat better than the global group on a syllable analysis task (72 percent correct versus 63 percent), but spectacularly better than the global group on a phoneme analysis task (60 percent correct versus only 16 percent correct for the global group). Thus, we see that awareness of phoneme segmentation is enhanced by a method of reading instruction that directs the child's attention to the internal structure of the word. We will have more to say about this later when we talk about instructional methods.

So much for linguistic awareness and its relation to reading an alphabetic language. We do not say that linguistic awareness is the only attribute needed for lexical access, just that it may be an important one. Another that should be mentioned is ability to do rapid automatic naming (RAN) (Denckla & Rudel, 1976). A recent study (Blachman, 1980) suggests that a three-part test that taps the language analysis skills of phoneme segmentation, the word retrieval ability of RAN, and the phonetic coding of oral memory tasks may provide a remarkably efficient predictor of future reading success. That brings us to our second major linguistic requirement of the reading process, namely, the requirement for phonetic coding in short-term memory.

Phonetic Coding in Short-Term Memory

It is obviously a characteristic of all language comprehension that the component words of a phrase or sentence must be held temporarily in memory so that the meaning of the whole phrase or sentence can be extracted. It is, of course, possible that in reading, some nonlinguistic representation--visual or semantic, perhaps--might be invoked (Kleiman, 1975). Such a strategy does appear to be used by the congenitally deaf (Locke, 1978), but they are notoriously poor readers.

At all events, we have assumed that in normal language processing, the use of phonetic structures is a particularly efficient way to meet the short-term memory requirements that all language comprehension imposes (Liberman, Mattingly, & Turvey, 1972). And that assumption was certainly reinforced in our minds by the abundant evidence in the psychological literature that when short-term memory is stressed, normal adults do rely on phonetic codes.

In view of these considerations, we were interested to learn whether beginning good and poor readers could be further distinguished by the degree to which they rely on a phonetic representation when short-term memory is

stressed. We assumed that good beginning readers of an alphabetic orthography would have the phonetic structure already available for use in short-term memory. As for the poor readers, we know that many have difficulty in going the analytic, phonetic route and might tend, therefore, to rely more heavily, perhaps, on representations of a visual or semantic sort.

To test that assumption, we carried out several experiments with children in the second year of elementary school. In these experiments, we used a procedure in which the subject's performance is compared on recall of phonetically confusable (rhyming) and nonconfusable (nonrhyming) material. Our expectation was that the rhyming items would generate confusions and thus penalize recall in subjects who use a phonetic representation in short-term memory.

The results showed that though the superior readers were better at recall of the confusable items, their advantage was virtually eliminated when the items were phonetically confusable. Phonetic similarity always penalized the good readers more than the poor ones. As can be seen in Figures 1, 2, and 3, these findings held true for recall of letters (Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979), words and sentences (Mann, Liberman, & Shankweiler, 1980) and obtained, moreover, whether the items to be recalled were presented to the eye or to the ear.

The longitudinal study mentioned before (Mann et al., Note 2) provides compelling evidence of the importance in beginning reading not only of linguistic awareness, as we reported above, but also of phonetic coding in short-term memory as well. In this study, kindergarteners were given the Corsi test of memory for the position of randomly scattered blocks (Corsi, 1972) and also tests for the memory of orally presented rhyming and nonrhyming sequences of words. The following year, as first graders, these same children were retested on those tasks, and in addition, were given a reading test by means of which they were grouped as good, average, or poor readers.

The findings are displayed in Table 1. As can be seen there, the performances of the three reader groups were quite undifferentiated on the Corsi memory test, which is nonverbal in nature. In contrast, the performances of the three groups on verbal memory tasks were strikingly and significantly differentiated. The difference related to how they were affected by rhyme: The good readers were strongly affected by it; the average readers less so; and the poor readers hardly at all. Thus once again, phonetic similarity penalized the better readers more than it did the poorer ones.

Recent studies by Byrne and Shea (1979) strongly support the finding that good readers tend to use phonetic representations in remembering linguistic materials. In addition, these studies provide compelling evidence that the poor readers, in contrast, may prefer a semantic strategy instead. Using a memory for repeated items design, these investigators first presented the subjects with foils that were either semantically or phonetically confusable with words on the antecedent list. They found that the poor reader in processing oral language favors a semantic coding strategy over the phonetic when the two are in competition, while the good reader does the opposite. In their second experiment, nonsense words were used and the foils were now

129
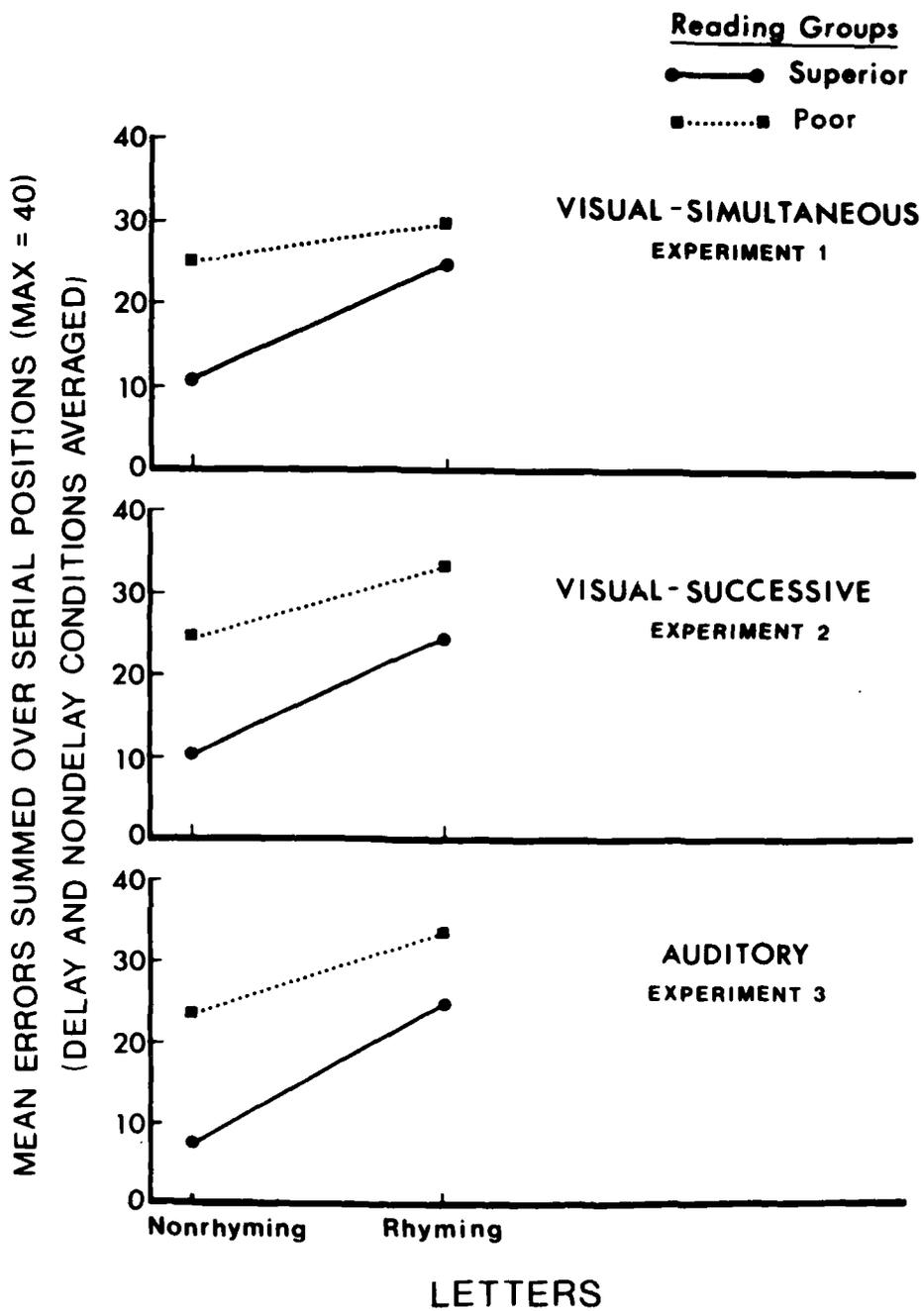
Figure 1. Mean errors of superior and poor readers on recall of letter strings, summed over serial positions. (Means from delay and nondelay conditions are averaged. Maximum = 40.)

130
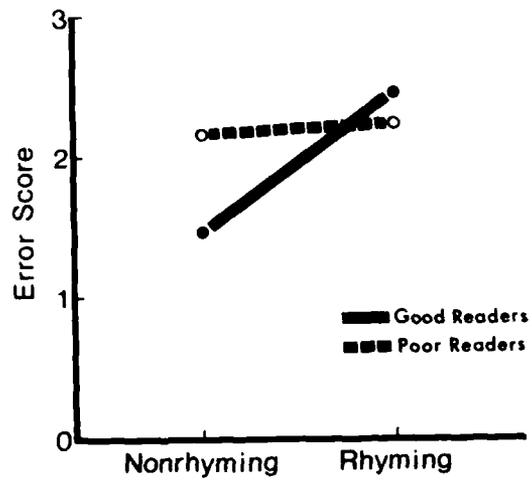
Figure 2. Mean error scores of good and poor readers on recall of word strings, in nonrhyming and rhyming conditions. (Maximum = 5.)
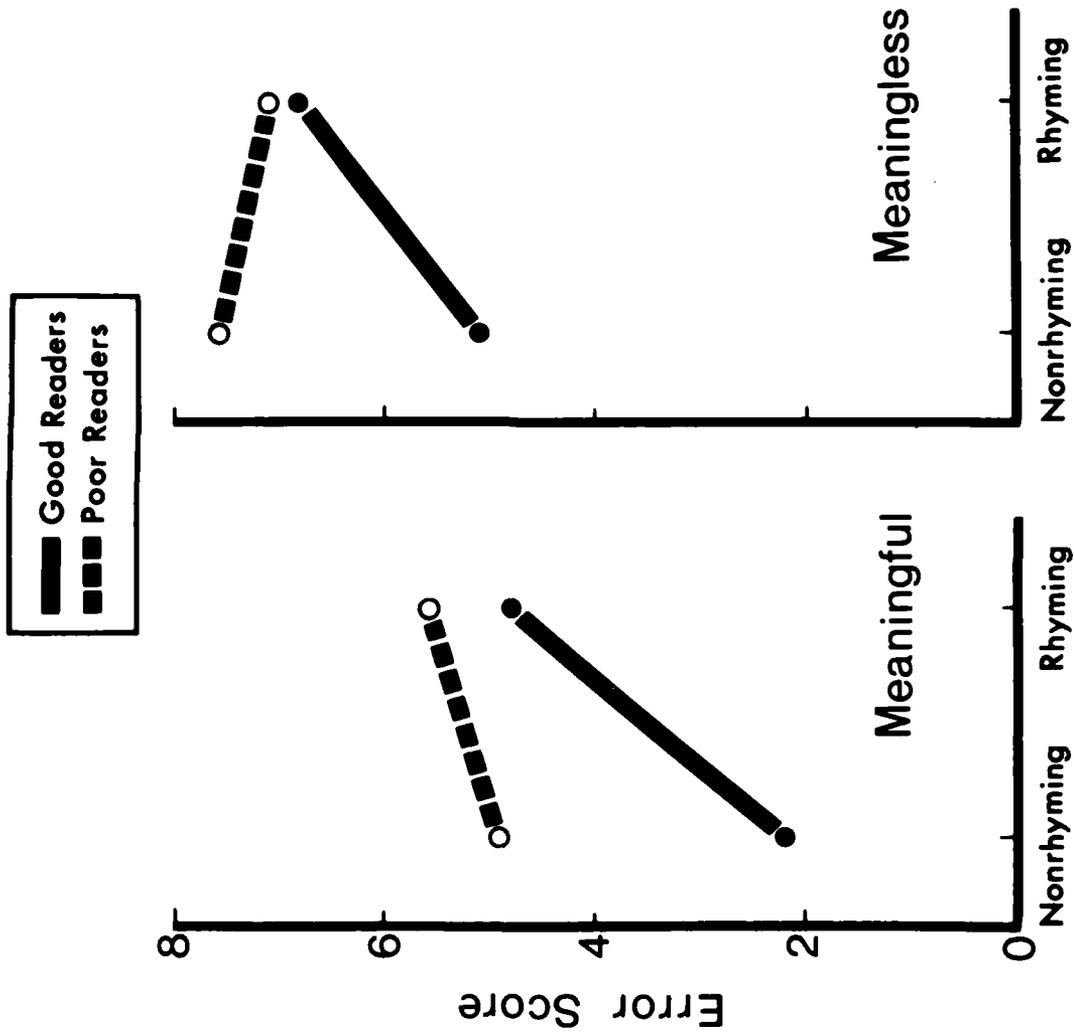
Figure 3. Mean error scores of good and poor readers on recall of meaningful and meaningless sentences in nonrhyming and rhyming conditions. (Maximum = 13.)

132

TABLE 1

Mean error scores of good, average and poor readers on memory tasks:
A longitudinal study (IQ determined in kindergarten, reading
achievement in first grade).

READING ABILITY	GRADE LEVEL	VERBAL MEMORY Max = 32		NONVERBAL MEMORY Max = 32	SYLLABLE SEGMENTATION TASK
		Nonrhyming Word Strings	Rhyming Word Strings	Corsi Blocks	(Percent passed in Kdgn.)
GOOD READERS N=26 IQ 114.7	KDGN	8.1	13.4	8.4	85%
	1st GRADE	5.5	12.1	8.7	
AVERAGE READERS N=19 IQ 114.7	KDGN	12.8	15.4	9.0	56%
	1st GRADE	9.2	11.3	8.1	
POOR READERS N=17 IQ 115.5	KDGN	13.2	15.0	10.1	24%
	1st GRADE	13.7	12.7	10.1	

133

either related or unrelated phonetically. Here they found that when the semantic mode was not available, the poor reader will use phonetic coding, but less well than the good reader.

It appears from all these findings that the difference between good and poor readers in recall of linguistic material will turn on their ability to use a phonetic representation, whether derived from print or speech. We see that, especially in the beginner, failure to establish a phonetic representation properly may be a cause as well as a correlate of poor reading. Moreover, the evidence thus far from the studies of phonetic coding in short-term memory certainly suggests that we may be dealing with a very general strategy used by the child in handling language, whatever its source.

To summarize our view, both linguistic awareness and phonetic coding in short-term memory are requirements for skilled reading, both appear to be deficient in the retarded reader, and both share the common trait that they require linguistic strategies for success.

SEX DIFFERENCES AND LINGUISTIC STRATEGIES

Given that good readers tend to use a linguistic strategy in both reading and listening whereas poor readers tend not to do so, the question we can now ask is whether girls and boys can be distinguished in this regard. We have not carried out any research ourselves to address this question directly, but for the purposes of this conference we recomputed by sex some of our longitudinal data on the linguistic performances of kindergarteners and first graders (Mann et al., Note 2). As expected, the nonverbal Corsi block test did not differentiate between good and poor readers. It also did not differentiate between boys and girls. Thus both samples were relatively well-matched in respect to general nonlinguistic memory. What we did find, however, was the usual strong interaction between reading ability and our linguistic measures, but no interaction between sex and the linguistic measures. As can be seen in Table 2, children who were good readers at the end of the first grade, whether boys or girls, tended to be strongly affected by rhyme in their memory performance. Thus, good readers, whether they were boys or girls, were apparently using phonetic strategies.

What about the poor readers? It is apparent from Table 2 that the children who were the poor readers at the end of first grade also performed similarly; whether they were boys or girls again made no difference. However, the performance of the poor readers was sharply different from that of the good readers: the poor readers, as usual, were hardly affected by rhyme at all.

Moreover, one sees from Table 3 that the same pattern of performance had obtained when all these children were kindergarteners. The future good readers, whether boys or girls, were affected by rhyme. They also could segment syllabically. In contrast, the future poor readers, whether boys or girls, were not affected by rhyme and could not segment syllabically. But none of the groups were differentiated in nonlinguistic memory.

134

TABLE 2

Mean error scores of first-grade good and poor readers
separated according to sex.

READING ABILITY \ SEX	VERBAL MEMORY Max = 32		NONVERBAL MEMORY Max = 32
	Nonrhyming Word Strings	Rhyming Word Strings	Corsi Blocks
GOOD READERS GIRLS N = 16	6.13	12.19	8.44
BOYS N = 10	4.36	12.00	8.50
POOR READERS GIRLS N = 6	15.33	14.50	10.67
BOYS N = 11	12.82	12.55	8.82

TABLE 3

Mean error scores of kindergarteners, separated according to sex and reading ability as first-graders (IQ determined in kindergarten).

READING ABILITY / SEX	VERBAL MEMORY Max = 32		NONVERBAL MEMORY Max = 32	SYLLABLE SEGMENTATION	
	Nonrhyming Word Strings	Rhyming Word Strings	Corsi Blocks	Raw Score	Percent Passed
GOOD READERS GIRLS N=16 IQ 113.5	9.44	13.81	8.13	12.69	88%
BOYS N=10 IQ 115.6	8.00	12.80	8.8	10.30	80%
POOR READERS GIRLS N=6 IQ 113.0	15.0	15.67	11.5	23.67	17%
BOYS N=11 IQ 115.6	12.18	15.55	9.27	20.82	27%

These findings will need to be replicated, of course, with experiments specifically addressed to this question of sex differences in reading, but these data certainly would make it seem as if differences in linguistic strategies, and not sex as such, will determine which children will have problems in reading.

It should be remarked at this point that a sex difference did appear in these data. That is, the poor readers in our sample tended more often to be boys, as is usually the case in clinic and school populations, while the good readers more often tended to be girls. We would interpret this to mean that at ages five and six, which is when the testing was done, more girls than boys have developed these basic abilities needed for reading. If the claim is that girls tend to mature earlier than boys, then it may be that girls develop more sophisticated linguistic strategies earlier than boys (Waber, 1977).

At all events, it is apparent that we need more information about the developmental progression of the various strategies available for dealing with language. We saw earlier that poor readers leap toward a semantic strategy in dealing with language when that option is available to them and turn to linguistic strategies only when other options are limited and, even then, do so reluctantly and inefficiently (Byrne & Shea, 1979). The semantic strategy in dealing with language is also typical of some kinds of aphasia, according to the interesting work of investigators at the Boston VA Hospital. Broca's aphasics apparently rely heavily on the content words for apprehending the meaning of sentences rather than dealing with the internal structure of the language, whether phonologic or syntactic (Caramazza & Zurif, 1976).

Nonlinguistic strategies appear also to be typical of younger children. Conrad (1972) found that in tasks stressing the short-term memory, younger children--those under six--appeared to be using nonphonetic strategies to hold information in memory. In contrast, children over six increasingly relied on a phonetic strategy. In fact, the older children preferred the phonetic strategy, just as adults do, even when it had a penalizing effect on their performance, as when they had to remember items that were phonetically confusing.

Thus we may say that the linguistic strategy as used by the good readers is a more mature strategy, akin to that used by normal adults, whereas the semantic strategy resorted to by poor readers is regressive, or at least less mature, and may be more akin to aphasic performance.

One may ask then whether the poor readers, regardless of sex, are constitutionally deficient in the abilities needed to grasp the formal or structural aspects of language, much as some aphasics are, or whether they are simply more immature and slower in developing these abilities. And in either case we may ask whether instruction will make a difference. And what kind of instruction would be most efficacious.

More research is needed in all these areas of concern before definitive answers can be given. We simply do not know whether the differences we find reflect a constitutional deficiency or a developmental lag or varying degrees of either or both. Until definitive answers are available, however, we must do the best we can.

137

Before presenting our suggestions for reading instruction and remediation, we should like to describe briefly three procedures in widespread use that appear to us to be misguided, and some of our reasons for believing them to be misguided. The first is a remedial procedure that makes the unfounded assumption that the difficulties of poor readers can typically be traced to deficits that are visual or motor in nature, presumably because the printed word is visually apprehended (Kephart, 1971; Lerner, 1971). This procedure ignores the fact that what the alphabetic writing system transcribes are the phonological segments of the spoken language and that what the child has to master are strategies for recovering the linguistic structure of the word from its encipherment in print. Moreover, there is abundant evidence that the problem of most poor readers is not in visual discrimination, visual sequencing, or visual-motor coordination but in the cognitive-linguistic sphere. So, remediation that concentrates on such tasks as visual matching of geometric figures, copying of beadstring patterns, visual-tracking and pursuit movements, and balance-beam walking is at best a waste of time if the goal is the improvement of reading skill. Such procedures may improve the child's ability to identify enemy aircraft, to follow the flight pattern of birds, or to ride a bicycle, but they will not improve his reading. One can point out, for instance, that even if the child's problem in reading really had to do with his eye movements, the visual treatment involving visual tracking and visual pursuit exercises could not help him. The eye movements in reading are well-known to be not tracking or pursuit movements at all, but rather saccadic movements or rapid jumps from fixation to fixation. The reading is done during the fixation, not during the saccadic jump. What is processed during the fixation and where the eye moves next is largely governed by cognitive and linguistic considerations (Rayner & McConkie, 1976), not optical considerations.

So much for the first misguided procedure. The second misguided procedure is of more recent vintage and was originally designed for developmental reading instruction, but has lately been recommended for remedial reading as well. Its originators call it the psycholinguistic guessing game (Goodman, 1969). In our view, this is an egregious misnomer because, far from encouraging the reader to use a linguistic approach, it encourages the child to try to bypass the linguistic structure of the word, and to go from the print directly to meaning. That is, the child is encouraged to rely heavily on guessing from the shape and context in lieu of using decoding skills. This procedure simply reinforces the same inefficient strategies that the poor reader already uses much to his disadvantage. We know from the extensive research of Perfetti and his associates (Perfetti, Goldman, & Hogaboam, 1979; Perfetti & Roth, in press) that it is the poor reader who relies most on context, not the skilled reader. Moreover, the poor reader uses context much less efficiently. We ourselves have shown (Shankweiler & Liberman, 1972) that a child's ability to read connected discourse is highly correlated not with guessing but with his ability to read individual words. In short, the skilled reader can read the individual words and uses guessing from context only when he must. Thus guessing can be useful on occasion when a word is difficult to decipher, but should not be the cornerstone of reading instruction and certainly not in the early stages of reading instruction or in the remediation of most reading disorders. So much for the so-called psycholinguistic guessing game approach.

138

The third procedure we consider to be misguided combines some aspects of the other two. That is, it treats the written word if it were a logogram, and encourages the child to rely on paired associate memory to relate the printed word with a particular spoken word and without regard to its internal segmental structure. This is the whole-word or look-say method. A corollary procedure draws the child's attention to the visual configuration of the word in terms of ascenders and descenders, or in relation to other special visual features ("remember this shape, it has a tail") and its associated meaning ("the one with the tail means monkey").

Having very briefly described what should not be done, we must now outline our own approach.

READING INSTRUCTION AND REMEDIATION

First, we should emphasize that our concern is with children who find it difficult to learn to read in an alphabetic writing system. We know that other orthographies are much easier for anyone to acquire at the outset. Take logographies, for example, the writing systems in which each character represents a word, instead of a letter as ours does. A recent study at the University of Connecticut (House, Hanley, & Magid, 1980) has shown that it is possible to teach retardates with a mental age of five or even less, who had never learned to read, to identify 200 or more pseudologograms and then to read off strings of the logograms correctly. They simply teach the retardates to pair a character with a word and to memorize the association between the two.

Very simple, very easy. In such an instructional procedure, a semantic strategy is all that is required for lexical access and no analysis below the level of the word is required.

Should we therefore use this as a model for instruction and remediation? Many educators today would say so. They would recommend that we forget about language analysis and encourage our children to treat alphabetically written words as if they were logograms. That is, they would, as we have said, teach the children to identify whole words by means of their shapes and other visual characteristics without regard to their linguistic components. The children would thus acquire a collection of word identifications by means of paired-association memory. Then, in reading connected text, the children would identify, as best they can, the words they have memorized, filling in the rest by guessing from context, again as best they can.

This kind of approach has been suggested as being especially appropriate for reading-disabled boys whose problem is said to be related to their particular cognitive style. Their cognitive style is said to be characterized by a tendency to apprehend stimuli as wholes, using a so-called right-hemisphere strategy, while girls are said to be more analytic in their cognitive style, using instead a left-hemisphere strategy. For this reason, the suggestion has been made that it might be desirable to teach boys by the whole-word method and girls by a more analytic method.

We need hardly point out two possible problems with this line of thinking. The first is that the boys' deficiency in analysis seems to be confined to linguistic matters and does not appear in the nonlinguistic tasks in which they apparently actually excel (see, for example, Symmes & Rapoport, 1972 on the dyslexic boys' excellence in block design). Thus the source of the boys' difficulties is not analysis as such, but rather _linguistic_ analysis. And the second problem is that it is precisely the whole-word, linguistic-analysis-be-damned approach that has been in widespread use in beginning reading programs over these many decades during which we have been amassing the frightening legions of reading-disabled boys in our schools. It certainly did not help them then and will not, in our opinion, help them now.

We would thus strongly disagree with the educators who in increasing numbers are suggesting that we ignore the alphabetic principle in teaching our children to read and that we concentrate instead on "reading for meaning," as they put it. It is true that some children, whether boys or girls, will learn to read even though the teaching method used initially by-passes the phonological structure of the word. The children achieve success in spite of the efforts of the reading establishment to keep the alphabetic principle a mystery to them, because the children themselves notice the relationships between how the words are written and how they are pronounced. The children themselves, in effect, discover and use the alphabetic principle on their own. We see this as testimony to the excellent native linguistic ability of those children, not to the method of instruction. There are, of course, wide individual differences in this trait as in any other.

We do not concede that because some children can pick up the principles of the orthography on their own, reading instruction should ignore this incredibly versatile and efficient symbol system. There will be too many children who will not make the discovery leap on their own, whether because of constitutional deficiency or maturational lag in linguistic abilities or whatever. Whether boys or girls, their strategies will be inefficient and hopeless. "That's one of the words with a tail, isn't it? Is it _baby_? _Funny_ was another one of those words with a tail, but that wouldn't make any sense. Oh, there's a dollar sign further down on the page. Maybe the word is _money_." The nonlinguistic whole-word method will provide would-be readers only with an ever-fading collection of words they recognize dimly, if at all, while they resort to incredibly inefficient visual or semantic strategies that prevent them from unlocking the alphabetic cipher and really learning to read.

If understanding of the phonological structure is desirable, as we believe, then the next question is whether it can indeed be taught to children who, for whatever reason, have not yet developed the knack. The Belgian research that we reported on above certainly suggests that reading instruction itself can be effective in the development of language analysis skills, at least at the first grade level. You will recall that their first graders who had been taught to read by a method emphasizing language analysis were strikingly better at phoneme segmentation tasks than children taught to read by the whole-word method. We can also report that teachers with whom we have worked over the years have all found that for most reading disabled children, prior training in the development of language analysis skills before formal reading instruction began was not only possible, but also extremely helpful in bringing about more successful reading in children previously resistant to

140

reading instruction. The Wallachs' study of inner city poor readers (Wallach & Wallach, 1980) and Isabel Beck's work with elementary school children (Beck & Mitroff, 1972) are two investigations that come to mind as providing more direct evidence of this in carefully devised research. Like them, we would attempt to meet the challenge of the alphabetic system by means of direct instruction and not leave it to chance discovery by the child.

The direct instruction of which we speak need not, as we implied earlier, be the letter-by-letter [də-ɔ-gə] "blend it, say it faster" procedure that has given phonics instruction such a bad name, though that might be better than no phonics instruction at all. There are many alternative ways of teaching children about the internal phonological structure of the word and how it relates to the orthography. These are limited only by the ingenuity and understanding of the teacher.1

REFERENCE NOTES

1. Treiman, R. A. Children's ability to segment speech into syllables and phonemes as related to their reading ability. Unpublished manuscript, Department of Psychology, Yale University, 1976.

2. Mara, V. A., Liberman, I. Y., & Shankweiler, D. A longitudinal study of some cognitive antecedents of reading proficiency. Manuscript in preparation, 1981.

REFERENCES

Alegria, J., Pignot, E., & Morais, J. Phonetic analysis of speech and memory codes in beginning readers. Journal of Psycholinguistic Research, in press.

Beck, I. L., & Mitroff, D. D. The rationale and design of a primary grades reading system for an individualized classroom. Pittsburgh: Learning Research and Development Center, University of Pittsburgh, 1972.

Blachman, B. The role of selected reading-related measures in the prediction of reading achievement. Unpublished doctoral dissertation, University of Connecticut, 1980.

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

Caramazza, A., & Zurif, E. B. Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. Brain and Language, 1976, 3, 572-582.

Conrad, R. The developmental role of vocalizing in short-term memory. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 521-533.

Corsi, R. M. Human memory and the medial temporal region of the brain. Unpublished doctoral dissertation, McGill University, 1972.

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

Fowler, C., Liberman, I. Y., & Shankweiler, D. On interpreting the error pattern of the beginning reader. Language and Speech, 1977, 20, 162-173.

Fowler, C. A., Shankweiler, D., & Liberman, I. Y. Apprehending spelling patterns for vowels: A developmental study. Language and Speech, 1979, 22, 243-252.

Goodman, K. Reading: A psycholinguistic guessing game. In K. S. Goodman & J. Fleming (Eds.), Selected papers from the IRA Preconvention Institute, Boston, April, 1968. Newark, Del.: International Reading Assoc., 1969.

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

House, B. J., Hanley, M. J., & Magid, D. F. Logographic reading by TMR adults. American Journal of Mental Deficiency, 1980, 85, 161-170.

Kephart, N. C. The slow learner in the classroom (2nd ed.). Columbus: Merrill, 1971.

Kleiman, G. M. Speech recoding in reading. Jou:nal of Verbal Learning and Verbal Behavior, 1975, 14, 323-339.

LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing. Cognitive Psychology, 1976, 6, 293-323.

Lerner, J. W. Children with learning disabilities (2nd ed.). Boston: Houghton-Mifflin, 1971, 136-169.

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

Liberman, A. M., Mattingly, I. G., & Turvey, M. Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), Coding processes and human memory. Washington, D.C.: Winston, 1972.

Liberman, I. Y. Basic research in speech and lateralization of language: Some implications for reading disability. Bulletin of the Orton Society, 1971, 21, 7-87.

Liberman, I. Y. Segmentation of the spoken word. Bulletin of the Orton Society, 1973, 23, 65-77.

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

Liberman, I. Y., & Shankweiler, D. Speech, the alphabet, and teaching to read. In L. Resnick & P. Weaver (Eds.), Theory and practice of early reading. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.

Liberman, I. Y., Shankweiler, D., Blachman, B., Camp, L., & Werfelman, M. Steps toward 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. Harris Sloan (Eds.), Auditory processing and language: Clinical and research perspectives. New York: Grune & Stratton, 1980.

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

Locke, J. L. Phonemic effects on the silent reading of hearing and deaf children. Cognition, 1978, 6, 175-187.

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

142

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

Perfetti, C. A., Goldman, S. R., & Hogaboam, T. W. Reading skill and the identification of words in discourse context. *Memory & Cognition*, 1979, 7, 273-282.

Perfetti, C. A., & Roth, S. Some of the interactive processes in reading and their role in reading skill. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading*. Hillsdale, N.J.: Lawrence Erlbaum Associates, in press.

Rayner, K., & McConkie, G. W. What guides a reader's eye movements? *Vision Research*, 1976, 16, 829-837.

Shankweiler, D., & Liberman, I. Y. 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, Mass.: MIT Press, 1972.

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

Symmes, J. S., & Rapoport, M. D. Unexpected reading failure. *American Journal of Orthopsychiatry*, 1972, 42, 82-91.

Waber, D. P. Sex differences in mental abilities, hemispheric lateralization, and rate of physical growth of adolescence. *Developmental Psychology*, 1977, 13, 29-38.

Wallach, M. A., & Wallach, L. Helping disadvantaged children learn to read by teaching them phoneme identification skills. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Zifcak, M. *Phonological awareness and reading acquisition in first grade children*. Unpublished doctoral dissertation, Department of Educational Psychology, University of Connecticut, 1977.

FOOTNOTES

[1] In a recent paper, we have set forth in greater detail some general guidelines for reading instruction and remediation (Liberman, Shankweiler, Blachman, Camp, & Werfelman, 1980).

WHEN A WORD IS NOT THE SUM OF ITS LETTERS:
FINGERSPELLING AND SPELLING*

Vicki L. Hanson

Abstract. In an experiment examining reading of fingerspelling, deaf signers of American Sign Language were asked to view finger- spelled words and nonwords. They then wrote the letters of the item just presented and made a judgment as to whether the item was a word or nonword. There was a large difference in ability to report the letters of words and nonwords. The letters of words tended to be accurately reported, while the letters of nonwords were much less accurately reported. Results indicated that these deaf subjects did not read fingerspelled words as individual letters. Rather, sub- jects made use of the underlying structure of words. Misspellings of words in this task and from free writing of deaf adults demonstrated a productive knowledge of English word structure, with striking similarities in error pattern being found from these two sources.

INTRODUCTION

Fingerspelling is a manual communication system in which there is a manual sign for each letter of the alphabet. Words are spelled out in this system. Fingerspelling is an important part of American Sign Language (ASL) as well as an integral part of manual systems based on English. As such, it is important to understand how fingerspelled words are processed by skilled users of the system. For this reason, an experiment was designed to examine the following questions: How are fingerspelled words read? Is reading words a letter-by-letter process of recognition? That is, is it necessary to

---

145

identify each letter of the word? Or, rather, when reading words is there recognition of letter groupings? And what kinds of errors are made when reading fingerspelling?

METHOD

Sixty fingerspelled items were presented, one at a time. Thirty were real words ranging in length from five to thirteen letters. Mean length was 8.3 letters per word. The following words were used: ADVERTISEMENT, AWKWARD-LY, BANKRUPTCY, BAPTIZE, CADILLAC, CAREFUL, CHIMNEY, COMMUNICATE, ELABORATE, FUNERAL, GRADUATE, HELICOPTER, HEMISPHERE, INTERRUPT, MOUNTAIN, PANTOMIME, PHILADELPHIA, PHYSICS, PREGNANT, PSYCHOLOGICAL, PUMPKIN, RHYTHM, SUBMARINE, SURGERY, THIRD, TOMATO, UMBRELLA, VEHICLE, VIDEO, VINEGAR. These thirty words were matched for average length with 30 nonwords. Twenty of these matched nonwords were pseudowords. Pseudowords were pronounceable, but they do not happen to be English words. The following pseudowords were used: BRANDIGAN, CADERMELTON, CHIGGETH, COSMERTRAN, EAGLUMATE, FREZNIK, FRUMHENSER, HANNERBAD, INVENCHIP, MUNGRATS, PHALTERNOPE, PILTERN, PINCKMOR, PRECKUM, RAPAS, SNERGLIN, STILCHUNING, SWITZEL, VALETOR, VISTARMS. The other ten nonwords were not possible English words. These orthographically impossible words were not pronounceable. The impossible words were as follows: CONKZMER, ENGKSTERN, FTERNAPS, HSPERACH, PGANTERLH, PIGTLANING, PKANT, RANGKPES, RICGH, VETMFTERN.

Stimulus words were recorded on videotape by a native ASL signer. Items were fingerspelled at a natural ASL rate of 354 letters per minute (see Bornstein, 1965). While words were fingerspelled at a slightly faster rate than nonwords, this difference in rate between words (mean rate of 369 letters per minute) and nonwords (mean rate of 339 letters per minute) was not statistically significant, $t_{(58)}=1.87$, $p > .05$. Real words, pseudowords, and impossible words were mixed throughout the list with each item followed by a 10 second blank interval to be used as a response period. Subjects were instructed that they would see many fingerspelled items and that for every item they were to do two things: First, write the letters they had just seen, and second, make a judgment as to whether that item was a word or nonword. The instructions, signed in ASL by the same person who fingerspelled the stimuli, were recorded on videotape.

Subjects were 17 congenitally deaf adults recruited through New York University and California State University, Northridge. Fifteen were native signers of ASL. The other two had learned ASL at age five and were considered by native signers to be fluent in ASL. There were eight men and nine women ranging in age from 17-53 years, mean age 31 years.

RESULTS AND DISCUSSION

Responses were analyzed for accuracy of letter report and correctness of word judgments. Shown in the first line of Table 1 are percentages of subjects' correct responses in the three conditions. These were trials on which both the letter report and word judgment decisions were correct. As can readily be seen, there were large performance differences for words, pseudo-words, and impossible words.

146

Table 1

Mean percentage of items correct in the three conditions.

	Words	Pseudowords	Impossible words
Total correct responses	61.0%	25.0%	11.2%
Correct word judgments	92.9%	83.5%	82.9%
Correct spelling following correct word judgment	62.9%	28.1%	12.9%

Response Accuracy

There are two possible sources of error in this experiment: recognition and letter report. It is possible that subjects recognized all the letters of an item correctly but later were unable to report the letters. Bearing on this issue, it is important to take note of the fact that subjects were accurate at making decisions as to whether a fingerspelled item was a word or nonword. As shown in Table 1, when words were presented, subjects correctly indicated that item was a word on more than 90% of the trials. The analysis of accuracy across conditions indicated, however, that accuracy was not constant across all stimulus types, $F(2,32)=3.84$, $p<.05$. Although word judgments were made more accurately for words than for nonwords (Newman-Keuls, $p<.05$), most likely indicating an expectancy for words, there was no difference in ability to respond that pseudowords were nonwords and ability to respond that impossible words were nonwords. If subjects were making decisions based simply on whether the fingerspelled nonwords were consistent with English orthography, there should have been more of a tendency to respond that pseudowords were English words than to respond that impossible words were English words. This was clearly not the case. It is reasonable to assume, therefore, that subjects generally recognized the words correctly when they responded that an item was a word, and to assume that they responded that an item was not a word when there was no recognition of an English word.

But while subjects were accurate at this word judgment task, they were not as accurate at letter re   ` a word was correctly recognized as an English word, what was the ʀᵣᵤᵤability that the word would be correctly spelled? As shown in the bottom line of Table 1, subjects correctly spelled 62.9% of the words following a correct word judgment. The fact that there were errors in letter report indicates that it is possible to recognize a word from its letters but not be able to use this knowledge productively to spell the words. Several times the experimenter noticed that when a fingerspelled word was presented, a subject produced the sign for the word, indicating that he or she recognized the word, but then was unable to spell the word. 147

In contrast to the accuracy in letter report for words following a correct word judgment, if pseudowords or impossible words were correctly identified as nonwords, accuracy of letter report was poor: 28.1% for pseudowords and 12.9% for impossible words. This difference in ability to report the letters of words, pseudowords, and impossible words is significant, $F(2,32)=82.59$, $p<.001$, with post hoc analysis revealing that letter report for words was significantly more accurate than letter report for nonwords (Newman-Keuls, $p<.01$). There was thus a word familiarity effect in this fingerspelling task. In addition, signers were more accurate at letter report for pseudowords than at letter report for impossible words (Newman-Keuls, $p<.01$). This greater accuracy for pseudowords than impossible words, consistent with effects in recognition of printed pseudowords and impossible words reported by Gibson, Shurcliff, and Yonas (1970) indicates that signers were able to make use of orthographic structure to read and remember letters of a new finger-spelled item.

The difference in ability to receive and report the different types of items suggests that much different processes are involved in reporting the different items. It suggests that subjects use orthographic structure to read and remember letters of words and pseudowords, while impossible words might have to be read on a letter-by-letter basis. Whether or not fingerspelled items are processed simply on a letter-by-letter basis can be ascertained by determining whether there is independence of letter report. To do this, words are scored for letter accuracy regardless of position. The probability of correctly reporting all of the letters in a word or nonword is compared with the probability of correctly reporting individual letters of the items. Independence of letter processing is indicated if the following equation holds:

$$p(\text{all letters of an item}) = p(\text{individual letters})^n$$

where n=number of letters in the word. Tests of letter independence were performed separately on words, pseudowords, and impossible words.

Analyzing probability (all letters vs. individual letters) by item length, it was found that for words and pseudowords the probability of correctly reporting all the letters of a word was greater than the probability of reporting the letters independently: for words, $F(1,16)=67.74$, $p<.001$; for pseudowords, $F(1,16)=27.82$, $p<.001$. This nonindependence of letter processing for these items indicates that words and pseudowords were not processed as individual letters. Rather, processing of a given letter was influenced by other letters of the item. This result is consistent with the idea that orthographic structure influenced recognition for words and pseudowords.

For impossible words, however, the probability of correctly reporting all the letters of an item was <u>not</u> greater than the probability of independently reporting each letter, $F(1,\overline{16})=1.82$, $p>.05$. Thus, for impossible words the letters were processed independently. These impossible words were not processed as groups of letters, but rather as letter strings. The reduced accuracy of letter report for impossible words in comparison to words and pseudowords indicates that subjects were not good at remembering fingerspelled items as unrelated letter strings.

148

The analyses above, therefore, indicate that subjects were more accurate at reporting words than pseudowords and were more accurate at reporting pseudowords than impossible words. This was due to differences in processing. While impossible words were processed as individual letters, letters of words and pseudowords were not processed independently. This nonindependence of letter processing suggests that the processings of these items are sensitive to orthographic structure. The word familiarity effect indicates additional processing benefits for actual English words.

Error Analysis

Incorrect responses were next subjected to an analysis of error type. Several determinations were made for each of the incorrectly reported words. First, were the written responses consistent with English orthography? Second, did the misspelling of a word preserve the pronunciation of the word presented, thus resulting in a phonetically accurate spelling? And third, what types of spelling errors were made?[1]

Orthography. It is clear that subjects were aware of the orthographi structure of English words. As shown in Table 2, for more than 70% of tl words and pseudowords the incorrect responses were consistent with Englir orthography. For impossible words, 60% of the incorrect responses w: ~e thus consistent, resulting in pronounceable letter strings. In fact, t e most frequent incorrect responses for impossible words were changes of this type. For example: FTERNAPS>ferntaps, PKANT>plant, VETMFTERN>vetfern, RICGK>rich, and RANGKPES>rangkes. These incorrect responses indicate a productive knowl-edge of English word structure.

Table 2

Classification of errors for the incorrect responses.

	Words	Pseudowords	Impossible words
Errors consistent with English orthography	76.8%	71.9%	60.4%
Phonetic misspellings	16.5%	(3.4%)	----

Phonetic misspellings. Did the misspellings of the English words preserve the pronunciation of the words presented? The majority did not. Errors that are pronunciation preserving may be called phonetic misspellings. Examples of common phonetic misspellings for hearing people are analisis (for analysis), bankrupcy (for bankruptcy), catagory (for category), and vidio (for video) (Masters, 1927; Sears, 1969). As shown in Table 2, only about 16% of the incorrect spellings for the English words in this experiment were

---

Table 3

Examples of incorrect responses in fingerspelling experiment. Word judgments were correct for all incorrect responses listed. Numbers in parentheses indicate duplicate responses.

Stimulus word	Deletions	Transpositions	Substitutions	Additions
ADVERTISEMENT	adverisement	adveristement		
BANKRUPTCY			bankrupacy	bankruptucy (2)
BAPTIZE		bapitze (3)		
CHIMNEY			chimmey	
FUNERAL		funreal	fuderal	
GRADUATE				grauduate
HEMISPHERE		hemipshere		
INTERRUPT	interupt			
PHILADELPHIA	Philadephia		Philalelphia	
SURGERY		surgrey (2)	surgury	
THIRD			thyrd	
UMBRELLA	umbella	umberlla		
VEHICLE	vehile	vechile (4)		
VIDEO	vido	viedo		
VINEGAR	vingar (3)	vineagr	vinigar	vineagar
BRANDIGAN		brandagin	brandigin	
CHIGGETH	chigeth (3)		chiggets	
COSMERTRAN		comsertran		
FREZNIK		frezink (3)		
HANNERBAD				hannerband (2)
MUNGRATS			mungrate (2)	
PILTERN				pilltern
RAPAS	raps (2)			
SWITZEL	swizel (2)	swiztel		
VALETOR				valentor
ENGKSTERN	engstern (4)			
FTERNAPS		ferntaps (2)		afternaps
RANGKPES	rangkes (2)	rangkeps		
RICGH	righ (3)			
VETMFTERN	vetfern (2)			

---

phonetic. Thus, while the misspellings were consistent with English orthography, for any given word the misspelling was not consistent with the pronunciation of that word.

Since by definition impossible words were not pronounceable, it was not possible to have pronunciation-preserving misspellings of the impossible words. Phonetic misspellings of the pseudowords are theoretically possible, but inspection of Table 2 reveals that pronunciation-preservings misspellings of these words were rare.

Types of errors. The types of errors in the incorrect responses were analyzed. The following categories were used for error classification: Letter deletions, additions, substitutions, and transpositions. Letter transpositions were incorrect orderings of the letters of an item. An error was counted as a substitution when an incorrect letter was written. Letter deletions and additions are self-explanatory. Examples of each of these error types are shown in Table 3.

In decreasing order of occurrence, the following kinds of errors were found in the present misspellings: letter deletions, transpositions, substitutions and additions. Percentages of occurrence for each kind of error are shown in Table 4. Notice that the occurrence for the different types of errors is similar for words and nonwords.

It is interesting to take notice of the error analysis for pseudowords. Since these items are possible English words, their analysis suggests the kind of errors people may make when learning a new word from fingerspelling. So, the kinds of errors to be expected in learning new words from fingerspelling would be predominantly letter deletions with letter transpositions and substitutions also fairly common.

Table 4

Percentage of each type of error for the incorrect responses examined in the analysis of error type.

	Words	Pseudowords	Impossible Words
Deletions	36.6%	34.7%	38.0%
Transpositions	31.4%	29.0%	23.9%
Substitutions	20.9%	24.5%	29.2%
Additions	10.9%	11.6%	8.8%

For each of the substitutions, a determination was made as to whether this was a substitution of a letter of similar handshape. This determination was based on the visual confusions of handshapes reported by Lane, Boyes-Braem, and Bellugi (1976). Since not all letters of the manual alphabet were included in that study of handshapes, it was necessary to extrapolate from their results for the present analysis. For example, in their study with moving signs the compact handshapes A, E, and O were found to be confusing. For purposes of the present analysis, the handshapes M, N, S, and T were included as compact handshapes that could be possible substitutions based on fingerspelling. Another fingerspelling substitution based on their study was the pair I and Y. The pair K and P were also counted as possible substitutions based on misreading of fingerspelling.

Using this system, it was found that many of the letter substitutions for words and pseudowords could be accounted for as misreading of fingerspelling based on handshape. The following are the percentages of substitution errors that may have been based on misreading of fingerspelling: 80.9% for words, 72.4% for pseudowords, 15.8% for impossible words. There is no apparent reason, however, why misreading of fingerspelled letters should be more common for words than for, say, impossible words. This pattern of substitution error therefore suggests a second alternative as to the basis for the substitutions. It is possible that substitutions were based on English word constraints. Inspection of the letters involved in the above analysis reveals that the analysis is confounded with vowel/vowel confusions and consonant/consonant confusions. In fact, analysis of the substitution errors revealed that subjects tended to substitute a vowel for a vowel or substitute a consonant for a consonant. This was true for 87.5% of the substitutions for words, for 69.0% of the substitutions for pseudowords, and for 68.4% of the substitutions for impossible words. Due to the confounding inherent in the letters examined here, it is not possible to state with certainty the basis for the substitution errors, although the error pattern is suggestive of the idea that letter substitutions were based on substitutions of a phonologically possible letter.

Error position. The position of the first error in each of the misspellings was also calculated. To make error position independent of word length, position was calculated as a proportion of the total word length. Mean position of first errors was as follows: words=.598, pseudowords=.602, impossible words=.538. Thus, the majority of incorrect responses did not occur until the second half of the word. Subjects were good at knowing the letters in the first half of the words with problems generally developing in the middle of the word. This finding is consistent with work showing that initial and final letters of fingerspelled words are identified better than medial letters (see Caccamise, Hatfield, & Brewer, 1978) and may be related to the fact that initial and final letters are held longer than medial letters (Reich, 1974).

Summary. In summary, analysis of the incorrect responses indicates that there were similar errors for words and nonwords. The majority of incorrect responses were found to be consistent with English orthography. The incorrect responses did not tend to preserve the pronunciation of the intended words. The errors tended to be letter deletions, transpositions, and substitutions occurring in the second half of the word.

Spelling

Spelling requires the ability to make productive use of English orthography. Hearing people tend to spell according to the pronunciation of words as evidenced in the frequency of phonetic misspellings they produce (Fischer, 1980; Masters, 1927; Simon & Simon, 1973). But reliance on pronunciation alone can lead to errors in spelling for a language with a complex orthography such as English. Simon and Simon (1973) have estimated that strict reliance on pronunciation will generate correct spellings for only about 50% of the words in English.

Deaf persons may not rely primarily, if at all, on word pronunciations when spelling. Hoemann, Andrews, Florian, Hoemann, and Jansema (1976) tested deaf children in a recognition test for spelling of common objects and found that no more than 19% of the errors for any age group were phonetic misspellings. In contrast, up to 83% of the misspellings made by hearing children in the same task were phonetic (Mendenhall, 1930). These results suggest that deaf children are not primarily relying on word pronunciations when spelling.[2]

To generate hypotheses as to the spelling processes used by deaf persons whose primary language is ASL, misspellings from the writing of deaf adults were collected. These misspellings, shown in Table 5, bear a striking resemblance to the spelling errors in the fingerspelling experiment. As in that experiment, the vast majority of misspellings are consistent with English orthography.

As in the results of Hoemann et al. (1976), the majority of errors did not preserve the pronunciation of the intended word. For these deaf persons, then, there does not seem to be reliance on word pronunciation when spelling. What process could be used? Inspection of error type may be of help in answering this question. Hoemann et al. found the most common type of spelling error to be letter deletions (42%), a finding that is consistent with the errors collected here from adults. Notice that this is also the most frequent type of misspelling in the fingerspelling experiment.

The pattern of errors for hearing and deaf persons is clearly different. For hearing persons, phonetic substitutions dominate the errors made (Fischer, 1980; Mendenhall, 1930). For deaf adults, the misspellings found in writing and the errors in the fingerspelling experiment were predominantly non-phonetic letter deletions. Also striking is that often in the misspellings of deaf persons all the correct letters for a word were found to be present, but the order of the letters was in error. As shown in Table 5, these transpositions occur not only within a syllable, but also across syllable boundaries, rendering misspellings that definitely are not phonetic. Again, this is consistent with the results of the fingerspelling experiment where transpositions were more common than even letter substitutions.

It would be too strong a statement to conclude from these observations that reliance on fingerspelling led to these misspellings found in free writing. These results, however, provide a basis for interesting speculation and further study.

Table 5

Examples of misspellings found in writing.

	Word spelled	Word intended
Letter deletions	bapist	baptist
	elborate	elaborate
	pinic	picnic
	psylogical	psychological
	stiring	stirring
Letter transpostions		
Within a syllable	thristy	thirsty
	umberlla	umbrella
Across syllable boundaries	bankcrupty	bankruptcy
	contuine	continue
Letter substitutions	chocalate	chocolate
	butch	dutch
	licinse	license
	mosquoto	mosquito
Letter additions	cancell	cancel
	frence	fence
	grazed	gazed
	preferre	prefer

REFERENCES

Bornstein, H. Reading the manual alphabet. Washington, D.C.: Gallaudet College Press, 1965.

Caccamise, F., Hatfield, N., & Brewer, N. Manual/simultaneous communication (M/SC) research: Results and implications. American Annals of the Deaf, 1978, 123, 803-823.

Cromer, R. F. Spontaneous spelling by language-disordered children. In U. Frith (Ed.), Cognitive processes in spelling. London: Academic Press, 1980.

Fischer, F. W. Spelling proficiency and sensitivity to linguistic structure. Unpublished doctoral dissertation, University of Connecticut, 1980.

Gibson, E. J., Shurcliff, A., & Yonas, A. Utilization of spelling patterns by deaf and hearing subjects. In H. Levin & J. P. Williams (Eds.), Basic studies in reading. New York:Basic Books, Inc., 1970.

Hoemann, H. W., Andrews, C. E., Florian, V. A., Hoemann, S. A., & Jansema, C. J. The spelling proficiency of deaf children. American Annals of the Deaf, 1976, 121, 489-493.

Lane, H., Boyes-Braem, P., & Bellugi, U. Preliminaries to a distinctive feature analysis of American Sign Language. Cognitive Psychology, 1976, 8, 263-289.

Masters, H. V. A study of spelling errors. Iowa City: University of Iowa Studies in Education, IV (4), 1st Series 138, September 1, 1927.

Mendenhall, J. E. An analysis of spelling errors: A study of factors associated with word difficulty. New York: Bureau of Publications, Teachers College, Columbia University, 1930.

Reich, P. A. In A. Makkai & V. B. Makkai, The first LACUS forum. Columbia, S.C.: Hornbeam Press, 1974, 348-356.

Sears, D. A. Engineers spell acoustically. College Compositions and Communication, 1969, 20, 349-351.

Simon, D. P., & Simon H. A. Alternative uses of phonemic information in spelling. Review of Educational Research, 1973, 43, 115-137.

Zakia, R. D., & Haber, R. N. Sequential letter and word recognition in deaf and hearing subjects. Perception & Psychophysics, 1971, 9, 110-114.

FOOTNOTES

[1]Not all incorrect responses could be classified in this way. Subjects' responses were often just a word judgment followed by a dash or the first letter or two of the stimulus item. If subjects failed to write at least 50% of the word, the word was not scored in the analysis of error type. In addition, there were responses that were so different from the target word that the origin of the error could not be determined. Combining these two sources, the following percentages of errors could not be counted in the analysis of error type: 16.0% for words, 45.4% for pseudowords, and 48.6% for impossible words.

[2]Cromer (1980) analyzed misspellings in the free writing of six orally educated deaf children in England (median age 10.5). By his analysis 67.5% of the misspellings could be described as phonetic. But it should be remembered that the strong oral tradition in England may have led to the phonetic misspellings he found.

A 'DYNAMIC PATTERN' PERSPECTIVE ON THE CONTROL AND COORDINATION OF MOVEMENT*

J. A. Scott Kelso,+ Betty Tuller,++ and Katherine S. Harris+++

1. INTRODUCTION

That speech is the most highly developed motor skill possessed by all of us is a truism, but how is this truism to be understood? Although the investigation of speech production and motor behavior have proceeded largely independently of each other, they are alike in sharing certain conceptions of how skilled movements are organized. Thus, regardless of whether one refers to movement in general or speech as a particular instance, it is assumed that for coordination to occur, appropriate sets of muscles must be activated in proper relationships to others, and correct amounts of facilitation and inhibition have to be delivered to specified muscles. That the production of even the most simple movement involves a multiplicity of neuromuscular events overlapping in time has suggested the need for some type of organizing principle. By far the most favored candidates have been the closed-loop servomechanism accounts provided by cybernetics and its allied disciplines, and the formal machine metaphor of central programs. The evidence for these rival views seems to undergo continuous updating (e.g., Adams, 1977; Keele, 1980) and so will not be of major concern to us here. It is sufficient to point out the current consensus on the issue: namely, that complex sequences of movement may be carried out in the absence of peripheral feedback, but that feedback can be used for monitoring small errors as well as to facilitate corrections in the program itself (e.g., Keele, 1980; Miles & Evarts, 1979).

But at a deeper level, none of these models offers a principled account of the coordination and control of movement. The arguments for this position have been laid out in detail elsewhere (Fowler, Rubin, Remez, & Turvey, 1980; Kelso, Holt, Kugler, & Turvey, 1980; Kugler, Kelso, & Turvey, 1980; Turvey, Shaw, & Mace, 1978) and will be elaborated here only inasmuch as they allow us to promote an alternative. To start, let us note that programs and the like-- though intuitively appealing--are only semantic descriptions of systemic behavior. They are, in Emmett's (1980) terms "externalist" in nature and are quite neutral to the structure or design characteristics of that which is being controlled. By assuming, a priori, the reality of a program account we

[HASKINS LABORATORIES: Status Report on Speech Research SR-65 (1981)]

impose from the outside a descriptive explanation that allows us to interpret motor behavior as rational and coherent. But it would be a categorical error to attribute to the concept _program_ causal status. Nevertheless, it is commonplace in the analysis of movement for·investigators to observe some characteristic of an animal's performance, such as the extent of limb movement, and conclude that the same characteristic is represented in the motor program (e.g., Taub, 1976). In like vein, the observation that lip rounding precedes the acoustic onset of a rounded vowel and therefore coarticulates with preceding consonants is explained by the presence of the feature [+ rounding] in the plan for a speech gesture (cf. Fowler, 1977). Such an interpretative strategy is akin to the observer of bee behavior who attributes the product of a behavior--honey arranged in hexagonal form--to a 'hexagon' program possessed by all bees. A more careful analysis would reveal that hexagonal tesselation or 'close packing' occurs whenever spherical bodies of uniform size and flexible walls are packed together. That is to say, 'close packing' is a consequence of dynamic principles that allow for the minimization of potential energy (least surface contact) and it is dynamics that determines the emergence of hexagonal patterns such as honeycombs (for further examples of complex form arising from dynamic principles, see D'Arcy Thompson, 1942; Kugler et al., 1980; Stevens, 1974).

The gist of the message here is that if we adopt a formal machine account of systemic behavior, we take out, in Dennett's (1978, p. 15) words, a "loan on intelligence," which must ultimately be paid back. Rather than focusing our level of explanation at an _order_ grain of analysis in which all the details of movement must be prescribed (see Shaw & Turvey, in press), a more patient approach may be to seek an understanding of the relations among systemic states as necessary _a posteriori_ facts of coordinated activity (see Rashevsky, 1960; Shaw, Turvey, & Mace, in press). In essence we would argue as Greene (Note 1) does that in order to learn about the functions of the motor system, we should first seek to identify the informational units of coordination.

Although the latter topic--coordination--has received some lip service in the motor control literature, a rigorous analysis of muscle collectives has (with few exceptions) not been undertaken as a serious scientific enterprise. We venture to guess that one of the reasons for such a state of affairs is that extant models of movement control (and skill learning) assume that the system is already coordinated. Thus, servomechanism accounts speak to the positioning of limbs or articulators in terms of, for example, some reference level or spatial target, but are mute as to how a set of muscles might attain the desired reference or target. Similarly, program descriptions of motor behavior assume that the program represents a coordinated movement sequence and that muscles simply carry out a set of commands (e.g., Keele, 1980; Schmidt, 1975). Any systemic organization of the muscles themselves is owing to the program--a _fait accompli_ that explains nothing.

But what does an adequate theory of movement coordination (and skilled behavior as well) have to account for? Fundamentally, the problem confronting any theorist of systemic behavior in living organisms is how a system regulates its internal degrees of freedom (cf. Bernstein, 1967; Boylls, 1975; Greene, 1972; Iberall & McCulloch, 1969; Tsetlin, 1973; Turvey, 1977; Weiss, 1941). A first step toward resolving this issue in motor systems is to

158

claim--following the insights of the Soviet school (e.g., Bernstein, 1967; Gelfand, Gurfinkel, Tsetlin, & Shik, 1971; Tsetlin, 1973)--that individual variables, say muscles, are partitioned into collectives or synergies where the variables within a collective change relatedly and autonomously. Combinations of movements are produced by changes in the mode of interaction of lower centers; higher centers of the nervous system do not command, rather they tune or adjust the interactions at lower levels (cf. Fowler, 1977; Greene, 1972, Note 1; Kelso & Tuller, in press; Tsetlin, 1973; Turvey, 1977). As Gelfand et al. (1971) suggest, learning a new skill (within the foregoing style of organization) consists of acquiring a convenient synergy, thus lowering the number of parameters requiring independent control (cf. Fowler & Turvey, 1978, for a skill learning perspective and Kugler, Kelso, & Turvey, in press, for a developmental analysis). Before going any further, we should note that the term "synergy" is used here in a way that is different from Western usage: A synergy (or coordinative structure, as we prefer to call it) is not limited to a set of muscles having similar actions at a joint, nor is it restricted to inborn reflex-based neurophysiological mechanisms (cf. Easton, 1972). Rather, synergies and coordinative structures connote the use of muscle groups in a behavioral situation: they are functional groupings of muscles, often spanning several joints that are constrained to act as a single unit. To paraphrase Boylls (1975), they are collections of muscles, all of which share a common pool of afferent and/or efferent information, that are deployed as a unit in a motor task.

In this paper we do not propose to continue the polemic for a coordinative structure style of organization. The evidence for coordinative structures in a large variety of activities is well documented (e.g., for speech, see Fowler, 1980; for locomotion, see Boylls, 1975; for postural balance, see Nashner, 1977; for human interlimb coordination, see Kelso, Southard, & Goodman, 1979a, 1979b) and the rationale for such an organizational style is compelling, though perhaps not accepted by all. Instead we want to focus first on the following question: When groups of muscles function as a single unit, what properties (kinematic and electromyographic) do they exhibit? We intend to show that there are certain features of neuromuscular organization that are common to many, if not all, modes of coordination including human speech. Second, and more important, we shall attempt to provide a principled rationale for why coordinative structures have the properties that they have. Such an account will not be in the algorithmic language of formal machines, where each aspect of the movement plan is explicitly represented. Rather we shall develop the argument based on dynamic principles that have their groundings in homeokinetic physics (cf. Iberall, 1977; Kugler et al., 1980; Yates & Iberall, 1973) and dissipative structure (dynamic pattern) theory (Katchalsky, Rowland, & Blumenthal, 1974; Prigogine & Nicolis, 1971)--that real systems (as opposed to formal machines) consist of ensembles of coupled and mutually entrained oscillators and that coordination is a natural consequence of this organization.

Although in previous work coordinative structures have been linked to dissipative structures (Kelso, Holt, Kugler, & Turvey, 1980; Kugler et al., 1980; see also Kugler et al., in press), here we shall prefer Katchalsky's term "dynamic pattern" (cf. Katchalsky et al., 1974). Traditionally, the word "structure" has referred only to static __spatial__ patterns that are at or near thermodynamic equilibrium. In contrast, the term "dissipative structure"

applies also to the temporal domain and refers to open nonequilibrium systems that require energy to maintain spatio-temporal patterns. Thus the term dynamic pattern is preferred not only because it removes the ambiguity between classical notions of the term structure and Prigogine's dissipative structures, but also because it captures the flavor of what is, in effect, a functional or dynamic organization. We are persuaded of the importance of dynamic patterns because they provide an accurate description of the appearance of qualitative change, or emergent properties, that cannot be understood with reference to quantitatively known component processes.

According to Katchalsky et al. (1974; see also Yates, 1980; Yates & Iberall, 1973) there are three essential ingredients for a system to display dynamic patterns. First, there should be a sufficiently large density of interacting elements or degrees of freedom. Second, the interactions should be non-linear in nature; and finally, free energy should be dissipated. As we shall see, the "stuff" of the motor system--synergies or coordinative structures--consists of precisely these ingredients.

The continuous dissipation and transformation of energy results in a fundamental property of living systems--cyclicity--and motivates the physical theory that complex systems are ensembles of non-linear, limit-cycle oscillators (homeokinetics; e.g., Iberall & McCulloch, 1969; Soodak & Iberall, 1978). This claim necessarily suggests that coordinated movement will be subject to particular kinds of constraints whose form we will attempt to elucidate shortly. But it is to the general issue of constraints that we first turn.

2. COORDINATIVE STRUCTURES AS CONSTRAINTS

As Mattingly (1980) points out in his review of Gödel, Escher, Bach: An Eternal Golden Braid (Hofstadter, 1979), it has long been recognized by linguistic theoreticians that a formal theory of grammar that allows an unrestricted use of recursive devices would be simply too powerful. Such a theory would permit the grammars that occur in natural languages, as well as an infinite number of grammars that bear no relation whatsoever to natural languages. Thus the claim that programs can be developed to model the human mind is vacuous: without incorporating constraints one program may be as good as any other, and neither may have anything to do with how real biological systems work.

In a similar vein, current theories of motor control fail to embody the concept of constraint: they do not capture the distinction between those acts that occur and those that are physically possible but never will occur. The motor program notion, for example, is a description of an act--specified in terms of the contractions of muscles--that is too powerful because it can describe acts that could never be performed by an actor. Theoretically, the motor program is as viable for unorganized convulsions as it is for coordinated movement (cf. Fowler, 1977). Boylls (1975) expresses an identical view of servomechanistic models. The concept of coordinative structure (in his terms, muscle linkages) "...by no means represents a conventional engineering approach to the control of motor performance, because the brain is not viewed as having the capacity to transfer an existing state of the musculature into

160

any other arbitrary state, however biomechanically sound. Most such unconstrained states would have no behavioral utility. Hence the linkage paradigm...naturally assumes that evolution has economized the motor system's task through constraints restricting its operation to the domain of behaviorally useful muscle deployments" (p. 168). If the proper unit of analysis for the motor system is indeed the coordinative structure, then the difference between coordinated and uncoordinated movement--between control and dyscontrol--is defined by what acts are actually performed, since the coordinative structure by definition is functional in nature.

We should clarify what we mean by "functional" here, for some may view it as a buzz word that glosses over underlying mechanisms. This would be a misunderstanding, for as Fentress (1976) has taken pains to point out, mechanism itself is a _functional_ concept and can only be considered in relative terms. Thus what constitutes a mechanism at one level of analysis becomes a system of interrelated subcomponents at a more refined level of analysis.[1] Questions pertaining to mechanisms (e.g., are coordinative structures mechanisms?) are only applicable when the context for the existence of a particular mechanism is precisely defined (cf. Kelso & Tuller, in press). This brings us to an important point: coordinative structures are functional units in the sense that the individual degrees of freedom constituting them are constrained by particular behavioral goals or effectivities (cf. Turvey & Shaw, 1979). Sharing the same degrees of freedom without reference to the effectivity engaged in by an actor would not constitute a functional unit.

Nowhere is this claim (insight?) more apparent than in modern ethological research where there is growing recognition that nervous systems are organized with respect to the relations among components rather than to the individual components themselves (cf. Bateson & Hinde, 1976; Rashevsky, 1960). Thus, in seeking to understand the nature of behavior, some ethologists consider it more appropriate to look for generalities across dimensions that are physically distinct but normally occur together (e.g., pecking and kicking during fights) rather than across dimensions that share the same physical form (e.g., pecking for food and pecking in fights [cf. Fentress, Note 2]). In our attempts to relate divergent levels of organization in biological systems (see below) we do well to keep the "functional unit" perspective to the forefront, for such units may well have been the focus of natural selection. Moreover, the implications for the acquisition of skill and motor learning are apparent. For example, if one were to ask whether speaking is a complex act, one answer is that it is complex for the child who is learning to speak but simple for the adult who has already acquired the necessary coordination to produce the sounds of the language. In the sense that the degrees of freedom of the speech apparatus are subject to particular constraints in the adult speaker (which it is our role to discover), then there is reason to believe that his/her neuromuscular organization is actually _simpler_ than that of the child _for the same act_ (cf. Yates, 1978, on complexity). Similarly, it is quite possible that so-called complex tasks that fit existing constraints may be much more easily acquired than the "simple" tasks we ask subjects to perform in a laboratory. We turn now to consider just exactly what form such constraints appear to take.

3. PROPERTIES OF COORDINATIVE STRUCTURES. LOCAL RELATIONS

If, as Gurfinkel, Kots, Paltsev, and Fel'dman (1971) argue, there are many different synergies or coordinative structures, then the key problem for a science of movement is to detect them and to define the context in which they are naturally realized. What should we be looking for and how should we be looking? If the constraint perspective is correct, then we may well expect to see--in any given activity--a constancy in the relations among components of a coordinative structure even though the metrical values of individual components may vary widely. For example, the temporal patterning of muscle activities may be fixed independent of changes in the absolute magnitude of activity in each muscle. Similarly, the temporal patterning of kinematic events may be fixed independent of changes in the absolute magnitude or velocity of individual movements.

One obvious strategy for uncovering relations among components is to change the metrical value of an activity (e.g., by increasing the speed of the action). In this fashion, we can observe which variables are modified and which variables, or relations among variables, remain unchanged. Notice that if one searches for canonical forms of an activity, then changing metrical properties obscures the basic form by altering properties of individual components that would otherwise remain stable. For example, in the study of speech, changes in speaking rate and syllable stress pose major problems for researchers looking for invariant acoustic definitions of phonemes. Alternatively, these changes may provide the major ways that invariance can be observed; some aspects of phonemes must change and other aspects must remain the same in order to preserve phonemic identity over changes in speaking rate and stress.

The properties of coordinative structures have been more fully articulated in a number of recent papers (Fowler, 1977; Kelso et al., 1980; Kugler et al., 1980; Turvey et al., 1978). Here we shall only present a small inventory of activities that reveal those properties. We shall try to show--at macroscopic and microscopic levels of behavior--that certain relations among variables are maintained over changes in others. In addition, a primary goal will be to extend this analysis, in a modest way, to the production of speech and beyond that to the intrinsic relations that hold across the systems for speaking, moving, and seeing.

Electromyographic investigations of locomotion illustrate the properties of coordinative structures discussed briefly above. For example, in freely locomoting cats (Engberg & Lundberg, 1969), cockroaches (Pearson, 1976), and humans (Herman, Wirta, Bampton, & Finley, 1976), increases in the speed of locomotion result from increases in the absolute magnitude of activity during a specific phase of the step cycle (see Grillner, 1975; Shik & Orlovskii, 1976), but the timing of periods of muscle activity remains fixed relative to the step cycle. In keeping with the notion of coordinative structures, the temporal patterning of muscle activities among linked muscles remains fixed over changes in the absolute magnitude of activity in individual muscles.

The literature on motor control of mastication offers an abundance of data understandable within a constraint perspective. For example, Luschei and Goodwin (1974) recorded unilaterally from four muscles that raise the mandible

162

in the monkey. The cessation of activity in all four muscles was relatively synchronous whether the monkey was chewing on the side ipsilateral or contralateral to the recorded side. In contrast, the amplitude of activity in each muscle was very sensitive to the side of chewing. In other words, the timing of activity periods of the four muscles remained fixed over large changes in amplitude of the individual muscle activities.

Similar timing relations have been reported in human jaw raising muscles. Møller (1974) observed that the timing of activity in the medial pterygoid and anterior temporalis muscles relative to each other remains unchanged during natural chewing of an apple, although the individual chews are of varying durations and amplitudes; the muscles acting synergistically to raise the jaw generally show fixed temporal patterns of activity over substantial changes in the magnitude of activity. Thexton's (1976) work suggests that this constancy of temporal relations holds for antagonistic muscle groups as well. Specifically, the timing of activity in the muscles that lower and raise the jaw is not sensitive to changes in consistency of the chewed food, although the amplitudes of activity in the muscles that raise the jaw decrease markedly as the food bolus softens.

The two activities discussed, locomotion and mastication, are easily described as fundamental patterns of events that recur over time. The observed pattern is not strictly stereotypic because it is modifiable in response to environmental changes, such as bumps in the terrain or changes in consistency of the food. This style of coordination--in which temporal relationships are preserved over metrical changes--may also hold for activities that are less obviously rhythmic and whose fundamental pattern is not immediately apparent. Examinations of kinematic aspects of two such activities, handwriting and typewriting, reveal these properties of coordinative structures.

At first blush, the control of handwriting does not appear to be in terms of a fundamental motor pattern that recurs over time. The linguistic constraints are considered primary, precluding the possibility of regularly occurring motor events. However, when individuals are asked to vary writing speed without varying movement amplitude, the relative timing of certain movements does not change with speed (Viviani & Terzuolo, 1980). Specifically, the tangential velocity records resulting from different writing speeds reveal that overall duration changed markedly across speeds. But when the individual velocity records are adjusted to approximate the average duration, the resulting pattern is invariant. In other words, major features of writing a given word occur at a fixed time relative to the total duration taken to write the word. The same timing relationships are preserved over changes in magnitude of movements, over different muscle groups, and over different environmental (frictional) conditions (cf. Denier van der Gon & Thuring, 1965; Hollerbach, 1980; Wing, 1978).

The control of typewriting, like handwriting, does not appear to be in terms of a fundamental motor pattern that recurs over time. But Terzuolo and Viviani (1979) looked for possible timing patterns in the motor output of professional typists and found that for any given word, the set of ratios between the times of occurrence of successive key-presses remained invariant over changes in the absolute time taken to type the word. When weights were

attached to the fingers, the temporal pattern of key-presses (the set of time ratios) was unaffected, although the time necessary to type the words often increased. Thus, temporal relationships among kinematic aspects of typewriting appear to be tightly constrained, although the time necessary to accomplish individual keystrokes may change.

A synergistic or coordinative structure style of organization appears to hold over diverse motor acts. The question remains as to whether this view can be applied to the production of speech. Specifically, do temporal relationships among some aspects of articulation remain fixed over metrical changes in the individual variables? Two obvious sources of metrical change in speech that have been extensively investigated are variations in syllable stress and speaking rate. If the view of systemic organization that we have elaborated here holds for speech production, we would expect to see a constancy in the temporal relationships among articulatory components (muscle activities or kinematic properties) over stress and rate variations. Allow us first to step back and examine briefly a general conception of how changes in stress and rate are accomplished.

Many current theories of speech motor control share the assumption that changes in speaking rate and syllable stress are independent of the motor commands for segmental (phonetic) units. Articulatory control over changes in speaking rate and syllable stress is considered as "...the consequence of a timing pattern imposed on a group of (invariant) phoneme commands" (Shaffer, 1976, p. 387). Lindblom (1963), for example, suggests that each phoneme has an invariant "program" that is unaffected by changes in syllable stress or speaking rate (tempo). Coarticulation results from the temporal overlap of execution of successive programs.[2] Thus, when a vowel coarticulates with a following consonant, it is because the consonant program begins before the vowel program is finished (see also Kozhevnikov & Chistovich, 1965; Stevens & House, 1963). According to these views, when speaking rate increases or stress decreases, the command for a new segment arrives at the articulators before the preceding segment is fully realized. The articulation of the first segment is interrupted, resulting in the articulatory undershoot and temporal shortening characteristic of both unstressed syllables and fast speaking rates. This scheme predicts that the relative temporal alignment of control signals for successive segments, and their kinematic realizations, will change as stress and speaking rate vary, a prediction contrary to the constancy in temporal relationships observed in locomotion, mastication, handwriting, and typewriting.

There exists electromyographic evidence, albeit quite limited, that the coordinative structure style of organization may hold for speech production, that is, that temporal relationships among aspects of intersegmental articulation remain constant over changes in stress and speaking rate. Experiments by Tuller, Harris, and Kelso (1981) and Tuller, Kelso, and Harris (1981) explored this question directly, by examining possible temporal constraints over muscle activities when stress and speaking rate vary. The five muscles sampled are known to be associated with lip, tongue, and jaw movements during speech.

When speakers were asked to increase their rate of speech, or decrease syllable stress, the acoustic duration of their utterances decreased as expected. The magnitude and duration of activity in individual muscles also

164

changed markedly. However, the relative timing of muscle activity was preserved over changes in both speaking rate and syllable stress. Specifically, the relative timing of consonant activity and activity for the flanking vowels remained fixed over suprasegmental change.

The preservation of relative timing of muscle activities is illustrated in Figure 1, which is essentially a 2 x 2 matrix of stress and rate conditions for the utterance /papip/. Each muscle trace represents the average of twelve tokens produced by one subject. Arrows indicate the onsets of activity for /a/ (anterior belly of digastric), /p/ (orbicularis oris inferior), and /i/ (genioglossus). Onset values, defined as the time when the relevant muscle activity increased to 10% of its range of activity, were determined from a numerical listing of the mean amplitude of each EMG signal, in microvolts, during successive 5 msec intervals.

As apparent from the figure, the onset of consonant-related activity occurred at an invariant time relative to the interval from onset of the first vowel to onset of the second vowel. That is, the following ratio remained fixed over suprasegmental changes in stress and rate:

$$\frac{V_1 \text{ to } C}{V_1 \text{ to } V_2} = k$$

where V_1 = onset of activity for production of the first vowel,
 C = onset of activity for production of the medial consonant,
 V_2 = onset of activity for production of the second vowel.

Activity for consonant articulation began at a constant phase position relative to the activity for the flanking vowels. This preservation of relative timing of consonant- and vowel-related muscle activity was observed for all utterances and muscle combinations sampled, and was independent of the large variations in magnitude and duration of individual muscle activity (for details see Tuller, Kelso, & Harris, 1981). These data fit the primary characteristic of coordinative structures outlined above; namely, there is a constancy in the relative temporal patterning of components, in this case muscle activities, independent of metrical changes in the duration or absolute magnitude of activity in each muscle.

In the brief review of locomotion, mastication, handwriting, and typewriting, we noted that these activities show temporal constraints at either an electromyographic or a kinematic level, constraints that fit a coordinative structure style of organization. Activities such as speech, handwriting, and typewriting, usually described as less stereotypic or repetitive than locomotion or mastication, can also be described within a synergistic or coordinative structure style of control (see also Kelso, Southard, & Goodman, 1979a, 1979b). In the next section we will attempt to extend this type of analysis to the relations that hold across different structural subsystems, such as the systems for speaking, moving and seeing.
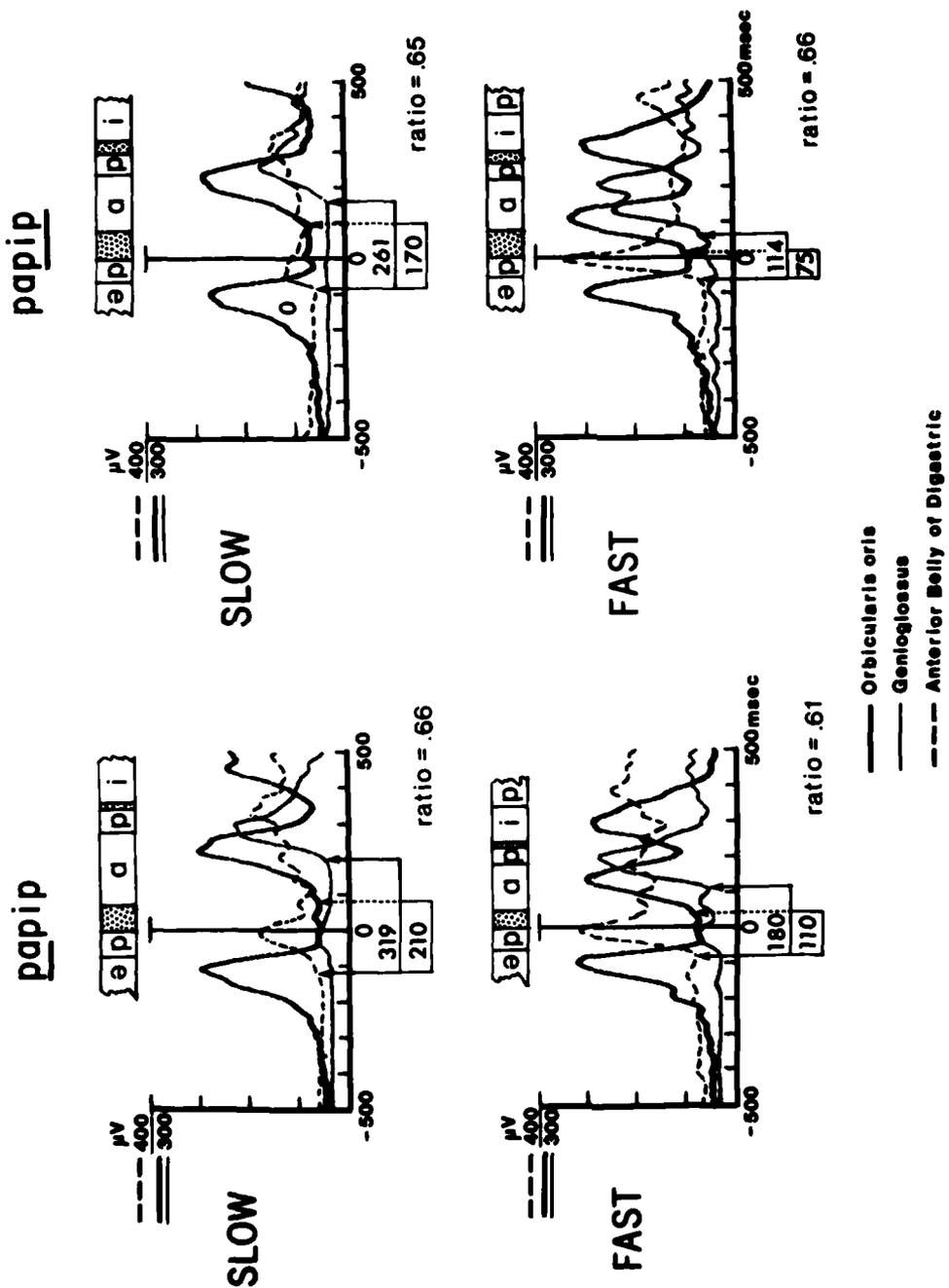
Figure 1. The utterance /papip/ spoken by one subject at two rates and with two stress patterns. Each muscle trace represents the average of twelve repetitions of the utterance. Arrows indicate onsets of activity for anterior belly of digastric (jaw lowering for /a/; the dotted line), orbicularis oris (lip movement for /p/; the thick line), and genioglossus (tongue fronting for /i/; the thin line). The ratio of the latency of consonant-related activity relative to the vowel-to-vowel period is indicated for each stress and rate condition.

4. PROPERTIES OF COORDINATIVE STRUCTURES: GLOBAL RELATIONS

The inventory presented above offers a view of motor systems that Gelfand and Tsetlin (1971) refer to as well-organized. Thus the working parameters of the system appear to fall into two distinct groups: essential parameters that determine the form of the function (also called the structural prescription, cf. Boylls, 1975; Kelso et al., 1979a, 1979b; Grimm & Nashner, 1978; Turvey et al., 1978) and nonessential parameters that lead to marked changes in the values of the function but leave its topology essentially unchanged. It is possible that a subdivision of the foregoing nature does not exist for every function; nevertheless, the distinction between essential and nonessential variables (between coordination and control, see Kugler et al., 1980) is apparent in a wide variety of activities.

As a historical note, we might remark that the distinction between variables of coordination and control is not entirely new (though there is little doubt of our failure to appreciate it). Over forty years ago von Holst (1937, English translation 1973), following his extensive studies of fish swimming behavior, hypothesized the presence of a duality between frequency and amplitude of undulatory movement (see also Webb, 1971). Invariably, amplitude of fin movement could be modulated (sometimes by as much as a factor of four) by, for example, the application of a brief pricking stimulus to the tail, without affecting frequency in any way. Von Holst (1937) concluded that this behavior may be explained as follows: "the automatic process (a central rhythm) determines the frequency, whilst the number of motor cells excited by *the process* at any one time defines--other things being equal--the amplitude of the oscillation" (pp. 88-89). There seems little doubt that neurophysiological research of the last decade has borne out von Holst's thesis--in general, if not in detail--with its discovery of numerous central rhythm generators (cf. Davis, 1976; Dellow & Lund, 1971; Grillner, 1975; Stein, 1978). We shall have much more to say about the nature of rhythmical activity in the next section; for the moment let us consider the possibility that the partitioning of variables into essential and nonessential is a basic design strategy for motor systems.

In the previous section we presented a brief inventory of activities that highlighted the nature of constraints on large numbers of muscles. Yet these activities illustrate the partitioning of variables within local collectives of muscles--muscles acting at single or homologous limbs or within a single structural subsystem. The arguments that a synergistic style of organization constitutes a design for the motor system would surely be strengthened if it could be shown that the same classification of variables into essential and nonessential holds for more than one structural subsystem. We turn then to examine a potential relationship that has intrigued numerous investigators, namely that between speaking and manual performance.

There is of course general agreement that language and speech are special functions of the left hemisphere, although there is little understanding as to why this should be so. It is beyond the scope of this paper to consider all the various hypotheses (perceptual, cognitive, etc.) that have been proposed for speech lateralization. Let us instead consider one approach to the problem stemming from the work of Kinsbourne and Hicks (1978a, 1978b; see also Kimura, 1976; Lomas & Kimura, 1976). Basically, and in brief, the argument

167

that Kinsbourne and others pursue is that language lateralization (productive and perceptual) arises as a result of the requirement for unilateral motor control of a bilaterally innervated motor apparatus (cf. Liberman, 1974). Kinsbourne and Hicks house a specific version of this notion in their well-popularized "functional cerebral space" model. They suggest that because the human operator has access to a limited amount of functional cerebral space, excitation from putative cortical control centers that are close together (e.g., for speaking and controlling the right hand) is likely to overflow and cause intrahemispheric interference. Conversely, the greater the functional distance between control centers, the less likely is contamination from one center to the other and the better is performance on simultaneous tasks. Experiments showing that right hand superiority in balancing a dowel on the index finger is lost when subjects are required to speak while doing the task (e.g., Kinsbourne & Cook, 1971; Hicks, 1975; Hicks, Provenzano, & Rybstein, 1975) all seem to support some type of functional space or intrahemispheric competition model.

These experiments also motivate a view of cerebral function in which speaking is considered dominant over the manual task. Unfortunately, the dependent measures employed--dowel balancing or number of taps on a key--do not allow us to examine possible interactions with speaking (e.g., whether pauses in tapping and pauses in speaking co-occur). This design deficiency is in part to blame for the focus on manual performance as it reflects intrahemispheric interference with little or no emphasis on possible comple-mentary effects on speech dynamics. Indeed, the failure to find effects on global measures of vocal performance (e.g., number of words generated in response to a target letter in 30 sec) has led some investigators to conclude that interference is a "one-way street," with "cognitive tasks having priority over motor systems" (Bowers, Heilman, Satz, & Altman, 1978, p. 555).

From our perspective it makes little sense to talk of interference, competition, and rigid dominance relations in a coordinated system. If speech and movement control systems are governed by the same organizational princi-ples, the issue for lateralization concerns the tightness of fit between these systems when control is effected by one limb or the other. Although we shall not speak to the laterality issue directly at this point, we do want to illustrate that apparent competition and interference between the subsystems for speaking and manual performance may be more correctly viewed as an effect of their mutual collaboration.

Consider the following experiment in which subjects[3] are asked to produce cyclical movements of a comfortable frequency and amplitude with their right index finger while simultaneously uttering a homogeneous string of syllables ("stock," "stock," etc.).[4] Obviously, subjects have no problem whatsoever in following these instructions. Now imagine that the subject is told to vary the stress of alternate syllables in a strong-weak manner (phonetically, /'stak, stak, 'stak, stak.../) while maintaining amplitude and frequency of finger movement constant. The waveform data for one such subject are shown in Figure 2. It is quite obvious that finger movements are modulated--in spite of instructions not to do so--such that they conform to the speech stress pattern; that is, longer finger movements accompany stressed syllables, and shorter finger movements accompany unstressed syllables. Is this the outcome of the speech system "driving," as it were, the motor system? A parallel

168

ALTERNATE STRESS OF SPEAKING

Finger Movement (Fp = 1.27Hz)

IRL

Flexion

+50°
0
-50

Speech

500msec
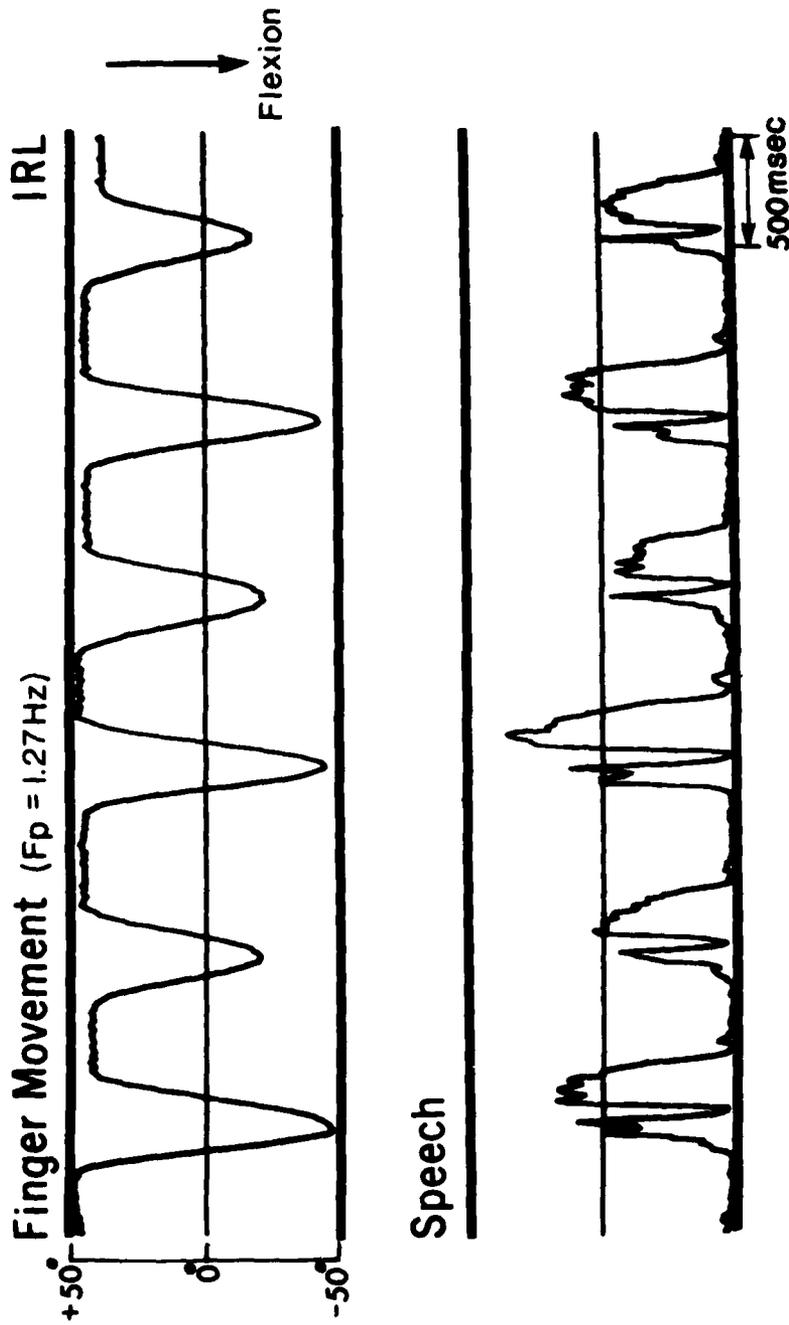
Figure 2. Simultaneous finger movement (top) and integrated speech waveform (bottom) produced by a subject when told to vary the stress of alternate syllables but maintain the amplitude and frequency of finger movements constant.

experiment in which subjects were asked to keep stress of speaking constant but to vary the extent of finger movement (i.e., alternating long and short excursions) suggests not. Often the result was that the change in amplitude of finger movement was accompanied by a change in the pattern of syllable production such that there was increased "stress"[5] with the longer finger movement. The waveform data for one such subject are shown in Figure 3.

These data speak to several issues. Of primary importance is the demonstration of <u>mutual interactions</u> among the subsystems for speaking and manual performance. Interestingly, this theme is also borne out in recent work on aphasic patients by Cicone, Wapner, Foldi, Zurif, and Gardner (1979). Speech and gesture seem to follow an identical pattern in aphasia: anterior (Broca's) aphasics seem to gesture no more fluently than they speak, and posterior (Wernicke's) aphasics (who generate much empty speech) gesture far more than normals.

But the broader impact of the present data on speaking and manual activity is not only their indication that the two activities share a common organizational basis (see also Studdert-Kennedy & Lane, 1980, for additional commonalities between spoken and signed language). Rather it is that the same design theme emerges in "coupled" systems as in "single" systems (such as those for walking, chewing, handwriting, typewriting, and speaking, reviewed in the previous section). When an individual speaks and moves at the same time, the degrees of freedom are constrained such that the system is parameterized as a total unit. The parameterization in this case, as in the case of single systems, takes the form of a distribution of force (as reflected in the mutual amplitude relations) among all the muscle groups involved.

An important property of collectives of muscles is their ability to establish and maintain an organization in the face of changes in contextual conditions. Thus Kelso and Holt (1980) show that human subjects can achieve invariant end-positions of a limb despite changes in initial conditions, unexpected perturbations applied during the movement trajectory, and both of these in the absence of awareness of limb position. The organization of limb muscles in this case appears to be qualitatively similar to a non-linear vibratory system (for more details and fur⁺ʰ ⁻ evidence see Bizzi, Dev, Morasso, & Polit, 1978; Cooke, 1980; Fel'dman, ʼ ⸴; Kelso, 1977; Kelso, Holt, & Flatt, 1980; Polit & Bizzi, 1978; ⸴ᵐ⸴ ˋ 1980; see also below). Similarly, in the well-known speech exⱼ ᵐᵉⁿᵗ ⸴ᵢ Folkins and Abbs (1975) loads applied to the jaw yielded "compensatory responses" in the lips to preserve ongoing articulation. In fact the movement of the jaw and lower lip covaried in such a way that the sum of their displacements tended to remain constant (but see Sussman, 1980, for possible methodological problems with compensation studies).

Is the preservation of such "equations-of-constraint" in the face of unexpected changes in environmental context also characteristic of coupled systems? In short the answer appears to be yes, at least if the following experiment is representative. Imagine that as an individual is synchronizing speech and cyclical finger movements (in the manner referred to earlier) a sudden and unexpected perturbation is applied to part of the system. In this case a torque load (of approximately 60 ounce-inch and 100 msec duration) is

ALTERNATE EXTENT OF FINGER MOVEMENTS

Finger Movement (Fp = 1.07 Hz)

IRL

Flexion

+45°

0

-45

Speech

500msec
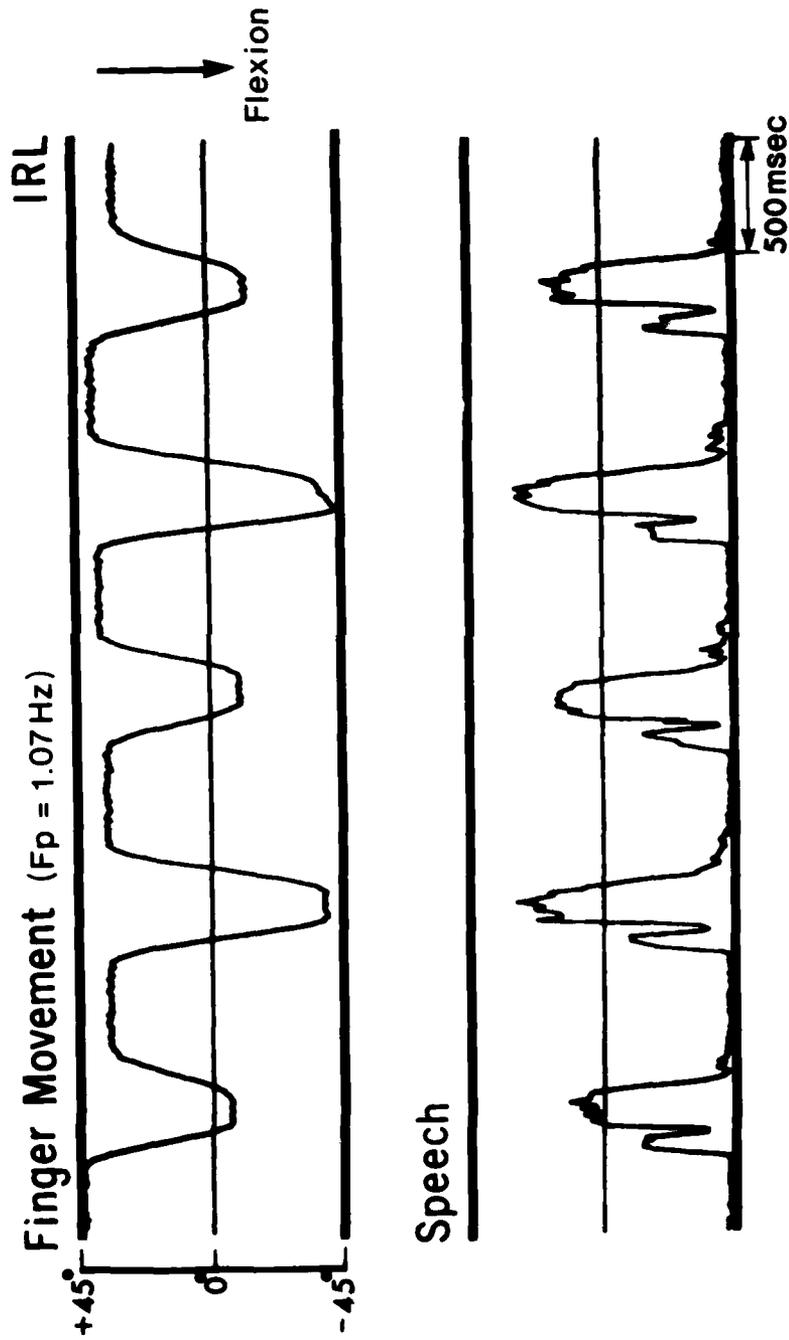
Figure 3. Simultaneous finger movement (top) and integrated speech waveform (bottom) produced by a subject when told to vary the extent of alternate finger movements but produce all syllables exactly like all other syllables.

171

added to the finger in such a way as to drive it off its preferred trajectory (see Kelso & Holt, 1980, for details of this technique). In order for the finger to return to its stable cycle, additional force must be supplied to the muscles. Qualitatively speaking, an examination of the movement waveform of Figure 4 reveals that the finger is back on track in the <u>cycle following the perturbation</u>. Of interest, however, is the speech pattern (again, the individual audio envelopes in Figure 4 correspond to the syllable /stak/ spoken at preferred stress and frequency). We see that the audio waveform is unaffected in the cycle in which the finger is perturbed: it is in the following cycle that a dramatic amplification of the waveform occurs. This result is compatible with the present thesis that systems, when coupled, share a mutual organization and that this organization may be preserved over efference (as in the stress-amplitude experiments) or afference (as in the present experiment). Thus a peripheral disturbance to one part of the system (requiring an additional output of force to overcome it) will have a correlated effect on other parts of the system to which it is functionally linked. Note that as in the previous experiments on speaking and moving, there is no support whatsoever for a one-way dominance of speech over manual performance. Were that the case, there is little reason to expect speaking to be modified in any way by finger perturbations.

Why then does the adjustment (maladjustment may be a more appropriate word) to speaking occur on the cycle <u>after</u> the perturbation? Some insight into this issue may be gleaned from a clever experiment on locomotion by Orlovskii and Shik (1965). Dogs were fitted with a force brake at the elbow joint and then were allowed to locomote freely on a treadmill. A brief application of the brake during the transfer-flexion phase not only retarded the movement of the elbow but also that of the shoulder, suggesting that both joints are constrained to act as a unit within the act of locomotion. Spinal mechanisms were implicated because the joints returned to their original velocities within 30 msec of the brake application. But of even greater interest was the <u>next</u> locomotory cycle, some 800-900 msec following the original perturbation. Here the transfer-flexion phase was delayed again as if the perturbation (along with an appropriate response) had reoccurred. Note that had the brake actually been applied, this "phantom braking response" (cf. Boylls, 1975) would have constituted an <u>adaptation</u>; indeed, this phenomenon of modifying current acts based on perturbations occurring in antecedent ones is called "next-cycle adaptation."

Although our understanding of such phenomena is still rather primitive (see Boylls, 1975, pp. 77-79 for one speculation of a neural type), the present "equations-of-constraint" perspective on coupled systems offers at least a descriptive account (see also Saltzman, 1979). From the mutual relations observed in the "stress" and "finger amplitude" experiments, we can generate the follc.ing simple constraint equation:

$$f(x,y) = k$$

where the variables x and y represent the set of muscles (subsystems) for speaking and manual activity, such that a specific change in x will be accompanied by a corresponding change in y to preserve the function, f, constant. Now imagine at time t_1 the variable y is altered via a peripheral perturbation such that a change in its value (in the form of an increase in

172

Figure 4. Simultaneous finger movement (top) and integrated speech waveform (bottom) produced during a sudden, unexpected finger perturbation. Notice the increase in amplitude of the syllable in the cycle following the perturbation (see text).

muscular force) is necessary to overcome the disturbance. As a consequence of "mechanical" constraints (e.g., neural conduction times, mechanical properties of muscles) the variable x cannot immediately adopt an appropriate value on the perturbed cycle. On the next cycle, however, the variable x takes on a complementary value as a necessary consequence of the fact that force is distributed between both systems.

Let us clarify one important aspect of this simple formulation. The interrelations observed here are not meaningfully described as "compensatory." That is, x is not incremented because it has to compensate for changes in y. The synergistic relations observed between speaking and manual activity are not based on a causal logic (because y, then x). Rather the coherency between systems is captured by an adjunctive proposition (since y is incremented, then x must necessarily also be incremented).[6] In the stress-finger amplitude experiment, x and y were simultaneously adjusted: In the perturbation experiment, as a consequence of inherent neuro-mechanical factors, x was not adjusted until the next cycle, even though y had returned to its preferred state. In both cases the basic notion is the same. That is, the complementary relations observed are a consequence of the total system functioning as a single, coherent unit.

The global relations between speaking and manual activity that we have identified above are, it seems, far from exotic, if we look for them through the right spectacles. Other systems with quite different structural designs appear to share the same style of coordination. Consider, as a final example, coordination between the eye and the hand. Imagine a situation in which the oculomotor system is partially paralyzed with curare and the subject asked to point ballistically at a target N degrees from visual center (Stevens, 1978). The typical result is that the limb overshoots the designated target—a phenomenon called "past pointing." A common explanation of this finding is that the subject estimates the movement as farther than N degrees because the intended eye movement (registered by an internal copy of the command or corollary discharge of N degrees) and the actual eye movement (N-k degrees) are discrepant. If the subject uses the mismatch information to adjust the limb movement, he will overshoot the target. But an alternative to this hypothesis is offered on the basis of a set of experiments on "past pointing" in patients with partial extra-ocular paralysis[7] (cf. Perenin, Jeannerod, & Prablanc, 1977).

While Perenin et al. argue that the mechanism leading to spatial mislocalization involves "the monitoring of the oculomotor output itself" rather than corollary discharge, we believe that their results can be explained within the present framework. We would argue that the actual amount of force required to move the partially paralyzed eye to a visual target accounts for "past pointing." Thus in a task involving the coupling of oculomotor and limb subsystems, parameterization occurs over the total, coupled system, so that the increase in force required to localize a partially paralyzed or mechanically loaded eyeball (cf. Skavenski, Haddad, & Steinman, 1972) is necessarily distributed to the system controlling the hand in a task that requires their coupled activity. There is no need to invoke a corollary discharge (Brindley, Goodwin, Kulikowski, & Leighton, 1976; Stevens, 1978) or an efference monitoring mechanism (Perenin et al., 1977); the eye-hand system is simply utilizing the design strategy that seems to work for many other activities that involve

174

large numbers of degrees of freedom. In short, the fascinating aspect of the data linking the eye, the speech apparatus, and the hand is that the relations observed apply to systems whose structural features are vastly different, just as these same coordinative structure properties apply to more 'local' collectives of muscles that share common structural elements.

5. RATIONALIZING COORDINATIVE STRUCTURES AS "DYNAMIC PATTERNS"8

We have seen in the previous sections that a ubiquitous feature of collectives of muscles is the independence of the force or power distributed into the collective and the relative timing of activities (electromyographic and kinematic) within the collective. In fact we have presented evidence suggesting that the motor system has a preferred mode of coordination: where possible, scale up on power but keep relative timing as constant as possible. The flexibility of the system is attained by adjusting the parametric values of inessential variables without altering the basic form of the function as defined by its essential variables. It remains for us now to rationalize why nature has adopted this strategy. In particular let us consider why it is that timing constraints are such a principal characteristic of coordinated movement. In fact this question could take a more general form: Why are humans inherently rhythmic animals?9 A short excursion into dynamics offers an answer to these questions in terms of physical principles. As we shall see, the physics of systems in flux defines living creatures as rhythmic; no new mechanisms need be introduced to account for the inherent rhythmicity (cf. Morowitz, 1979).

Dynamics--the physics of motion and change--has not been considered particularly appropriate for an analysis of biological systems because, until quite recently, it has dealt almost exclusively with linear conservative systems. In simple mechanical systems such as a mass-spring, the equation of motion describes a trajectory towards an equilibrium state. Thus a linear system represented by the following second order differential equation:

$$m\ddot{x} + c\dot{x} + kx = 0 \tag{1}$$

will decay in proportion to the magnitude of its viscous (frictional) term (c) and o. 'latory motion will cease. All this is predicated on the second law of ther...odynamics--time flows in the direction of entropy. Yet living systems are characterized by sustained motion and persistence; as Schroedinger (1945) first remarked, they "accumulate negentropy." Living systems are not statically stable; they maintain their form and function by virtue of their <u>dynamic stability</u>.

How might we arrive at a physical description of biological systems that does not violate thermodynamic law? Consider again the familiar mass-spring equation, but this time with a forcing function, F(t):

$$m\ddot{x} + c\dot{x} + kx = F(t) \tag{2}$$

Obviously it is not enough to supply force to the system; to guarantee persistence (and to satisfy thermodynamic principles) the forcing function

must exactly offset the energy lost in each cycle. Real systems meet this requirement by including a function--called an escapement--to overcome dissipative losses. The escapement constitutes a non-linear element that taps some source of potential energy (as long as it lasts) to compensate for local thermodynamic losses. Thus, a pulse or "squirt" of energy is released via the escapement such that, averaged over cycles, the left hand side of equation (2) equals the right hand side and sustained motion is thereby assured.

The foregoing description is of course the elementary theory of the clock (see Andranov & Chaiken, 1949; Iberall, 1975; Kugler et al., 1980; Yates & Iberall, 1973, for many more details), but it draws our attention to some fundamentally important concepts: First, stability can only be established and maintained if a system performs work; second, work is accomplished by the flow of energy from a high source of potential energy to a lower potential energy "sink"; third, stated as Morowitz's theorem, the flow of energy from a source to a sink will lead to at least one cycle in the system (Morowitz, 1979).

That cyclical phenomena abound in biological systems is hardly at issue here (see Footnote 9, the chronobiology literature [Aschoff, 1979] and also reviews by Oatley & Goodwin, 1971; Wilke, 1977). Nor is the notion--favored by investigators of movement over the years--that 'clocks,' 'metronomes' or rhythm generators may exist for purposes of timing (e.g., Keele, 1980, for recent discussion; Kozhevnikov & Chistovich, 1965; Lashley, 1951). However, we might emphasize that the many extrinsic "clock" mechanisms are not motivated by thermodynamic physical theory. The view expressed here--which can only mirror the emphatic remarks of Yates (1980)--is that cyclicity in complex systems is ubiquitous because it is an obligatory manifestation of a universal design principle for autonomous systems.

Such a foundation for complex systems leads us, therefore, away from more traditional concepts. The Barnard-Cannon principle of homeostasis, for example, which provides the framework on which modern control theory--with its reference levels, comparators, error correction mechanisms and so on--is built, is obviated by a dynamic regulation scheme in which internal states are a consequence of the interaction of thermodynamic engines (cf. Soodak & Iberall, 1978). The latter scheme, appropriately termed homeokinetic, conceives of systemic behavior as established by an ensemble of non-linear oscillators that are entrained into a coherent harmonic configuration. For homeokinetics, many degrees of freedom and the presence of active, interacting components is hardly a "curse" in Bellman's (1961) terms; rather it is a necessary attribute of complex systems.

That the constraints imposed on coordinated activity--whether it be of speech or limbs (or both)--should take the form of a dissociation between power and timing is now less mysterious within this framework than before. Coordinative structures are non-linear oscillators (of the limit cycle type, see below) whose design necessarily guarantees that the timing and duration of "squirts" of energy will be independent of their magnitude within a fixed time frame (a period of oscillation, see Kugler et al., 1980). Referring back to equation (2), the magnitude of the forcing function will be some proportion of the potential energy available, but the forcing function itself is not dependent on time (cf. Iberall, 1975; Yates & Iberall, 1973). Non-

176

conservative, non-linear oscillators are truly autonomous devices in a formal mathematical sense; time is nowhere represented in such systems (Andranov & Chaiken, 1949) and energy is provided in a "timeless" manner.

An example may be helpful at this point. It comes from a fascinating experiment by Orlovskii (1972) on mesencephalic locomotion in the cat. If one selectively stimulates the hindlimb areas of Red and Dieters nuclei in a stationary cat, the flexor and extensor synergies (corresponding to swing and stance phases, respectively) can be energized. During induced locomotion, however, continuous stimulation of one site or the other has an effect only when the respective synergies were actually involved in the step cycle. Supraspinal influences (the energy supply) are only tapped in accordance with the basic design of the spinal circuitry. It is the latter--as in real clocks--that determines when the system receives its pulse of energy as well as the duration of the pulse (see also Boylls, 1975, for a discussion of spinal "slots," and Kots' 1977 analysis of the cyclic "quantized" character of supraspinal control, pp. 225-229).

The organization realized by coordinative structures--as we have noted-- is not obtained without cost; non-linear "dynamic patterns" emerge from the dissipation of more free energy than is degraded in the drift toward equilibrium. Thus the stability of a collective is attained by the physical action of an ensemble of "squirt" systems in a manner akin to limit cycle behavior (cf. Katchalsky et al., 1974; Prigogine & Nicolis, 1971; Soodak & Iberall, 1978). It remains for us now to illustrate--albeit briefly and in a very preliminary way--some of the behavioral predictions of the dynamic perspective on coordinated movement. These necessarily fall out of the properties of non-linear limit cycles--a topic that we can address here only in a rather terse way.

Homeokinetic theory characterizes biological systems as ensembles of non-linear oscillators coupled and mutually entrained at all levels of organization. It predicts the discovery of numerous cyclicities and evidence of their mutual interaction. As noted above, the only cycles that meet the non-linear, self-sustaining, dynamic stability criteria that homeokinetics demands are called limit cycles (cf. Goodwin, 1970; Soodak & Iberall, 1978; Yates & Iberall, 1973) and it is their properties from which insights into behavior might emerge. Here we give a sampling of current work in progress (Kelso, Holt, Rubin, & Kugler, in press). By and large, the research involves cyclical movements of the hand alone or in combination with speech (see Section 4).

(a) Response to perturbations/changes in initial conditions:

As Katchalsky et al. (1974) note, the essential difference between linear or non-linear conservative oscillators and limit cycle oscillators (which obey non-linear dissipative dynamics) is that perturbations applied to a conservative oscillator will move it to another orbit or frequency, whereas a limit cycle oscillator will maintain its orbit or frequency when perturbed. An examination of Figure 5 helps clarify this point. In Figure 5A, we show the position versus time, and velocity versus position, functions for linear and non-linear types of oscillators. In Figure 5B the spiral trajectory in the

PHASE PLANE TRAJECTORY

POSITION-TIME FUNCTION
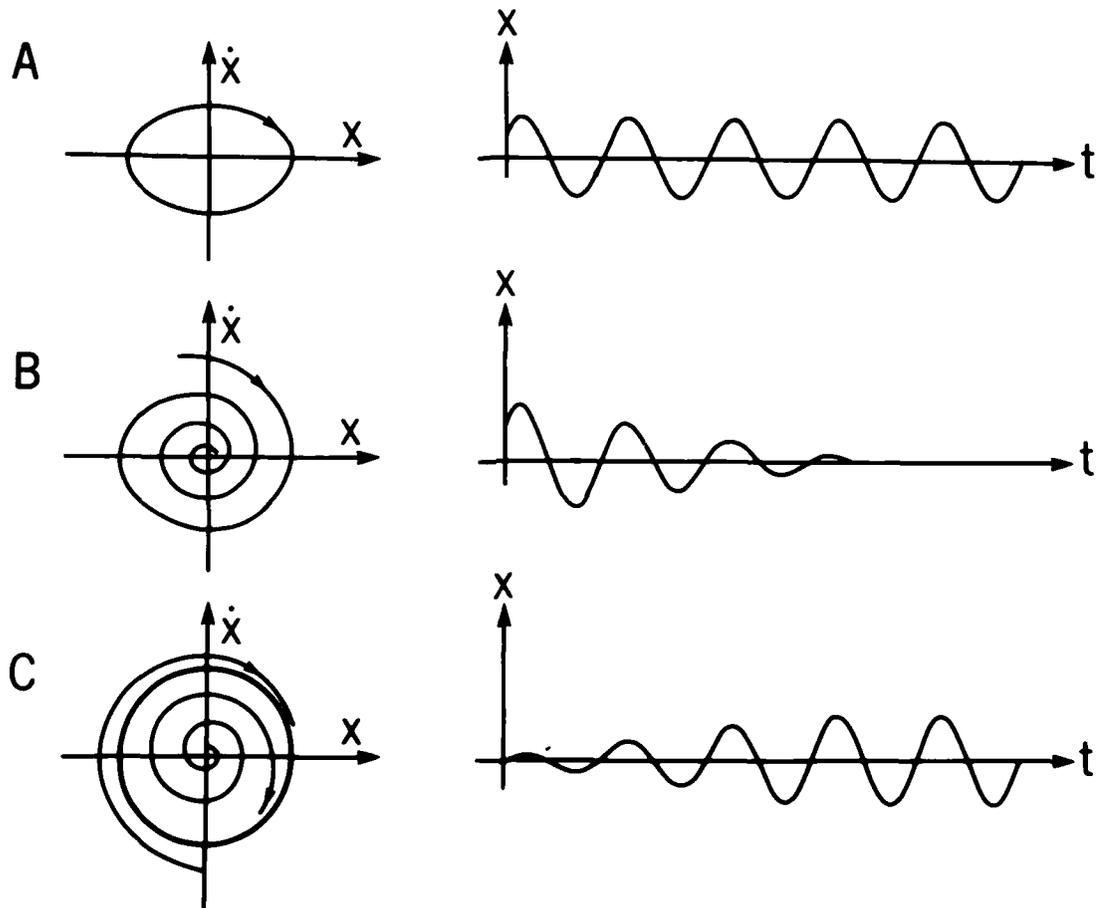
Figure 5.   Phase plane trajectories and corresponding position-time functions
for three different types of oscillation.

A. Idealized harmonic motion
B. Damped harmonic motion
C. Limit cycle oscillatory motion (see text for details).

178

phase plane represents an oscillation that continuously decreases in amplitude until it comes to a standstill. This is the phase trajectory (velocity vs. position relation) of a stable, damped oscillation. A change in any parameter in the equation describing this motion, for example, the damping coefficient, would drastically change the form of the solution and thus the phase trajectory. In such linear systems there is then no _preferred_ set of solutions in the face of parameter changes. In sharp contrast, non-linear oscillators of the limit cycle type possess a family of trajectories that all tend asymptotically towards a _single_ limit cycle despite quantitative changes in parameter values (see Figure 5C). Thus, a highly important property of limit cycle oscillators is their _structural stability_ in the face of variations in parameter values.

We have shown in a set of experiments on two-handed cyclical movements (Kelso et al., in press), that the limbs (in this case the fingers) maintain their preferred frequency and amplitude relations no matter how they are perturbed. Perturbations took the form of brief (100 msec) or constant (applied at a variable point during the cycle and maintained throughout) torque loads unexpectedly applied to one hand or the other via DC torque motors situated above the axis of rotation of the metacarpophalangeal joints. In all four experiments there were no differences in amplitude or duration (1/f msec) pre- and post-perturbation (for many more details, see Kelso et al., in press). Moreover, the fact that non-linear oscillators must degrade a large amount of free energy to offset the energy lost during each cycle suggests that they will be quickly resettable following a perturbation. This was precisely the case in our experiments. The fingers were in phase in the cycle immediately following the perturbation as revealed by cross-correlations between the limbs as a function of phase lag and by individual inspection of displacement-time waveforms. This capability to return to a stable, bounded phase trajectory despite perturbations, predicted by limit cycle properties, is an extension of our previous work (and that of others) on single trajectory movements (see Section 4 above). The latter, it will be remembered, display the "equifinality" property in the face of perturbations, changes in initial conditions and deafferentation (see Bizzi, in press). The organization over the muscles is qualitatively like a non-linear oscillatory system, regardless of whether one is speaking of discrete or cyclical movements (cf. Fel'dman, 1966; Fowler et al., 1980; Kelso & Holt, 1980; Kelso, Holt, Kugler, & Turvey, 1980).

(b) Entrainment properties

We have characterized coordination in biological systems as arising from cooperative relationships among non-linear oscillator ensembles. As already intimated, the chief mode of cooperation among self-sustaining oscillators is entrainment or synchronization. Strictly speaking the latter terms are not synonymous: synchronization is that state which occurs when both frequency and phase of coupled oscillators are matched exactly; entrainment refers to the matching of frequencies, although one oscillator may lead or lag the other.

When coupled oscillators interact, _mutual entrainment_ occurs (the 'magnet' effect of von Holst, 1937, English translation 1973) with only a

small frequency detuning (cf. Minorsky, 1962). Another form of mutual inter-
action occurs if the frequency of one oscillator is an integer multiple of
another to which it is coupled, a property termed subharmonic entrainment or
frequency demultiplication. These preferred relationships are ones that
coupled oscillators assume under conditions of maximal coupling or phase
locking. Years ago, von Holst discovered coordinative states in fish fin
movements that correspond to the different types of entrainment discussed here
(see von Holst, 1973, for English translation). The most common mode of
coordination he termed absolute coordination, a one-to-one correspondence
between cyclicities of different structures. The second and much less common
interactive mode he called relative coordination. Here the fins exhibit
different frequencies, although at least one corresponds to that seen in the
absolute coordination state. In more recent times, Stein (1976, 1977) has
elaborated on von Holst's work using the mathematics of coupled oscillators to
predict successfully patterns of neuronal activity for interlimb coordination.
The oscillator theoretic approach to neural control, as Stein (1977) remarks,
is still in an embryonic state. In our experiments we have taken a step in
what we hope is a positive direction by examining the qualitative predictions
of the theory without immediate concern for its neural basis. The results are
intuitively apparent to any of us who have tried to perform different cyclical
movements of the limbs at the same time. Thus the cyclical movements of each
limb operating singly at its own preferred frequency mutually entrain when the
two are coupled together (von Holst's M-effect). When an individual is asked
to move his/her limbs at different frequencies, low integer subharmonic
entrainment occurs. An example of the waveforms of both limbs shown in Figure
6 also suggests amplitude modulation (von Holst's _superimposition_ effect).
Thus on some coinciding cycles a "beat" phenomenon can be observed (particu-
larly in the 2:1 ratio) in which the amplitude of the higher frequency hand
increases in relation to non-coincident cycles. These preferred relationships
are _emergent_ characteristics of a system of non-linear oscillators; the
collection of mutually entrained oscillators functions in a single unitary
manner.

Entrainment properties are not restricted to movements of the limbs, but
are also evident (as predicted by the principles of homeokinetic physics) in
systems that share little or no common structural similarity. Returning to
our analysis of the interrelationships between speaking and manual activity,
we have shown that subjects, when asked to speak (again the familiar syllable
/stak/) at a different rate from their preferred finger rate, do so by
employing low integer sub- or super-harmonics (see Figure 7). The situation
is reversed (though not necessarily symmetrically) when the individual is
asked to move the finger at a different rate from speaking. The ratios chosen
are always simple ones (e.g., 2:1 or 3:1 or 3:2; see Figure 8). The strict
maintenance of cyclicity as predicted by homeokinetic theory is abundantly
apparent. Entrainment ensures a stable temporal resolution of simultaneous
processes throughout the whole system. Moreover, entrainment of oscillators
is limited to a relatively restricted frequency range captured in Iberall and
McCulloch's poetics as an "orbital constellation."

Homeokinetic theory requires a dynamic system analysis that, to be used
optimally, requires a research decision as to the likely limiting conditions
for the spectrum of effects of interest. In the continuum of cyclical
processes, coherency is determined by the longest period over which "thermody-

180
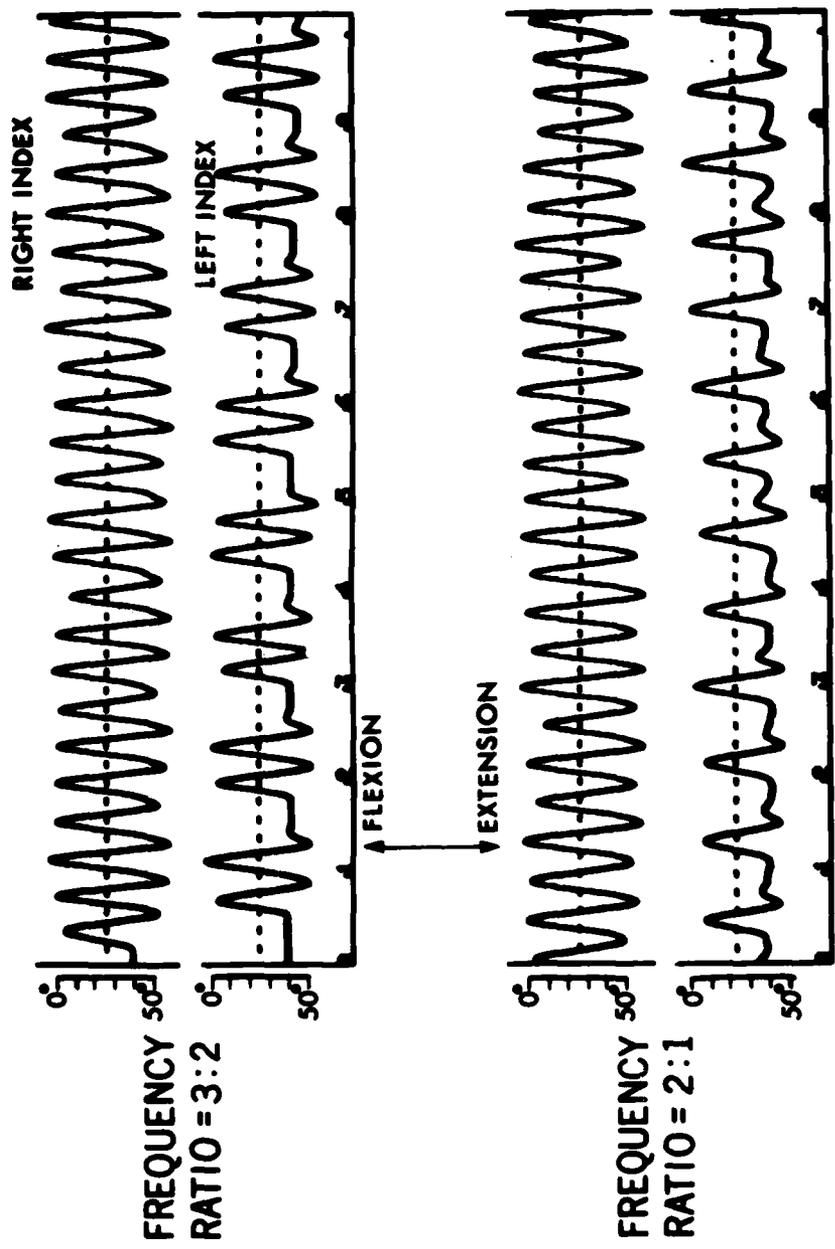
Figure 6. An example of one subject's response to instructions to move the fingers at different frequencies. On some coinciding cycles, a "beat" phenomenon can be observed in which the amplitude of the higher frequency hand increases in relation to non-coincident cycles (see especially 2:1 ratio).

CHANGE RATE OF SPEAKING
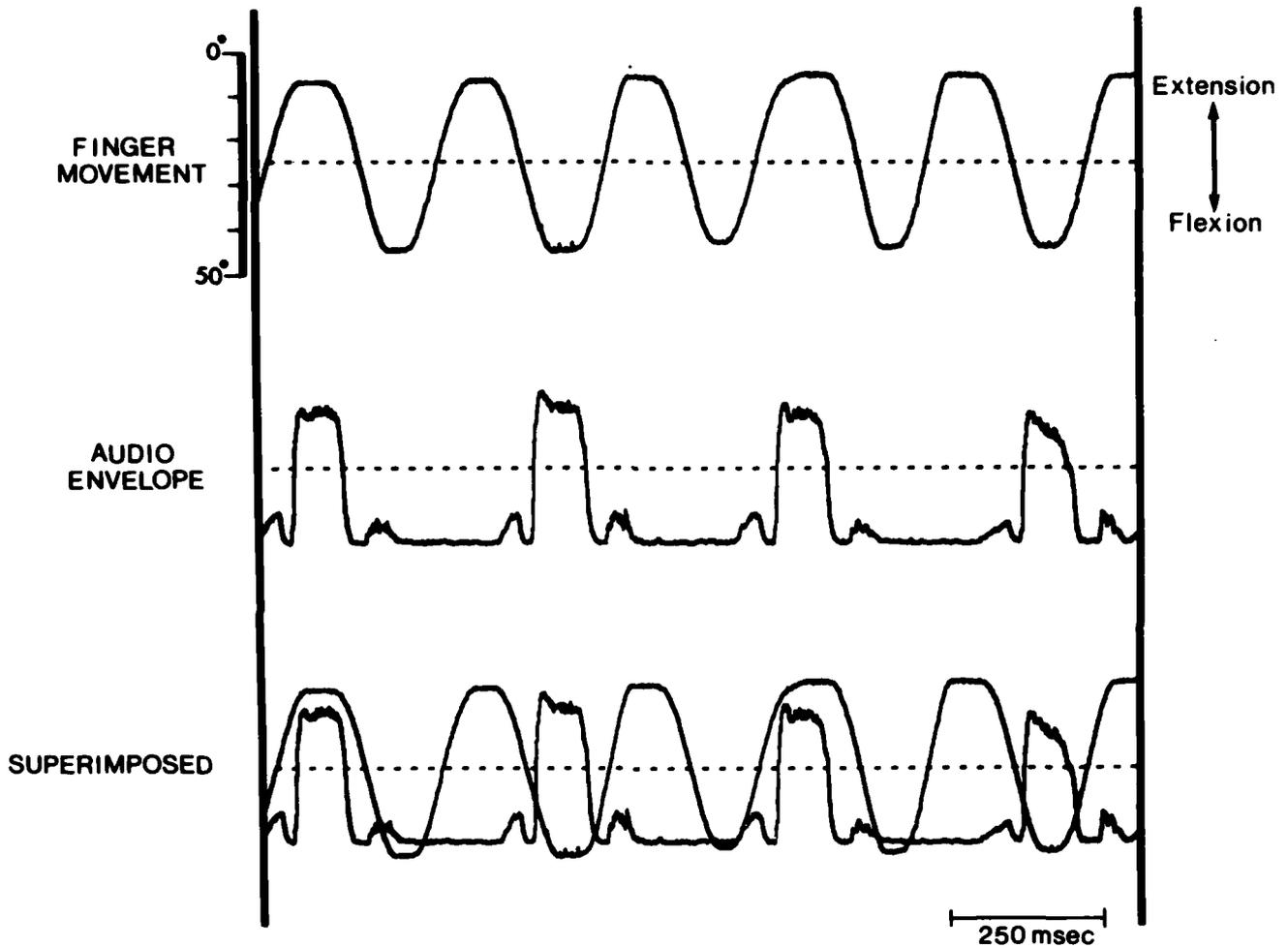
Figure 7. Simultaneous finger movement (top) and integrated speech waveform (bottom) produced by a subject asked to speak at a different rate from finger movement. The subject shown considered each flexion and extension as a separate finger movement. Thus, the finger to speech ratio is 3:1.

CHANGE RATE OF FINGER MOVEMENT
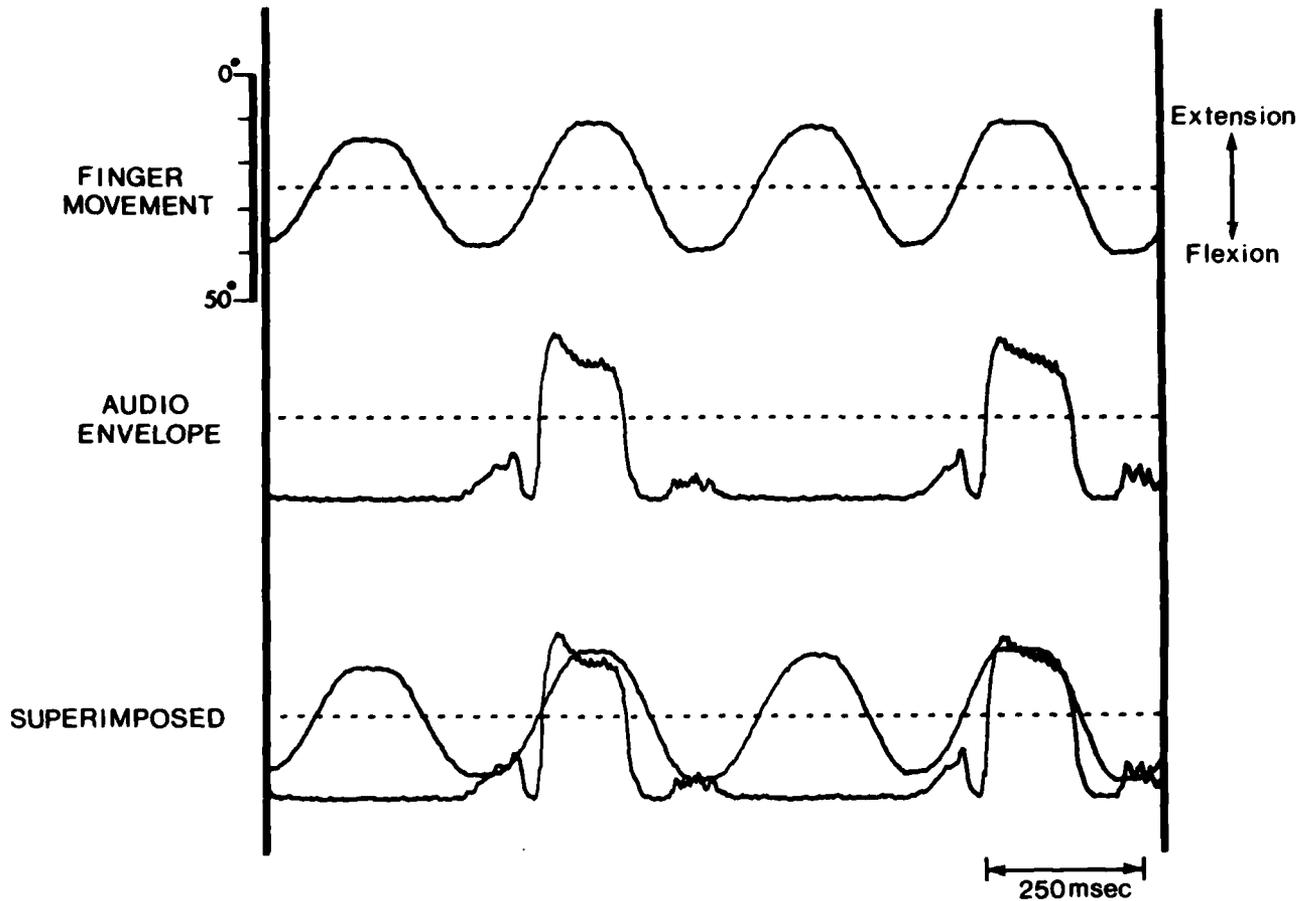
Figure 8. Simultaneous finger movement (top) and integrated speech waveform (bottom) produced by a subject when asked to move his finger at a different rate from his speaking. This subject shows a 2:1 ratio of finger movement to speech, each syllable synchronized with every second finger extension.

183

namic bookkeeping" is closed. For those interested in the production of speech a possible candidate oscillation over which articulatory cycles of shorter periods may cohere is the "breath group" (cf. Lieberman, 1967) or more globally the respiratory cycle (Fowler, 1977; Turvey, 1980). The latter, tied as it is to metabolic processes, may well be the organizing period for all the activity patterns of an animal. It is well known, for example, that during exercise, respiration is often synchronized with movements of body parts (Astrand & Rodahl, 1970). But even when metabolic demands are not altered from a resting state, preliminary data indicate entrainment between breathing and limb movements (see also Wilke, Lansing, & Rogers, 1975).

In Figure 9, we see data from the now familiar task of speaking and performing cyclical finger movements. In the first case the subject is instructed to move the left index finger at a different rate from speech. The finger wave form is highly regular except at one particular point where a pause is evident. From the acoustic signal it is obvious that the pause in finger movement coincides perfectly with respiratory inhalation. In a parallel condition in which the subject is instructed to speak at a different rate from finger movement, we see exactly the same co-occurrence of breathing and a pause in the finger movements (see Figure 10). Aside from the fact that these data provide further and perhaps the most compelling evidence of entrainment in coupled systems, there is also the suggestion that both systems cohere to the longer time-scale activity, namely breathing. Since the flow of oxygen constitutes a sustained temporal process in the system (the "escapement" for the thermodynamic power cycle), it seems reasonable to suppose that the respiratory cycle may play a cohering role around which other oscillations seek to entrain. But at this point the question is hypothetical in the face of nonexistent data.

We do not wish to give the impression, however, that the cohering role of the respiratory cycle gives it dominant status. On the contrary, it is well known that the respiratory cycle itself changes character to accommodate the demands of speech (e.g., Draper, Ladefoged, & Whitteridge, 1960). In fact, the entrainment of these systems cannot be explained solely on the basis of metabolic demands. When subjects read silently (Conrad & Schönle, 1979), or when finger movements required are of minimal extent (Wilke, 1977), respiratory rhythms change to be compatible with the other activity. The point is that in an oscillator ensemble there is no fixed dominance relation. There are different modes of interaction (e.g., frequency and amplitude modulation) and there may be preferred phase relationships, as in the extreme case of maximal coupling or phase-locking between two oscillators. A wide variety of behavioral patterns emerge from these interactions; there is structure and a complex network of interconnections but, strictly speaking, no dominance relation.

6. CONCLUDING REMARKS

The major problem confronting a theory of coordination and control (whether it be of speech or limbs) is how stable spatiotemporal organizations are realized from a neuromuscular basis of very many degrees of freedom. Here we have offered the beginnings of an approach in which solutions to the degrees of freedom problem may lie--not in machine-type theories--but in the

MOVE FINGER AT A DIFFERENT RATE FROM SPEAKING

Finger Movement (Fp = 1.70Hz)
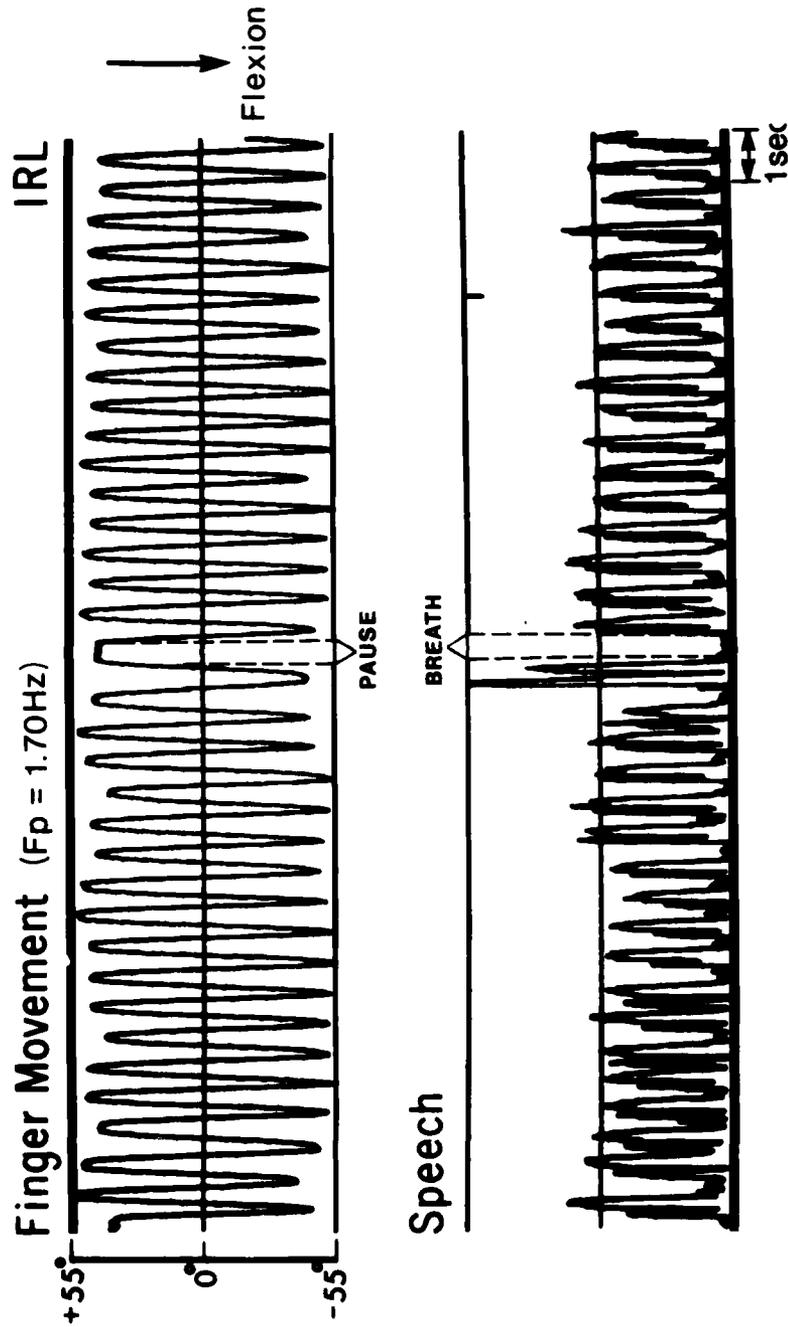
Speech

Figure 9. Simultaneous finger movement (top) and integrated speech waveform (bottom) produced by a subject when told to move her finger at a different rate from speaking. Pause in the finger movement and the simultaneous inhalation are indicated.

185

SPEAK AT A DIFFERENT RATE FROM FINGER MOVEMENT
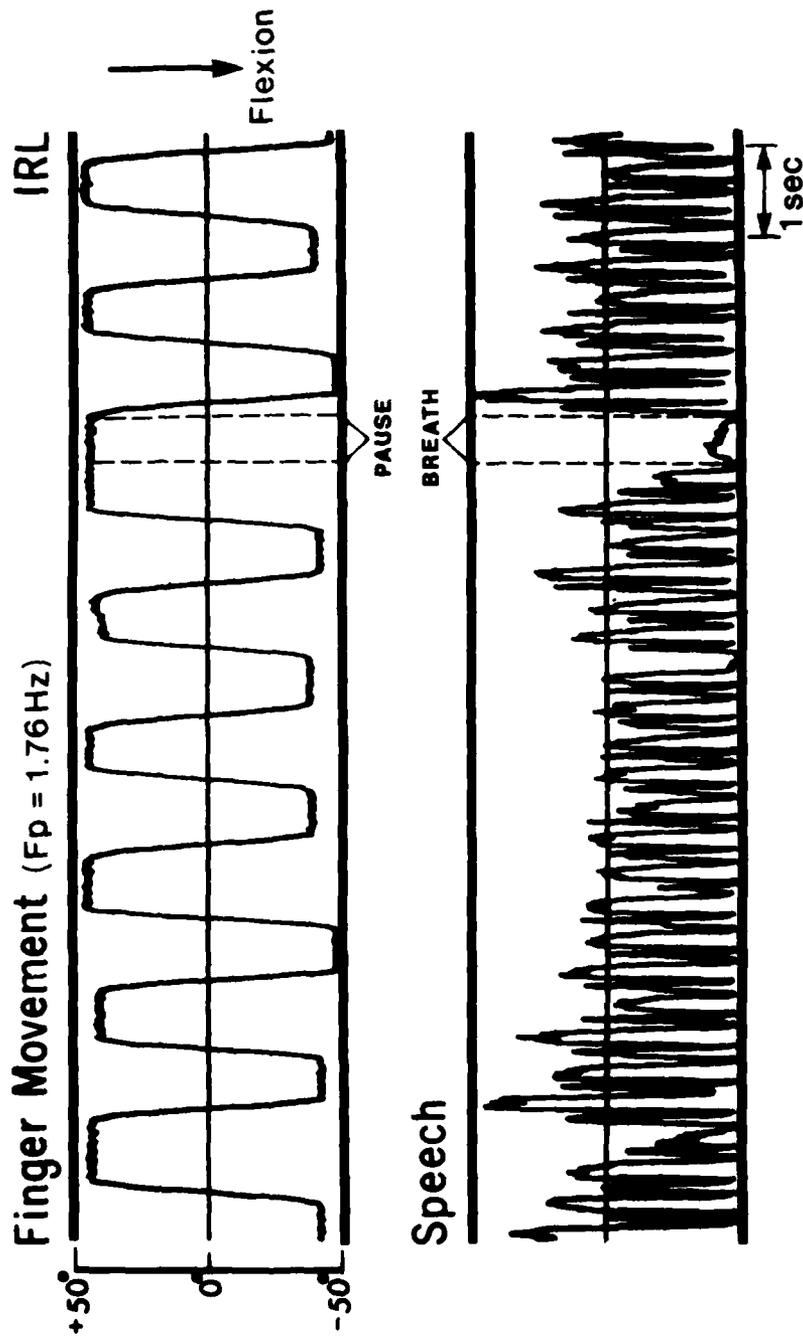
Figure 10. Simultaneous finger movement (top) and integrated speech waveform (bottom) produced by a subject when told to speak at a different rate from finger movement. A pause in the finger movement, and the simultaneous inhalation are indicated.

contemporary physical theories of dissipative structures and homeokinetics. A central characteristic of such theories is that complex systems consist of collectives of energy-flow systems that interact in a unitary way and, as a consequence, exhibit limit cycle oscillation. Many of the motor behaviors discussed in this paper can be rationalized according to limit cycle properties. Common to all of them--including speech--is that certain qualitative properties are preserved over quantitative changes in the values of individual components (muscles, keypresses, kinematic attributes). This feature of coordinated activity exists across all scales of observation; it is as applicable to the microscale (e.g., physiological tremor) as it is to the gross movement patterns of locomotion. We suspect that the functional similarities observed across levels of analysis index the design of the motor system. Thus, even though the material composition varies dramatically from level-to-level, certain qualitative properties, like cycling, remain invariant (cf. Kugler et al., in press; and for a similar view, Mandell & Russo, 1980).

Central to the view expressed here (see also Kelso, in press; Kugler et al., 1980, in press; Yates & Iberall, 1973) is that new forms of spatiotemporal organization are possible when scale changes and nonlinearities are present, and an energy supply is available. When a stable system is driven beyond a certain critical value on one of its parameters, bifurcation occurs and qualitatively new structures emerge (cf. Guttinger, 1974). There are many examples of such phase transition phenomena in nature (see Haken, 1977; Prigogine, 1980; Winfree, 1980, for examples) and probably in movement as well. We know, for example, that at low velocities quadrupeds locomote such that limbs of the same girdle are always half a period out of phase. But as velocity is scaled up, there is an abrupt transition from an asymmetric to symmetric gait (Shik & Orlovskii, 1976). The phase relations of the limbs change, but we doubt if a new "program" is required (Shapiro, Zernicke, Gregor, & Diestel, in press) or that one needs to invoke a "gait selection" process (Gallistel, 1980). Emergent spatiotemporal order, in the view expressed here, is not owing to an a priori prescription, independent of and causally antecedent to systemic behavior. Rather it is an a posteriori fact of the system's dynamical behavior. As Gibson (1979) remarked, behavior is regular without being regulated.

The present perspective--with appropriate extensions (e.g., to a reconceptualization of 'information' in naturally developing systems; Kugler et al., in press)--is less antireductionistic than it is an appeal for epistemological change. Contemporary physics as characterized here does not assign priority to any privileged scale of analysis: There is no "fundamental unit" out of which one can construct a theory of systemic phenomena (see Buckley & Peat, 1979; Yates, 1978). Instead, homeokinetics and dissipative structure/dynamic pattern theory offer a single set of physical principles that can be applied at all levels of analysis. If there is reductionism, it is not in the analytical sense but rather to a minimum set of principles.

REFERENCE NOTES

1. Greene, P. H. Strategies for heterarchical control--an essay. I. A style of controlling complex systems. Unpublished manuscript, Department of Computer Science, Illinois Institute of Technology, 1975.

2. Fentress, J. C. *Order and ontogeny: Relational dynamics*. Paper given at Interdisciplinary Study of Behavioral Development, Bielefeld, Germany, March, 1978.
3. Patten, B. C. *Environs: Relativistic elementary particles for ecology*. Paper presented at the dedication of Environmental Sciences Laboratory Building, Oak Ridge National Laboratory, Oak Ridge, Tennessee, February 26-27, 1979.

REFERENCES

Adams, J. A. Feedback theory of how joint receptors regulate the timing and positioning of a limb. *Psychological Review*, 1977, 84, 504-523.

Andranov, A., & Chaiken, C. E. *Theory of oscillations*. Princeton, N.J.: Princeton University Press, 1949.

Aschoff, J. Circadian rhythms: General features and endocrinological aspects. In D. Krieger (Ed.), *Endocrine rhythms*. New York: Raven Press, 1979.

Astrand, P. O., & Rodahl, K. *Textbook of work physiology*. New York: McGraw-Hill, 1970.

Bateson, P. P. G., & Hinde, R. A. (Eds). *Growing points in ethology*. Cambridge: Cambridge University Press, 1976.

Bellman, R. *Adaptive control processes: A guided tour*. Princeton, N.J.: Princeton University Press, 1961.

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

Bizzi, F. Paper to appear in *The production of speech*, P. MacNeilage (Ed.). New York: Springer-Verlag, in press.

Bizzi, E., Dev, P., Morasso, P., & Polit, A. Effect of load disturbances during centrally initiated movements. *Journal of Neurophysiology*, 1978, 41, 542-555.

Bowers, D., Heilman, K. M., Satz, P., & Altman, A. Simultaneous performance on verbal, non-verbal and motor tasks by right-handed adults. *Cortex*, 1978, 14, 540-556.

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

Brindley, J. S., Goodwin, G. M., Kulikowski, J. T., & Leighton, D. Stability of vision with a paralyzed eye. *Journal of Physiology*, 1976, 258, 65-66.

Buckley, P., & Peat, F. D. *A question of physics: Conversations in physics and biology*. Toronto: University of Toronto, 1979

Cicone, M., Wapner, W., Foldi, N., Zurif, E., & Gardner, H. The relation between gesture and language in aphasic communication. *Brain and Language*, 1979, 8, 324-329.

Conrad, B., & Schönle, P. Speech and respiration. *Archiv fuer Psychiatrie und Nervenkrankheiten*, 1979, 226, 251-268.

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

Davis, G. Organizational concepts in the central motor networks of invertebrates. In R. M. Herman, S. Grillner, P. S. G. Stein, & D. G. Stuart (Eds.), *Neural control of locomotion*. New York: Plenum, 1976.

188

3 or 3
AD A
099958

END
DATE
FILMED
6-81
DTIC

Dellow, P. G., & Lund, J. P. Evidence for central timing of rhythmical mastication. Journal of Physiology, 1971, 215, 1-13.

Denier van der Gon, J. J. & Thuring, J. Ph. The guiding of human writing movements. Kybernetik, 1965, 4, 145-147.

Dennett, D. C. Brainstorms: Philosophical essays on mind and psychology, Montgomery, Vt.: Bradford Books, 1978.

Desmedt, J. E. (Ed). Physiological tremor, pathological tremor and clonus. Progress in Clinical Neurophysiology, Vol. 5. Basel: Karger, 1978.

Draper, M. H., Ladefoged, P., & Whitteridge, D. Expiratory pressure and air flow during speech. British Medical Journal, 1960, 1, 1837-1843.

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

Emmett, K. Intentional systems: Dennett's philosophy of psychology. Cognition and Brain Theory, 1980, 3, 109-111.

Engberg, I., & Lundberg, A. An electromyographic analysis of muscular activity in the hindlimb of the cat during unrestrained locomotion. Acta Physiologica Scandia, 1969, 75, 614-630.

Fant, G., Stalhammar, U., & Karlsson, I. Swedish vowels in speech material of various complexity. In Speech Communication Seminar, Stockholm, 1974, Uppsala: Almqvist and Wiksell, 1974.

Fel'dman, A. G. 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, 1966, 11, 766-775.

Fentress, J. C. Dynamic boundaries of patterned behavior: Interaction and self-organization. In P. P. G. Bateson & R. A. Hinde (Eds.), Growing points in ethology. Cambridge: Cambridge University Press, 1976.

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

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

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

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

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

Gallistel, C. R. The organization of action: A new synthesis. Hillsdale, N.J.: Erlbaum, 1980.

Gay, T. Effect of speaking rate on vowel formant movements. Journal of the Acoustical Society of America, 1978, 63, 223-230.

Gay, T., Ushijima, T., Hirose, H., & Cooper, F. S. Effect of speaking rate on labial consonant-vowel articulation. Journal of Phonetics, 1974, 2, 47-63.

Gelfand, I. M., Gurfinkel, V. S., Tsetlin, M. L., & Shik, M. L. Some problems in the analysis of movements. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge, Mass.: MIT Press, 1971.

Gelfand, I. M., & Tsetlin, M. L. Mathematical modeling of mechanisms of the central nervous system. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge, Mass.: MIT Press, 1971.

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

Goodwin, B. Biological stability. In C. H. Waddington (Ed.), Towards a theoretical biology. Chicago: Aldine, 1970.

Greene, P. H. Problems of organization of motor systems. In R. Rosen & F. Snell (Eds.), Progress in theoretical biology. New York: Academic Press, 1972.

Grillner, S. Locomotion in vertebrates. Physiological Reviews, 1975, 55, 247-304.

Grimm, R. J. & Nashner, L. M. Long loop dyscontrol. In J. E. Desmedt (Ed.), Cerebral motor control in man: Long loop mechanisms, Progress in Clinical Neurophysiology, Vol. 4. Basel: Karger, 1978.

Gurfinkel, V. S., Kots, Y. A., Paltsev, E. I., & Fel'dman, A. G. The compensation of respiratory disturbances of the erect posture of man as an example of the organization of interarticular interaction. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge, Mass.: MIT Press, 1971.

Guttinger, W. Catastrophe theory in physics and biology. In M. Conrad, W. Guttinger, & M. Dalcin (Eds.), Lecture notes in biomathematics, Vol. 4, Physics and mathematics of the nervous system. Berlin: Springer-Verlag, 1974.

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

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

Herman, R., Wirta, R., Bampton, S., & Finley, R. Human solutions for locomotion: Single limb analysis. In R. M. Herman, S. Grillner, P. S. G. Stein, & D. G. Stuart (Eds.), Neural control of locomotion. New York: Plenum Press, 1976.

Hicks, R. E. Intrahemispheric response competition between vocal and unimanual performance in normal adult human males. Journal of Comparative and Physiological Psychology, 1975, 89, 50-60.

Hicks, R. E., Provenzano, F. J., & Rybstein, E. D. Generalized and lateralized effects of concurrent verbal rehearsal upon performance of sequential movements of the fingers by the left and right hands. Acta Psychologica, 1975, 39, 119-130.

Hofstadter, D. R. Gödel, Escher, Bach: An eternal golden braid. New York: Basic Books, 1979.

Hollerbach, J. M. An oscillation theory of handwriting. Cambridge, Mass.: MIT Artificial Intelligence Laboratory, 1980.

Holst, E. von. The behavioral physiology of animal and man: The collected papers of Erich von Holst (Vol. 1) (R. Martin, trans.). London: Methuen and Co., Ltd, 1973.

Iberall, A. S. On nature, man and society: A basis for scientific modeling. Annals of Biomedical Engineering, 1975, 3, 344-385.

Iberall, A. S. A field and circuit thermodynamics for integrative physiology: I. Introduction to general notion. American Journal of Physiology, 1977, 2, R171-R180.

Iberall, A. S., & McCulloch, W. S. The organizing principle of complex living systems. Transactions of the American Society of Mechanical Engineers, June, 1969, 290-294.

Katchalsky, A. K., Rowland, V., & Blumenthal, R. Dynamic patterns of brain cell assemblies. Neurosciences Research Program Bulletin, 1974, 12(1).

Keele, S. W. Behavioral analysis of motor control, In V. Brooks (Ed.), Handbook of physiology: Motor control. Washington, D.C.: American Physiological Society, 1980.

Kelso, J. A. S. Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology, 1977, 3, 529-543.

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

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

Kelso, J. A. S., Holt, K. G., & Flatt, A. E. The role of proprioception in the perception and control of human movement: Toward a theoretical reassessment. Perception & Psychophysics, 1980, 28, 45-52.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. On the concept of coordinative structures as dissipative structures: II. Empirical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North Holland, 1980.

Kelso, J. A. S., Holt, K., Rubin, P., & Kugler, P. Patterns of human interlimb coordination emerge from the properties of non-linear oscillators: Theory and data. Journal of Motor Behavior, in press.

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

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

Kelso, J. A. S., & Tuller, B. Toward a theory of apractic syndromes. Brain and Language, in press.

Kimura, D. The neural basis of language qua gesture. In H. Whitaker & H. A. Whitaker (Eds.), Studies in neurolinguistics (Vol. 3). New York: Academic Press, 1976.

Kinsbourne, M., & Cook, J. Generalized and lateralized effects of concurrent verbalization on a unimanual skill. Quarterly Journal of Experimental Psychology, 1971, 23, 341-345.

Kinsbourne, M., & Hicks, R. E. Functional cerebral space: A model for overflow, transfer and interference effects in human performance: A tutorial review. In J. Requin (Ed.), Attention and performance VII. Hillsdale, N.J.: Erlbaum, 1978. (a)

Kinsbourne, M., & Hicks, R. E. Mapping cerebral functional space: Competition and collaboration in human performance. In M. Kinsbourne (Ed.), Asymmetrical function of the brain. New York: Cambridge University Press, 1978. (b)

Kots, Ya. M. The organization of voluntary movement. New York: Plenum, 1977.

Kozhevnikov, V., & Chistovich, L. Speech: Articulation and perception. Moscow-Leningrad, 1965 (English translation: J.P.R.S., Washington, D.C. No. JPRS 30543).

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On the concept of coordinative structures as dissipative structures. I. Theoretical lines of

convergence. In G. E. Stelmach (Ed.), Tutorials in motor behavior. Amsterdam: North-Holland, 1980.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), The development of human movement control and coordination. New York: John Wiley, in press.

Lashley, K. The problem of serial order in behavior. In L. A. Jeffress (Ed.), Cerebral mechanisms in behavior. New York: Wiley, 1951.

Liberman, A. M. The specialization of the language hemisphere. In F. O. Schmitt & F. G. Worden (Eds.), The neurosciences: Third study program. Cambridge, Mass.: M.I.T. Press, 1974, 43-56.

Liberman, M., & Prince, A. On stress and linguistic rhythm. Linguistic Inquiry, 1977, 8, 249-336.

Lieberman, P. Intonation, perception, and language. Cambridge, Mass.: M.I.T. Press, 1967.

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

Lindblom, B. E .F. Motor control mechanisms. Papers from the Institute of Linguistics, University of Stockholm, 1974, 26, 1-19.

Lomas, J., & Kimura, D. Intrahemispheric interaction between speaking and sequential manual activity. Neuropsychologia, 1976, 14, 23-33.

Luschei, E. S., & Goodwin, G. M. Patterns of mandibular movement and jaw muscle activity during mastication in the monkey. Journal of Neurophysiology, 1974, 37, 954-966.

Mandell, A. J., & Russo, P. V. Stochastic periodicity: Latent order in variance. Totus Homo, 1980, 12, 23-36.

Mattingly, I. G. Epimenides at the computer. The Yale Review, Winter 1980.

Miles, F. N., & Evarts, E. V. Concepts of motor organization. Annual Review of Psychology, 1979, 30, 327-362.

Minorsky, N. Nonlinear oscillations. Princeton, N.J.: Van Nostrand, 1962.

Møller, E. Action of the muscles of mastication. In Y. Kawamura (Ed.), Physiology of mastication. Basel: S. Karger, 1974.

Morowitz, H. J. Energy flow in biology. Woodbridge, Conn.: Oxbow Press, 1979.

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

Oatley, K., & Goodwin, B. W. The explanation and investigation of biological rhythms. In W. P. Colquhoun (Ed.), Biological rhythms and human performance. New York: Academic Press, 1971.

Orlovskii, G. N. The effect of different descending systems on flexion and extensor activity during locomotion. Brain Research, 1972, 40, 359-371.

Orlovskii, G. N., & Shik, M. L. Standard elements of cyclic movement. Biophysics, 1965, 10, 935,944.

Patten, B. C., & Auble, G. T. Systems approach to the concept of niche. Synthese, in press.

Pearson, K. G. The control of walking. Scientific American, 1976, 235, 72-79.

Perenin, M. K., Jeannerod, M., & Prablanc, C. Spatial localization with paralyzed eye muscles. Ophthalmologica, 1977, 175, 206-214.

Polit, A., & Bizzi, E. Processes controlling arm movements in monkeys. Science, 1978, 201, 1235-1237.

Prigogine, I. From being to becoming. San Francisco: W. H. Freeman, 1980.

192

Prigogine, I., & Nicolis, G. Biological order, structure and instabilities. Quarterly Review of Biophysics, 1971, 4, 107-148.

Rashevsky, N. Mathematical biophysics, physico-mathematical foundations of biology (Vol. 2). New York: Dover, 1960.

Saltzman, E. Levels of sensorimotor representation. Journal of Mathematical Psychology, 1979, 20, 92-163.

Schmidt, R. A. A schema theory of discrete motor skill learning. Psychological Review, 1975, 82, 225-260.

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

Schroedinger, E. What is life? London: Cambridge University Press, 1945.

Shaffer, L. H. Intention and performance. Psychological Review, 1976, 83, 375-393.

Shapiro, D. C., Zernicke, R. F., Gregor, R. J., & Diestel, J. D. Evidence for generalized motor programs using gait pattern analysis. Journal of Motor Behavior, in press.

Shaw, R., & Turvey, M. T. Coalitions as models for ecosystems: A realist perspective on perceptual organization. In M. Kubovy & J. Pomerantz (Eds.), Perceptual organization. Hillsdale, N.J: Erlbaum, in press.

Shaw, R. E., Turvey, M. T., & Mace, W. Ecological psychology: The consequences of a commitment to realism. In W. Weimer & D. Palermo (Eds.) Cognition and symbolic processes (Vol. 2). Hillsdale, N.J.: Erlbaum, in press.

Shik, M. L., & Orlovskii, G. N. Neurophysiology of locomotor automatism. Physiological Reviews, 1976, 56, 465-501.

Skavenski, A. A., Haddad, G., & Steinman, R. M. The extraretinal signal for the visual perception of direction. Perception & Psychophysics, 1972, 11, 287-290.

Soodak, H., & Iberall, A. S. Homeokinetics: A physical science for complex systems. Science, 1978, 201, 579-582.

Stein, P. S. G. Mechanisms of interlimb phase control. In R. M. Herman, S. Grillner, P. S. G. Stein, & D. G. Stuart (Eds.), Neural control of locomotion. New York: Plenum Press, 1976.

Stein, P. S. G. Application of the mathematics of coupled oscillator systems to the analysis of the neural control of locomotion. Federation Proceedings, 1977, 36, 2056-2059.

Stein, P. S. G. Motor systems, with special reference to the control of locomotion. Annual Review of Neuroscience, 1978, 1, 61-81.

Stevens, J. R. The corollary discharge: Is it a sense of position or a sense of space? The Behavioral and Brain Sciences, 1978, 1, 163-165.

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

Stevens, P. S. Patterns in nature. Boston: Little Brown, 1974.

Studdert-Kennedy, M., & Lane, H. Clues from the differences between signed and spoken language. In U. Bellugi & M. Studdert-Kennedy (Eds.), Signed and spoken language: Biological constraints on linguistic form. Weinheim: Verlag Chemie, 1980.

Sussman, H. M. Methodological problems in evaluating lip/jaw reciprocity as an index of motor equivalence. Journal of Speech and Hearing Research, 1980, 23, 699-702.

Taub, E. Movements in nonhuman primates deprived of somatosensory feedback. *Exercise Sports Sciences Review*, 1976, 4, 335-374.

Terzuolo, C. A., & Viviani, P. The central representation of learned motor patterns. In R. E. Talbott & D. R. Humphrey (Eds.), *Posture and movement*. New York: Raven Press, 1979.

Teuber, H. L. Alteration of perception after brain injury. In J. C. Eccles (Ed.), *Brain and conscious experience*, New York: Springer-Verlag, 1966.

Thexton, A. T. To what extent is mastication pre-programmed and independent of peripheral feedback? In D. J. Anderson & B. Matthews (Eds.), *Mastication*. Bristol: Wright, 1976.

Thompson, D. A. W. *On growth and form* (2nd Ed.). Cambridge, England, 1942.

Tsetlin, M. L. *Automata theory and modeling in biological systems*. New York: Academic Press, 1973.

Tuller, B., Harris, K. S., & Kelso, J. A. S. Articulatory motor events as a function of speaking rate and stress. *Haskins Laboratories Status Report on Speech Research*, 1981, SR-65, this volume.

Tuller, B., Kelso, J. A. S., & Harris, K. S. Phase relationships among articulator muscles as a function of speaking rate and stress. *Haskins Laboratories Status Report on Speech Research*, 1981, SR-65, this volume.

Turvey, M. T. 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, N.J.: Erlbaum, 1977.

Turvey, M. T. Clues for the organization of motor systems. In U. Bellugi & M. Studdert-Kennedy (Eds.), *Signed and spoken language: Biological constraints on linguistic form*. Weinheim: Verlag Chemie, 1980.

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

Turvey, M. T., Shaw, R. E., & Mace, W. Issues in the theory of action: Degrees of freedom, coordinative structures and coalitions. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Erlbaum, 1978.

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

Webb, D. W. The swimming energetics of trout. I. Thrust and power output at cruising speeds. *Journal of Experimental Biology*, 1971, 55, 489-520.

Weiss, P. Self-differentiation of the basic patterns of coordination. *Comparative Psychology Monographs*, 1941, 17(4).

Wilke, J. T. Ultra radian biological periodicities in the integration of behavior. *International Journal of Neuroscience*, 1977, 7, 125-143.

Wilke, J. T., Lansing, R. W., & Rogers, C. A. Entrainment of respiration to repetitive finger tapping. *Physiological Psychology*, 1975, 3, 345-349.

Winfree, A. *The geometry of biological time*. New York: Springer Verlag, 1980.

Wing, A. M. Response timing in handwriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning*. New York: Academic Press, 1978.

Yates, F. E. Complexity and the limits of knowledge. *American Journal of Physiology: Regulatory, Integrative, and Comparative*, 1978, 234, R201-R204.

Yates, F. E. Physical causality and brain theories. *American Journal of Physiology*, 1980, 238, R277-R290.

Yates, F. E., & Iberall, A. S. Temporal and hierarchical organization in biosystems. In J. Urquart & F. E. Yates (Eds.), _Temporal aspects of therapeutics_. New York: Plenum, 1973, 17-34.

FOOTNOTES

[1]For example, the structure DNA can be taken as mechanism at one level of analysis, but at another level DNA is more appropriately described as a set of interacting components such as proteins and enzymes.

[2]Although Lindblom's later work does not adhere to the originally described model (e.g., Lindblom, 1974), it has strongly influenced recent experimental work (e.g., Fant, Stalhammar, & Karlsson, 1974; Gay, 1978; Gay, Ushijima, Hirose, & Cooper, 1974; Harris, 1978) and, we believe, is representative of a class of theories of speech motor control.

[3]We have tested a total of seven subjects in a number of different experimental situations. Although we shall not present averaged data here, the figures shown are representative of the performance of all of our subjects. In fact, some subjects show greater effects than those illustrated here.

[4]The apparatus for recording finger movements has been described in detail elsewhere (Kelso & Holt, 1980). Basically, the finger slips into a sleeve whose axis of rotation is coupled to a potentiometer, thus enabling us to obtain a full component of kinematic characteristics. Both finger and speech waveforms were recorded on FM tape for later off-line analysis on a PDP 11/45 computer.

[5]We use the word "stress" here guardedly because we have not yet performed listener tests on subjects' productions. It is clear, however, that the amplitude of the audio waveform is modulated according to what the finger is doing.

[6]The idea that adjunctive logic rather than conditional or causal logic is necessary to capture the mutual compatabilities among system components is owing to Shaw and Turvey (e.g., Shaw & Turvey, in press; Turvey & Shaw, 1979). There is growing acceptance of this view in ecological science (cf. Patten, Note 3; Patten & Auble, in press).

[7]We are indebted to Edward Reed for bringing these data to our notice. Reed properly argues that the integration of experiments on extraocular paralysis favoring corollary discharge theory (cf. Teuber, 1966) is based on an "argument from exclusion." That is, all other possible accounts are excluded, therefore corollary discharge theory is correct. We concur with Reed, and offer a simpler account of the data.

[8]Parts of this section (pp. 28-33) also appear—with minor modifications—in Kelso (in press).

[9]We do not believe this to be a trivial question. Even "at rest," man is operating periodically (cf. Desmedt, 1978, for review on normal "resting" tremor). At more macroscopic levels we are subject to circadian phenomena (e.g., Aschoff, 1979). Even the structure of language--if recent generative theories are a yardstick (e.g., Liberman & Prince, 1977)--is inherently rhythmic.

MOTIVATING MUSCLES: THE PROBLEM OF ACTION*

J. A. Scott Kelso+ and Edward S. Reed++

How do you get motives into muscles? Psychology by and large has avoided this question like a plague. Theories of motive states, like the grand theories of biology (such as the molecular theory of the genetic code) are "just so" theories; a quick wave of the hand and sexual urges are translated into muscle potentials. But, as the physiological psychologist C. R. Gallistel is quick to point out, the story is not that simple. In fact, a major problem in modern psychology is the conceptual chasm between what we know about muscles and what we know about motivational processes. In short, there is a need for a theory of action.

According to Gallistel, the guts of the theory have been in the literature all the time just waiting to be organized in a way that would satisfy the palate of the modern psychologist. Gallistel's approach is, by his own admission, plagiaristic: He places in front of the reader some of the classic, but infrequently cited papers that he believes provide a conceptual basis upon which to build a theory of action. These range from a chapter in Sherrington's "Integrative Action of the Nervous System" (1906) to von Holst's "Nature of Order in the Central Nervous System" (1938) to Weiss's insightful treatise on the problem of coordination (1941). Along the way he provides summaries and discussions of the newer data showing, more or less, how well recent findings fit the insights of these forerunners to modern neurobiology. Few would argue with Gallistel's selections and he should be commended for bringing them together for students of movement.

Of course the intent of the book goes far beyond reminding us of the writings of Sherrington et al.--interesting though they are. By drawing concepts and examples from the neurobehavioral study of animal activity and linking them to some recent work on cognitive psychology (such as Cooper and Shepard's work on mental rotation), the author proposes--in recognition of its roots in behavioral neurobiology and ethology--a "neuroethological theory of action" (p. 361). It is on the achievement of this admittedly lofty goal--not on the achievements of others--that one must evaluate this book. Gallistel's basic claim is that it is possible to bridge the chasm between motives and muscles by means of lessons learned in physiological psychology. In our opinion this may be somewhat premature. We suspect that the physiological

psychologist's foundation for a theory of action is, as of now, more modest than the author thinks.

The basic building blocks of action?

Part of the problem in Gallistel's theory stems from his identification of the "elementary units of behavior." There are three of them in the author's view--reflexes, servomechanisms, and oscillators--all of which, when combined in particular ways, yield complex behaviors. The principle central to creating purposive actions is called selective potentiation. According to this principle, elementary units are not ordered directly by central programs, but rather subsets of them are "selectively potentiated" to fit prevailing circumstances. Selective potentiation, in a sense, specifies "viable options" and, in so doing, provides the animal with flexible control. As an example, at the highest level of a hierarchically structured system, central programs are thought to control the potential for action in lower level reflex arcs, ensuring that reflex action is consonant with certain specific environmental events. By merely controlling the potential for action one can account for why the same stimulus--a tap to the paw of a locomoting cat--facilitates the flexion reflex during the swing phase and the extension reflex during the stance phase. Both are adaptive responses and "selective potentiation is the agent of behavioral harmony" (p. 279).

Buy why--we may ask--should a reflex or any other putative element constitute a building block of motivated behavior? And on what grounds would we select (or potentiate) one unit over another. Consider as a test case the work of Sherrington, which the author uses to promote the reflex unit. Sherrington's reflex hypothesis was an attempt to describe a type of mechanism to explain how the central nervous system accomplished some of its integrative function (see Swazey, 1969). However, Gallistel does not tell us about the reflex hypothesis: rather the reflex is characterized as one of the elementary units of behavior. Apparently the author agrees with Skinner (1938) that a "reflex is not, of course, a theory. It is a fact. It is an analytical unit which makes the investigation of behavior possible" (p. 9). This is odd, for Sherrington himself asserted that reflexes do not exist, except for a very few non-functional cases such as the patellar reflex. In fact, Gallistel's book contains the relevant quote: "The simple reflex is a convenient, if not probable fiction" (Sherrington, in Gallistel, p. 22). If reflexes are one of the units of behavior and if, as Gallistel claims, more complex behaviors are constructed out of them, then reflexes had better exist, for if the building blocks of something do not exist, then that something cannot exist. Of course the concepts of reflex, servomechanism, and oscillator have been, and probably will remain, useful for developing intuitions about the way motor systems work. But that is not to say they are the stuff out of which organisms construct actions, or psychologists should construct theories of action.

A basic assumption behind the author's perspective is that the organization of action can be explained by physically realizable principles and processes (p. 6). Later on he castigates the information processing approach to cognitive psychology, with its emphasis on computer metaphors, as failing to come to grips with the problem of action: "The structure of overt computer action bears little if any interesting resemblance to the structure of animal

action" (p. 360). Gallistel is not alone in this view, but does he practice what he preaches? Not if his extensive use of computer terminology is anything to go by. "Central programs," for example, are "complex units of behavior" that figure heavily in Gallistel's explanations of purposive action. It is "The structure of these complex units of action and the structures that interconnect them [that] delimit the animal's behavioral options" (p. 391). There is not much internal consistency here: programs constitute the language of formal symbol manipulating machines (computers) not the language of physical principles. The failures of physiological connectionism are patched up with computer-metaphor connectionism; the old gap between muscles and motivation is simply replaced by a new gap between the physiologically irrelevant language of symbol manipulations and the physiologically embodied processes of action. Gallistel recognizes this problem, but his attempts to resolve it (as in his discussion of Deutsch's work) do not go far enough.

Units of action versus Units in action

It is interesting in this regard that physics, unlike biology and psychology, has largely abandoned the language of unitary mechanism and has replaced it with the concept of systems of interlocking dimensions. This is a necessary development, for what constitutes a unit at one level of analysis is merely a system of interrelated parts at finer grains of analysis. The concept of interlocking dimensions allows for physically realizable models that cut across several grains of analysis, whereas the units of action proposed by Gallistel are, at best, functional units of action at a single grain, losing their relevance at higher or lower levels of analysis. It is precisely this focus on understanding the systemic "relational dynamics" (to use Fentress's term) that motivated Bernstein (whose work is not discussed by Gallistel) and, later, Greene and Turvey (whose work is reviewed in Chapter 12) to promote the idea of "coordinative structures" as functional groupings of muscles constrained to act in a unitary fashion. Unlike reflexes, servomechanisms and the like, but like oscillatory systems, coordinative structures are units of action at any level of analysis, not merely units in actions. Evolution, development, and learning all play a role in economizing the tasks of the motor system via constraints that limit its operations to ranges of activity that can be behaviorally useful. In short, questions of mechanism (which Gallistel addresses) are not ontologically separate from questions of origin (which Gallistel, like most of psychology, chooses to ignore).

Much of Gallistel's synthesis of the locomotion literature fits the coordinative structure paradigm rather well, yet on the surface he is quite critical of the Bernsteinian approach as espoused by Greene and Turvey. On the one hand, Greene's mathematical development of Bernstein's idea is seen as "largely schematic," and Turvey's use of mathematical metaphors "opaque." On the other hand, the author recognizes that "the Turvey conceptualization has much in common with the one presented here" (p. 361). This is evident for all to see and it is a pity that some of the derogatory remarks (as well as some of the confusion) could not have been avoided, as perhaps would have been the case had the author consulted some of the later work of Turvey and his colleagues.

Towards the end of the book the author offers a self-indictment of his efforts that perhaps is too harsh: "I began" the author says "by trumpeting my commitment to a physically realizable account of the principles that organize animal action. I end by babbling about my mental image of New York" (p. 388). But the oscillator concept elaborated in Chapters 4, 5, and 12 is very elegant and stimulating indeed, and it may touch base with physically realizable principles more closely than Gallistel recognizes. Thus the newly emerging physical biology of Iberall and Yates recognizes living systems as composed of ensembles of coupled and mutually entrained oscillators. In this view, termed homeokinetic (cf. Iberall, 1978), the oscillatory behavior so common in biological systems is not owing to special mechanisms (like pacemaker neurons), but is a general physical property of systems undergoing energy flux. The beauty of an oscillatory design, of course, and its appeal to the theorist of action, is that a wide diversity of behavioral outputs (and kinematic detail) emerges from coupling processes, such as phase modulation, among interacting oscillators.

Since the link from physics to biology and psychology is still being forged (and resisted by some), one suspects that Gallistel's commitment to physical principles--admirable though it may be--will not be realized for a while. In fact, given psychology's rather limited efforts to actively develop any theory (never mind a theory) of action, it is not surprising that Gallistel's synthesis falls short of the mark. But, if this book motivates psychology to pick up the gauntlet, then Gallistel can claim no little success.

REFERENCES

Iberall, A. S. Cybernetics offers a (hydrodynamic) thermodynamic view of brain activities. An alternative to reflexology. In F. Brambilla, P. K. Bridges, E. Endroczi, & G. Henser (Eds.), Perspectives in endocrine psychobiology. London: John Wiley, 1978.

Skinner, B. F. The behavior of organisms. New York: Appleton-Century-Crofts, 1938.

Swazey, J. Reflexes and motor integration: Sherrington's concept of motor integration. Cambridge, Mass.: Harvard University Press, 1969.

SOME REFLECTIONS ON SPEECH RESEARCH*

Franklin S. Cooper

INTRODUCTION

It is a privilege indeed to give the introductory paper at this Conference on the Production of Speech. The topic is an important one, at the cutting edge of present-day speech research, so it is not surprising that several divergent paths are being followed. This meeting gives us an opportunity not only to compare recent findings but also to reexamine our research goals--to ask again what it is we are looking for.

In his letter of invitation, Peter MacNeilage suggested that I include a retelling--for his students, since the rest of you know the story--of how Haskins Laboratories became involved in speech research and how the initial work on perception developed into parallel research on speech production. Since the story starts from a conceptual context that is no longer familiar or is but dimly remembered, it seemed useful to go back to the still earlier events and ideas from which acoustic phonetics emerged some thirty odd years ago. So, in the first half of this talk, I have tried to cover very briefly the contributions of linguists and of engineers to concepts of speech that were current at the beginning of the fifties, and then to turn to events at Haskins Laboratories as a case history of how those concepts continued to evolve.

Who would not be tempted to push on from history to prognostication? I have tried to avoid that trap in the second half of the talk and, instead, to look at present-day research from a little distance--to reflect on where it seems to be going and how this follows from current concepts about the nature of speech. In doing so, I have found it instructive to think about the orientation of the research effort with respect to the processes by which speech flows from speaker to listener. But more of that later.

SPEECH RESEARCH TO THE NINETEEN SIXTIES

It is obvious, I suppose, that the topics we choose to talk about at conferences such as this depend on what we currently know and believe about speech. It was always so, but what was known was different twenty, or fifty, or a hundred years ago. Furthermore, what was known at any given time consisted of concepts as well as facts; indeed, only those facts agreeable to the concepts were likely to have been discovered or to have survived.

If, then, we want to understand the basis for our own research undertakings--the sometimes shaky ground on which we build--it may be better to trace back through the ideas that were held about speech rather than try to find our way through the forest of facts that surrounded them. But first, some words of warning: The trip will be a sketchy one. You must expect gaps, biases, and disproportionate attention to personal experience; also, less attention to credits and priorities than in a proper review.

EARLY IDEAS ABOUT SPEECH: Linguistic and Phonetic[1]

Now at any given time, ideas about speech depended on who held them. As of a hundred years ago, linguists and phoneticians were about the only people interested in speech and their concerns were with historical and family relationships among languages. Since they dealt mainly with written language, it is not surprising that the study of spoken language put emphasis on ways to "write" speech sounds. Thus, the IPA transcription system drew heavily on Henry Sweet's Broad Romic notation which, in turn, was indebted to Melville Bell's Visible Speech, a system of descriptive symbols to show deaf students how to articulate the sounds of speech. So, very early--and even earlier for Sanskrit--speech came to be thought about as a string of symbols. This view followed naturally from the way phoneticians dealt with speech, that is, by listening carefully and discovering by trial and error how to produce acceptable imitations. Thus, perception and production shared about equally in shaping the phonetician's concept of speech: perception gave irreducible units, production identified them with gestures, and the use of a notational system legitimized an underlying invariance, despite ubiquitous variation in the actual sounds. There have, of course, been changes in emphasis and genuine refinements of these ideas, but the framework remains.

One of the refinements dealt with the problem of variability by distinguishing among the kinds of variability: those that were distinctive and so made a difference in meaning, those that were systematic but not distinctive, and those that seemed just to happen. But even within these categories there was further variation when one considered actual speech sounds and this made it necessary to assume idealized entities, phonemic in nature, as counterparts of the erstwhile phonetic symbols. A further refinement attributed internal structure to the phoneme and came to characterize it as a bundle of distinctive features.

The interest of phoneticians and linguists in the production of speech very soon led to physiological experiments. These deserve our admiration for the ingenuity, even heroism, with which kymograph and tambours, Helmholtz resonators, and manometric flames were used to test and refine impressionistic ideas about specific sounds and how they were made. But the tools were then too crude to let experimental phonetics develop along lines of its own, and the better instruments that came with the nineteen twenties and thirties were mainly in the hands of engineers, who had quite different ideas about speech, as we shall see.

202

EARLY IDEAS ABOUT SPEECH: Communications Engineering[2]

Let us turn to the years following the First World War and to the revolution in communications technology that occurred in the twenties. Many things were new then that we now take for granted: radio broadcasting, talking movies, the rebirth of the phonograph, and even primitive attempts at television. Much of this was due to the vacuum tube amplifier, for the ability to amplify signals as weak as speech had many practical consequences.

One consequence was that speech itself became of interest to engineers: that is, there was a practical need for telephone engineers to know more about speech as a signal, since that is what a telephone must transmit. At the beginning of the twenties, speech was commonly viewed as a kind of "acoustic stuff"--complex in detail but essentially homogeneous on average: "a continuous flow of distributed energy, analogous to total radiation from an optical source. This idea of speech is a convenient approximation, useful in the study of speech reproduction by mechanical means" (Crandall, 1917).

But ideas changed as better tools became available. In the late twenties, a new high-speed oscillograph focused interest briefly on the waveform of speech (Fletcher, 1929). This soon gave way to interest in spectral representations and to the possibility that all speech sounds--not just vowels--could be described in terms of their "characteristic bands," that is, their prominent steady-state frequency components (Collard, 1930).

The conceptual shift from static components to a dynamically changing spectrum came rather slowly. In 1934, Steinberg published what is, in retrospect, the first speech spectrogram. But this one crude, schematic "spectrogram" of a single short sentence had required several hundred hours of hand measurement and computation, so it is easy to see why this way of representing speech--and of thinking about it--remained a curiosity for so long.

By the beginning of the next decade, a different way of thinking about speech--much closer to the views of phoneticians, but still rooted in engineering--was being proposed by Homer Dudley (1940). He explained speech by drawing an analogy with radio waves, which are not themselves the message, but only its carrier. So with speech: the message is the subaudible articulatory gestures that are made by the speaker; the sound stuff is only an acoustic carrier modulated by those gestures. This remarkable insight was obscured, for purely technical reasons, when it was embodied in hardware-- voder and vocoder--since the gestural component became a set of fixed filters and the point of view shifted from gestures back to spectra.

The influence of instruments on ideas is nowhere better illustrated than by the unveiling of the sound spectrograph (Potter, 1946). Now that spectrograms could be made in minutes, they had a profound effect on speech research. They provided, quite literally, a new way to look at speech, as well as new ways to think about it. One way, of course, was the familiar description in spectral terms, but with a new richness of detail. A second way was to view the spectrogram as a road map to the articulation. A third way was to view spectrograms simply as patterns. The richness of detail then became just a nuisance, since it obscured the underlying, simpler design.

203

You will have noticed that engineering ideas about speech, as of the late forties, treated it as primarily an acoustic phenomenon, an ongoing stream that is complex, variable in structure, and continually changing. This contrasts with phonetic ideas that viewed speech as a sequence of discrete entities. These phonetic units were of an ambivalent acoustic-articulatory nature, but they were unitary nevertheless and their symbols stood for some kind of underlying idealized entities.

ACOUSTIC PHONETICS: the Forties and Fifties.

This is about how things stood at the beginnings of the new science of acoustic phonetics. It is difficult to recapture either the conceptual currents or the sense of adventure of the late forties and early fifties. A few happenings from that period were the publication of Visible Speech with its catalog of spectrograms by Potter, Kopp, and Green (1947), and a classic interpretive account by Martin Joos (1948). At one of the early MIT Speech Conferences--happenings in their own right--Jakobson, Fant, and Halle (1951) circulated a draft of Preliminaries to Speech Analysis. This sought to round out the concept of Distinctive Features by showing their correlates in spectrographic as well as in articulatory and impressionistic terms. Then, too, there were new instruments, notably the speech synthesizers, and the ideas they fomented. More of this later.

First, who were the people at the speech conferences and what were their interests? Half at least came from engineering backgrounds and were interested in how the speech signal could be manipulated for practical communications purposes. Experimental psychologists were becoming interested in the perception of speech. Phoneticians, the few there were, were of course much interested in the new possibilities for describing speech sounds, but most linguists, especially of the American School, found little that seemed relevant to their concerns with theory and formal structures. One result of the imbalance, especially between linguists and engineers, was that the term "phoneme" lost its precision in discussions of speech research and was misused more often than not. Another consequence was that almost everyone, but especially the engineers, adopted without reservation the view that speech in its very nature was a succession of unitary sounds and that the invariances implied by phonemic symbols were actually there in the acoustic signals, if only one could find them. This idea was implicit--often explicit--in most of the research of that period, and is not unfamiliar to this day.

There was also, in the research of the forties and fifties, a preoccupation with the acoustic and receptive aspects of speech.[3] I recall rather little work, other than that of Stetson (1951), on physiological aspects of speech production, though there was much excellent research on the relationship of articulatory configurations to acoustic output (Fant, 1960; Stevens & House, 1955, 1956).

PERCEPTION TO PRODUCTION: a Case History

I should like now to abandon all attempts to trace the full range of ideas about speech into the sixties and seventies and turn to a more nearly

personal account of how one sequence of ideas evolved--between the forties and sixties--from a non-speech concern with sensory aids, via work on speech perception, to physiological research on speech production. Again, I beg your indulgence for retelling a story that is familiar to many of you.

Alvin Liberman and I discovered speech shortly after World War II. We were trying to build a reading machine for blinded veterans by turning letter shapes into distinctive acoustic shapes. In fact, that was fairly easy. The resulting acoustic alphabets were learnable, but they were essentially useless because reading with them was intolerably slow (Cooper, 1950). The irony of the situation finally came home to us: in talking about our problem, we were using with great facility a complex, high-rate sound system to ask why it was so hard to make a simple sound system work at all, even at moderate rates. Maybe the real problem was to find out how speech is perceived, and why so fast? We did two things that proved to be important: we built a speech synthesizer and with it we lured Pierre Delattre into working with us (Liberman & Cooper, 1972).

The Pattern Playback converted spectrograms back into sound--not quality speech but a fairly faithful rendering of the spectrum. The device was based on the very simple idea that spectrograms appeal to the eye because they reveal important spectral patterns in spite of a lot of acoustic clutter. So, if one could abstract the simple underlying patterns--by tracing them from spectrograms--and then play them back as sound, he could know by listening whether or not he had captured the essence of the speech. In the simplest case, the pattern elements that served as acoustic cues would be the invariants that correspond to the phonemes.

It was, in fact, possible to tease out sets of acoustic cues and even, by the mid-fifties, to use them in synthesizing speech "by rule" (from a phonemic text) rather than by copying spectrograms. But two things were puzzling: for one, the cues were rarely, if ever, truly invariant: for another, though they were indeed cues in the acoustic domain, they were not easy to describe or classify in conventional acoustic terms; rather, they seemed to fall naturally into articulatory categories. One reason why this might be so--an essentially trivial reason--is that the phonemic classification used in discovering the cues is itself based on articulation. Another more interesting reason could be that the perception of speech sounds is in fact based on the gestures by which speech is produced rather than on the sounds as acoustic entities (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Studdert-Kennedy, 1978).

A variety of mechanisms can be imagined by which this might happen. The particular hypothesis that led the Haskins group into research on speech production had its roots in Donald Hebb's ideas about neural nets (Hebb, 1949) and possible interactions between sensory and motor networks, though precise mechanisms have not been a feature of what soon came to be called a motor theory of speech perception. Actually, neither the theory nor the possible mechanisms were directly involved in the rationale for the research on speech production--only the hypothesis that the underlying units of speech are articulatory in their natures. If they are, then the chances that these units will emerge in recognizable form get better and better the farther one can go experimentally toward the origins of the neuromotor signals that drive

205

articulation.4 This led us to use electromyography for the study of muscle activity and to supplement it with analyses of movement (mostly by cineradiography) and, of course, spectrographic analyses of the acoustic signal.

This was the rationale for our research. Actually, in its early stages when Katherine Harris and Peter MacNeilage joined in the work, the ideas we were talking about were more concrete. The working hypothesis was that, if things were really simple, then features and phonemes might be characterizable by motor commands to those particular muscles mainly involved in the respective articulations, and also that EMG signals would reveal those motor commands. Various qualifications were built into what we said about these expectations: thus, no one could be sure about how much higher-level restructuring there might be between linguistic unit and explicit neuromotor signal. For the very simple situation we first studied--lip closure for the bilabial stops--even the simple hypothesis seemed adequate; further studies, though, showed context dependence and the need for a less simplistic explanation (Cooper, 1966; Harris, 1974; MacNeilage, 1970; MacNeilage & DeClerk, 1969; MacNeilage & Sholes, 1964). Invariance, like the Holy Grail, seems always to remain just out of reach.

The experience of the Haskins group in studying speech _perception_ explains one, though only one, of the reasons for a general shift toward research on speech _production_ and particularly toward attempts to provide a basis in motor organization for understanding the communicative role of speech. It would be interesting, if time allowed, to review various models that have been proposed for speech perception and production and for the relationships between them. Fortunately, this is not necessary for production models since an excellent review of just this topic has recently been published and its author is here with us (Kent, 1976).

Let me say again that this brief look backward at speech research was not intended as a review of the subject, not even a sketchy one; rather, it is my impression of how some of the important ideas about speech developed and, especially, how a new interest in speech production developed out of research on speech perception. Other people would have other views, but I think we might agree in a general way as to where we stand now, at the beginning of the eighties.

SOME REFLECTIONS ON CURRENT CONCEPTS AND OTHER MATTERS

We have by now amassed much factual knowledge about speech production. We have developed the tools for learning even more. But we do not yet have a satisfactory model, or an understanding, of how speech conveys language. Why should this be? Do the difficulties and complexities inhere in the problem-- that is, in the nature of speech processes--or rather in the ways we have chosen to think about the problem? The organizer of our conference has given me leave to reflect on some of these basic issues--at my own peril, of course. One hazard is being dogmatic--which brings to mind the moral of Thurber's fable about a city dog who visited his cousin in the country. The city dog, know-it-all that he was, ignored his country cousin's willingness to answer questions about the animals of the forest. So, from a porcupine, he learned about guided missiles--though not about discretion--and he learned about

chemical warfare from a little black and white animal that seemed only to be waving its tail in surrender. The country dog reflected, as his city cousin limped back to the safety of the alleys, that "sometimes it is better to ask some of the questions than to know all of the answers" (Thurber, 1940).

Even questions, if they are about fundamental issues, may lead one into talking about things so familiar that they seem altogether obvious. But, the obvious--that which you see when you see it--can sometimes be that which you do not see, really see, until it jumps out at you. So perhaps there are insights to be had even from questioning things long familiar.

Let us look first at coarticulation--surely as familiar a topic as one could find; next, at some consequences of differing orientations to this problem; and then at the role of timing in speech.

COARTICULATION: Problem or Pseudoproblem?

Coarticulation has been so much with us that it seems almost to have become an independent entity. Indeed, such comments as that certain speech behaviors "are due to coarticulation" seem even to imply that coarticulation caused them to happen. As a working definition, let us start with Hammarberg's view (1976) that "Coarticulation is...a process whereby the properties of a segment are altered due to the influences exerted on it by neighboring segments." The central implication is that the successive segments intended by a speaker will reappear in the acoustic signal, but with their ideal acoustic shapes changed to adapt them to the local context. The adaptations are not trivial; they are not mere smoothings at the boundaries, but often amount to complete restructuring of segments and clusters of segments. So it is not surprising that much effort has gone into accounting for these effects, or that coarticulation is commonly regarded as a central problem for research in speech production.

But the explanations one has to contrive for his data, using coarticulation as a conceptual framework, are becoming ever more complex, and there has been a growing unease about this over the past several years. Are the difficulties of data interpretation due, perhaps, to faulty conceptions? If so, where did we go astray? There are several possibilities, some of which I should like to consider with you.

One view puts the blame on choosing the wrong size of linguistic unit as the input segments of speech production. Phonemes or bundles of features have been the usual choices. Perhaps larger units such as the syllable or stress group would allow more felicitous explanations, though this has yet to be demonstrated.

A second view also puts the blame on units, in particular, that the units chosen were linguistic units. Rather, according to this view, there is need for units of a different kind--for production units that are inherent in the articulatory process, just as comparable units inhere in other skilled motor behaviors. In this vein, MacNeilage and Ladefoged (1976) comment on the "inappropriateness of conceptualizing the dynamic processes of articulation itself in terms of discrete, static, context-free linguistic categories, such

as 'phoneme' and 'distinctive features'." They go on to say, "...there has arisen a need for new concepts to characterize articulatory function, concepts more appropriate to the description of movement processes than of stationary states."

Yet another view focuses on the properties of linguistic units, whether they be phoneme, feature bundle, or other canonical form. This view has been taken as a point of departure by Carol Fowler and her colleagues (Fowler, Rubin, Remez, & Turvey, 1980) in considering speech production in terms of coordinative structures. Although they do not challenge the use of units that are of the linguistic kind, they point out that the properties usually attributed to such units--that they are discrete and static--are in fact irrelevant to their linguistic function. This leaves the way open "to discover some way to characterize these units that preserves their essential linguistic properties, but also allows them to be actualized unaltered in a vocal tract and in an acoustic signal."

Let us, instead of following this line of argument, consider further the properties "discrete and static." Even if we do not challenge the attribution of such properties to abstract linguistic units, should we not question the assumption that these properties will survive intact all the transformations that are involved in the act of speaking, and emerge at the end of that process as properties of the articulatory and acoustic entities? We know from experience that speech entities do not have these properties, but was there really any basis for supposing that they would? or even that input units of whatever kind would reappear as output units of the same general size and kind?

Nevertheless, it is just these assumptions about the survival of segments that have trapped us into viewing speech as a succession of entities that ought to have retained their canonical forms, but could not for the merely practical reasons to which we give the name "coarticulation."

RESEARCH ORIENTATIONS AND THEIR CONSEQUENCES

A consequence of all the attention given to coarticulation has been to focus experimental work on the relationships between one stage and the next of the production process, i.e., on successive causes and effects as one looks downstream, following the flow of messages from their inception by a speaker to their acoustic realization as speech and to their eventual assimilation by a listener. Thus, much attention is being given to careful measurement of forces, motions, mechanical linkages and properties of the articulatory mechanism as a way to predict articulatory outcomes.

Such concerns have a long history, but it seems to me that the emphasis has shifted increasingly over the past several years toward this downstream orientation and away from an earlier upstream orientation. For that earlier orientation, i.e., looking upstream, the problems were different and so were the experimental paradigms--necessarily so, since theoretical orientation affects what one looks for in Nature quite as much as observations about Nature affect theory. Now, looking upstream means trying to guess what causes were responsible for the effects that one is now observing; for example, what

kind of neuromotor pattern would bring tongue tip to alveolar ridge regardless of jaw opening? and, for a longer leap, in what degree would such a neuromotor pattern reflect phonetic or phonemic units?

I am inclined to take seriously this distinction between upstream and downstream orientations toward speech research,[5] i.e., to consider it a real dichotomy, since it has consequences for both theory and practice. Let us consider some of these consequences, but without making value judgments or disparaging one research orientation merely because another may be in fashion.

<u>Differences of Method</u>. The obvious difference between the two orientations is one of method: downstream, one works from known cause to predicted effect; upstream, from known effect to a plausible cause. Now, guessing at causes is much chancier than figuring out effects just as in football passing is more venturesome than line-bucking, though it has more potential for yardage. The case can be made on historical grounds that upstream methods have contributed most of the advances to our knowledge of speech, though the method was most successful when the inferential leaps were small. The failures, when the attempted leap was all the way to a linguistic unit, were more spectacular, but even so they provoked good research and some careful thinking about theories and models.

<u>Differences in Models and Theories</u>. The nature of theories and models about speech is in fact much affected by the upstream vs. downstream orientation of the research. This is due in part to what we expect of a good model, in particular, the demand we make that it should have <u>both</u> predictive power and explanatory power. The former includes, of course, the capability to account for <u>all</u> effects in terms of their causes, not merely those more esteemed effects that were foretold. Also, predictive power implies an accounting that is as quantitative and as precise as may be--in the limit, a mathematical model.

Explanatory power seems intuitively desirable, though just what one means by "explanation" is not immediately evident. Perhaps the way Bridgman (1936) put it will meet our need: "Explanation consists merely in analyzing our complicated systems in such a way that we recognize in the complicated system the interplay of elements already so familiar to us that we accept them as not needing explanation."

Physics offers many examples of how models and theories differ in predictive and explanatory power: the Bohr atom was understandable, even believable, but in predictive power it was inferior to the much more opaque wave- and quantum-mechanical models. In optics, two distinct models were needed to achieve both prediction and explanation. Perhaps the classic extreme in predictive power is Einstein's formulation: $e = mc^2$. It predicts with precision, and it is admirably simple and parsimonious as well, but it <u>explains</u> absolutely nothing about how or why energy and matter can be interconverted.

There is, it would seem, an inherent incompatibility--perhaps a trading relation--between predictive power and explanatory power. Moreover, this characteristic of theories and models interacts with the orientation of research efforts. Thus, downstream efforts to account for effects and to do

so reliably and accurately leads almost inevitably to models that predict, but are often wanting in explanatory power. Sometimes this imbalance results from devising rules or formulae without due concern for a rationalizing mechanism; sometimes, it follows from complicating the mechanism past all understanding with more and more parameters and linkages. Of course, common sense should keep such efforts at realism from leading to a model so complex that it approximates the organism itself.

An upstream orientation is likely to depend heavily on analogies with known mechanisms for its inspired guesses, and so its models can be expected to explain better than they predict. But when rule systems are substituted for concrete mechanisms--a choice not excluded by upstream orientation-- explanatory power is retained only to the extent that the rules are well motivated. A more serious hazard, judging from experience, is the "black box" model, usually a block diagram. Models of this kind can "explain" almost anything--so long as one does not enquire too closely into the inner workings of certain components.

If there is a moral to be drawn from these observations about models, I suppose it is that one should remember the biases inherent in his own research orientation and try a little harder for a reasonable balance between explanation and prediction; also, that one should try to accept philosophically that he cannot expect both virtues in full measures from either his own model or those of his colleagues.

Orientation and the Problem of Relevance. The bias toward one or another kind of model is not the only consequence that follows from research orientation. Upstream from where we now are in studying speech production-- and I take our present stance to be at the level of observing neuromuscular and movement events--there is not much room left for direct physiological assessment of the causes for the events we observe, and so we must fall back on behavioral indicators. True, there is much yet to be done to complete the representation of speech at the neuromuscular-movement level, especially when feedback loops are included. Nevertheless, the main upstream goal is to find out how neural signals are put together to drive the motor events of speech. This forces one, however reluctantly, to think about those patternings of neural activity in relation to the structure of the speech message. We are, after all, attempting to account for purposeful motor behavior, and that can hardly be done without taking account of the purpose, namely, to convey a message. It might help if we knew the nature and properties of the entities that make up a message--though we might then fall into the error of expecting these entities to survive the downstream transformations into neuromuscular, configurational and acoustic representations of that message!

But if upstream research is obliged to be message oriented, that same compulsion keeps it from wandering away from the goal of understanding speech as communication. Does this restraint apply also to downstream research? Is it similarly constrained and guided? Not by its own nature, I think, since all manner of neuromuscular movement, and even acoustic events, challenge us to explore their cause-effect relationships. But only a limited set of these challenges lie on the critical path to an understanding of how speech conveys messages. It is no derogation of, say, motor behavior to assert that not all of it is relevant to speech, and especially to speech as communication.

210

Where can one find guidance? Probably not--as both logic and experience would warn us--by looking within a particular representation for entities and/or properties that properly belong to the message itself in its original form. Since this warning applies also to the terminal representation--the acoustic signal--all we have left are perceptual criteria; that is, if we wish to assess the relevance of a production event, we must ask a listener whether it does or doesn't make a difference in the message--a difference at some linguistic level. All this does not imply, of course, that perceptual tests should regularly be incorporated into production research; rather, that thinking about perceptual relevance when planning production experiments will help to keep the research on target. It may seem ironic that whether we try to go downstream or upstream we do not escape linguistic units, or some entities very like them. Perhaps we must learn to live with them.

Coarticulation again--and Relevance. It was, you may remember, coarticulation that led us into these reflections on research orientations and their consequences. Are there consequences for coarticulation itself? It had already been found suspect as a conceptual framework because it depended so heavily on the reincarnation of presumed input units, entities which were not themselves above suspicion. It now seems necessary to look carefully even at those phenomena that are loosely called "coarticulation effects." To what extent are they still a central concern of speech research, or even relevant to it, if one hews to the line of communicative function? The intent of the question is not to imply a negative answer, but rather to suggest that such phenomena should be scrutinized as to relevance before they are investigated in detail, at least under the banner of speech research.

TIMING OF SPEECH EVENTS

Let me turn to another topic--timing--in some of its several aspects. Relative timing is generally considered an important aspect of speech production. Indeed, some of the recent approaches such as Action Theory give it a central place. Also, in some recent experiments--as well as in many older ones--we see anew how close is the relationship between production and perception.

Duration. It is an easy step, by equally easy assumptions, from the relative timing of speech events to the durations of individual events. There is in fact a considerable literature about durations, much of it flawed by the easy assumptions I have just mentioned. The most transparently questionable one is that the durations of individual phones is to be found by subdividing the total duration of the string into successive intervals--which is the same as supposing that phones do not overlap along the time axis. To put the same point another way, paralleling questions about coarticulation, is it reasonable to suppose that whatever inherent duration a phoneme might have would survive all the transformations between its central and its acoustic embodiments? Even if it did, could one expect that just those acoustic segments that are easiest to measure would be those that truly "belong" to the consonants and vowels?

Relative Timing. But the relative times at which events are initialized is a feature of almost every model of speech production. Are there ways to

observe what this initial timing might be? Could we, for example, get people to tell us when things happen? Some recent--and very neat--experiments follow on from the observation of Morton, Marcus, and Frankish, (1976), that listeners hear acoustically isochronous digit sequences as anisochronous. In these follow-on experiments, talkers were asked to produce isochronous sequences of syllables with the same, and also with alternating, initial consonants. Even though those sequences that were spoken with alternating initial consonants were not isochronous by acoustic measures, they were judged by listeners to be evenly paced. "The findings," to quote Carol Fowler and her colleagues (Fowler, 1979; Tuller & Fowler, 1980), "suggest that listeners judge isochrony on the basis of acoustic information about articulatory timing rather than on some articulation-free acoustic basis." It will not surprise you to hear that electromyographic measures support this idea. They show that talkers are indeed pacing their gestures, not the sounds they make.

Such uses of electromyography to get at the relative timing of articulatory events has some noteworthy advantages as compared with measures of movement and acoustic output, though all these measures in combination are essential to fully specify an articulatory gesture. Arguments in support of electromyographic measures are that the onset of electrical activity in a muscle is usually easier to detect with precision than the onset of the consequent movement; also, the electrical activation of several different muscles that participate in a single movement can be sorted out and timed separately, and so more easily and accurately than the components of the movement can be timed. Acoustic events, although some of those due to occlusions and releases can be timed with precision, are as a class only loosely coupled to the onsets of the motor events of articulation, and so provide only indirect information about the organization of motor control.

There is, in addition to these pragmatic considerations, a persuasive rationale for the use of electromyography in studying the relative timing of articulatory events, namely, that electromyography marks rather directly the time of execution--though not the magnitude--of motor commands from which the happenings downstream eventuate. To put it another way, measures of timing that are taken downstream (on movement and acoustic events) will often be less reliable or interpretable since they are likely to be contaminated by factors that operate after--and so do not affect--electromyographic measures of timing.

Even so, it is sometimes argued that one cannot safely make inferences upstream without full knowledge of all downstream consequences because these consequences may affect what one is observing at any given level and attempting to explain from above. This is a very general, almost philosophical, point which one cannot totally reject--because sometimes it has merit--but cannot fully accept either, because it counsels the despair of indefinite delay: the dismal prospect that one cannot even look upstream until he has learned all about everything downstream. Perhaps a practical approach is to examine carefully how speech is represented at the particular level under study. Is the representation reasonably complete? Are its parts reasonably independent of each other? and of subsequent representations? For EMG, the relative timing part of the representation seems to meet these criteria--with one proviso--though the relative magnitude part often does not. The proviso has to do with feedback loops that might introduce differential delays between observed and presumed timing.

212

In commenting on timing as a part of Action Theory, I can be quite brief because that topic will be dealt with later in this conference. Let me mention only one point: If timing is taken to be an inherent part of the central representation of speech units--whatever they are--then the problem of serial ordering (as it was put by Lashley) simply disappears and with it the special machinery required to actualize the units on schedule. These issues are developed in an incisive way in a recent article in the Journal of Phonetics (Fowler, 1980). Even if that view of timing proves to have other, equally troublesome, problems, at least it is a move away from complex timing mechanisms as the stuff from which models of speech production are made.

Surely there are many other questions that ought to be asked about other topics, but let me bring to a close these reflections of an old country dog, and thank you for your attention.

REFERENCES

Bridgman, P. W. The nature of physical theory. Princeton, N.J.: Princeton University Press, 1936 (NY: Dover Publications, p. 63).

Collard, J. Calculation of the articulation of a telephone circuit from the circuit constants. Electronic Communication, 1930, 8, 141-163.

Cooper, F. S. Research on reading machines for the blind. In P. A. Zahl (Ed.), Blindness: Modern approaches to the unseen environment. Princeton, N.J.: Princeton University Press, 1950.

Cooper, F. S. Describing the speech process in motor command terms. Haskins Laboratories Status Report on Speech Research, 1966, SR-5/6, 2.1-2.27.

Cooper, F. S. Acoustics in human communication: Evolving ideas about the nature of speech. Journal of the Acoustical Society of America, 1980, 68, 18-21.

Cooper, F. S., Liberman, A. M., Harris, K. S., & Grubb, P. M. Some input-output relations observed in experiments on the perception of speech. In Proceedings of the Second International Congress on Cybernetics. Namur, Belgium: Association Internationale de Cybernetique, 1958.

Crandall, I. B. The composition of speech. Physical Review, 1917, 10, 74-76.

Dudley, H. The carrier nature of speech. Bell System Technical Journal, 1940, 19, 495-515.

Fant, C. G. M. Acoustic theory of speech production. The Hague: Mouton, 1960.

Fletcher, H. Speech and hearing. New York: Van Nostrand, 1929, 26-46.

Fowler, C. A. "Perceptual centers" in speech production and perception. Perception & Psychophysics, 1979, 25, 375-388.

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

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

Fry, D. B. Phonetics in the twentieth century. In T. A. Sebeok (Ed.), Current trends in linguistics (Vol. 12, Part 4, 2201-2239). The Hague: Mouton, 1974.

Hammarberg, R. The metaphysics of coarticulation. Journal of Phonetics, 1976, 4, 353-363.

Harris, K. S. Physiological aspects of articulatory behavior. In T. A. Sebeok (Ed.), Current trends in linguistics (Vol. 12, Part 4, 2281-2302). The Hague: Mouton, 1974.

Hebb, D. O. The organization of behavior. New York: Wiley, 1949.

Jakobson, R., Fant, C. G. M., & Halle, M. Preliminaries to speech analysis: The distinctive features and their correlates. Cambridge, Mass.: MIT Press, 1951.

Joos, M. Acoustic phonetics. Language Monograph No. 23, Vol. 24, Suppl., 1948.

Kent, R. D. Models of speech production. In N. J. Lass (Ed.), Contemporary issues in experimental phonetics. New York: Academic Press, 1976.

Liberman, A. M., & Cooper, F. S. In search of the acoustic cues. In A. Valdman (Ed.), Papers in linguistics and phonetics to the memory of Pierre Delattre. The Hague: Mouton, 1972, 329-338.

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

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

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

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

MacNeilage, P. F., & Ladefoged, P. 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, 1976, 90.

MacNeilage, P. F., & Sholes, G. N. An electromyographic study of the tongue during vowel production. Journal of Speech and Hearing Research, 1964, 7, 209-232.

Morton, J., Marcus, S., & Frankish, C. Perceptual centers (P-centers). Psychological Review, 1976, 83, 405-408.

Potter, R. K. Introduction to technical discussions of sound portrayal. Journal of the Acoustical Society of America, 1946, 18, 1-3. See also the five related articles that follow this introduction.

Potter, R. K., Kopp, G. A., & Green, H. C. Visible speech. New York: Van Nostrand, 1947.

Steinberg, J. C. Application of sound measuring instruments to the study of phonetic sounds. Journal of the Acoustical Society of America, 1934, 6, 16-24.

Stetson, R. H. Motor phonetics. Amsterdam: North Holland Publishing Company, 1951.

Stevens, K. N., & House, A. S. The development of a quantitative description of vowel articulation. Journal of the Acoustical Society of America, 1955, 27, 484-493.

Stevens, K. N., & House, A. S. Studies of formant transitions using a vocal tract analog. Journal of the Acoustical Society of America, 1956, 28, 578-585.

Thurber, J. The scottie who knew too much. In Fables for our time and famous poems illustrated. New York: Harper, 1940.

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

214

FOOTNOTES

[1]For a broad-ranging review of this topic, see D. B. Fry, Phonetics in the twentieth century. In T. A. Sebeok (Ed.), Current trends in linguistics (Vol. 12, Part 4). The Hague: Mouton, 1974, 2201-2239.

[2]Condensed from a brief review presented at the 50th Anniversary Celebration of the Acoustical Society of America, June 12, 1979. See Cooper, 1980.

[3]Thus, Jakobson, Fant, and Halle (1951, p. 12) comment in their "Preliminaries..." that "the closer we are in our investigation to the destination of the message (i.e. its perception by the receiver), the more accurately can we gage the information conveyed by its sound shape. This determines the operational hierarchy of levels of decreasing pertinence: perceptual, aural, acoustical and articulatory (the latter carrying no direct information to the receiver). The systematic exploration of the first two of these levels belongs to the future and is an urgent duty."

[4]This is just the opposite of the strategy described in the quotation from Jakobson, Fant, and Halle (Footnote 3). For an early account of the production-oriented strategy, see Cooper et al., 1958.

[5]The parallels with inductive and deductive inference will be obvious; however, these terms imply an emphasis on method, per se, whereas I wish to stress the vector relationships between method and process, i.e., the orientation of research aims to speech flow.

ON LEVELS OF DESCRIPTION IN SPEECH RESEARCH*

Bruno H. Repp

Abstract. Many researchers use linguistic category names (conso-
nants, vowels, syllables) to refer to observations and measurements
made in records of the acoustic speech signal. The present paper
serves as a reminder that linguistic categories are abstract and
have no physical properties, and that, therefore, their physical
correlates in the speech wave are appropriately described in acous-
tic terms only.

Every branch of science needs a precise terminology to describe the
phenomena it is investigating. If there are different levels of observation,
different terms must be applied at each level in order to avoid confusion.
For example, the psychologist must distinguish the perceptual category "red"
from the neurophysiological processes that lead to the percept; and they in
turn must be distinguished from the energy and wavelength of the light that
impinges on the retina. If redness were a physical property of the light
wave, it would be difficult to explain why, for example, a certain wavelength
is called "red" by one viewer but "orange" by another and "gray" by a third
(who happens to be color-blind).

Scientists concerned with speech must be especially careful because there
are at least six different levels of description, each requiring its own
separate set of terms: articulation, acoustic waveform, neurophysiological
processes, conscious percept, nonlinguistic auditory impressions, and abstract
linguistic theory. Unfortunately, the mixing of terms from different levels
is a common practice of speech scientists. In particular, perceptual-
cognitive (phonetic, linguistic) categories are often applied to acoustic
observations. It is the purpose of the present paper to discourage this
usage, as far as possible.

Terms such as "vowel duration", "fricative amplitude", "syllable onset",
"/p/ duration", etc. abound in the literature. The measurements referred to
by these terms are made on spectrograms or oscillograms, i.e., on graphic
records of an acoustic waveform. Thus, they concern (the visual correlates
of) acoustic segments, such as periods of periodicity, noise, or silence. Why

---

217

do so many researchers use linguistic categories (vowels, consonants, syllables) to describe these acoustic segments? Is it just carelessness, or does it reflect some incorrect assumptions about the nature of phonetic segments?

One possibility is that underlying this usage of terms is a theory of speech segmentation that considers linguistic categories as a classification system for acoustic segments that are arranged like beads on a string. This view was widely held until the advent of the sound spectrograph; however, it has long been proven to be false. There is no one-to-one (or even many-to-one) correspondence between acoustic and linguistic segments; rather, the acoustic information for successive linguistic units overlaps and interacts. This fact has been referred to as "encodedness" or "parallel transmission of information" (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). It is a consequence of the complex dynamics of articulation. Although the input to the articulatory system may consist of a sequentially arranged string of abstract linguistic units (this is a hypothesis, not a fact), the articulatory movements corresponding to these units are no longer strictly sequential, and they are subject to passive as well as planned contextual variation. While discontinuities in the acoustic output may directly reflect changes in the state of the articulators and of the larynx, it is a serious mistake to consider them as boundaries of linguistic segments (cf. Fant, 1962).

Since these facts are by now generally accepted, it seems unlikely that any serious researcher would still espouse a naive beads-on-a-string theory. However, it is important to keep in mind that this conception remains the natural choice of anyone who reflects upon the structure of speech without ever having inspected a record of its acoustic waveform. Lax use of terms by professional scientists encourages such misconceptions and impedes the task of getting the facts across to students and the interested public.

Being aware of these facts, many speech scientists nevertheless use linguistic terms (consonants, vowels, syllables) as if they were acoustic categories--a classification of speech sounds. Perhaps, this malpractice originated with the time-honored but quite misleading term, speech sounds. For, patently, we do not normally perceive a sequence of sounds when we listen to speech but a linguistic message in which phonetic segments are the smallest units. These units are abstractions. They are the end result of complex perceptual and cognitive processes in the listener's brain, and it is likely that, excluding certain laboratory tasks, they are in fact not perceptual primitives but are derived by cognitive analysis from larger units, such as syllables or words (cf. Foss & Blank, 1980). Moreover, it appears that their conscious perception presupposes familiarity with an alphabetic writing system (Morais, Cary, Alegria, & Bertelson, 1979). That is, except for the rare preliterate individual who arrives at some rough approximation through intense reflection upon the nature of speech and language (witness the uniqueness of the invention of the alphabet!), awareness of the linguistic segment inventory generally derives from the experience of learning to read and write alphabetically (Lüdtke, 1969), and thus is heavily influenced by the spelling system of a language. Linguistic segments are important concepts for describing and explaining language structure. However, whether units corresponding to these abstract categories play any role at a subconscious level in ongoing speech perception is an open question; certainly, they could not do so as abstract categories which are, by definition, post-perceptual. It seems likely that

the structures utilized by the perceptual system require an entirely different (and novel) set of descriptors.

Abstract linguistic segments (the traditional "speech sounds") must be distinguished from the actual sounds of speech. These sounds can be described only in auditory terms, such as "hiss", "buzz", "silence", etc. Our vocabulary to describe these auditory impressions is rather limited (see, however, Pilch, 1979, for an attempt to organize and enrich it). These auditory qualities of the speech wave usually go unnoticed because the listener's attention is focused on the linguistic message. Considerable attention and experience are required to gain access to the auditory properties of speech, particularly to those aspects that support phonetic perception (as contrasted with suprasegmental characteristics such as intonation or voice quality that are more readily brought into awareness). Psychologists have been interested in this fact, as shown by the numerous studies of "categorical perception" which assess the (in)ability of listeners to discriminate speech stimuli on an auditory basis.

Acoustic aspects of the speech waveform do have a rather close relation to the auditory qualities perceived by a careful listener, but the relationship between acoustic segments and phonetic percepts (i.e., linguistic categories) is more complex. In general, several acoustic segments are relevant to the perception of a single phonetic segment, and each individual acoustic segment typically contains information about more than one phonetic segment. A phonetic category is not just a label attached to a particular combination of acoustic segments; for example, stop consonants in initial, medial, and final position have quite different acoustic correlates. Nor is it a label attached to the particular auditory qualities of the relevant acoustic segments, singly or in combination. Nor is it, strictly speaking, a classification of articulatory maneuvers or positions. Rather, a phonetic category is a perceptual-cognitive state resulting from the integration of diverse acoustic information into a unitary percept according to principles that are specific to phonetic perception and are best explained by reference to the articulatory origin of the speech signal. Alternatively, and perhaps more commonly, awareness of phonetic segments follows lexical access and thus results from cognitive analysis following primary perception (cf. Foss & Blank, 1980). That is to say that special perceptual and cognitive processes intervene between the acoustic signal and the phonetic percept. Therefore, phonetic categories--consonants, vowels, and even syllables--cannot be said to be in the acoustic signal. They have no physical properties--such as duration, spectrum, and amplitude--and, therefore, cannot be measured. (The properties they do have, such as distinctive features, are equally abstract; see Parker, 1977, for an excellent discussion of this issue.) The acoustic signal only contains the information that supports their perception; this information can be described (e.g., in terms of acoustic segments or "cues") and measured along acoustic dimensions.

Some might want to argue that vowels and consonants are in the signal but in a shingled, interwoven fashion. In other words, a phonetic segment could be defined as the totality of all acoustic cues that support its perception. Such an operational definition, while reasonably unambiguous, still commits a category error because it ignores the perceptual and cognitive processes that intervene between acoustic cues and phonetic percept. For example, if one

(e.g., in a study of the "phoneme restoration effect"--Warren, 1970; Samuel, in press) "removes a consonant" from an utterance by gating out certain portions of a speech signal, what is eliminated is the information that supports perception of the consonant. To state that the consonant has been removed from the waveform would not be proper; indeed, it might be misleading because it suggests (incorrectly) that only information pertaining to the consonant has been removed.

It would be unrealistic to demand that terms such as "vowel duration" and "fricative amplitude" be banned forever. However, I would like to urge researchers (1) to avoid them whenever possible, and (2) if they are to be used, to define precisely in acoustic terms what they are intended to refer to. It is by no means true that a seemingly innocuous term such as "vowel duration" has a generally agreed-upon interpretation in every context (see Lisker, 1974). Only if a vowel occurs in isolation is there no ambiguity. In the utterance /ba/, on the other hand, does vowel duration include the initial formant transitions which support the perception of the stop consonant? In /pa/, does it include the period of aspiration following the labial release? (If vowel duration is treated as a perceptual, not acoustic, quantity, these become legitimate empirical questions--cf. Raphael, Dorman, & Liberman, 1980.) In most cases, only terms such as "periodicity", "aspiration noise", "release burst", and "formant transitions" (including a suitable criterion for their beginning or end) permit an unambiguous specification of what is being measured. Once such a specification is provided by an author, and only then, the term "vowel duration" may be acceptable for the sake of convenience, although "duration of periodicity" (or whatever acoustic term is appropriate in a given context) would be preferable.

There are differences in the degree to which various misapplications of linguistic terms are inappropriate. This degree roughly parallels the dimension of "encodedness". For example, "fricative duration" will in most cases be unambiguously understood as referring to the duration of the noise (frication) portion of a stimulus, although the formant transitions in the surrounding acoustic segments contribute to the fricative percept (Harris, 1958; Whalen, 1981) and thus are part of the set of relevant cues. However, the noise is not "the fricative", and to call it so is awkward, at the least. Much more confusion is created by a term such as "stop consonant duration". While, in medial position, many will understand the term to refer to the period of relative silence resulting from oral closure (even though this is only one of several relevant acoustic cues), in utterance-initial position it might refer to the release burst alone, or the burst plus aspiration, or the burst plus aspiration plus formant transitions; in utterance-final position, it might refer to the formant transitions only (if the stop is unreleased) or to the period of silence with or without the release burst and/or the transitions (if the stop is released); and in an utterance such as /ækt/, with the first stop unreleased, it is not clear at all where the first stop ends and the second stop begins. Therefore, this term should not be used at all, not even after describing exactly what is being measured; instead, specific acoustic terms should be used throughout.

This request is not nearly as radical as it may seem. Definition of acoustic segments in purely physical terms can be cumbersome, e.g., "the periodic portion following the fricative noise". It is quite legitimate,

therefore, to name the linguistic segment for which a given acoustic segment is the primary cue, as long as the main term is physical in nature, e.g., "the 'u' periodic portion", "the 'p' silence", or "the 's' noise". Consistent use of such a terminology should place only a minor burden on researchers accustomed to speak loosely of "/p/ duration" or "/s/ amplitude"; however, it would greatly increase the clarity of many research reports.

Clearly, many of these arguments have been presented before (see especially Fant, 1962; Lisker, 1957, 1974; Parker, 1977; Pilch, 1974; Zwirner & Zwirner, 1970). However, they seem to have had little impact and, therefore, are worth repeating. Examples of terminological carelessness still abound in the literature. To quote just one recent example from an otherwise excellent paper: Mills (1980) states, referring to utterance-initial consonants (and without further qualification), that ".../s/ has a lower amplitude than /b/" and ".../s/ is longer in duration than /b/" (p. 82). Similarly awkward or outright misleading statements can also be found in the pages of this Journal* (see, e.g., Umeda, 1977). Although there are, of course, many authors who take great care to avoid such terminological confusion, I suspect that they are not in the majority. I hope the present note will draw attention to this problem and contribute to its gradual elimination.

REFERENCES

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

Foss, D. J., & Blank, M. A. Identifying the speech codes. Cognitive Psychology, 1980, 12, 1-31.

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

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

Lisker, L. Linguistic segments, acoustic segments, and synthetic speech. Language, 1957, 33, 370-374.

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

Lüdtke, H. Die Alphabetschrift und das Problem der Lautsegmentierung. Phonetika, 1969, 20, 147-176.

Mills, C. B. Effects of context on reaction time to phonemes. Journal of Verbal Learning and Verbal Behavior, 1980, 19, 75-83.

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

Parker, F. Distinctive features and acoustic cues. Journal of the Acoustical Society of America, 1977, 62, 1051-1054.

Pilch, H. Phonemtheorie (3rd ed.). Basel: Karger, 1974.

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

Raphael, L. J., Dorman, M. F., & Liberman, A. M. On defining the vowel duration that cues voicing in final position. Language and Speech, 1980, 23, 297-307.

---

*The Journal of the Acoustical Society of America.

Samuel, A. G. Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, in press.

Umeda, N. Consonant duration in American English. *Journal of the Acoustical Society of America*, 1977, *61*, 846-858.

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

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

Zwirner, E., & Zwirner, K. *Principles of phonometrics*. University of Alabama Press, 1970.

A NOTE ON THE BIOLOGY OF SPEECH PERCEPTION*

Michael Studdert-Kennedy+

The goal of a biological psychology is to undermine the autonomy of whatever it studies. For language, the goal is to derive its properties from other, presumably prior, properties of the human organism and its natural environment (cf. Lindblom, 1980). This does not mean that we should expect to reduce language to a mere collection of non-linguistic capacities in the individual, but it does mean that we should try to specify the perceptual and motor capacities out of which language has emerged in the species. The likelihood that this endeavor will go far with syntax in the near future is low, because we still know very little about the perceptuomotor principles that might underlie syntactic capacity--that is why current study of syntax is, from a biological point of view, descriptive rather than explanatory. But the prospects are better for phonology, because phonology is necessarily couched in terms that invite us to reflect on the perceptual and motor capacities that support it.

As we come to understand the extralinguistic origins of the sound pattern of language, we may also come upon hypotheses as to its perceptuomotor mechanisms. Those hypotheses must be compatible with (and may even derive from) our hypothesis as to phylogenetic origin. If we forget this, we risk offering tautology as explanation, because we are tempted to attribute descriptive properties of language to the organism rather than functional properties of the organism to language (cf. Turvey, 1980). I believe that this happens at several points in the otherwise excellent discussions of infant and adult speech perception by Eimas (in press) and of hemispheric specialization by Morais (in press). Both authors, at some point, take a descriptive property of language, its featural structure, and attribute a matching mechanism of featural analysis to the language perceiver. This, of course, is mere tautology. Plausible hypotheses as to the nature of the perceptual mechanism must await a deeper understanding of the functions and extralinguistic origins of linguistic structure.

Consider, in this light, the data and inference that have led to current interest in features and the perceptual mechanisms that supposedly extract them from the signal. The story begins with early studies intended to define the acoustic boundaries of phonetic categories (e.g., Cooper, Liberman, Delattre, & Gerstman, 1952). The experimental paradigm entailed synthesizing a consonant-vowel syllable, varying some property, or set of properties, along an acoustic continuum from one phonetic category to another, and then calling on listeners to identify or to discriminate between the syllables. Since the end-point syllables typically differed from each other by a single phonetic feature, such as manner or place of consonant articulation, the procedure served to specify an acoustic correlate of that feature.

As is well known, listeners typically divide such a continuum into sharply defined categories and, when asked to discriminate between syllables, do well if the syllables belong to different categories, badly if they belong to the same category, so that a peak appears in the discrimination function at the boundary between categories. This phenomenon, termed "categorical perception," was of interest for several reasons. First, it was believed to be peculiar to speech; second, it was assumed to be the laboratory counterpart of the process by which listeners categorize the acoustic variants of natural speech; third, the sharp categories and poor within-category discrimination hinted at some specialized mechanism (such as analysis-by-synthesis or a feature detecting device) for transforming a physical continuum of sound into the abstract, opponent categories that are the stuff of phonetic and phonological systems.

In due course, the experiments of Eimas and his colleagues, using "high amplitude sucking" with infants and selective adaptation with adults, led to an explicit model of categorical perception, in particular, and of phonetic perception, in general. This work has already stimulated almost a decade of invaluable research from which there has emerged a preliminary taxonomy of the infant's perceptual capacities for speech. However, the model that the research has inspired is weak on several counts. In its early versions, the model invoked devices for extracting abstract, phonetic features; later versions, faced with accumulating evidence of contextual dependencies in selective adaptation (e.g., Bailey, 1973), not to mention the unexpected skills of the chinchilla (Kuhl & Miller, 1978,), substituted acoustic for phonetic feature detectors (Eimas & Miller, 1978).

But consider the difficulties. First, we now know that categorical perception is not peculiar to speech, nor even to audition (e.g., Pastore, Ahroon, Baffuto, Friedman, Puleo, & Fink, 1977), so that students of speech perception are excused from postulating a specialized mechanism to account for it. Second, we have no grounds for supposing that the laboratory phenomenon of categorical perception has anything more important in common with the categorizing processes of normal listening than that they both involve classifying variants. The acoustic variations within categories of natural speech are either prosodic variants associated with a particular phone in a particular segmental context (e.g., [d] before [a]), spoken at different rates, with different stress and so on, or segmental variants, intrinsic to the production of a particular phone in different contexts (e.g., [d] before [a] or [i]). These are the types of variant that the listener has to categorize in natural speech, and neither of them is known to be mimicked by

the continua of synthetic speech. Indeed, acoustic variants that surround a phonetic boundary on a synthetic continuum (where all the interesting experimental effects appear, such as discrimination peaks and adaptive shifts in identification) may not only never occur in natural speech, but may even be literally unpronounceable (as in a synthetic series from [b] to [d], for example). They can hardly therefore operate as psychologically effective barriers to ensure a "quantal" percept (Stevens, 1972).

The third and most serious weakness is with the presumed role of acoustic feature-detecting devices in speech perception. As we have noted, the categorical perception paradigm typically manipulates a single dimension of the signal at a time to assess its contribution to a particular phonetic contrast. However, virtually every phonetic contrast so far studied can be cued along several distinct dimensions, and the various cues then enter into trading relations. The precise position of the boundary along a synthetic continuum for a given cue varies with the values assigned to other contributing cues. The most familiar instance comes from trading relations among cues to the voicing of syllable-initial stop consonants (e.g., Lisker & Abramson, 1964; Summerfield & Haggard, 1977), to which burst energy, aspiration energy, first formant onset frequency, fundamental frequency contour and the timing of laryngeal action all contribute. Other instances are provided by cues to the fricative-affricate distinction (Repp, Liberman, Eccardt, & Pesetsky, 1978), to stops in English fricative-stop-liquid clusters (Fitch, Halwes, Erickson, & Liberman, 1980) and in fricative-stop clusters (Bailey & Summerfield, 1980), and so on (for a preliminary review, see Liberman & Studdert-Kennedy, 1978). Are we to assign a new pair of opponent feature detectors (with contextually dependent, "tuneable" boundaries) to each new dimension that we discover? This may be difficult since, as several authors have remarked (e.g., Lisker, 1978; Bailey & Summerfield, 1980; Remez, Cutting, & Studdert-Kennedy, 1980), the number of isolable dimensions, relevant to any particular perceptual distinction, may have no limit.

We cannot escape from this reductio ad absurdum by positing fewer and higher order detectors, because the absurdity lies in the detectors, not in their proliferation. For example, the goal of Stevens' work (e.g., Stevens, 1975; Stevens & Blumstein, 1978) is to arrive at an integrated, summary description of the cue complex associated with each phonetic feature contrast. Thus, in his work on stops, Stevens describes various general properties of the whole spectrum, using the terminology of distinctive feature theory (e.g., grave-acute, diffuse-compact), and posits a matching set of acoustic "property detectors." This ensures that the number of supposed detectors will be no more than exactly twice the number of distinctive feature contrasts. However, by adopting the terminology of phonological theory, it also makes plain that we are dealing with tautology, not explanation.

The error in postulating detectors does not lie therefore in the claim that the signal undergoes analysis along several channels--that might even be true. Rather, the error lies in offering to explain phonetic capacity by making a substantive physiological mechanism out of a descriptive property of language. The error is attractive, because the feature or property detector has a veneer of biological plausibility: it promises to link language with ethology, on the one hand, through the trigger features of Tinbergen (1951; Mattingly, 1972) and the bird-song templates of Marler (1970), and with

physiology, on the other, through the selectively responsive cells of the bullfrog (Capranica, 1965), the cat (Whitfield & Evans, 1965), and the squirrel monkey (Wollberg & Newman, 1972). Yet, whatever the importance of this single-cell work to physiology, its psychological import is nil, since it merely supports the truism that some isolable and distinctive physiological event corresponds to every isolable and distinctive property of the physical world to which an organism is sensitive. The notion of innate song or call templates has even less to offer for an understanding of human language ontogeny. Such devices may ensure species recognition and successful repro- duction among organisms, such as the chaffinch and the bullfrog, which have brief or non-existent periods of parental care, and therefore, little or no opportunity to discover the marks of their species. But this is not the human condition. And, given the varied solutions to the problem of learning a species-specific song, even among closely related species of songbird (Kroods- ma, 1981), it is implausible to suppose that we can explain language ontogeny by invoking mechanisms proper to animals with a different ecology and for which we have no evidence in the human (for elaboration, see Studdert-Kennedy, 1981). What we should be asking instead is: What function does the capacity for perceptual analysis fulfill? Or, a little differently, what properties of the human organism force language into a featural structure?

Before I suggest an approach to this question, let me comment on another area of research where we run into a dead end, if we do not raise the question of biological function: hemispheric specialization. Morais (in press) brings together an impressive body of experimental findings from laterality studies, and shows conclusively that we simplify and gloss over discrepancies, when we characterize the left hemisphere as linguistic, the right as non-linguistic. He proposes to resolve the discrepancies by superordinate classification of the tasks at which the hemispheres excel, terming the left hemisphere "analytic," the right "holistic."

These descriptions certainly provide a fair partition of the reported data. But there are two objections to the proposal. First, it is too narrow, because it confines itself to the supposed perceptual modes of the hemis- pheres. Yet we act no less than we perceive: perception is controlled by, and controls, action. Therefore, it is the joint perceptuomotor processes that we should try to capture in a description of a hemispheric mode. Second, the proposal is too broad, because it does not consider the question of phylogenetic origin. Presumably, a behavioral mode (if there be such) does not evolve without a behavior to support. But Morais has no suggestions as to what that behavior might be. For my part, I am inclined to suppose that it might be language.

In any event, the linguistic capacities of the left hemisphere, in most individuals, are attested to by a mass of clinical and experimental data (e.g., Milner, 1974; Zaidel, 1978; Zurif & Blumstein, 1978). These capacities call for more than mere classification with supposedly kindred skills: they call for explanation. That is, they raise the question: What property of the left hemisphere predisposed it to language? Three items of evidence converge on a possible answer. First is the dominance of the left hemisphere in the motor control of speech for some 95% of the population. Second is the dominance of the left hemisphere in manual praxis for some 90% of the population. Third is the recent demonstration that American Sign Language

(ASL), the first language of some 100,000 deaf individuals in the United States, has a defining property of primary, natural languages: a dual pattern of formational structure ("phonology") and syntax (Klima & Bellugi, 1979). Presumably ASL uses the hands rather than, say, the feet, because the hand has the speed and precision to support a rapid, informationally dense signaling system of the kind that a language demands.

Taken together, these facts almost force the hypothesis that the primary specialization of the left hemisphere is motoric rather than perceptual. Language would then have been drawn to the left hemisphere because the left hemisphere already possessed the neural circuitry for control of fingers, wrists, arms and for unilateral coordination of the two hands in the making and use of tools--precisely the type of circuitry needed for control of larynx, tongue, velum, lips and of the bilaterally innervated vocal apparatus. (Perhaps it is worth remarking that the only other secure instance of cerebral lateralization is also for control of a complex bilaterally innervated vocal apparatus--in the canary [Nottebohm, 1977]).

The general hypothesis is not new. Semmes (1968), for example, proposed such an account of the cerebral link between speech and manual control. She argued from a study of the effects of gunshot lesions that the left hemisphere was focally organized for fine, sequential, sensorimotor control, while the right was diffusely organized for holistic perception and action. Recently, Kimura (e.g., Kimura & Archibald, 1974; Kimura, 1979) and Kinsbourne (e.g, Kinsbourne & Hicks, 1978) have carried the hypothesis further, looking for evidence of competition and facilitation between speaking and manual action. Current research is developing procedures and paradigms to increase the precision and rigor of such work (Kelso, personal communication).

What insight can this motoric view of language and hemispheric speciali- zation lend into the origins of phonetic features? Note, first, that the signs of ASL, no less than the syllables and segments of spoken language, can be economically described in terms of features (Klima & Bellugi, 1979). Moreover, the articulators of both vocal tract and hands are relatively few: most are engaged, even if only passively, in the production of every sign or syllable. An ample repertoire of units therefore calls for repeated use of the same gesture by the same articulator in combination with different actions of other articulators. These recurrent gestures are, we may surmise, the instantiation, alone or in combination, of phonetic features (Studdert-Kennedy & Lane, 1980). However, the features are not detachable entities; rather, they are recurrent properties or attributes of the signs and segments (Fowler, Rubin, Remez, & Turvey, 1980; Turvey, 1980; Bladon & Lindblom, in press). This view sits comfortably with recent evidence that metathesis tends to involve unitary phonetic segments rather than features (Shattuck-Hufnagel & Klatt, 1979). And from this we may well infer that, just as they are not put in, features are not taken out. That is to say, the perceived feature is an attribute, not a constituent, of the percept, and we are absolved from positing specialized mechanisms for its extraction.

None of what I have said above should be taken to imply that speech is not the peculiar and peculiarly efficient acoustic carrier of language. On the contrary, speech is peculiar and distinctive precisely because its processes of production and perception must have evolved _pari passu_ with

language itself. Just how speech gives the listener access to his language is still a puzzle, and not one that seems likely to be solved by bare psychoacoustic principle.

Let me illustrate with two recent experiments. First is a study by Fitch, Halwes, Erickson, and Liberman (1980), demonstrating the perceptual equivalence, in a speech context, of two distinct cues to a voiceless stop in a fricative-stop-liquid cluster: silence and rapid spectral change. These investigators constructed two synthetic syllables, [plɪt] and [lɪt], the first differing from the second only in having initial transitions appropriate to a labial stop. If a brief bandpassed noise, sufficient to cue [s], was placed immediately before these syllables, both were heard as [slɪt], but if a small interval of silence (long enough to signal a stop closure) was introduced between [s] and the vocalic portion, both were heard as [splɪt]. What is of interest is that the silent interval necessary to induce the stop percept was shorter when the vocalic portion carried transitions than when it did not. By systematically manipulating the duration of the silent interval before each of the two syllables, Fitch et al. titrated the effect of the initial transition and found it equivalent to roughly 25 msec of silence. Moreover, they demonstrated that these two diverse cues--silence and spectral shift--were additive (or multiplicative) in the sense that discrimination between [slɪt] and [splɪt] was close to chance when the cues were in conflict (e.g., a short interval + [plɪt], or a long interval + [lɪt]), but was facilitated when they worked together: a long interval + [plɪt] was usually perceived as [splɪt], a short interval + [lɪt], as [slɪt]. Presumably, the grounds of this spectral-temporal equivalence are simply that the duration of stop closure and the extent of a following formant transition covary in the articulation of a natural utterance. Certainly, there are no psychoacoustic grounds for expecting the equivalence, and we may therefore fairly conclude that it is peculiar to speech.

In fact, Best, Morrongiello, and Robson (in press) have demonstrated just this in an ingenious experiment using "sine-wave speech" (cf. Remez, Rubin, Pisoni, & Carrell, in press). Best and her colleagues constructed a sound from three sine waves modulated to follow the path of the center frequencies of the three formants of a naturally spoken syllable, [deɪ], in two forms: one form had a relatively long initial F_1 transition ("strong" [deɪ]), one had a relatively short initial F_1 transition ("weak" [deɪ]). Given a perceptual set for speech, some listeners identify these sounds as [deɪ] and [eɪ], while others hear them as different non-speech chords. If a suitable patch of noise is placed immediately before these sounds, they can be heard as [seɪ]; if a sufficient silent interval is introduced between noise and sine waves, a "speech" listener will hear [steɪ], and he will hear it with a shorter interval before "strong" [deɪ] than before "weak" [deɪ].

On this basis, Best et al. constructed two continua, analogous to those of the earlier experiments, varying silent interval in combination with one or other of the [deɪ] "syllables." To obtain identification functions without an explicit request for identification, they used an A X B procedure. In this procedure A and B are endpoints of a synthetic continuum. The task of the listener on each trial is to judge X as "more like A" or "more like B." Thus, despite the bizarre quality of their stimuli, Best et al. were able to obtain identification functions and to assess the perceptual equivalence of silence

and formant transitions in a manner analogous to that of the earlier /slɪt-splɪt/ studies. Their fifteen listeners divided themselves neatly into three groups of five. Two of these groups never heard the sounds as speech and demonstrated no perceptual equivalence between silence and spectral change: one group was sensitive to variations in silence, but not in frequency, the other to variations in frequency, but not in silence. Only the five listeners who heard the sounds as /seɪ/ or /steɪ/ demonstrated a trading relation between silence and spectral change.

The burden of this elegant study matches the conclusion drawn by Jusczyk (in press) from his review of infant research and by my colleague, Donald Shankweiler, and me some years ago from a dichotic study: "...the peculiarity of speech may lie not so much in its acoustic structure as in the phonological information that this structure conveys. There is therefore no reason to expect that specialization of the speech perceptual mechanisms should extend to the mechanisms by which the acoustic parameters of speech are extracted" (Studdert-Kennedy & Shankweiler, 1970, p. 590).

If this conclusion is correct, we may review the goals of those who hope to advance our understanding of the biological foundations of language by studying infants. Their proper task is not so much to establish psychoacoustic capacity as to track the process by which infants discover the communicative use and linguistic organization of the sounds they hear and the signs they see (cf. MacKain, Note 2). This is the species-specific, epigenetic process for which we shall find no counterpart in the chinchilla.

REFERENCE NOTE

1. Bailey, P. J. Perceptual adaptation for acoustical features in speech. Speech Perception (Department of Psychology. The Queens University of Belfast), 1973, Series 2, 29-34.
2. MacKain, K. S. On assessing the role of experience in infant speech perception. (Forthcoming)

REFERENCES

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

Best, C. T., Morrongiello, B., & Robson, R. The perceptual equivalence of two acoustic cues for a speech contrast is specific to phonetic perception. Perception & Psychophysics, in press.

Bladon, A., & Lindblom, B. Modeling the judgment of vowel quality differences. Journal of the Acoustical Society of America, in press.

Capranica, R. R. The evoked response of the bullfrog. Cambridge, Mass.: M.I.T. Press, 1965.

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

Eimas, P. D. Speech Perception: A view of the initial state and perceptual mechanisms. In Proceedings of C.N.R.S. Conference on Cognition (Abbaye de Royaumont, France, June 15-18, 1980), in press.

Eimas, P. D., & Miller, J. L. Effects of selective adaptation on the perception of speech and visual patterns: Evidence for feature detectors. In R. D. Walk & H. L. Pick, Jr. (Eds.), Perception and experience. New York: Plenum, 1978.

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

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), Language production. New York: Academic Press, 1980, 373-420.

Jusczyk, P. Infant speech perception. In Proceedings of C.N.R.S. Conference on Cognition (Abbaye de Royaumont, France, June 15-18, 1980), in press.

Kimura, D. Neuromotor mechanisms in the evolution of human communication. In H. D. Steklis & M. J. Raleigh (Eds.), Neurobiology of social communication in primates. New York: Academic Press, 1979, 197-219.

Kimura, D., & Archibald, Y. Motor functions of the left hemisphere. Brain, 1974, 97, 337-350.

Kinsbourne, M., & Hicks, R. E. Mapping cerebral functional space: Competition and collaboration in human performance. In M. Kinsbourne (Ed.), The asymmetrical function of the brain. New York: Cambridge University Press, 1978, 267-273.

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

Kroodsma, D. E. Ontogeny of bird song. In G. B. Barlow, K. Immelmann, M. Main, & L. Petrinovich (Eds.), Behavioral development: The Bielefeld interdisciplinary project. New York: Cambridge University Press, 1981.

Kuhl, P. K., & Miller, J. D. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. Journal of the Acoustical Society of America, 1978, 63, 905-917.

Liberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. Leibowitz, & H.-L. Teuber (Eds.), Handbook of sensory physiology, Vol. VIII: Perception. Heidelberg: Springer-Verlag, 1978, 143-178.

Lindblom, B. The goal of phonetics, its unification and application. Phonetica, 1980, 37, 7-26.

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

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

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

Mattingly, I. G. Speech cues and sign stimuli. American Scientist, 1972, 60, 327-337.

Milner, B. Hemispheric specialization. In F. O. Schmitt & F. G. Worden (Eds.), The neurosciences: Third study program. Cambridge, Mass.: M.I.T. Press, 1974.

Morais, J. The two sides of cognition. In Proceedings of C.N.R.S. Conference on Cognition (Abbaye de Royaumont, France, June 15-18, 1980), in press.

Nottebohm, F. Asymmetries in neural control of vocalization in the canary. In S. Harnad, R. W. Doty, L. Goldstein, J. Jaynes, & G. Krauthamer (Eds.), Lateralization in the nervous system. New York: Academic Press, 1977, 23-44.

230

Pastore, R. E., Ahroon, W. A., Baffuto, K. J., Friedman, C., Puleo, J. S., & Fink, E. A. Common factor model of categorical perception. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 686-696.

Remez, R. E., Cutting, J. E., & Studdert-Kennedy, M. Cross-series adaptation using song and string. Perception & Psychophysics, 1980, 27, 524-530.

Remez, R. E., Rubin, P. A., Pisoni, D. B., & Carrell, T. Speech perception without traditional speech cues. Science, in press.

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

Semmes, J. Hemispheric specialization: A possible clue to mechanism. Neuropsychologia, 1968, 6, 11-26.

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

Stevens, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), Human communication: A unified view. New York: McGraw Hill, 1972, 51-66.

Stevens, K. N. The potential role of property detectors in the perception of consonants. In G. Fant & M. A. A. Tatham (Eds.), Auditory analysis and perception of speech. New York: Academic Press, 1975, 303-330.

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

Studdert-Kennedy, M. The beginnings of speech. In G. B. Barlow, K. Immelmann, M. Main, & P. Petrinovich (Eds.), Behavioral development: The Bielefeld Interdisciplinary Project. New York: Cambridge University Press, 1981.

Studdert-Kennedy, M., & Lane, H. The structuring of language: Clues from the differences between signed and spoken language. In U. Bellugi & M. Studdert-Kennedy (Eds.), Signed language and spoken language: Biological constraints on linguistic form. (Dahlem Konferenzen). Weinheim/Deerfield Beach, Florida/Basel: Verlag Chemie, 1980, 29-39.

Studdert-Kennedy, M., & Shankweiler, D. P. Hemispheric specialization for speech perception. Journal of the Acoustical Society of America, 1970, 48, 579-594.

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

Tinbergen, N. A study of instinct. Oxford: Oxford University Press, 1951.

Turvey, M. T. The structuring of language. Clues from the organization of motor systems. In U. Bellugi & M. Studdert-Kennedy (Eds.), Signed language and spoken language: Biological constraints on linguistic form. (Dahlem Konferenzen). Weinheim/Deerfield Beech, Florida/Basel: Verlag Chemie, 1980, 41-56.

Whitfield, I. C., & Evans, E. F. Responses of auditory cortical neurons to stimuli of changing frequency. Journal of Neurophysiology, 1965, 28, 655-672.

Wollberg, Z., & Newman, J. D. Auditory cortex of squrrel monkey: Response patterns of single cells to species-specific vocalizations. Science, 1972, 175, 212-214.

Zaidel, E. Lexical organization in the right hemisphere. In P. A. Buser &
    A. Rougeul-Buser (Eds.), Cerebral correlates of conscious experience.
    Amsterdam: Elsevier/North Holland Biomedical Press, 1978, 177-197.
Zurif, E. B., & Blumstein, S. E. Language and the brain. In M. Halle,
    J. Bresnan, & G. A. Miller (Eds.), Linguistic theory and psychological
    reality. Cambridge, Mass.: M.I.T. Press, 1978, 229-245.

MORE ON DUPLEX PERCEPTION OF CUES FOR STOP CONSONANTS

Brad Rakerd+, Alvin M. Liberman++, and David Isenberg+++

Abstract. In an earlier experiment (Liberman & Isenberg, 1980) it was shown that when the vocalic formant transitions (appropriate for the stops in a synthetic approximation to [spa] or [sta]) were presented to one ear, and the remainder of the synthetic pattern to the other, listeners reported a duplex percept. One side of the duplexity was the same coherent syllable ([spa] or [sta]) that is perceived when the pattern is presented in its original, undivided form; the other was a nonspeech chirp that corresponds to what the transitions sound like in isolation. It was also shown that a period of silence between the fricative noise and the vocalic portion of the syllable was essential to the perception of the transitions when, on the speech side of the percept, they supported identification of the stops; but the silence had no measurable effect on those same transitions when they were discriminated as nonspeech chirps. There was, however, no comparison of the effect of silence on the speech and nonspeech percepts when the subjects had to perform the same task in response to both. In the experiment reported here, the subjects did perform the same task: they discriminated, not only the chirps, but also the speech. It was found that the silence cue had a large effect on the speech side of the percept, but had little effect on the nonspeech side. This result, taken together with those obtained in the earlier experiment, strongly implies that the effect of silence as a cue for stop consonants is owing primarily to phonetic (rather than auditory) processes.

The experiment reported here is an extension of an earlier one (Liberman & Isenberg, 1980) that exploited the phenomenon of duplex perception to determine why silence is an important cue for stop consonants. Shortly, we will discuss these two experiments in detail. Before that, however, we should look closely at just what duplex perception is and what it might represent.

[HASKINS LABORATORIES: Status Report on Speech Research SR-65 (1981)]

An example of duplex perception, appropriate for purposes of explication, is found in a recent study of the perceived contrast between [ra] and [la] (Isenberg & Liberman, 1978; Liberman, 1979). The procedure for obtaining the phenomenon was like that of Rand (1974). First, the syllables [ra] and [la], shown schematically in the top half of Figure 1, were synthesized so as to make the perceived distinction depend entirely on the transition of the third formant. Then, as shown in the bottom half of the figure, these patterns were divided into two constituents. One, labeled 'base' and shown at the left, included all aspects of the pattern that were identical in the two syllables. When presented by itself, this common core was perceived as a syllable, almost always as [ra]. The other constituent, shown to the right, was one or the other of the third-formant transitions that, in the undivided syllable, critically distinguished [ra] from [la]. In isolation, these transitions were perceived variously, but in no case did they sound the same as when, in the undivided patterns, they were essential to the difference between the syllables; by most listeners, indeed, they were thought to be not-very-speechlike, but discriminably different, 'chirps.' The last, and critical, step was to put the base into one ear and one or the other of the isolated transitions into the other, being careful, of course, to make the temporal relation between the dichotically presented constituents the same as it had been in the undivided patterns.

The result was a duplex percept. One component was a syllable that listeners 'correctly' perceived as [ra] or [la] according to the nature of the third-formant transition. The other component, perceived at the same time as the syllable, was a not-very-speechlike chirp. This percept corresponded to the one that had been produced by the third-formant transition in isolation. The two percepts were not only phenomenally distinct but also dissociable, as could be inferred from the further finding that listeners were able to report changes in the loudness of the syllable or the chirp according as the intensity of the base or the third-formant transition was varied.

What interests us here is not so much that the dichotically presented constituents were fused in perception, but rather that one of them was also perceived as if it had not fused. This is the more interesting because the constituent that both fused and did not fuse is the one of the two that, in isolation, did not sound like speech. Thus, given the third-formant transition appropriate for [l] but perceived in isolation as a chirp, and given also the base that was perceived by itself as [ra], listeners did not perceive only the result of fusion: the syllable [la]. Had they perceived only [la], we should have supposed that they were experiencing an effect no different from the one that is obtained in ordinary dichotic fusion, as, for example, when all of the first and second formant is put into one ear and all of the third formant into the other (Broadbent, 1955; Broadbent & Ladefoged, 1957; Halwes, 1969; Rand, 1974; Darwin, Howell, & Brady, 1976; Turek, Dorman, Franks, & Summerfield, 1980). Neither did the listeners perceive all possibilities: the 'fused' [la], the 'unfused' [ra], and the 'unfused' chirp. Had they so perceived the dichotically presented stimuli, we might have supposed that there were, somehow, two consciously available stages (fused and unfused) of auditory processing, or, alternatively, an auditory stage (the two unfused percepts) followed by a phonetic stage (the fused percept). What the listeners did, in fact, perceive was the 'fused' [la] and the 'unfused' chirp. Thus, perception was not, as it might have been, either unitary or triplex.
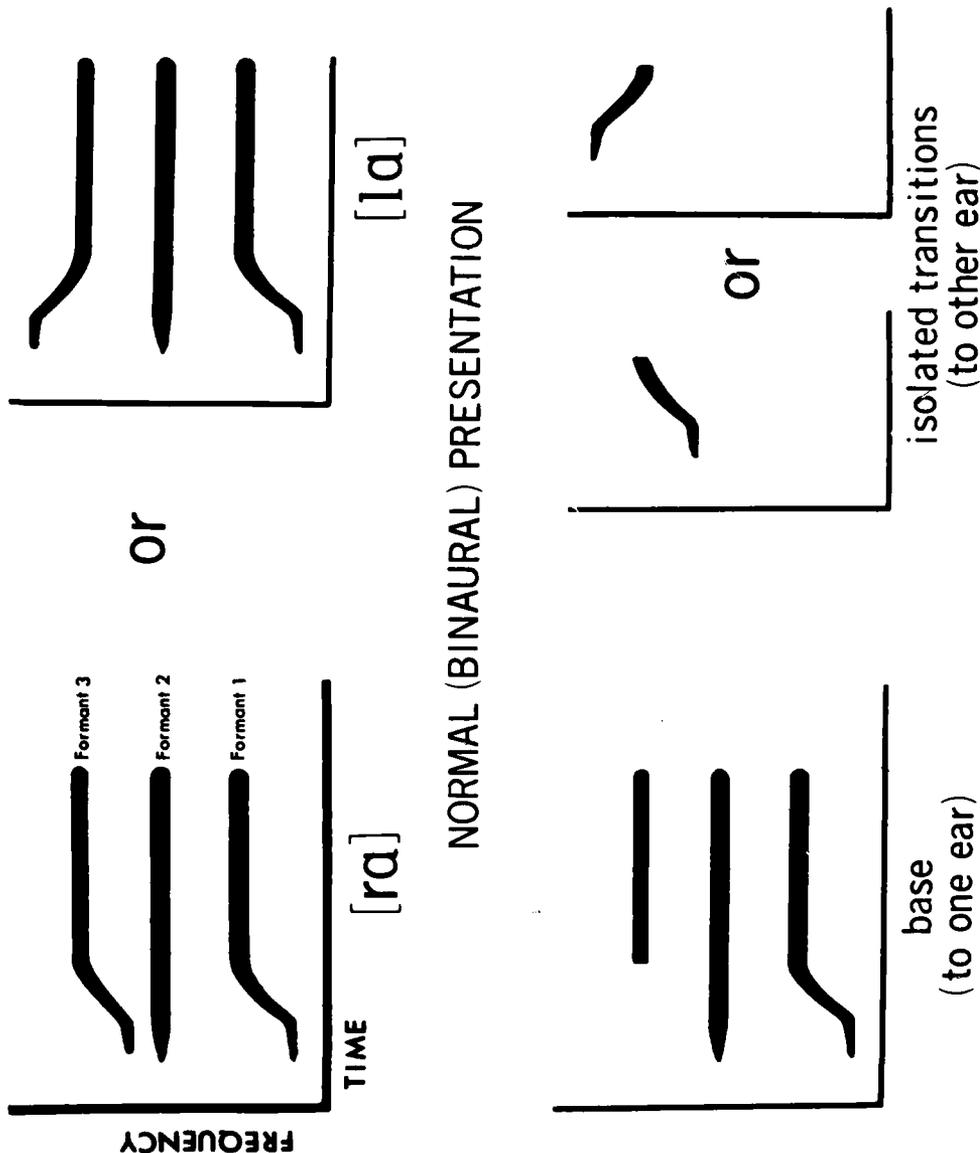
234

Figure 1. Schematic representations of patterns appropriate for duplex perception of [ra] and [la].

Quite remarkably, it was duplex, which is to say that it represented two ways of processing the stimuli: as speech and as nonspeech. More to the point, the two ways of perceiving, and the duplex percept that resulted, turned on the [1] transition. On the 'chirp' side of the percept, that transition was perceived in a way we will call 'auditory,' because the conscious impression was of sound but not speech; moreover, it had those characteristics that psychoacoustic considerations would have led us to expect. On the other side, the same transition was perceived as having the singularly different quality, hard to describe in auditory terms, that distinguishes [la] from [ra]. We take that different percept to result from correspondingly different processes; in our view, the mode which those processes serve deserves the name 'phonetic,' because its percepts have just those characteristics we can be aware of when we listen to consonants and vowels.

Let us return now to a consideration of the current experiment and the earlier one that motivated it. In the earlier experiment (Liberman & Isenberg, 1980) the phenomenon of duplex perception was extended to the case of fricative-stop-vowel syllables ([spa], [sta]) in which perception of the stop depends on an interval of silence positioned between the noise of the fricative and the (appropriate) vocalic transitions. To obtain the duplex percept, patterns like those shown in Figure 2 were used. In the top row are the synthetic syllables from which the patterns were derived. Shown there is the silent interval that serves as a necessary condition for the perception of either of the stop consonants [p] or [t]. Shown also are the contrasting formant transitions that underlie the distinction between these stops. In the bottom row we see how the syllables were divided into constituents for dichotic (and duplex-producing) presentation. The constituent shown at the bottom right of the figure is simply the transitions of the second and third formants, the only cues in these patterns that distinguish [spa] from [sta]. The other constituent is displayed at the lower left of Figure 2 as the pattern labeled 'base.' This is what remains of the original syllables when the second- and third-formant transition cues have been removed and the transition of the first formant straightened. It consists of a patch of fricative noise, followed by a brief period of silence, and then by three steady-state formants. We straightened the first formant because, in the duplex percept, the rising transition seen in the pattern at the top of the figure is important but not absolutely necessary for the perception of a stop consonant. The result of this maneuver was to make the isolated second- and third-formant transitions carry, not only the distinction between [p] and [t], but also more of the information about stop-consonant manner.

The principal conclusion from this experiment was that duplex perception did occur: the formant transitions simultaneously supported speech and nonspeech percepts. On the speech side, the transitions were essential to the perceived distinction between [spa] and [sta], but only when there was an appropriate period of silence in the base consituent; without silence in the base, listeners perceived the 'stopless' [sa], though the same transitions had been presented. On the nonspeech side, the transitions were perceived as chirps and were accurately discriminated as same or different according as the transitions that produced them were the same or different.

Secondarily, the results provided some evidence relevant to the question: does silence affect the transitions differently on the two sides of the duplex
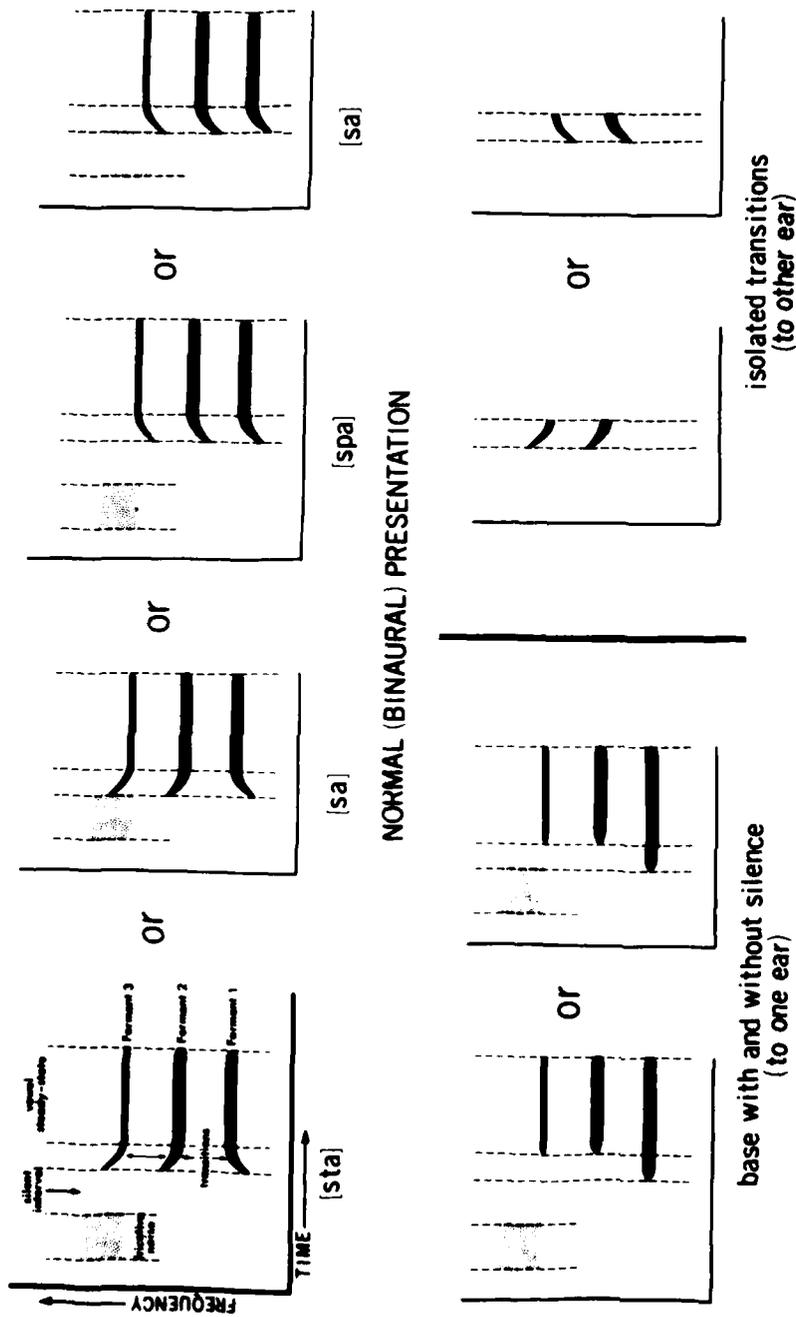
Figure 2. Schematic representations of patterns used as stimuli.

percept? In that connection, it should be noted that silence did have a gross effect on the transitions when they were processed as speech, in which case they were critical to the perceived distinction between [spa] and [sta], though the same silence had no measurable influence on those same transitions when, simultaneously, they were being discriminated as nonspeech chirps. This implied that the effect of the silence cue is not owing to auditory mechanisms of masking or interaction, but should rather be seen as the outcome of a distinctive phonetic process, specialized to treat the presence or absence of silence as phonetically relevant information. Such information reveals that the talker's vocal tract closed, as it must to produce the stop in [spa] or [sta], or that it did not, as it does not when the talker articulates the 'stopless' [sa]. Though that conclusion is supported by the results of the experiment, the support is not so strong as it might be, since the two sides of the duplex percept were measured in different ways: by identification on the speech side (because only identification could establish that the stimuli were, in fact, heard as speech), but by discrimination on the nonspeech side (because identification of the chirps is rather difficult and also not necessary for the purpose of proving that the subjects did, in fact, perceive the nonspeech appropriately). There was, then, no comparison of the effect of silence on speech and nonspeech percepts when the subjects had to perform the same task in response to both. The purpose of this experiment is to repair that omission. Accordingly, the subjects will be required to discriminate, not only the chirps, but also the speech. Given that duplex perception of the transitions was demonstrated in the earlier experiment (Liberman & Isenberg, 1980), these discrimination measures should provide a further test of the hypothesis that, in the perception of these stops, the effect of silence is phonetic rather than auditory.

METHOD

Stimuli

The stimuli of this experiment were identical to those shown in Figure 2 and described in detail in the earlier experiment (Liberman & Isenberg, 1980).

Procedure

As in the earlier experiment, a single experimental trial consisted of the presentation of one dichotic stimulus followed, after 420 msec, by presentation of another. In other respects, however, the procedure of this experiment differed from that of the earlier one. Most importantly, it differed in the task set for the subjects and in the combinations of dichotic stimuli that were used in the various experimental trials.

Consider, first, the subjects' task. It was, on both sides of the percept, to try to discriminate the successively presented stimuli of each trial. Subjects were asked to listen for a difference in these stimuli and then to report how confident they were that a difference had been detected. In rating confidence, they were instructed to use the following scale: '1' if "not confident" that a difference had been detected, '5' if "completely confident," and '2,' '3,' or '4' for intermediate degrees of confidence. It was strongly emphasized to all subjects that they were to base their ratings

238

on any difference they could detect. Indeed, subjects were given explicitly to understand that even though two dichotic stimuli might appear to them as tokens of the same type (for example, as tokens of [sa]), they were nevertheless to listen carefully for any difference they might hear and, if confident a difference (of any kind) had been detected, to assign an appropriately high confidence rating.

As for the combinations of dichotic stimuli in the experimental trials, they were so composed as to exhaust all possible pairings of silence - no silence and 'p' - 't' transitions. Thus, a single experimental trial had in its two base constituents one of the following three combinations: silence in both, silence in neither, or silence in one but not the other. As for the combinations of transitions, they were, on each experimental trial, either the same (both 'p' or both 't') or different (one 'p,' the other 't'). There were, then, three combinations of the base times two combinations of the transitions, making a total of six combinations overall. These six are the fundamental conditions of this experiment and will hereafter be so called.

For each of the conditions described above, we made several types of experimental trials. This was done in order to take into account that there were two ways in which the transitions could be the same (both could be 'p' or both 't'), and also to counterbalance for order whenever the two dichotic stimuli of a trial were different (silence vs. no silence in the base constituents, or 'p' vs. 't' in the transition constituents). The result was a total of 16 types of experimental trials. These were recorded onto a test tape in four different randomizations. With this procedure, the experimental conditions with silence in both base constituents were represented on the tape eight times each, as were those with silence in neither base. As a result of counterbalancing, the conditions with silence in one base constituent but not the other were represented 16 times each.

Having satisfied ourselves in the earlier experiment that subjects could, on each experimental trial, judge both sides of the duplex percept, we decided in this experiment to set them the simpler task of judging but one side of the percept at a time. The tape was presented four times. On two of those presentations subjects were asked to judge the speech side of the percept; on the remaining two they judged the nonspeech side, the order of speech and nonspeech judgments having been counterbalanced. There were, then, 16 speech and 16 nonspeech judgments made in each experimental condition that had silence in both base constituents or in neither; in the conditions with silence in one base constituent but not the other, 32 speech and 32 nonspeech judgments were made. The dichotic arrangement of the stimuli--the pairing of constituent (base or transitions) with ear (right or left)--was half the time one way and half the other. The order of these arrangements was counterbalanced.

Subjects

Ten college students were in the initial pool of subjects. All were native speakers of English, none had any known hearing loss, and all were naive with respect to the nature of the stimuli and the purpose of the experiment.

These subjects were screened on the basis of two tests: having been presented (binaurally) with the electronically fused constituents, they were first asked to identify the resulting stimuli as [spa], [sta], or [sa]; then, having been presented (binaurally) with the isolated transitions, they were asked to identify them as patterns that "glided up" or "glided down." On the basis of these tests, two of the ten subjects were eliminated: one because she could not identify the syllables, the other because she could not identify the 'chirps.'

There was also a brief training session, aimed at getting the subjects accustomed to the dichotically presented pairs and to perceiving the two sides of the duplex percept. In this session, the patterns were presented dichotically, and the subjects, having been asked to attend to the speech on some trials and to the nonspeech on others, identified the stimuli as in the screening test. All subjects performed well with the speech stimuli, but two of the eight managed to perform only slightly above chance with the nonspeech chirps. Nevertheless, these two subjects were not eliminated from the experiment.

RESULTS AND DISCUSSION

The aim of this experiment, it will be remembered, was to determine whether the silence cue has a different effect on the discriminability of the formant transitions when, on the one side of the duplex percept, they are critical for the perception of stop consonants and when, on the other, they are perceived as nonspeech chirps. In Figure 3 we see the mean confidence ratings that constitute the results of the experiment. These ratings reflect the subjects' confidence that they detected differences in the pairs of dichotic stimuli presented on each experimental trial. (The scale on which those ratings were ordered ranged from 1 to 5.) Plainly, there is a difference in the mean ratings according as the subjects were judging the speech or the nonspeech sides of the percept.

Consider, first, the leftmost panel of the figure, which displays the results for the condition in which there was no silence in either of the base constituents. Though such a combination was never presented as such in the earlier experiment, we should infer from the results obtained there that the speech side of the duplex percept would have sounded more or less like [sa], regardless of the transitions. Accordingly, we should expect that the transitions would be relatively hard to discriminate when perceived as part of the speech pattern. On the nonspeech side, however, we should suppose that, as in the earlier experiment, discriminability would be relatively little affected by the absence of silence. The results of this second experiment confirm these expectations. Given no silence in either base constituent, the speech percepts were not well discriminated, though the ratings were somewhat higher when the transitions were, in fact, different.[1] On the nonspeech side the results stand in contrast. There, the transitions were relatively well discriminated when they were, in fact, different, though not, of course, when they were the same. A two-way analysis of variance (with the factors speech – nonspeech and same – different transitions) confirmed that silence did, indeed, affect the discriminability of the transitions differently on the speech and nonspeech sides of the percept, $F(1,7) = 26.17$, $p < .01$.

240
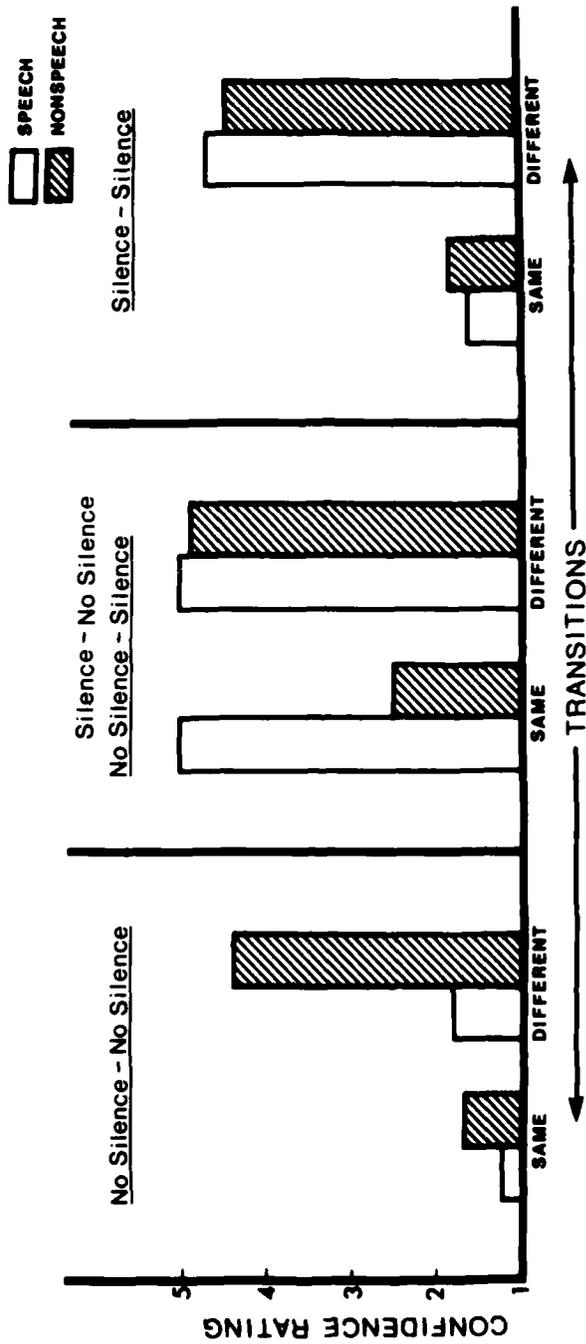
Figure 3. Mean ratings assigned in the conditions of the experiment. Ratings were assigned by the eight subjects to reflect their confidence that the two stimuli of each experimental trial were different.

241

Table 1

Confidence Ratings Assigned by the Individual Subjects

Experimental Conditions	Subjects	Same		Different	
		Speech	Non-Speech	Speech	Non-Speech
No Silence –	1	1.00	2.00	1.00	4.50
No Silence[a]	2	1.51	1.57	1.57	4.69
	3	1.38	1.44	1.51	4.44
	4	1.32	1.07	3.32	4.75
	5	1.00	1.88	1.00	4.63
	6	1.26	1.00	1.75	4.94
	7	1.25	2.25	3.00	4.32
	8	1.25	2.13	1.25	2.63
	X	1.25	1.67	1.80	4.36
Silence –	1	5.00	2.88	5.00	4.97
No Silence	2	5.00	2.04	5.00	4.85
	3	5.00	2.69	5.00	4.69
No Silence –	4	5.00	1.23	5.00	5.00
Silence[b]	5	5.00	2.72	5.00	4.91
	6	4.63	1.00	5.00	4.94
	7	5.00	3.91	5.00	5.00
	8	5.00	3.38	5.00	4.38
	X	4.95	2.48	5.00	4.84
Silence –	1	1.50	2.13	4.94	4.94
Silence[a]	2	1.82	1.57	4.38	4.63
	3	1.75	1.69	4.82	4.44
	4	1.63	1.00	4.75	5.00
	5	1.38	2.94	5.00	4.75
	6	1.83	1.00	3.88	4.94
	7	1.50	2.25	4.75	4.32
	8	1.57	2.13	5.00	2.63
	X	1.62	1.84	4.69	4.46

[a]Each of these scores is the mean of 16 judgments.

[b]Each of these scores is the mean of 32 judgments.

Consider, next, the center panel, where we see the results for the condition in which there was silence in one of the base constituents but not in the other. This is the same as the condition that was used throughout the earlier experiment, where subjects identified the pattern with silence as [spa] or [sta] (depending on the nature of the transitions in the other ear), while identifying the pattern without silence as [sa] (regardless of the transitions). We are not surprised, therefore, to see that when subjects discriminated the speech percepts they confidently perceived a difference between the 'silence' and 'no silence' dichotic stimuli, and they did so whether the transitions were the same or different. (Presumably, they perceived a stop in the one case but not in the other.) The result on the nonspeech side is different. There, the stimuli were readily discriminated when the transitions were different but not when they were the same, notwithstanding the fact that silence was always present in one of the dichotic stimuli but not in the other. That silence affected the discriminability of the transitions differently for speech and nonspeech in this condition is confirmed by analysis of variance, $F(1,7) = 40.93$, $p < .01$.

Finally, there is the condition in which there was silence in both base constituents. Though this condition was not presented as such in the earlier experiment, we can infer from the results obtained there that all stimuli would have been perceived, on the speech side, as containing stops. What is more, stops would have been perceived to be the same or different depending on whether the transitions were the same or different. Not surprisingly, we see this inference supported in the results of the present experiment: subjects discriminated the speech percepts as different when the transitions were different, but not when the transitions were the same. On the nonspeech side, we should expect the same result, and we see that it was, in fact, obtained. That discriminability of the transitions was not significantly different on the speech and nonspeech sides of the percept was confirmed by analysis of variance, $F(1,7) < 1.0$.

To see how fairly the group data, as shown in Figure 3 and discussed above, represent the performances of individual subjects, we should examine Table 1. There, we see that seven of the eight subjects conformed quite well to the group result. The single exception (Subject 8) is one of the two subjects who, as noted under Method, performed poorly with the chirps during the training session that preceded the experiment proper.

The results can be summarized quite simply: the silence cue had a different effect on discrimination of the formant transitions depending on whether they supported the perception of stop consonants or whether, alternatively, they were perceived as nonspeech chirps. Putting these results together with those obtained in the earlier experiment, we conclude that the effect of silence on the perception of the formant transitions is primarily phonetic rather than auditory.

REFERENCES

Broadbent, D. E. A note on binaural fusion. Quarterly Journal of Experimental Psychology, 1955, 7, 46-47.
Broadbent, D. E., & Ladefoged, P. On the fusion of sounds reaching different sense organs. Journal of the Acoustical Society of America, 1957, 29, 708-710.

Darwin, C. J., Howell, P., & Brady, S. A. Laterality and localization: A right ear advantage for speech heard on the left. Haskins Laboratories Status Report on Speech Research, 1976, SR-48, 257-277.

Halwes, T. G. Effects of dichotic fusion on the perception of speech. Unpublished doctoral dissertation, University of Minnesota, 1969. (Issued as Supplement to Haskins Laboratories Status Report on Speech Research, September, 1969)

Isenberg, D. S., & Liberman, A. M. Speech and nonspeech percepts from the same sound. Journal of the Acoustical Society of America, 1978, 64(S1), S20. (Abstract)

Liberman, A. M. Duplex perception and integration of cues: Evidence that speech is different from nonspeech and similar to language. In E. Fisher-Jørgensen, J. Rishel, & N. Thorsen (Eds.), Proceedings of the IXth International Congress of Phonetic Sciences (Volume II). Copenhagen: University of Copenhagen Press, 1979.

Liberman, A. M., & Isenberg, D. S. Duplex perception of acoustic patterns as speech and nonspeech. Haskins Laboratories Status Report on Speech Research, 1980, SR-62, 47-57.

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

Turek, S. V., Dorman, M. F., Franks, J. R., & Summerfield, Q. Identification of synthetic /bdg/ by hearing-impaired listeners under monotonic and dichotic formant presentation. Journal of the Acoustical Society of America, 1980, 67, 1031-1040.

FOOTNOTE

[1]Just how discriminable patterns of this sort will be depends, in our experience, on several factors. When silence is removed from a pattern containing a 't' transition, the resulting percept is not likely to be very different from a perfectly normal [sa], if only because the places of production (hence the second- and third-formant transitions) for 't' and 's' are virtually the same (alveolar). The 'p' transitions, on the other hand, are appropriate to a different place of production (bilabial); hence they are not so readily 'absorbed' into the fricative percept when, in the absence of silence, perception of the stop vanishes. If the 'p' transitions are of very low intensity, it is possible that the listener will simply perceive [sa]. But if perception is affected by the 'p' transitions, then we can expect any one of the following consequences: (1) the perceived fricative takes on the place of production of the 'p' transitions, in which case the percept becomes [fa]; (2) a semivowel appropriate to the place of the 'p' transitions is introduced, in which case the percept becomes [swa]; or (3) the transitions are rejected as speech yet remain audible, in which case the listener is aware of a nonspeech 'chirp' or 'thump.' At all events, we do not expect—at least not in all cases—that the 't' and 'p' transitions will be perfectly indiscriminable when they are heard as speech in the no-silence condition.

THE CONTRIBUTION OF AMPLITUDE TO THE PERCEPTION OF ISOCHRONY

Betty Tuller+ and Carol A. Fowler++

Abstract. Previous studies (e.g., Fowler, 1977, 1979; Morton, Marcus, & Frankish, 1976) have shown that listeners' judgments of isochrony in speech are not based on the intervals between onsets of acoustic energy of successive syllables. An alternative proposal is that the perception of isochrony involves computations based on aspects of the amplitude contour of each syllable (Marcus, 1976). The present experiment used the technique of "infinite peak clipping" to assess the importance of the syllable's amplitude contour, particularly the peak increment in spectral energy, to listeners' judgments of isochrony. Infinite peak clipping gives all syllables, regardless of phonetic makeup, the same amplitude contour; only the durations vary. The results indicate that listeners' judgments of isochrony are unaffected by infinite peak clipping and thus are not based on the amplitude contour of syllables.

Sequences of digits presented at acoustically regular intervals are perceived to occur with unequal spacing. Moreover, when allowed to adjust the intervals between successive digits until they sound isochronous, subjects introduce systematic departures from acoustic isochrony (Morton, Marcus, & Frankish, 1976). These departures are such that the temporal alignment of a word relative to its neighboring words varies with the duration of acoustic energy prior to the acoustic onset of its vowel. Thus, for example, the acoustic onset-to-onset time, or "syllable-onset-asynchrony," for a word pair such as "eight-six" tends to be shorter than for "six-eight."

These findings indicate that listeners' judgments of rhythmicity in speech are not based on the intervals between the onsets of acoustic energy of successive syllables. Morton et al. proposed that, instead, listeners judge the timing of word sequences based on reference points, termed "P-centers," within each word. The "P-center" is described as the "psychological moment of occurrence" of a word. Other investigators have identified what is probably the same reference point and have called it a "stress beat" (Allen, 1972; Rapp, 1971). We will use this more descriptive term.

Further investigation by Morton et al. failed to reveal any obvious acoustic markers of stress beats. Specifically excluded as markers were the

---

[HASKINS LABORATORIES: Status Report on Speech Research SR-65 (1981)]

acoustic onset of the word, the acoustic onset of the stressed vowel, and the peak intensity of the word or vowel.

Two other experimental investigations were designed to pinpoint the locus of the stress beat in a word (Allen, 1972; Rapp, 1971), although neither study discovered how a stress beat is marked acoustically. Allen's subjects tapped their fingers "on the beat" of a designated syllable in a sentence, whereas Rapp's subjects repeated disyllabic nonsense utterances "on the beat" of a regularly occurring pulse. In both studies, the tap or pulse was located near the acoustic onset of the stressed vowel, but preceded it by a variable duration that correlated positively with the acoustic duration of the prevocalic consonant or consonant cluster.

Marcus (1976), using Rapp's data, evaluated an acoustic model of isochrony in which combinations of simple acoustic cues determine the location of a syllable's stress beat. The duration of the syllable-initial consonant (or cluster) prior to vowel onset in fact predicted the location of stress beats rather well. Notice, however, that this model does not involve vowel duration or the duration of consonant(s) following the stressed vowel, both factors that may influence stress beat location (Marcus, 1976). Thus, Marcus (1976) proposed a model for determining P-center or stress beat location that weights segment durations occurring before and after vowel onset.

Both the Rapp model and the Marcus model entail demarcating the vowel onset--a determination that is difficult to make reliably. In an attempt to reduce the subjective quality of determining vowel onset, Marcus tested a set of parameters suggested by Sambur and Rabiner (1974) for the automatic extraction of vowel onset from the speech waveform. The time of occurrence of one of these parameters, the peak increment in spectral energy in the first and second formants, was considered the most appropriate acoustic correlate of vowel onset. That is, the well-defined measure of peak increment of spectral energy closely approximated the more subjective measure of vowel onset and was therefore substituted for vowel onset in Marcus's equation for determining stress beat location. In sum, Marcus proposed a generalization of Rapp's model using the variable of peak increment in spectral energy instead of vowel onset and including the duration of acoustic segments following the point of peak increment.

The experiment described here assessed the importance of the syllable's amplitude contour, particularly of the peak increment in spectral energy, to listeners' perception of isochrony. To this end, we used the procedure of infinite peak clipping to control changes in spectral energy. Infinite peak clipping reduces the speech waveform to a series of rectangular waves of equal amplitude in which the discontinuities correspond to the crossing of the time axis in the original speech signal. Considerable information is retained in infinitely peak-clipped speech; conversation may be perceived with little or no difficulty, although the perception of the phonetic composition of isolated words may be impaired (Licklider & Pollack, 1948).

The location within a syllable of the peak increment in spectral energy will shift when the syllable is infinitely peak-clipped. Infinitely peak-clipped syllables have their peak increment at syllable onset. Thus, if the perception of isochrony depends in any way on the location of the peak

increment, the intervals between syllables that subjects require in order to hear the sequence as isochronous should not be the same when the syllables are infinitely peak-clipped as when they are not. Specifically, the method of infinite peak clipping gives all syllables, regardless of phonetic composition, the same initial contour; only the durations vary. Thus, sequences with onset-to-onset times that are measured to be isochronous should be more *nearly* perceptually isochronous when they are infinitely peak-clipped than when they are not.

Method

Subjects

The subjects were eight adult females and five adult males. All of the subjects were naive to the purposes of the experiment and none of the subjects had previously heard infinitely peak-clipped speech.

Stimuli

One male speaker, naive to the purpose of the experiment, was asked to produce a series of nonsense-syllable sequences. Each sequence was composed of two monosyllables repeated in alternation five times. The monosyllables all rhymed with /ad/ but differed in initial consonant or consonant cluster. Combinations of syllables were devised to maximize the expected acoustic anisochrony. Sequences contained the syllable /stad/, /shad/ or /strad/, each produced in alternation with /ad/; the syllables /stad/, /shad/, /chad/, and /strad/ were each produced in alternation with /tad/; /skad/ and /chad/ were produced in alternation with /nad/; and /stad/ and /strad/ were alternated with /sad/. Thus, eleven sequences were produced in all.

The speaker was asked to produce these utterances at a comfortable rate, *stressing every syllable as equally as possible*, and to produce the sequences "as if speaking in time to a metronome." The utterances were tape recorded and subsequently input into a Honeywell DDP-224 computer for waveform editing using the pulse code modulation (PCM) system at Haskins Laboratories.

Editing proceeded by first excising the central eight syllables from each sequence, in order to minimize the effects of initial and final lengthening. Four versions of each sequence were then constructed. One version of each sequence consisted of the middle eight syllables of the original sequence with the syllable-onset-asynchronies of the naturally-spoken sequence and with the amplitude envelope unaltered. The second version was constructed from the first so that the acoustically-defined onset-to-onset times were equal in duration. This acoustic isochrony was achieved by determining the longest interval from version one of each sequence, then electronically splicing silence onto all the shorter intervals in the sequence. The largest asynchrony between adjacent intervals in the natural sequences ranged from 19 msec in /stad, sad, stad, sad.../ to 338 msec in /strad, ad, strad, ad.../.

Two more versions of each sequence were created. They corresponded to the natural and adjusted versions just described, but were infinitely peak-clipped. "Silent" durations between syllables were electronically reduced in

247

amplitude so that any background hum, or machine noise, would not be increased in amplitude and be distracting to the listener (cf. Licklider & Pollack, 1948). Syllables were infinitely peak-clipped by electronically increasing the amplitude of each syllable until all points within the syllable were of sufficient amplitude to exceed hardware limitations and were thus "clipped."

When the sequences were output onto magnetic tape, they were filtered so that high frequencies were attenuated. Thus, the stimuli were not strictly rectangular. High-frequency attenuation "rounds the edges" of each syllable. However, as in stimuli that have been infinitely peak-clipped but not filtered, all syllables result in the same initial acoustic contour, although the syllable durations vary (see Figure 1).

Infinitely peak-clipped (C) and not peak-clipped (NC) sequences were presented in a blocked design. Half the subjects heard C sequences first, and half heard NC sequences first. On each trial within a block, subjects heard two eight-syllable sequences presented two seconds apart. In one of the sequences, the intervals between syllables were as naturally spoken; in the other sequence, the intervals were altered to be acoustically equal. The order of the two sequence types was randomized within each block. Both sequences were then repeated in the same order, with two seconds between them.

The subjects' task was to judge which of the two sequences sounded more "rhythmic." Subjects were instructed that in the context of the experiment, "rhythmic" meant "as if the syllables were spoken in time to a metronome." One practice trial was given at the start of each block.

Thus, the eleven sequence types were randomly ordered twice--once for the NC versions and once for the C versions. The temporally-normal and temporally-altered versions were presented and then repeated. The subject had to indicate which of the two versions sounded more rhythmic.

If a subject judges rhythmicity by using the point of peak increment in spectral energy, the pattern of results for C and NC stimuli should differ. Specifically, based on previous studies (Fowler, 1977, 1979), we expect subjects to choose the temporally-normal version of an NC sequence as being more rhythmic than the temporally-altered version of the same sequence. For C stimuli, the peak increment in spectral energy occurs at the onset of, or at least very early in, the syllable so that the peak increment will occur at more nearly isochronous intervals when the sequences are temporally altered to produce acoustic isochrony.

Results and Discussion

In both the NC and C conditions, subjects chose the natural, acoustically anisochronous version of each sequence pair with far greater than chance frequency. On the eleven sequences, the natural version was chosen a mean of 10.15 ($\underline{sd}$=1.1) and 9.92 ($\underline{sd}$=1.6) times, NC and C versions, respectively. These values both differ significantly from the chance value of 5.5 [paired t-tests: $t(12) = 15.71$, $p < .0001$ and $t(12) = 9.93$, $p < .0001$, for NC and C, respectively].

248
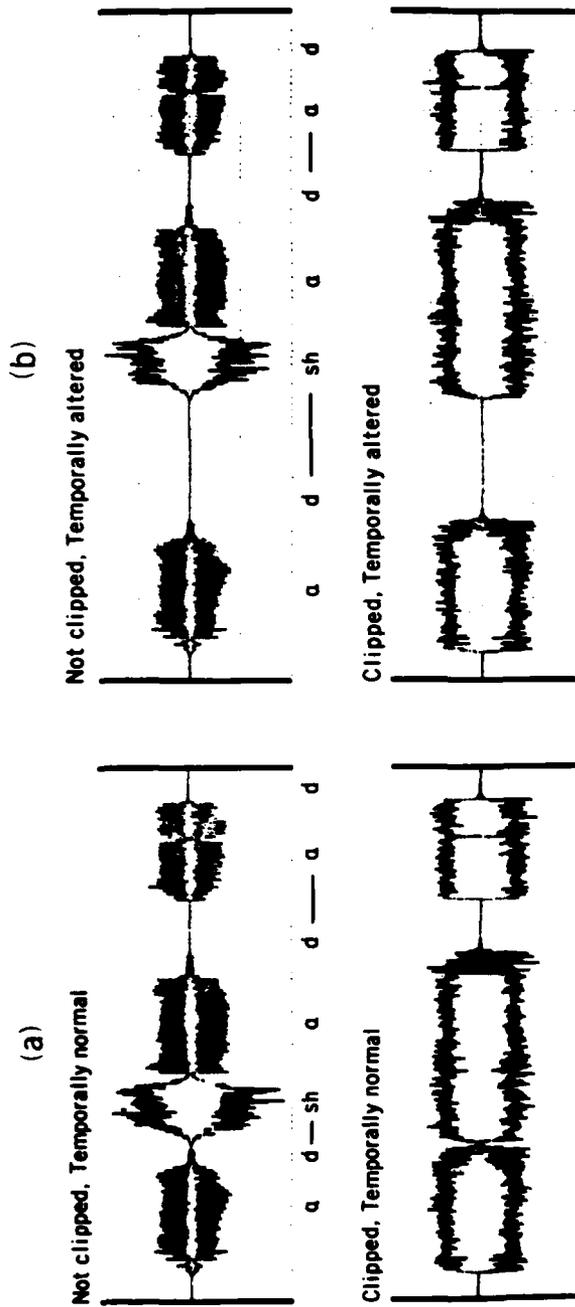
Figure 1. Sections of four versions of the sequence /ad, shad, ad, shad.../. a) The stimulus-onset-asynchronies as naturally spoken, with the amplitude envelope unaltered (top) and infinitely peak-clipped (bottom). b) The onset-to-onset times of syllables adjusted to be equal in duration with the amplitude envelope unaltered (top) and infinitely peak-clipped (bottom).

The number of times that subjects chose the natural version of each sequence did not differ between conditions (NC vs. C), as shown by a paired t-test [t(12) = .43, p > .1].

The results of this experiment do not support the hypothesis that peak increment of spectral energy plays a primary role in the perception of isochronous speech. Indeed, they tend to rule out any explanation of subjects' timing judgments in these studies that invokes the amplitude contour of the syllables. Subjects' judgments of isochrony were unaffected by the infinite peak clipping of syllables.

The results replicate earlier findings that listeners judge sequences of syllables with naturally-produced syllable onset asynchronies as more isochronous than sequences of syllables with acoustically-defined isochronous onsets (Fowler, 1977, 1979). In addition, the results indicate that these judgments are unaffected by the amplitude characteristics of the acoustic waveform.

These results do not signify necessarily that the onset of the stressed vowel is unimportant to the perception of isochrony. They do suggest that peak increment of spectral energy is not a perceptual correlate of vowel onset insofar as its manipulation had no effect on the perception of isochrony.

REFERENCES

Allen, G. The location of rhythmic stress beats in English: An experimental study. Language and Speech, 1972, 15, 72-100.

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

Fowler, C. A. "Perceptual centers" in speech production and perception. Perception & Psychophysics, 1979, 25, 375-388.

Licklider, J. C. R., & Pollack, I. Effects of differentiation, integration, and infinite peak clipping upon the intelligibility of speech. Journal of the Acoustical Society of America, 1948, 20, 42-51.

Marcus, S. M. Perceptual centres. Unpublished doctoral dissertation, King's College, Cambridge, 1976.

Morton, J., Marcus, S., & Frankish, C. Perceptual centers (P-centers). Psychological Review, 1976, 83, 405-408.

Rapp, K. A study of syllable timing. Papers from the Institute of Linguistics, University of Stockholm, 1971, 14-19.

Sambur, M. R., & Rabiner, L. R. A speaker-independent digit-recognition system. Bell System Technical Journal, 1974, 54, 81-102.

ON GENERALIZING THE RABID-RAPID DISTINCTION BASED ON SILENT GAP DURATION*

Leigh Lisker+

Abstract. Several studies have reported that the durations of silent gaps affect listeners' decisions in identifying an auditory stimulus as rabid or rapid. It appears to be accepted that silent gap duration is a cue to stop voicing. Several implications of this asserted connection deserve some discussion. First of all, since the voicing feature is commonly said to distinguish the two phoneme sets /bdg/ and /ptk/, we should like some assurance that silent gap duration operates for all stop places of articulation. Data exist which indicate that the effectiveness of this feature is far from uniform for /b/-/p/, /d/-/t/, and /g/-/k/. In the second place, if a short silent gap elicits rabid responses, and /b/ is said to be voiced--i.e., characterized by glottal signal during closure--then we might suppose that listeners cannot distinguish between presence and absence of such signal when short silent gaps are reported as /b/. In fact, listeners can detect this difference within short closures, and some can indeed give it a phonetic interpretation. Third, we may inquire whether the variation in silent gap duration needed to effect a shift in linguistic identification falls within the range observed in natural speech. A comparison of experimentally determined category boundaries with measurements of natural speech shows that the connection is not always close.

Several studies have reported that in English words such as rabid and rapid the lips are closed longer for /p/ than for /b/ (Lisker, 1957; Sharf, 1962; Suen & Beddoes, 1974; Umeda, 1977). Some have also presented experimental data to show that the presence of laryngeal buzz during closure is not a necessary condition for hearing medial /b/, and that the duration of a silent closure interval affects its interpretation as /b/ or /p/ (Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Lisker, 1957; Port, 1979). The boundary value between /b/ and /p/ is not some fixed duration of silent gap, however; among other things it depends on the duration of an immediately preceding voiced interval--in rabid vs. rapid on the duration of the [ae] vowel (Port, 1979). The longer the vowel (within limits), the longer the silent gap must be for rapid rather than rabid to be heard. Since phonological considerations dictate that these words be spelled with different consonant symbols and

---

[HASKINS LABORATORIES: Status Report on Speech Research SR-65 (1981)]

identical vowels, we also say that vowel duration too is a cue to the consonantal feature of voicing that is said to distinguish /b/ from /p/. It has, in fact, been asserted that the relevant temporal measure is not closure duration, but the ratio of that quantity to the duration of an immediately preceding vowel or sonorant interval (Port, 1979). In this discussion, however, attention will be restricted to the role of closure duration.

To say that closure duration is a cue to stop voicing raises several questions. First of all, if closure duration is a stop voicing cue, then it presumably helps to distinguish not only /p/ from /b/, but /t/ from /d/ and /g/ from /k/. Is this in fact the case? Second, we may ask whether closure duration is effective generally, or only under certain special conditions. If the latter is true, then what are those conditions, and how likely are they to be satisfied in natural speech? It might possibly be the case that only under the peculiar circumstance where other features, commonly found in nature, have been carefully "neutralized" in synthetic speech patterns, does closure duration emerge as a factor with a measurable effect on word identification. Third, if a silent gap sometimes yields <u>rabid</u>, is this because listeners are unable to detect presence vs. absence of buzz within intervals shorter than those that elicit <u>rapid</u> judgments?

In answering such questions the first point to be made is that varying closure duration affects the <u>rabid-rapid</u> pair only when the closure is acoustically zero; if the closure is buzz-filled, only <u>rabid</u> is reported. Figure 1 shows the effects on listeners' labeling behavior of adding and subtracting closure buzz and varying closure durations in two natural tokens of <u>rabid</u> and <u>rapid</u>. These tokens were recorded by a single male talker, digitized and stored in computer memory by means of the Haskins Laboratories' pulse code modulation system (PCM) at a 10 kHz sampling rate, and the computer-assisted editing was performed on the digitized waveforms. Silencing and prolonging the /b/ closure transformed <u>rabid</u> to <u>rapid</u>. On the other hand, shortening the /p/ closure reduced the number of <u>rapid</u> judgments, but even for the shortest duration imposed (30 msec) the addition of buzz had some effect on word identification. The particular crossover values exhibited by these data, 75 msec for /b/ > /p/ and 35 msec for /p/ > /b/, are in themselves of no great significance: the same operations performed on other natural tokens of these words have often failed to turn up similar crossover durations, and have in fact sometimes failed to effect any decisive shift at all in word identity (Lisker, 1978). What we can say is that, in general, <u>rabid</u> tokens tend, with increasing duration of silence closure, to elicit an increasing percentage of <u>rapid</u> responses. Original <u>rapids</u>, which have naturally silent closures, are less reliably transformed to convincing <u>rabids</u> by shortening their closures. In nature intervocalic /b/ closures are regularly filled with laryngeal buzz, so that it is only when buzz is deleted from a signal that presumably includes other /b/ cues that we are likely to achieve a signal sufficiently ambiguous as between /b/ and /p/ for closure duration to take on a decisive role. On the other hand, an incoherent mix of cues is in itself not enough, since the combination of closure buzz with all the extra-closure features of an original <u>rapid</u> is often not ambiguous enough to allow closure duration much scope as a cue to the /b/-/p/ contrast.

Most of the work on closure duration as a stop voicing cue has dealt with the labial stops. Have we, by luck or by design, chosen the place of

# RAPID vs RABID
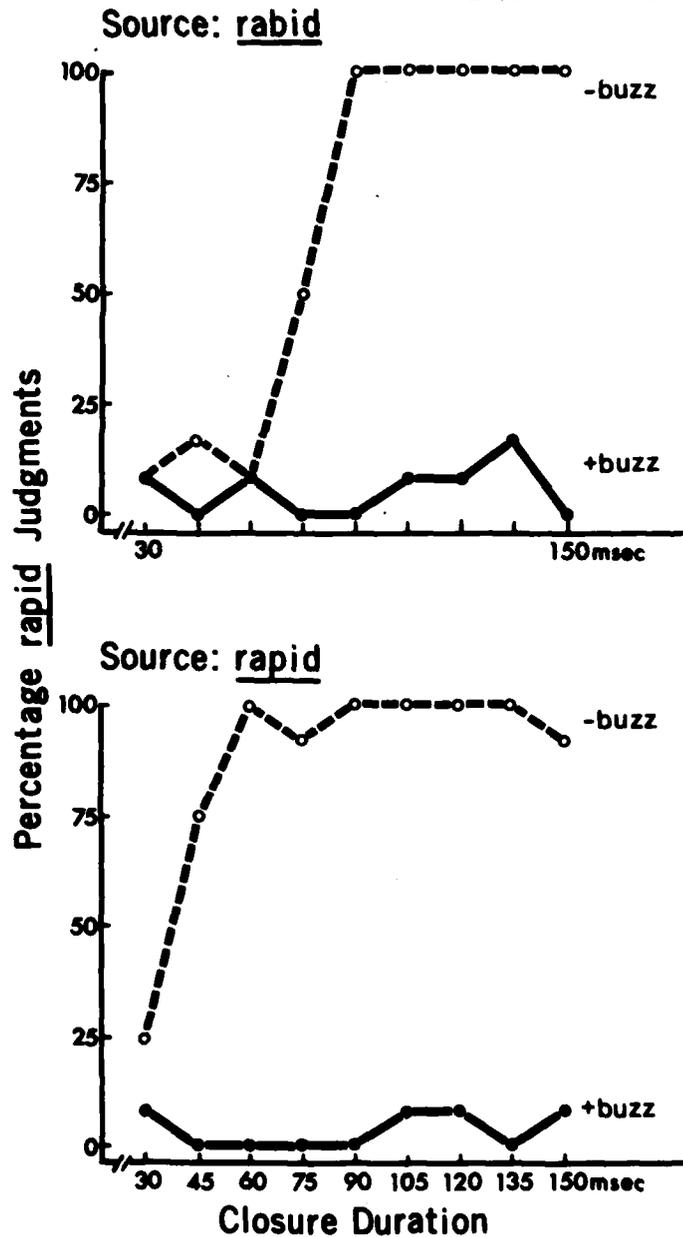
Figure 1. Labelings of edited natural tokens of <u>rabid</u> and <u>rapid</u>. Closure intervals varying in 15 msec steps from 30 to 150 msec were either silent or filled with naturally produced glottal buzz. Six phonetically naive listeners made two judgments of each of the 36 acoustically distinct stimuli presented in random order. Items were identified as either <u>rabid</u> or <u>rapid</u>.

253

articulation where closure duration "works best?" When we turn to the apicals, /t/ and /d/, we encounter in American English the notorious effect of the "flapping rule," which erases the phonetic difference in word pairs such as betting-bedding. Since the flaps in the two words show no consistent difference in the duration of constriction (Fox & Terbeek, 1977), the fact that contrast is reduced (very possibly to zero) may be said to follow from the hypothesis that closure duration is an important cue to the distinction between the /ptk/ and /bdg/ phoneme sets in medial position within trochaic words. However, a /t/-/d/ distinction is maintained in trochaic words such as center and sender, in which the medial closures are initially nasalized. In dialects for which the first word is phonetically ['sɛnt'ɚ] the closure is longer than in sender, but the procedure of silencing and prolonging the /nd/ closure is as ineffective in changing sender to center as reducing the /nt/ closure is in shifting center to sender. Thus silencing and prolonging the /nd/ closure does not yield /nt/, nor does shortening the /nt/ closure result in /nd/. But if we reduce the closure of sender, a shift in word identity is achieved: listeners report hearing ['sɛɾɚ'], that is, a form of center with a medial flap rather than a voiceless stop. Figure 2 presents data to show the effect of reducing the duration of the /nd/ closure, which, it should be noted, was buzz-filled. This relation between closure duration and membership in /ptk/ vs. /bdg/ is not what we should immediately predict from the rabid-rapid case.

The velar stops, /g/ and /k/, appear to be, from the data of Figure 3, more like the labials than the apicals, although /g/ shifts to /k/ less surely than /b/ goes to /p/ with silencing and lengthening of closure.

From the foregoing it seems that in speech signals, i.e., speechlike signals of natural origin, silent gap duration works most reliably as a stop voicing cue in shifting /b/ to /p/, less effectively for the velars, and quite anomalously for the apicals. But even for the labials the effectiveness of this single feature is limited. If we imagine a listener, whether a human or some automatic recognition system, that relied entirely on closure duration, then data of the kind shown in Figure 4 (/b/ and /p/ closure durations measured from five talkers) suggest that the probability of correctly separating these categories would not be spectacularly high. For each talker /b/ durations are less than /p/, though usually with some overlap in their ranges, but the intertalker variation is large enough to indicate a serious need of time normalization before one could put much reliance on closure duration as a sole criterion in recognition. Moreover, the data of Figure 4 derive from productions of isolated words, for which the durational differences between /b/ and /p/ are greater than they are for the same words in sentences. (We may note that the very shortest /b/ closure measured was about 45 msec, a value rather greater than the /p/ > /b/ crossover of 35 msec shown in Figure 1.)

Finally we may ask whether the evaluation of stimuli with short silent gaps as forms containing /b/ depends on an inability to discriminate between stimuli differing only with respect to the acoustic nature of the closure interval, i.e., whether silent or buzz-filled. To test this hypothesis a set of stimuli was derived from a natural token of rabid that had previously been found to go to rapid when its closure was silenced and prolonged to a duration exceeding 75 msec. Sixteen stimuli were prepared: eight closure durations,

254
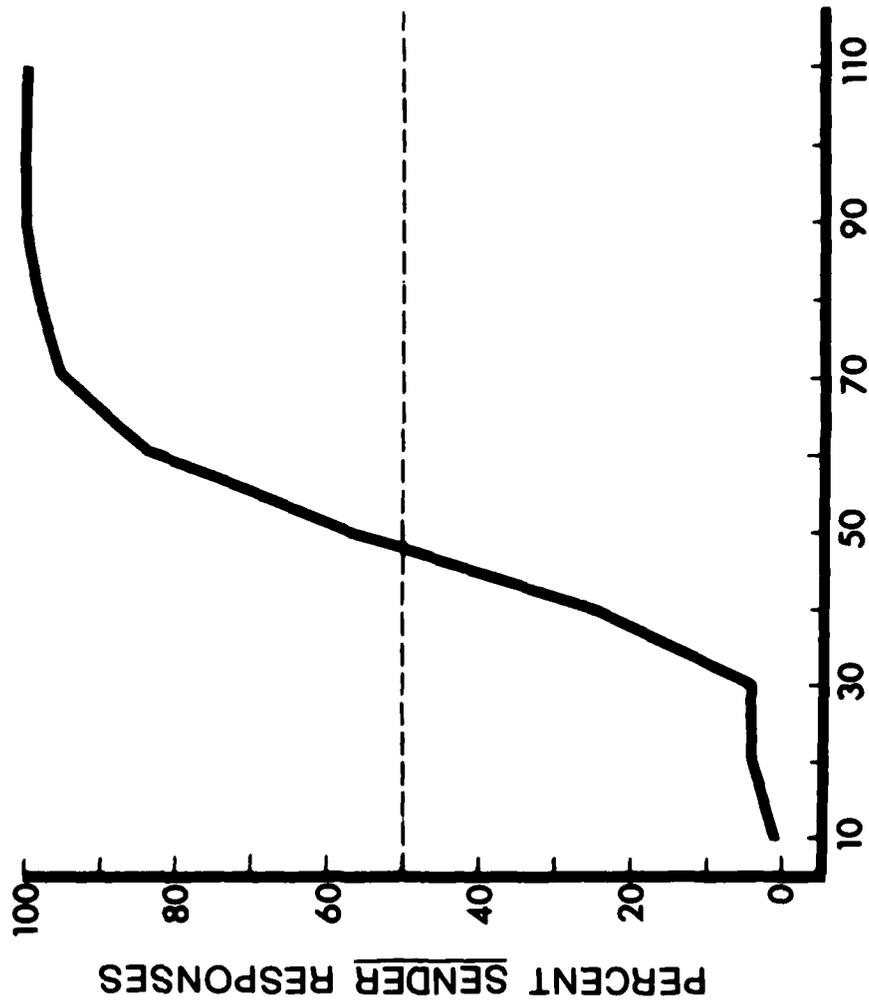
Figure 2. The voiced and largely nasalized closure of a naturally produced sender was reduced in 10 msec steps from an original duration of 110 msec. The word center was most often reported for closures shorter than 50 msec; seven naive listeners made nine independent judgments of each test stimulus.

255

"locker" vs "lager"
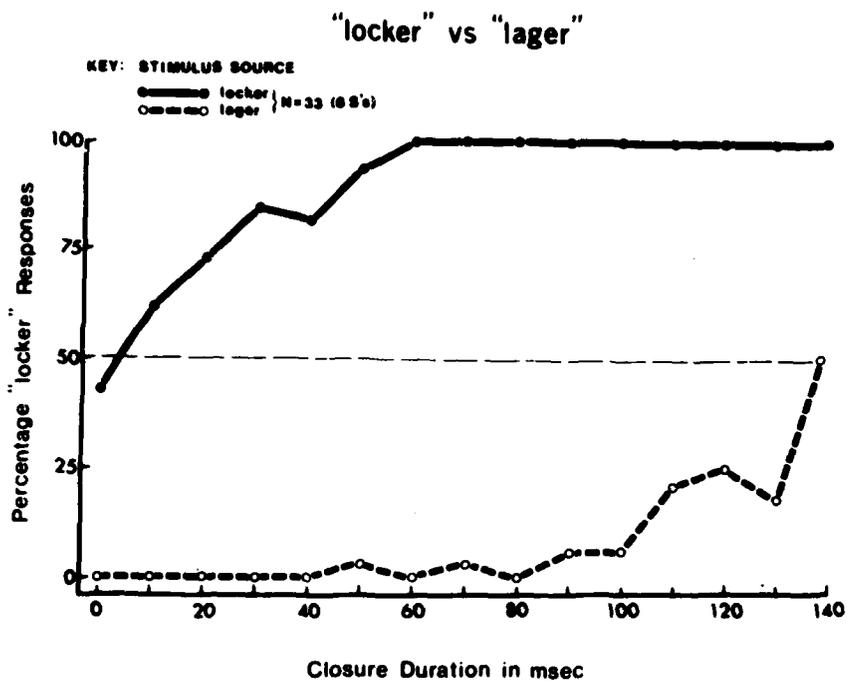
Figure 3. Six listeners made a total of 33 responses to stimuli derived from natural tokens of <u>lager</u> and <u>locker</u> ([ˈlagɚ]-[ˈlakɚ]), whose closures were silenced and varied in 10 msec steps from 0 to 140 msec.

RABID and RAPID

CLOSURE DURATIONS IN ISOLATED PRODUCTIONS
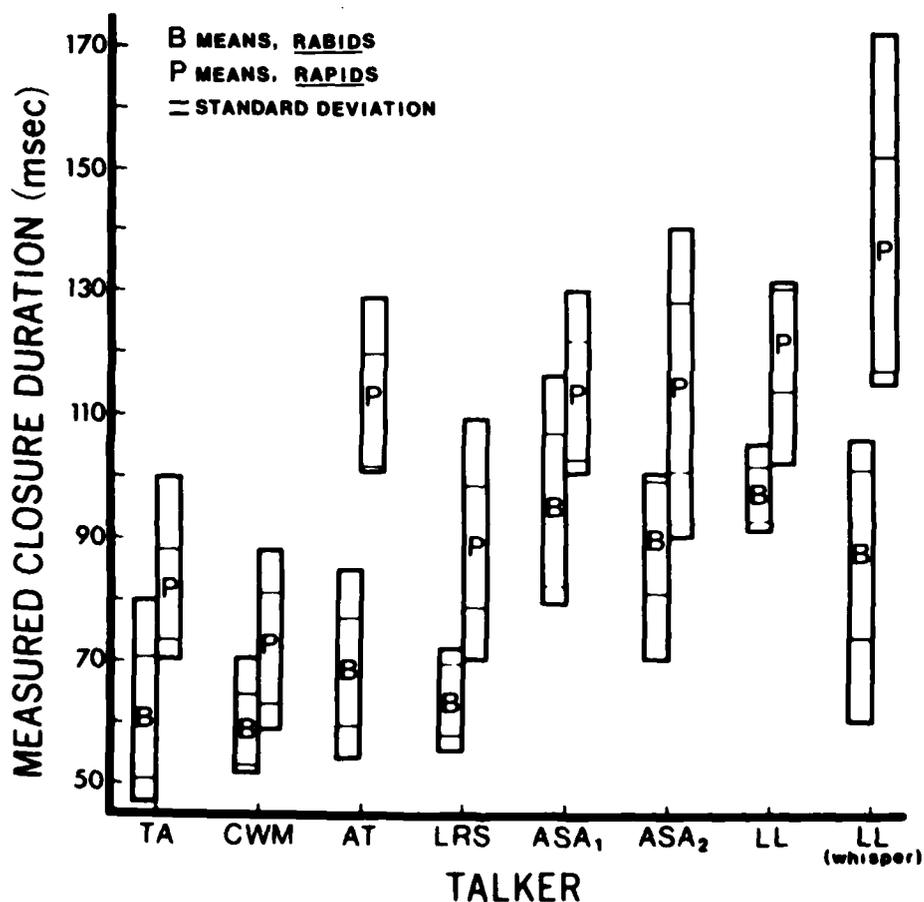

Figure 4. Closure durations measured from spectrograms of tokens of <u>rabid</u> and <u>rapid</u> produced as isolated items read from a randomized list. Each talker produced 11 tokens of each word per reading. Talker ASA read the list on two occasions, while speaker LL read the list once with normal voice and once with whisper.

257

## DISCRIMINATION OF BUZZED AND NON-BUZZED CLOSURES
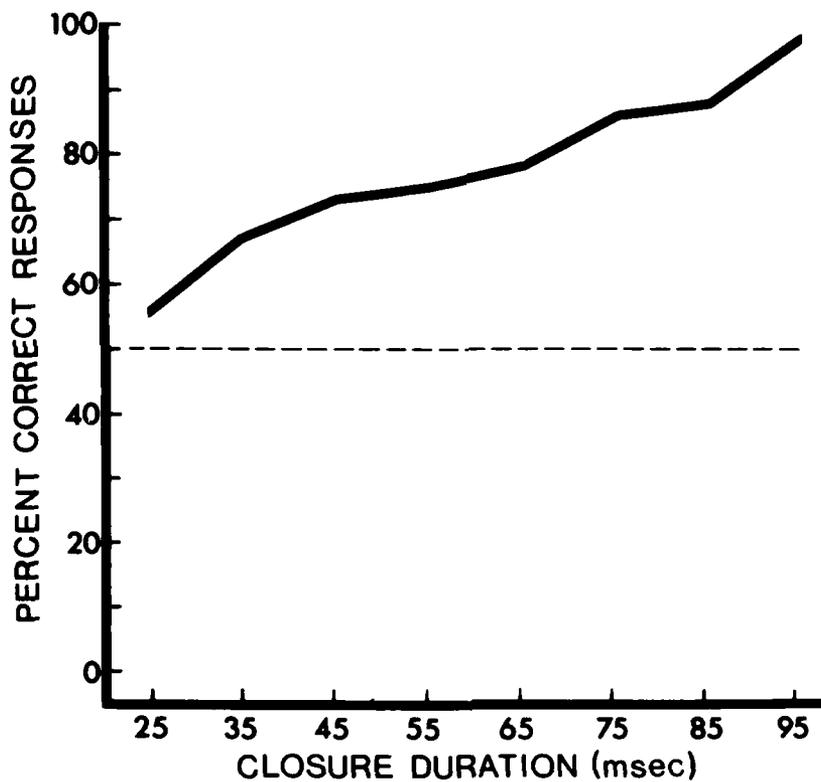
Figure 5. Discrimination of stimuli differing with respect to buzzed
vs. silent medial closures. All stimuli were derived from a natural
token of rabid, and presented to ten subjects in AXB triads. Each
point represents percentage corred "oddity" judgments of twenty per
subject.

258

ranging in ten msec steps from 25 to 95 msec, with each closure being either acoustically silent or filled with naturally produced laryngeal buzz derived from the original rabid token. These stimuli were arranged in AXB triads such that in each triad A and B stimuli differed only with respect to the nature of the closure signal, while the X stimulus was identical with either A or B. Figure 5 shows how well listeners performed when they were asked to identify the "odd" member of each of the test triads. With 200 trials for each pair of stimuli tested it is clear from the data that for durations down to about 50 msec the ten listeners who performed the task distinguished between closure silence and closure buzz at better than a chance level.

It may be concluded from all the preceding that silent gap duration can serve as a sufficient cue to stop voicing only under very special conditions: 1) it works with some reliability only for medial labial stops, 2) it is further limited to signals containing other features that normally accompany laryngeal buzz. If the silent gap whose duration can signal /b/ or /p/ must be located in a context in which only a buzzed closure occurs in nature, this amounts to saying that its usefulness as a cue is restricted practically to acoustic patterns generated only in the laboratory. In nature a brief silent closure involving the lips will most probably be heard as /p/, while a long buzzed closure will undoubtedly be reported as a /b/.

REFERENCES

Fox, R. A., & Terbeek, D. Dental flaps, vowel duration and rule ordering in American English. Journal of Phonetics, 1977, 5, 27-34.

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., & Bastian, J. An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. Language and Speech, 1961, 4, 175-195.

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

Lisker, L. The reliability of closure features as cues to medial stop voicing in English. Journal of the Acoustical Society of America, 1978, 64(S1), S18. (Abstract)

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

Sharf, D. J. Duration of post-stress inter-vocalic stops and preceding vowels. Language and Speech, 1962, 5, 26-30.

Suen, C. J., & Beddoes, M. P. The silent interval of stop consonants. Language and Speech, 1974, 17, 126-134.

Umeda, N. Consonant duration in American English. Journal of the Acoustical Society of America, 1977, 61, 846-858.

II. PUBLICATIONS

III. APPENDIX

PUBLICATIONS

Bell-Berti, F.   A spatial-temporal model of velopharyngeal function.   In
    Norman J. Lass (Ed.), Speech and language: Advances in basic research
    practice (Vol. 4). New York: Academic Press, 1980, 291-316.

Carden, G., Levitt, A., Jusczyk, P., & Walley, A.   Evidence for phonetic
    processing of cues to place of articulation: Perceived manner affects
    perceived place. Perception & Psychophysics, 1981, 29(1), 26-36.

Hanson, V. L. Processing of written and spoken words: Evidence for common
    coding. Memory & Cognition, 1981, 9, 93-100.

Healy, A. F., & Repp, B. H. Context sensitivity and phonetic mediation in
    categorical perception. Journal of Experimental Psychology: Human
    Perception and Performance, in press.

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

Repp, B. H. On levels of description in speech research. Journal of the
    Acoustical Society of America, in press.

Repp, B. H., & Mann, V. A. Perceptual assessment of fricative-stop coarticu-
    lation. Journal of the Acoustical Society of America, 1981, 69, forth-
    coming.

Studdert-Kennedy, M., & Shankweiler, D. Hemispheric specialization for
    language processes. Science, 1981, 211, 960-961.

Verbrugge, R. R. The primacy of metaphor in development. New Directions for
    Child Development, 1979, 6, 77-84.

Verbrugge, R. R. Transformations in knowing: A realist view of metaphor. In
    R. P. Honeck & R. R. Hoffman (Eds.), Cognition and figurative language.
    Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980, 87-125.

Verbrugge, R. R. Two feasts of metaphor. [Review of Metaphor and thought by
    A. Ortony (Ed.), and On metaphor by S. Sacks (Ed.).] Contemporary
    Psychology, 1980, 25, 827-828.

Wexler, B. E., & Halwes, T. Right ear bias in the perception of loudness of
    pure tones. Neuropsychologia, 1981, 19, 147-150.

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

APPENDIX

DTIC (Defense Technical Information Center) and ERIC (Educational Resources Information Center) numbers:

Status Report		DTIC	ERIC
SR-21/22	January - June 1970	AD 719382	ED-044-679
SR-23	July - September 1970	AD 723586	ED-052-654
SR-24	October - December 1970	AD 727616	ED-052-653
SR-25/26	January - June 1971	AD 730013	ED-056-560
SR-27	July - September 1971	AD 749339	ED-071-533
SR-28	October - December 1971	AD 742140	ED-061-837
SR-29/30	January - June 1972	AD 750001	ED-071-484
SR-31/32	July - December 1972	AD 757954	ED-077-285
SR-33	January - March 1973	AD 762373	ED-081-263
SR-34	April - June 1973	AD 766178	ED-081-295
SR-35/36	July - December 1973	AD 774799	ED-094-444
SR-37/38	January - June 1974	AD 783548	ED-094-445
SR-39/40	July - December 1974	AD A007342	ED-102-633
SR-41	January - March 1975	AD A013325	ED-109-722
SR-42/43	April - September 1975	AD A018369	ED-117-770
SR-44	October - December 1975	AD A023059	ED-119-273
SR-45/46	January - June 1976	AD A026196	ED-123-678
SR-47	July - September 1976	AD A031789	ED-128-870
SR-48	October - December 1976	AD A036735	ED-135-028
SR-49	January - March 1977	AD A041460	ED-141-864
SR-50	April - June 1977	AD A044820	ED-144-138
SR-51/52	July - December 1977	AD A049215	ED-147-892
SR-53	January - March 1978	AD A055853	ED-155-760
SR-54	April - June 1978	AD A067070	ED-161-096
SR-55/56	July - December 1978	AD A065575	ED-166-757
SR-57	January - March 1979	AD A083179	ED-170-823
SR-58	April - June 1979	AD A077663	ED-178-967
SR-59/60	July - December 1979	AD A082034	ED181-525
SR-61	January - March 1980	AD A085320	ED-185-636
SR-62	April - June 1980	**	**
SR-63/64	July - December 1980	**	**

Information on ordering any of these issues may be found on the following page.

**DTC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

ED numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ERIC Document Reproduction Service
Computer Microfilm International
   Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

<u>Haskins Laboratories Status Report on Speech Research</u> is abstracted in <u>Language and Behavior Abstracts</u>, P.O. Box 22206, San Diego, California 92122.

AD-A099 958

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)*	2a. REPORT SECURITY CLASSIFICATION
Haskins Laboratories 270 Crown Street New Haven, Connecticut 06510	UNCLASSIFIED
	2b. GROUP N/A

3. REPORT TITLE

Haskins Laboratories Status Report on Speech Research, No. 65, January-March, 1981.

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Interim Scientific Report

5. AUTHOR(S) *(First name, middle initial, last name)*

Staff of Haskins Laboratories, Alvin M. Liberman, P.I.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
March, 1981	278	561

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
HD-01994 NS13617 N01-HD-1-2420 G-80-0178 RR-05596 MCS79-16177 PRF8006144 NS13870	SR-65 (1981)
	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	12. SPONSORING MILITARY ACTIVITY
N/A	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:
-Distinguishing temporal information for speaking rate from temporal information for intervocalic stop consonant voicing
-Articulatory motor events as a function of speaking rate and stress
-Phase relationships among articulator muscles as a function of speaking rate and stress
-Interarticulator programming in obstruent production
-An electromyographic-cinefluorographic-acoustic study of dynamic vowel production
-Should reading instruction and remediation vary with the sex of the child?
-When a word is not the sum of its letters: Fingerspelling and spelling
-A 'dynamic pattern' perspective on the control and coordination of movement
-Motivating muscles: The problem of action
-Some reflections on speech research
-On levels of description in speech research
-A note on the biology of speech perception
-More on duplex perception of cues for stop consonants
-The contribution of amplitude to the perception of isochrony; and
-On generalizing the <u>rabid-rapid</u> distinction based on silent gap duration.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

Speech Perception:
   rate, stop consonants, voicing
   research, reflections
   levels, description
   biological basis
   duplex perception, cues
   isochrony, amplitude
   silence, duration, phonetic distinctions
 	
Speech Articulation:
   rate, stress, motor events
   muscles, phase relationships
   obstruents, articulator programming
   vowels, dynamic
 	
Reading:
   instruction, remediation, sexual variation
   fingerspelling, spelling, words
 	
Movement:
   pattern, dynamic, coordination, control
   muscles, theory, action						