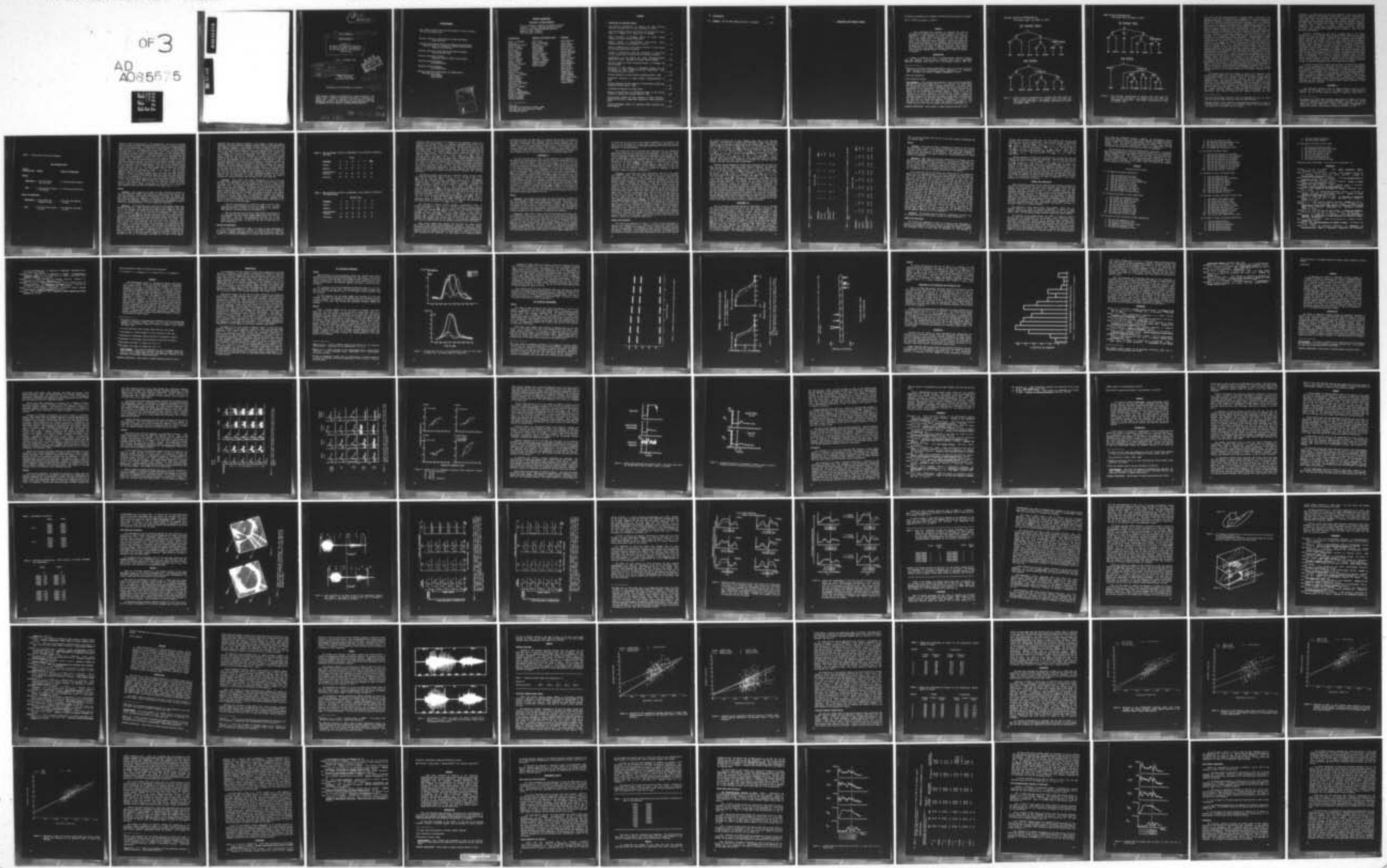
AD-A065 575   HASKINS LABS INC NEW HAVEN CONN   F/G 5/7

SPEECH RESEARCH. A REPORT ON THE STATUS AND PROGRESS OF STUDIES--ETC(U)

DEC 78 A M LIBERMAN   PHS-HD-01994

UNCLASSIFIED   SR-55/56(1978)   NL



sentences that were positive instances of the semantic concept

12 14 SR-55/56 (1978)

9 Status Report, ~~on~~ 1 Jul-31 Dec 78,

6

SPEECH RESEARCH.

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

1 July - 31 December 1978

15 PHS-HD-Ø1994,
NSF-BNS76-82Ø33

10 Alvin M. Liberman
Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

11 31 Dec 78

Distribution of this document is unlimited.

12 228p.

(This document contains no information not freely available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.)

JOB

79 03 12 057

406 643

ACKNOWLEDGMENTS

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

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

Assistant Chief Medical Director for Research and Development, Research Center for Prosthetics, Veterans Administration
Contract V101(134)P-342

National Institute of Child Health and Human Development
Contract N01-HD-1-2420

National Institutes of Health
Biomedical Research Support Grant RR-5596

National Science Foundation
Grant BNS76-82023

National Science Foundation
Grant MCS76-81034

National Institute of Neurological and Communicative Disorders and Stroke
Grant NS13870
Grant NS13617

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

Arthur S. Abramson*
Thomas Baer
Fredericka Bell-Berti+
Catherine Best+
Gloria J. Borden*
Guy Carden*
Robert Crowder*
Michael Dorman*
Donna Erickson*
William Ewan*
Carol A. Fowler*
Jane H. Gaitenby
Thomas J. Gay*
Katherine S. Harris*
Alice Healy*
David Isenberg+
Leonard Katz*
Scott Kelso
Andrea G. Levitt*
Isabelle Y. Liberman*
Leigh Lisker*
Anders Löfqvist[2]
Virginia Mann+
Charles Marshall
Ignatius G. Mattingly*
Nancy McGarr*
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Donald P. Shankweiler*
Michael Studdert-Kennedy*
Michael T. Turvey*
Robert Verbrugge*
Hirohide Yoshioka[1]

Technical and Support Staff

Todd Allen*
Eric L. Andreasson
Elizabeth P. Clark
Donald Hailey
Terry Halwes
Po-Chia Hsia*
Elly Knight*
Sabina D. Koroluk
Agnes M. McKeon
Nancy R. O'Brien
William P. Scully
Richard S. Sharkany
Leonard Szubowicz
Edward R. Wiley
David Zeichner

Students*

David Dechovitz
Laurel Dent
Laurie Feldman
Hollis Fitch
Carole E. Gelfer
Robb Gilford
Janette Henderson
Charles Hoequist
Kenneth Holt
Robert Katz
Morey J. Kitzman
Peter Kugler
Roland Mandler
Karen Marcarelli
Leonard Mark
Suzi Pollack
Patti Jo Price
Brad Rakerd
Abigail Reilly
Arnold Shapiro
Janet Titchener
Emily Tobey-Cullen
Betty Tuller
N. S. Viswanath
Douglas Whalen

*Part-time
[1] Visiting from University of Tokyo, Japan
[2] Visiting from Lund University, Sweden
+NIH Research Fellows

CONTENTS

I. Manuscripts and Extended Reports

I. MANUSCRIPTS AND EXTENDED REPORTS

The Relative Accessibility of Semantic and Deep Structure Syntactic Concepts*

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

ABSTRACT

Three experiments were conducted to determine the relative accessibility of semantic and deep structure syntactic concepts. In Experiment I, which employed a concept formation task, subjects learned the concept "deep structure subject" more slowly than the case concept "experiencer." In Experiments II and III, which employed a new recognition memory procedure, subjects performed more poorly when the sentences to be remembered were differentiated on the basis of deep structure syntactic relations than when they were differentiated on the basis of semantic relations. These results favor Fillmore's case grammar, or another semantically-based theory, rather than the "standard theory" of Chomsky in a model of linguistic behavior.

INTRODUCTION

A number of different versions of transformational generative grammar have been proposed in recent years, including, among others, "generative semantics" (McCawley, 1968; Lakoff, 1971), case grammar (Fillmore, 1968, 1970,

*Portions of this paper concerning Experiments I and II were read at the 18th Annual Meeting of the Psychonomic Society, Washington, D. C., November, 1977. This paper also appears in Memory & Cognition 6, 518-526.

+Also Yale University.

++Also Wellesley College.

Acknowledgment: This research was supported in part by PHS Grants MH26573 and RR07015, NSF Grant BNS77-00077 to Yale University and NICHD Grant HD01994 and BRSG Grant RR05596 to Haskins Laboratories. The senior author was supported by a Junior Faculty Fellowship from Yale University during the preparation of this manuscript. The authors are deeply indebted to Guy Carden for many informative discussions, Robert G. Crowder for a careful reading of an earlier version of this manuscript, George A. Miller for inspiring us to perform Experiment III and for other valuable suggestions, Jonathan Livingston for help with the design of Experiment I and help with the construction of the sentences used in Experiments I and II, Wendy Brown for help with the conduct and analyses of Experiments I and II, Oren Pataschnik for help with the conduct and analyses of Experiment I, and Loretta Polka for help with the conduct and analyses of Experiment III.

Surface Structure Representation:
"The doctor gave the books to John."

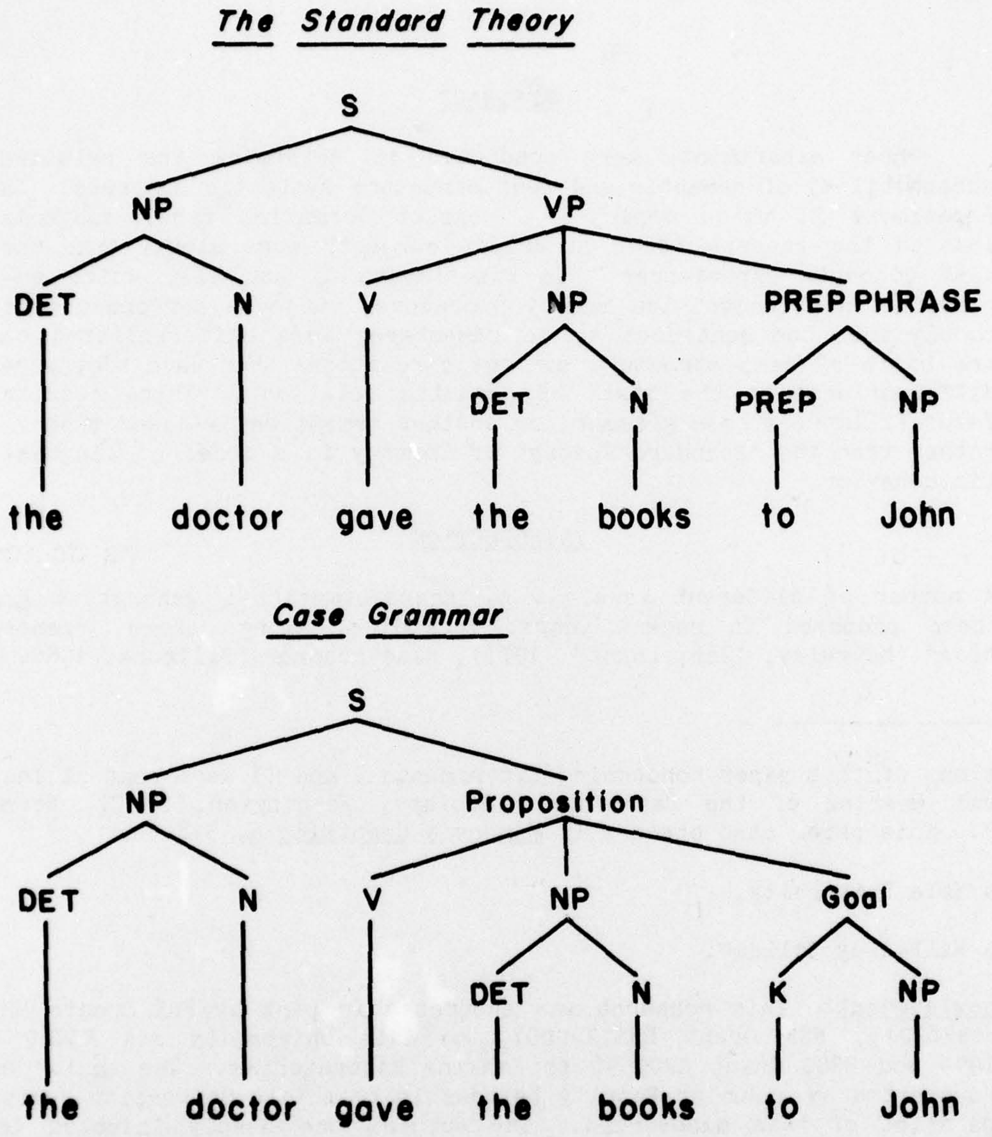


Figure 1: Surface structure representation of sentence "The doctor gave the books to John" according to the standard theory (top panel) and case grammar (bottom panel). Case grammar representation is based on Fillmore (1968).

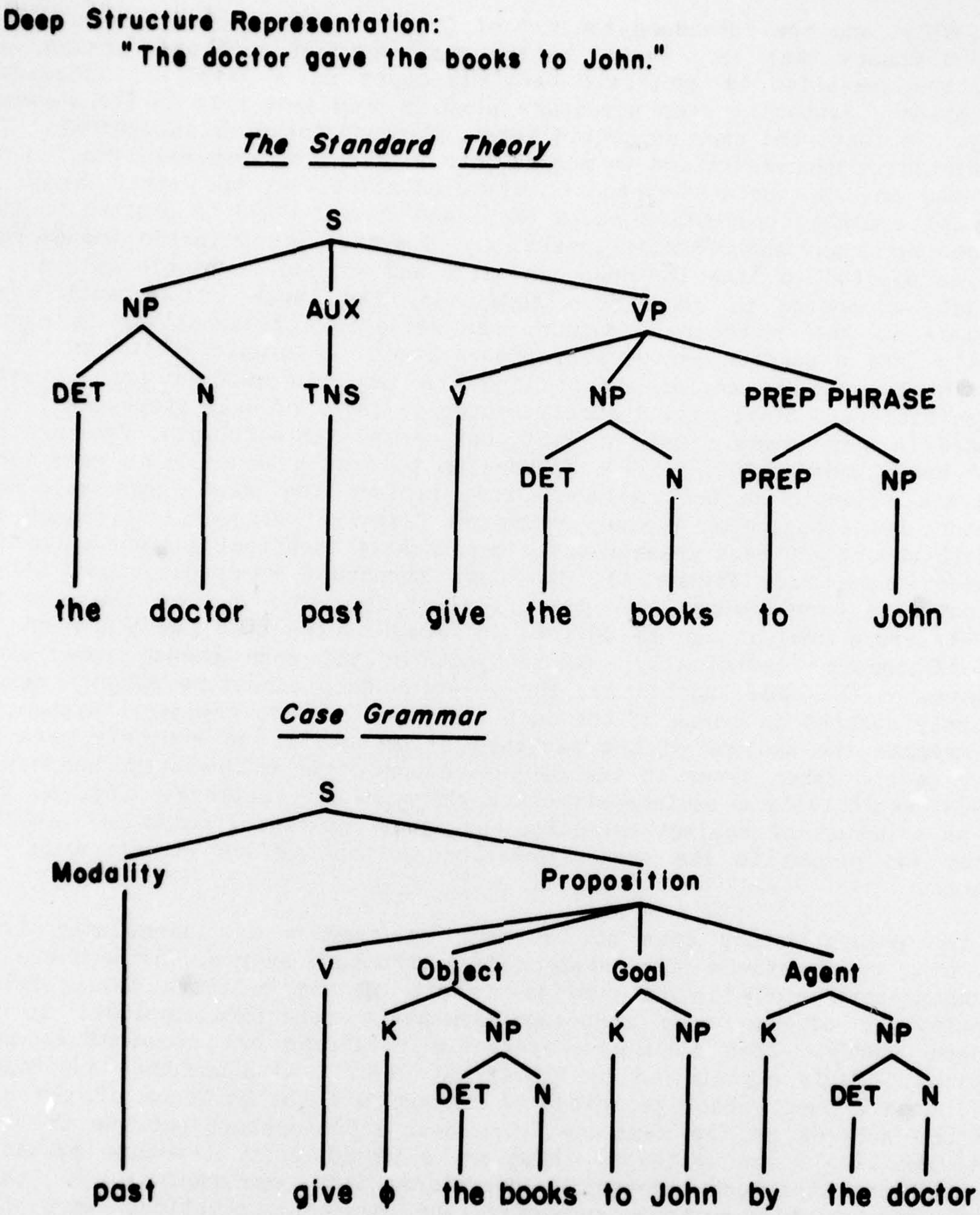


Figure 2: Deep structure representation of sentence "The doctor gave the books to John" according to the standard theory (top panel) and case grammar (bottom panel). Case grammar representation is based on Fillmore (1968).

1971, 1977), and the "standard theory" of Chomsky (1965). One aspect of the standard theory that is attacked by the proponents of both case grammar and generative semantics is syntactic deep structure as a level of linguistic description. Syntactic deep structure plays a prominent role in the standard theory. In fact, the deep syntactic level of representation is central. The deep syntactic representation is mapped into a semantic representation, on one hand, and into a surface syntactic representation, on the other hand. A system of semantic projection rules (Katz and Fodor, 1963) is posited to link the deep syntactic and semantic levels, and a system of syntactic transformations is posited to link the deep syntactic and surface syntactic levels. In contrast, according to generative semantics, the level of syntactic deep structure is not necessary. Rather, the semantic representation is mapped directly into a surface syntactic representation. A single system of transformational rules is envisioned to link the semantic and surface syntactic representations. Similarly, a purely syntactic level of deep structure is not included in case grammar and, in fact, was deemed "an artificial intermediate level" by Fillmore (1968, p. 88). According to case grammar, case relations, which are semantic as well as syntactic, replace the purely syntactic deep structure relations, such as deep structure (logical) subject.[1] Although the standard theory and case grammar posit essentially identical surface structure representations (see Figure 1), the deep structure representations differ substantially (see Figure 2).[2] Note in particular that whereas the relation deep structure subject can be defined in terms of the deep structure of the standard theory (technically, the subject is the noun phrase immediately dominated by the node labeled S), the relation deep structure subject cannot be simply defined in terms of the deep structure of case grammar. Rather, in case grammar the subject of the sentence is treated in an entirely parallel manner to the other cases in the deep structure, and the relation subject is "seen as exclusively a surface-structure phenomenon" (Fillmore, 1968, p. 17). Only as a result of subject selection and transformational rules is a subject created and placed in its proper location in the surface structure of the sentence.

The present study does not attempt to resolve the linguistic issue concerning the existence of a level of deep structure syntax. Rather, the aim of the present study is to provide a test of the relative psychological accessibility of the deep structure syntactic relations included in the standard theory. Deep structure syntactic relations are compared to case relations. Early experiments by Blumenthal (1967) and Blumenthal and Boakes (1967) used a cued recall technique to demonstrate the salience of the deep structure subject of the sentence. The deep structure subject was the best cue to recall a given sentence. However, although deep structure syntactic and surface structure syntactic relations were unconfounded in these experiments, deep structure syntactic and semantic relations were left

[1] In a more recent paper, Fillmore (1977) has recognized the need for purely syntactic deep structure relations as well as case relations.

[2] Although Fillmore (1968) originally represented deep structures in terms of tree structures, as in Figure 2, more recently Fillmore (1971) has announced a preference for a different type of notation.

confounded. (In other words, cue words that differed in their deep structure syntactic categories also differed in their semantic roles.) In contrast, in a more recent study also employing cued recall of sentences, Perfetti (1973) unconfounded deep structure syntactic and semantic relations but left confounded deep structure syntactic and surface structure syntactic relations. (In other words, cue words that differed in their deep structure syntactic categories also differed in their surface structure syntactic categories.) The present study successfully unconfounds for the first time all three types of relations--surface structure syntactic, deep structure syntactic, and semantic.

Instead of employing the cued recall technique used in the studies reviewed above, in the present study we use a concept formation task, which enables us to assess the extent to which subjects are able to learn by example various syntactic and semantic concepts. This technique was employed by Baker, Prideaux and Derwing (1973) to study surface structure syntactic concepts and by Shafto (1973) to study the semantic concepts of case grammar. Although Shafto successfully studied the ease of learning various case relations (and found "agent" easiest, followed by "experiencer," followed by "instrument" and "object"), he did not compare case concepts to any other linguistic concepts. In Experiment I we compare the ease of learning a case relation and a deep structure syntactic relation. Two baseline conditions are also included in this experiment, one to provide information about the upper limit of performance and the other to provide information about the lower limit of performance. In the first baseline condition subjects learn a surface structure syntactic relation, expected to be relatively trivial, and in the second baseline condition, subjects learn an arbitrary relation, defined in a manner analogous to that of the other three concepts.

It should be noted that although case relations are specifically manipulated in this investigation, this study will not allow us to discriminate among various semantically-based grammars, since differences in case relations are necessarily confounded with differences in other semantic variables. Likewise, this investigation will not allow us to discriminate among different models of sentence memory and comprehension that assert structures of a semantic variety (for example, Rumelhart, Lindsay and Norman, 1972; Schank, 1972; Anderson and Bower, 1973; Kintsch, 1974). However, this study will enable us to discriminate between such models based primarily on semantic relations and any plausible alternative models based primarily on deep structure syntactic relations.

EXPERIMENT I

This experiment employed lists of simple sentences, each of which included the word John in one of two semantic roles ("experiencer" or "goal"[3]), one of two deep structure syntactic categories (deep structure

[3] We employed the definitions for these cases given by Fillmore (1971) and restricted ourselves to the benefactive meaning of the goal case: "Where there is a genuine psychological event or mental state verb, we have the Experiencer;...where there is a transfer or movement of something to a person, the receiver as destination is taken as the Goal." (p. 42)

TABLE 1: Eight sentence types with examples.

		Deep Structure Role	
Surface Structure Role		Subject	Object of Preposition
Subject			
	Experiencer	1. John was sleepy near the fire.	7. John was assured misery.
	Goal	4. John was the recipient of the grant.	5. John was given the book.
Object of Preposition			
	Experiencer	2. The accident was imagined by John.	3. The roar was deafening to John.
	Goal	8. The fruit was obtained by John.	6. The property was leased to John.

subject or deep structure object of the preposition), and one of two surface structure syntactic categories (surface structure subject or surface structure object of the preposition). Although there are eight possible combinations of these three kinds of relations, only six of them (types 1 - 6) were employed here (see Table 1). Two combinations (types 7 and 8)--deep structure object of preposition/surface structure subject/experiencer, and deep structure subject/surface structure object of preposition/goal--were not employed since fully satisfactory examples of these types could not be generated (see the introduction to Experiment III below for a more complete discussion of this problem). Since the two missing combinations included one of each of the deep structure syntactic concepts, one of each of the semantic concepts, and one of each of the surface structure syntactic concepts, their exclusion should not bias the learning of any of these concepts. For each of the concepts learned by the subjects, three of the sentence types were instances of the given concept and three were not. For the semantic concept, the three positive types were those with experiencer (types 1, 2, 3, Table 1) and the three negative types were those with goal (4, 5, 6); for the deep structure syntactic concept, the three positive types were those with deep structure subject (1, 2, 4) and the three negative types were those with deep structure object of the preposition (3, 5, 6); for the surface structure syntactic concept, the three positive types were those with surface structure subject (1, 4, 5) and the three negative types were those with surface structure object of the preposition (2, 3, 6); for the arbitrary concept the three positive types had no regular relationship to each other (2, 4, 6) and the three negative types also were not related in any regular way (1, 3, 5).

Method

Subjects. Forty young men and women, who were recruited by posters on the Yale University campus, participated as subjects and were paid $1 for their participation, which lasted approximately 20 minutes. No subject had any formal training in linguistics. There were four conditions--Arbitrary, Deep Structure Syntactic, Semantic and Surface--with ten subjects in each condition. The assignment of subjects to conditions was determined by time of arrival for testing according to a fixed rotation of conditions.

Materials. Sixty sentences, ten of each of the six sentence types, were employed as stimuli. These sentences are shown in the Appendix. Note that for all the sentences where John was the surface subject, John was the first word in the sentence, and for all sentences where John was the surface object of the preposition, John was the last word in the sentence. Hence surface location of the word John was perfectly confounded with its surface structure category but unconfounded with its deep structure category and semantic role. Each of the sixty sentences was typed in the center of four 4 X 6 cards. Four decks of cards were constructed, one for each of the four conditions. Each deck included all sixty sentences; only the order of the sentences varied across decks. In each deck the order of the sentences was pseudo-random with the constraint that each 12-sentence block included two sentences of each type. The sentences within a given one of the five blocks were the same in the four decks, but the order of sentences within a block differed across decks. The order of sentences in a given block in the Semantic condition was the same as that in the Deep Structure Syntactic condition, except for the placement of four of the sentences, including two of each of two types--

sentences that were positive instances of the semantic concept but negative instances of the deep structure syntactic concept (type 3), and sentences that were negative instances of the semantic concept but positive instances of the deep structure syntactic concept (type 4). The two sentences of the first type in the deep structure syntactic deck were replaced by the two sentences of the second type, and vice versa, to form the semantic deck. However, the two sentences of a given type maintained their position relative to each other. Similarly, the order of sentences in a given block in the Surface condition was the same as that in the Deep Structure Syntactic condition except for the placement of four of the sentences, including two of each of two types--sentences that were positive instances of the surface concept but negative instances of the deep structure syntactic concept (type 5) and sentences that were negative instances of the surface concept but positive instances of the deep structure syntactic concept (type 2). The two sentences of the first type in the deep structure syntactic deck were replaced by the two sentences of the second type, and vice versa, to form the surface deck. The deck for the Arbitrary condition was analogously related to that of the Deep Structure Syntactic condition. These relationships among the four decks insured that the sequences of correct responses (positive and negative instances of the given concept) were the same in all four conditions.

Procedure. Each subject was tested individually on one of the four concepts. The experimenter, who sat across a table from the subject, showed the subject all the cards from the appropriate deck, one at a time in the prescribed order. The subject was allowed to view only the sentence currently under test at any given instant. The subject was to respond orally "yes" or "no" to each sentence depending on whether he thought it was a positive or negative instance of the concept. After the subject responded, the experimenter supplied immediate oral feedback by telling him whether he was correct and what the correct answer was. The experimenter recorded the subject's responses on an answer sheet, with an indication of whether a given response was an error. Sentence presentation was experimenter paced, dependent on the subject's speed of responding. The following instructions were read to the subject at the start of the experiment:

> "You will be presented with a series of cards on each of which is printed a simple sentence with the word John in it. The word John has some kind of relation to the other words in the sentence. Your task is to determine what the relationship is.
>
> When you see each card you are to judge whether the sentence on it illustrates the test relation between John and the other words in the sentence. Say 'yes' if you think that it does or 'no' if you think that it does not. The experimenter will tell you if your answer is right or wrong. You may not look back at cards that you have already seen."

Results and Discussion

The results are summarized in Table 2 in terms of mean percentages of errors per 12-sentence block as a function of block position and condition. As expected, subjects performed best in the Surface condition, where the concept to be learned was assumed to be trivial, and worst in the Arbitrary

TABLE 2: Mean percentage of errors in Experiment I as a function of Condition and Block.

	Block					
Condition	1	2	3	4	5	Mean
Surface	18	9	5	4	3	8
Semantic	42	27	22	13	9	23
Deep Structure Syntactic	33	38	28	24	25	30
Arbitrary	49	43	44	35	34	41

TABLE 3: Mean percentage of errors in Experiment I as a function of Condition and Sentence Type.

	Sentence Type					
Condition	1	2	3	4	5	6
Surface	8	10	7	9	9	5
Semantic	19	29	26	32	13	16
Deep Structure Syntactic	11	44	42	30	31	20
Arbitrary	26	48	56	22	68	27

condition, where the concept to be learned was not meaningful (except as a disjunction of sentence types). In addition, performance was better in the Semantic condition than in the Deep Structure Syntactic condition, but this difference was not statistically significant. Two analyses of variance were performed on these data, one with subjects (F_1) and one with sentences (F_2) as the random effect. The statistic $\underline{min}$ F' (Clark, 1973) was computed on the basis of these analyses. These analyses yielded a significant main effect of condition, $\underline{min}$ F' (3,46) = 9.8, $p < .001$ [F_1 (3,36) = 11.1, MS_e = 5190, $p < .001$; F_2 (3,90) = 82.4, MS_e = 140, $p < .001$]. Newman-Keuls tests, based on the analysis with subjects as the random effect, revealed significant differences between the Deep Structure Syntactic condition and the Surface condition and between the Arbitrary condition and the Semantic and Surface conditions, all at the .01 level, as well as a significant difference between the Semantic and Surface conditions at the .05 level. No other differences among conditions were statistically significant. In particular, these analyses did not allow us to distinguish between the critical Semantic and Deep Structure Syntactic conditions or between the Deep Structure Syntactic and Arbitrary conditions.

Learning was evident across the five 12-sentence blocks; the main effect of blocks was significant, $\underline{min}$ F' (4,84) = 6.7, $p < .001$ [F_1 (4,144) = 15.5, MS_e = 818, $p < .001$; F_2 (4,30) = 11.9, MS_e = 212, $p < .001$]. Furthermore, there was more learning evident across the five blocks in the Semantic condition than in the Deep Structure Syntactic condition. Although the overall analyses of variance did not reveal a significant interaction of condition by blocks, $\underline{min}$ F' (12,128) < 1 [F_1 (12,144) = 1.6, MS_e = 818, p = .105; F_2 (12,90) = 1.8, MS_e = 140, p = .054], planned analyses of variance with only the critical Semantic and Deep Structure Syntactic conditions did reveal a significant interaction of condition by blocks in both the test with subjects as the random effect, F_1 (4,72) = 3.9, MS_e = 715, p = .006, and the test with sentences as the random effect, F_2 (4,30) = 2.9, MS_e = 192, p = .037, but not in the more conservative test combining them, $\underline{min}$ F' (4,74) = 1.7, p = .163. In addition, the planned analysis for the two critical conditions yielded a significant main effect of condition with sentences as the random effect, F_2 (1,30) = 8.0, MS_e = 192, p = .008.

The analyses of variance further revealed a significant effect of sentence type, $\underline{min}$ F' (5,131) = 4.6, $p < .001$ [F_1 (5,180) = 7.9, MS_e = 1500, $p < .001$; F_2 (5,30) = 11.1, MS_e = 212, $p < .001$], and a significant interaction of condition by sentence type, $\underline{min}$ F' (15,270) = 2.6, p = .001 [F_1 (15,180) = 3.8, MS_e = 1500, $p < .001$; F_2 (15,90) = 8.1, MS_e = 140, $p < .001$]. Table 3 presents the mean percentages of errors as a function of condition and sentence type. Clearly certain sentence types caused more trouble for learning some concepts than others, but the nature of the interaction of condition and sentence type did not appear to be completely comprehensible and, as will be shown below, was not entirely consistent across experiments.

These results suggest that semantic case relations are indeed learned more rapidly than deep structure syntactic relations. Since the largest difference between the Semantic and Deep Structure Syntactic conditions was at the last block of training, where learning was greatest, a more sensitive test comparing these two conditions seems to be one where all testing is conducted after training has been completed. For that reason a new recognition memory paradigm was devised for Experiment II to compare the learning of the semantic and deep structure syntactic relations with all testing conducted after the completion of training. This paradigm, like the concept formation task, was designed to test the psychological accessibility of various linguistic concepts. Whereas the concept formation task allowed us to determine whether subjects could learn the given concepts, the recognition memory task allows us to determine whether the given concepts will be discovered by subjects in their attempts to learn a list of sentences for a subsequent memory test.

EXPERIMENT II

The same relations were tested in this experiment as in Experiment I. Furthermore, the same sentences were employed in the two experiments with the following important exception: There were two versions of each of the sixty sentences, with one version containing the word John, as earlier, and one version containing the word Sam instead of John. Every subject was shown each of the sixty sentences, half of which were in the version with John and half in the version with Sam. As in Experiment I, there were four groups of subjects, the groups in this case differing in the rule used to assign John or Sam to each sentence. The assignment was made on the basis of either the deep structure syntactic category of the word John or Sam, the semantic role, the surface structure syntactic category, or the arbitrarily defined rule employed in Experiment I. The subjects' task in this experiment was first to study the given sentences and later, on a subsequent recognition memory test, to decide whether a given sentence which had been studied earlier included John or Sam. In contrast to Experiment I, subjects were not specifically told about the existence of a consistent relation between John (or Sam) and the rest of the words in each sentence. Therefore this experiment allowed us to determine how readily the given relations were discovered by subjects in the course of memorizing a list of sentences, rather than whether the subjects could learn the given relations when required to do so.

Method

Subjects. Forty young men and women, who were recruited by posters on the Yale University campus, participated as subjects and were paid $1 for their participation, which lasted approximately 20 minutes. There were four conditions--Arbitrary, Deep Structure Syntactic, Semantic, and Surface--with ten subjects in each condition. For each condition, there were two subgroups of subjects (A and B) with five subjects in each subgroup. The assignment of subjects to conditions and subgroups was determined by time of arrival for testing according to a fixed rotation of conditions and subgroups.

Materials. Eight decks of cards, each card containing one sentence, were constructed for training, two decks for each of the four groups of subjects. For a given group, the sentences in each deck were identical to those employed in Experiment I except that in one deck (Deck A) all the sentences that were positive instances of the concept tested in Experiment I included the word John and all the sentences that were negative instances of the concept included the word Sam, and in the other deck (Deck B) the opposite assignment of John and Sam was employed. In each condition, one subgroup of subjects (A) was given Deck A and the other subgroup (B) was given Deck B. This method of counterbalancing assured that across subjects the words John and Sam would not be confounded with the positive and negative instances of the concept. The order of the sentences in a given deck varied across subjects and was determined by the experimenter's thoroughly shuffling the deck of cards before handing it to the given subject.

Four typewritten lists of sentences were constructed for the recognition memory test, one list for each of the four conditions. The list for a given condition included the same sentences as in Experiment I, in the same order. The only differences between the form of the sentences as they appeared on the cards in Experiment I and as they appeared on the test lists in Experiment II were that the sentences were numbered (from 1-60) in Experiment II but not in Experiment I, and the word John in each sentence in Experiment I was replaced by the pair of words John/Sam in Experiment II. As a result of these constraints, the order of correct answers (John or Sam) was the same for subjects in all four subgroups A, and was the same for subjects in all four subgroups B, but the correct answers for subjects in subgroups A were the reverse of those for subjects in subgroups B.

Procedure. Each subject was tested individually with one of the eight study decks of cards. The first eight subjects run, one in each subgroup, were given eight minutes to study the deck of sentences (timed by the experimenter with a stopwatch). The remaining 32 subjects were given five minutes to study the deck of sentences, since the performance of the first subjects seemed to approach the ceiling. Subjects were in no way restricted in their method of studying the sentences. They were allowed to sort the sentences into piles, and they were allowed to look at a given sentence any number of times. The subjects were not encouraged to use any particular strategy in studying the sentences; however, they were given written instructions describing exactly what their task would be during the recognition memory test: "You will be presented with a stack of cards. On each card is a sentence which involves either John or Sam. You are to study these sentences for five (eight) minutes. At the end of that time you will be given two sheets of paper which include each of the sentences on the cards with the words John and Sam replaced by John/Sam. Your task will be to recall for each sentence whether John or Sam was involved in that sentence as it appeared on the card. You are to indicate your response by circling one of the two words John or Sam in the given sentence on the sheet of paper." After studying the sentences on the cards, subjects were reminded of their task on the recognition memory test. Subjects were then given the appropriate test list of sentences and responded by circling the word John or Sam in each sentence, depending on which word they thought occurred in the sentence when it appeared on the card. Subjects were required to respond to every test sentence; they were not allowed to leave blanks. Subjects were given as much time as they needed to complete the recognition memory test.

Results and Discussion

The results are summarized in Table 4 in terms of mean percentages of errors on the recognition test as a function of condition and sentence type. The data were averaged over subgroups (A and B) since that factor was not found to be significant. The difference between the Deep Structure Syntactic and Semantic conditions in this experiment was striking. Performance on the deep structure syntactic relations was considerably worse than on the semantic and surface relations and, in fact, somewhat worse than on the arbitrary relations. Two analyses of variance were performed on these data, one with subjects (F_1) and one with sentences (F_2) as the random effect. According to these analyses, there was a significant effect of condition, *min* F' (3,46) = 5.5, p = .003 [F_1 (3,36) = 6.2, MS_e = 598, p = .002; F_2 (3,162) = 48.5, MS_e = 77, $p < .001$], an effect of sentence type which approached significance, *min* F' (5,205) = 2.1, p = .069 [F_1 (5,180) = 3.4, MS_e = 174, p = .006; F_2 (5,54) = 5.3, MS_e = 112, $p < .001$], as well as a significant interaction of these two factors, *min* F' (15,307) = 1.8, p = .029 [F_1 (15,180) = 2.7, MS_e = 174, p = .001; F_2 (15,162) = 6.0, MS_e = 77, $p < .001$]. As in Experiment I, certain sentence types caused more trouble for some conditions than for others. However, the nature of the interaction of condition and sentence type was somewhat different from that found in Experiment I.

Separate planned analyses of variance were conducted with the data from just the two critical conditions (Deep Structure Syntactic and Semantic). These analyses also revealed a significant effect of condition, *min* F' (1,21) = 6.8, p = .016 [F_1 (1,18) = 7.4, MS_e = 724, p = .014; F_2 (1,54) = 87.8, MS_e = 61, $p < .001$], but the effect of sentence type, *min* F' (5,144) = 1.2, p = .337 [F_1 (5,90) = 1.8, MS_e = 158, p = .129; F_2 (5,54) = 3.3, MS_e = 84, p = .011], and the interaction of condition and sentence type, *min* F' (5,138) = 1.8, p = .113 [F_1 (5,90) = 2.5, MS_e =158, p = .035; F_2 (5,54) = 6.5, MS_e = 61, $p < .001$], were not significant. Furthermore, Newman-Keuls tests, based on the overall analysis conducted with subjects as the random effect, revealed significant differences ($p < .05$) between the Deep Structure Syntactic and Semantic conditions, between the Arbitrary and Semantic conditions, and between the Arbitrary and Surface conditions, a significant difference ($p < .01$) between the Deep Structure Syntactic and Surface conditions, and nonsignificant differences between the Semantic and Surface conditions and between the Deep Structure Syntactic and Arbitrary conditions. It is clear from these data that subjects easily discovered the semantic case relations discriminating the *Sam* and *John* sentences when studying the sentences for a subsequent recognition memory test, and the subjects were not able to discover the deep structure syntactic relations so easily. These latter concepts were discovered no more easily than purely arbitrarily defined relations.

EXPERIMENT III

Only six sentence types were employed in Experiments I and II although there are eight possible combinations of the three kinds of relations. The two missing sentence types (7 and 8) had been excluded because it was difficult to find satisfactory examples of them. However, examples of these sentence types do exist and one of each of these two types is shown in Table 1. The problem with sentences of these types is that the case role of *John* seems to be ambiguous. Specifically, in type 7 sentences *John* seems to be in both the roles of experiencer and goal, although the experiencer role does seem more salient. In some type 8 sentences, it is not entirely clear whether *John* is in the role of goal or agent, a problem which exists for several sentences of other types as well. Despite these difficulties, all eight sentence types were employed in Experiment III, which was otherwise a replication of Experiment II. (Type 7 was selected to be a positive instance of the arbitrary concept, and type 8 was selected to be a negative instance of the arbitrary concept.) This experiment enabled us to test our contention

TABLE 4: Mean percentage of errors in Experiment II as a function of Condition and Sentence Type.

	Sentence Type						
Condition	1	2	3	4	5	6	Mean
Surface	4	3	2	1	1	5	2.7
Semantic	5	4	9	5	4	2	4.8
Deep Structure Syntactic	6	28	14	19	25	17	18.2
Arbitrary	4	17	7	11	26	33	16.3

TABLE 5: Mean percentage of errors in Experiment III as a function of Condition and Sentence Type.

	Sentence Type								
Condition	1	2	3	4	5	6	7	8	Mean
Surface	3.75	5.00	1.25	2.50	1.25	3.75	2.50	6.25	3.28
Semantic	5.00	7.50	7.50	7.50	6.25	10.00	6.25	11.25	7.66
Deep Structure Syntactic	21.25	26.25	20.00	30.00	23.75	22.50	16.25	28.75	23.59
Arbitrary	13.75	31.25	37.50	30.00	31.25	37.50	15.00	30.00	28.28

that the previous results were not due to any bias created by employing only six sentence types.

Method

Subjects. Forty male and female undergraduate students of Yale College who were taking a course in introductory psychology participated as subjects, receiving course credit. As in Experiment II, there were eight subgroups of subjects with five subjects in each subgroup. The assignment of subjects to subgroups was determined by time of arrival for testing according to a fixed rotation of subgroups.

Materials. Eight decks of cards were constructed for training, one deck for each of the eight subgroups of subjects. The decks were constructed in a manner strictly analogous to that employed for Experiment II except that there were 64, rather than 60, sentences in each deck, including 8, rather than 10, of each sentence type. The sentences employed are shown in the Appendix. The two sentences of each of the original six types included in Experiment II but not in Experiment III have an asterisk beside them in the Appendix.

Four typewritten lists of sentences were constructed for the recognition memory test in an analogous manner to the lists constructed for Experiment II. On each list the order of the sentences was pseudo-random, with the constraint that each 16-sentence block included two sentences of each type. The sentences within a given one of the four blocks were the same on the four lists, but the order of the sentences within a block differed across lists. In particular, the order of the sentences in a given block in the Semantic condition was the same as that in the Deep Structure Syntactic condition, except for the placement of eight of the sentences, including two of each of four types--sentences that were positive instances of the semantic concept but negative instances of the deep structure syntactic concept (types 3 and 7), and sentences that were negative instances of the semantic concept but positive instances of the deep structure syntactic concept (types 4 and 8). The two sentences of type 3 in the deep structure syntactic deck were replaced by the two sentences of type 4, and the two sentences of type 7 in the deep structure syntactic deck were replaced by the two sentences of type 8, and vice versa, to form the semantic deck. However, the two sentences of a given type maintained their position relative to each other. The lists for the Surface and Arbitrary conditions were analogously related to the list for the Deep Structure Syntactic condition. As in the earlier experiments, the relationships among the four lists insured that the sequence of correct answers (John or Sam) was the same for subjects in all four subgroups A, and was the same for subjects in all four subgroups B.

Procedure. The procedure was the same as in Experiment II except that all subjects were given five minutes to study the deck of sentences.

Results and Discussion

The results are summarized in Table 5 in terms of mean percentages of errors on the recognition test as a function of condition and sentence type. Despite the difference in the sentence types included, the pattern of results was strikingly similar to that found in Experiment II. Two analyses of variance were performed on these data, one with subjects (F_1) and one with sentences (F_2) as the random effect. These analyses yielded a significant main effect of condition, $\min F'$ $(3,46) = 10.1$, $p < .001$ [F_1 $(3,36) = 11.4$, $MS_e = 1024$, $p < .001$; F_2 $(3,168) = 84.6$, $MS_e = 111$, $p < .001$], but neither the main effect of sentence type, $\min F'$ $(7,149) = 1.1$, $p = .373$ [F_1 $(7,252) = 2.6$, $MS_e = 179$, $p = .013$; F_2 $(7,56) = 1.9$, $MS_e = 198$, $p = .090$], nor the interaction of condition and sentence type, $\min F'$ $(21,418) < 1$ [F_1 $(21,252) = 1.2$, $MS_e = 179$, $p = .238$; F_2 $(21,168) = 1.6$, $MS_e = 111$, $p = .063$], was significant.

Separate analyses of variance were conducted with the data from just the Semantic and Deep Structure Syntactic conditions. These analyses also yielded a significant effect of condition, $\min F'$ $(1,21) = 7.1$, $p = .015$ [F_1 $(1,18) = 7.6$, $MS_e = 1337$, $p = .013$; F_2 $(1,56) = 98.7$, $MS_e = 82$, $p < .001$]. In addition, Newman-Keuls tests, based on the overall analysis conducted with subjects as the random effect, revealed significant differences ($p < .01$) between the Deep Structure Syntactic and Semantic conditions, between the Arbitrary and Semantic conditions, between the Arbitrary and Surface conditions, and between the Deep Structure Syntactic and Surface conditions; but the differences between the Deep Structure Syntactic and Arbitrary conditions and between the Semantic and Surface conditions were not significant by these tests. The conclusions reached on the basis of Experiment II are clearly supported by this pattern of results.

SUMMARY AND CONCLUSIONS

These experiments indicate that the deep structure syntactic relations we studied were both learned more slowly and discovered less readily than the semantic case concepts we studied. These results suggest that the deep structure syntactic relations included in the standard theory are less accessible than semantic relations and, in fact, are no more accessible than arbitrarily defined relations. An implication of these findings is that the cued recall studies purporting to demonstrate the importance to sentence memory of the relation deep structure subject (Blumenthal, 1967; Blumenthal and Boakes, 1967) were misleading because of their confounding deep structure syntactic and semantic relations. The present study suggests that deep structure syntactic relations, when unconfounded from semantic relations, do not play a major role in sentence memory.

More generally, these results favor case grammar, rather than the standard theory, in a model of linguistic behavior. It should be noted, however, that although these results are clearly difficult for the standard theory, they do not discriminate among different semantically-based grammars.

Our rejection of the standard theory in a model of linguistic behavior may not seem very consequential for two reasons. First, there have been a proliferation of proposed revisions of the standard theory; however, some influential theorists have recently argued (see Bever, Katz and Langendoen, 1976) that the dismissal of the standard theory may have been too rash. It is also relevant to note (see footnote 1) that Fillmore, who initially rejected purely syntactic deep structure relations, now includes them, along with case relations, in his case grammar (Fillmore, 1977). Second, our results follow a number of others that have failed to provide support for the standard theory as a basis for linguistic behavior. However, the experimental evidence against the standard theory has been evidence against the grammatical operations, not against the structural descriptions of the theory (Fodor, Bever and Garrett, 1974). The present evidence against the standard theory pertains instead to the structural descriptions. Hence, our demonstration of the inaccessibility of deep structure syntax has definite consequence.

This study also has an important methodological implication. In studying the relative accessibility of various linguistic concepts, the new recognition memory paradigm developed here seems to be more sensitive than the traditional concept formation task. Two factors may have been responsible for the increase in sensitivity: (1) All testing occurred after training was completed in the memory paradigm. (2) The memory paradigm tested whether a given concept would be discovered by the subject rather than whether the concept could be learned by the subject.

APPENDIX

Sentences Used as Stimuli

(1.) DS Subject--SS Subject--Experiencer

1. John was sleepy near the fire.
2. John was warm near the radiator.
3. John was comfortable near the window.
4. John was nervous next to the swimming pool.
5. John was cool near the stream.
6. John was content on the balcony.
7. John was confident at the wheel.
8. John was speechless in the gallery.
*9. John was unhappy near the stage.
*10. John was at ease in the motor boat.

(2.) DS Subject--SS Object of Prep.--Experiencer

1. The accident was imagined by John.
2. The concept was visualized by John.
3. The concert was enjoyed by John.
4. The odor was savored by John.
5. The director was feared by John.
6. The story was believed by John.
7. The theory was respected by John.
8. The teacher was despised by John.
*9. The result was foreseen by John.
*10. The earthquake was felt by John.

(3.) DS Object of Prep.--SS Object of Prep.--Experiencer

1. The roar was deafening to John.
2. The nap was refreshing to John.
3. The suggestion was disturbing to John.
4. The wasp was annoying to John.
5. The mask was frightening to John.

6. The play was amusing to John.
7. The conclusion was astonishing to John.
8. The crime was puzzling to John.
*9. The voyage was exciting to John.
*10. The message was comforting to John.

(4.) DS Subject--SS Subject--Goal

1. John was the recipient of the grant.
2. John was the beneficiary of the allowance.
3. John was the inheritor of the mansion.
4. John was the borrower of the bicycle.
5. John was the thief of the porcelain.
6. John was the buyer of the refrigerator.
7. John was the acquirer of the painting.
8. John was the consignee of the suitcase.
*9. John was the receiver of the prize.
*10. John was the catcher of the ball.

(5.) DS Object of Prep.--SS Subject--Goal

1. John was given the book.
2. John was bequeathed the inheritance.
3. John was tossed the paper.
4. John was handed the spatula.
5. John was mailed the record.
6. John was awarded the medal.
7. John was dealt the ace.
8. John was paid the bribe.
*9. John was sent the instructions.
*10. John was assigned the duty.

(6.) DS Object of Prep.--SS Object of Prep.--Goal

1. The property was leased to John.
2. The bottle was passed to John.
3. The football was kicked to John.
4. The jewels were entrusted to John.
5. The materials were supplied to John.
6. The money was allotted to John.
7. The scholarship was granted to John.
8. The reward was presented to John.
*9. The book was returned to John.
*10. The results were communicated to John.

(7.) DS Object of Prep.--SS Subject--Experiencer

1. John was assured misery.
2. John was permitted remorse.
3. John was authorized exuberance.
4. John was provided serenity.
5. John was offered happiness.
6. John was guaranteed anxiety.

7. John was promised tranquility.
8. John was allowed timidity.

(8.) DS Subject--SS Object of Prep.--Goal

1. The fruit was obtained by John.
2. The merchandise was recovered by John.
3. The taxes were collected by John.
4. The prize was received by John.
5. The crop was gathered by John.
6. The ball was caught by John.
7. The donuts were taken by John.
8. The frisbee was retrieved by John.

*Sentences used in Experiments I and II but not in Experiment III.

REFERENCES

Anderson, J. R. and G. H. Bower. (1973) Human Associative Memory. (Washington, D.C.: Winston).

Baker, W. J., G. D. Prideaux and B. L. Derwing. (1973) Grammatical properties of sentences as a basis for concept formation. Journal of Psycholinguistic Research 2, 201-220.

Bever, T. G., J. J. Katz and D. T. Langendoen. (1976) An Integrated Theory of Linguistic Ability. (New York: Crowell).

Blumenthal, A. L. (1967) Prompted recall of sentences. Journal of Verbal Learning and Verbal Behavior 6, 203-206.

Blumenthal, A. L. and R. Boakes. (1967) Prompted recall of sentences. Journal of Verbal Learning and Verbal Behavior 6, 674-676.

Chomsky, N. (1965) Aspects of the Theory of Syntax. (Cambridge, Mass.: M.I.T. Press).

Clark, H. H. (1973) The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior 12, 335-359.

Fillmore, C. J. (1968) The case for case. In Universals in Linguistic Theory, ed. by E. Bach and R. T. Harms. (New York: Holt, Rinehart, & Winston), pp. 1-88.

Fillmore, C. J. (1970) Subjects, speakers, and roles. Synthese 21, 251-274.

Fillmore, C. J. (1971) Some problems for case grammar. In Report of the Twenty-Second Annual Round Table Meeting on Linguistics and Language Study, ed. by R. O'Brien. (Washington, D.C.: Georgetown University Press), pp. 35-56.

Fillmore, C. J. (1977) The case for case reopened. In Syntax and Semantics, Vol. 8, ed. by P. Cole and J. M. Sadock. (New York: Academic Press).

Fodor, J. A., T. G. Bever and M. F. Garrett. (1974) The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar. (New York: McGraw-Hill).

Katz, J. J. and J. A. Fodor. (1963) The structure of a semantic theory. Language 39, 170-210.

Kintsch, W. (1974) The Representation of Meaning in Memory. (Hillsdale, New Jersey: Erlbaum).

Lakoff, G. (1971) On generative semantics. In Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology, ed. by D. Steinberg and L. A. Jakobovits. (Cambridge: Cambridge University Press), pp. 232-296.

McCawley, J. D. (1968) The role of semantics in a grammar. In Universals in Linguistic Theory, ed. by E. Bach and R. T. Harms. (New York: Holt, Rinehart, & Winston), pp. 124-169.

Perfetti, C. A. (1973) Retrieval of sentence relations: Semantic vs. syntactic deep structure. Cognition 2, 95-105.

Rumelhart, D. E., P. H. Lindsay and D. A. Norman. (1972) A process model for long-term memory. In Organization of Memory, ed. by E. Tulving and W. Donaldson. (New York: Academic Press), pp. 197-246.

Schank, R. C. (1972) Conceptual dependency: A theory of natural language understanding. Cognitive Psychology 3, 552-631.

Shafto, M. (1973) The space for case. Journal of Verbal Learning and Verbal Behavior 12, 551-562.

Some Relationships between Articulation and Perception[*]

F. Bell-Berti,[+] L. J. Raphael,[++] D. B. Pisoni[+++] and J. R. Sawusch[++++]

ABSTRACT

Electromyographic studies of the American English vowel pairs /i-I/[**] and /e-E/[***] reveal two different production strategies: some speakers appear to differentiate the members of each pair primarily on the basis of tongue height; for others the basis of differentiation appears to be tongue tension. To determine if these differences in production might correspond to differences in perception, two vowel identification tests were given to the EMG subjects. Subjects were asked to label the members of a seven-step vowel continuum, /i/ through /I/. In one condition each item had an equal probability of occurrence. The other condition was an anchoring test: the first stimulus, /i/, was heard four times as often as any other stimulus. Compared with the equal-probability test labeling boundary, the boundary in the anchoring test was displaced toward the more frequently occurring stimulus. The magnitude of the shift of the labeling boundary was greater for subjects using a production strategy based on tongue height than for subjects using tongue tension to differentiate these vowels, suggesting a possible link between strategies used in speech production and aspects of perceptual analysis of speech sounds.

[*] Preliminary versions of this paper were presented at the 9th International Congress on Acoustics, Madrid, Spain, 4-9 July 1977, and at the 95th Meeting of the Acoustical Society of America, Providence, Rhode Island, 16-19 May, 1978.

[+] On leave, Montclair State College, Upper Montclair, New Jersey.

[++] Also Herbert H. Lehman College, The City University of New York.

[+++] Department of Psychology, Indiana University, Bloomington, Indiana.

[++++] Department of Psychology, State University of New York at Buffalo.

[**] Throughout this paper /I/ stands for I.P.A. /ɪ/.

[***] Throughout this paper /E/ stands for I.P.A. /ɛ/.

__Acknowledgment__: This work was supported in part by NINCDS grants NS-13617, NS-05332; NICHHD grant HD-01994; BRSG grant RR-05596; NIDR grant DE-01774; and by NINCDS grant NS-12179 and NIMH grant MH-24027 to the Department of Psychology, Indiana University.

[HASKINS LABORATORIES: Status Report on Speech Research SR-55/56 (1978)]

INTRODUCTION

It is generally true that studies of human speech communication have been directed to questions of how speech is perceived or how speech is produced, with little direct investigation, and much speculation, about how these two events may be linked. While different schools have posited different cause-effect directions between these events (for example, the motor theory of speech perception and the acoustic theory of speech production), few, if any, suggest that the two are completely independent. For, whichever direction the relationship may go, it is the acoustic signal produced by the human vocal tract and perceived by the human auditory system for which the theories must account (cf. Liberman, Cooper, Harris and MacNeilage, 1962; Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967; Ladefoged, DeClerk, Lindau and Papçun, 1972; Stevens and Halle, 1967; Stevens and House, 1972; Cooper, 1972).

Experimental support for the view that speech production and perception are mediated by a common mechanism may be found in several recent studies by Cooper (1974; Cooper and Lauritsen, 1974; Cooper and Nager, 1975) dealing with perceptuo-motor adaptation. These studies have shown that immediately after listeners are presented with many repetitions of a voiceless stop consonant, their voice onset time (VOT) values decrease as they produce voiceless stop-plus-vowel sequences. These results were similar to the results of selective adaptation experiments in which repetitive listening to an adaptor altered the perception of a test series varying in VOT (Eimas and Corbit, 1973). The perceptuo-motor adaptation studies suggest that the interaction between speech production and perception can be directly demonstrated. This might be accomplished by discovering already existing (as opposed to experimentally induced) differences between the production strategies of two (or more) populations of speakers for a given class of sounds, and then showing that those differences were isomorphic with the differing perceptual behaviors of the populations for the same class of sounds. The experiments reported here were designed to investigate the possibility of such a perceptuo-productive isomorphism for several members of the class of English vowel sounds.

The members of the front series of English vowels /i-I-e-E/ have been variously described in the phonetics literature either as differing among themselves in both tongue height and duration, or as differing within the pairs /i-I/ and /e-E/ in tongue tension with consequent differences in duration and differing between pairs in tongue height. In an earlier study of the production of these and other vowels (Raphael and Bell-Berti, 1975), we obtained electromyographic (EMG) recordings from the extrinsic tongue muscles of three speakers of American English to discover which, if any, of these muscles displayed a difference in overall amount of activity corresponding to the traditional tense-lax distinction between members of the English vowel pairs /i-I/, /e-E/ and /u-ʊ/. The data we gathered provided support for the notion that tension is a necessary, or sufficient, differentia of production for some speakers, but not for others. In the present study we hoped to determine whether differences in vowel production might in some way be related to differences in vowel perception, particularly vowel identification. To this end, we collected EMG as well as vowel-identification data from a group of eight subjects.

THE PRODUCTION EXPERIMENT

Method

EMG potentials from the genioglossus muscle were recorded with bipolar hooked-wire electrodes inserted percutaneously into the muscle. The action of this muscle is to bunch the mass of the tongue and draw it forward in the oral cavity, especially for the high- and mid-front vowels (Smith[1]; Harris, 1971; Raphael and Bell-Berti, 1975; Kakita[2]; Raphael, Bell-Berti, Collier and Baer, in press).

The utterances used in this experiment included the vowels /i/, /I/, /e/, and /E/, produced in a əpVp frame. The utterances were placed in random lists that were read by each subject until 18 to 30 tokens of each utterance type were recorded.

The EMG potentials and the speech signal were recorded on an FM tape recorder. The onset of voicing of the stressed vowel was identified visually for each syllable. The repetitions of each utterance type were aligned with reference to this point, and the EMG data subsequently computer sampled and averaged.

Results

Figure 1 contains examples of the two patterns of muscular contraction found for the front vowels. In one pattern, (talker KSH) there is a decreasing order of activity corresponding to the traditional articulatory description of tongue height for the front vowels (Figure 1a): peak activity decreases through the vowel series /i-I-e-E/. Further, contrasting the EMG curves of /i/ and /e/ with those of /I/ and /E/, it is evident that the former are bimodal, perhaps reflecting the diphthongization of the "tense" vowels. In the second (LJR) pattern of muscular contraction (Figure 1b) there are greater, and almost identical, levels of muscle activity for the two tense vowels /i/ and /e/, and a considerably lesser degree of activity for the two lax vowels, /I/ and /E/; that is, there is a decreasing order of activity through the vowel series /i-e-I-E/.[3] Further, the EMG curves are smooth and unimodal, compared with those of the other pattern, described above, and perhaps reflecting a different strategy for diphthongization of the tense vowels for this speaker.

[1] Smith, T. St. J. (1970) A Phonetic Study of the Function of the Extrinsic Tongue Muscles. Unpublished Ph. D. dissertation, U. C. L. A.

[2] Kakita, K. M. (1976) Activity of the Genioglossus Muscle During Speech Production: An Electromyographic Study. Unpublished D. M. S. dissertation, University of Tokyo.

[3] It must be remembered, however, that the differences in absolute microvolt potential and duration of activity separating /i/ from /e/ and /I/ from /E/ are quite small.

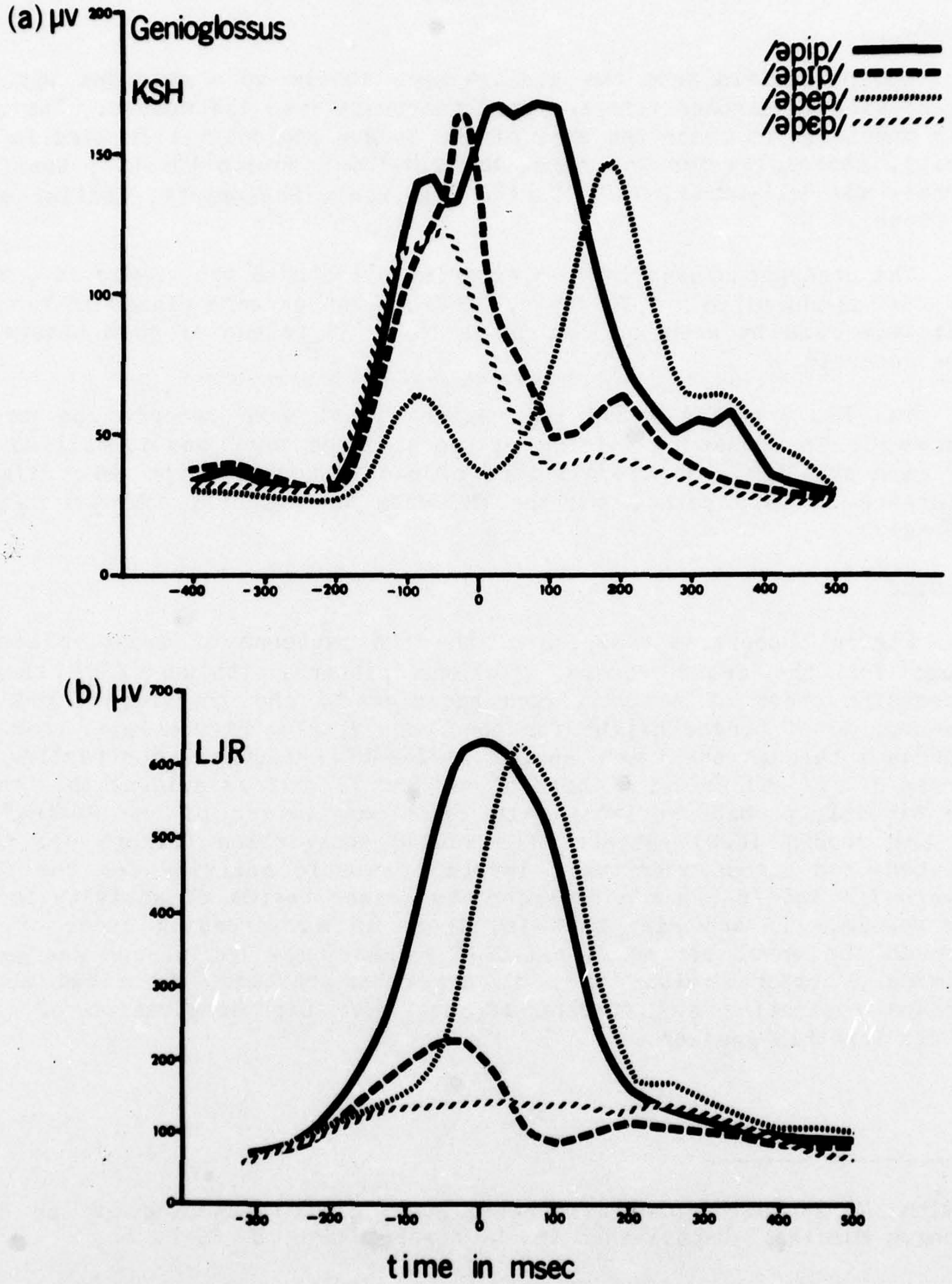


Figure 1: Averaged EMG activity of the genioglossus muscle for the vowels /i,I,e,E/, for: (a) subject KSH; (b) subject LJR.

Although both these patterns of muscular contraction preserve the tense-lax distinction between /i-I/ and /e-E/, they do so in markedly different ways. In the first pattern, corresponding to the traditional picture of tongue height for front vowels, the traditionally lax /I/ is characterized by the data as being more tense than the traditionally tense /e/, and being very close to /i/ with regard to tension--although not with regard to duration. That is, peak EMG activity decreases through the series /i-I-e-E/, suggesting a distinction among vowels that reflects the usual description of a tongue height or tongue bunching continuum. The second pattern, however, does not correspond to the usual description of tongue height, with /i/ and /e/, both traditionally described as tense vowels, showing considerably more genioglossus activity than lax /I/ and /E/. That is, for this pattern EMG activity decreases through the series /i-e-I-E/.

The question posed by these production data for theories of speech perception is whether apparent differences in the necessary differentiae for production are reflected in differentiae employed in perception. That is, do talkers who rely on different mechanisms for producing vowels also rely on different properties or strategies in perceiving them? In order to answer this question we turned to tests of perception.

THE PERCEPTION EXPERIMENTS

Method

Two types of vowel perception tests were administered to each of two groups of listeners.[4] The tests were composed from a continuum of vowel stimuli ranging from /i/ to /I/ in seven steps. The vowels were synthesized on the vocal tract analogue synthesizer at the Research Laboratory of Electronics at M. I. T. (Figure 2). The frequencies of the first three formants were varied in equal logarithmic steps, while those of the fourth and fifth formants were held constant at 3500 Hz and 4500 Hz, respectively, for all seven stimuli. Vowel duration was 300 msec, with rise and decay times of 50 msec. The fundamental frequency fell linearly from 125 Hz to 80 Hz across the duration of each vowel.

These seven vowel stimuli were recorded on magnetic tape as two test series. In the control series, each of the stimuli occurred ten times; in the anchor series, stimulus 1 (/i/) occurred 40 times and each of the other stimuli occurred ten times. In both series, stimuli were presented one at a time with a 4-second pause between successive items. Subjects were asked to identify the stimuli as either /i/ or /I/. All subjects listened to two presentations of the control series followed by the anchor series.

[4] The first group of listeners consisted of 137 students at either Indiana University or the State University of New York at Buffalo. The second group of 13 listeners consisted of students or research associates at Haskins Laboratories at the time of their participation in this study. Except for the three subjects whose EMG data were reported by Raphael and Bell-Berti (1975), the subjects were not told the purpose of the perception or production experiments until after both sets of data had been collected.

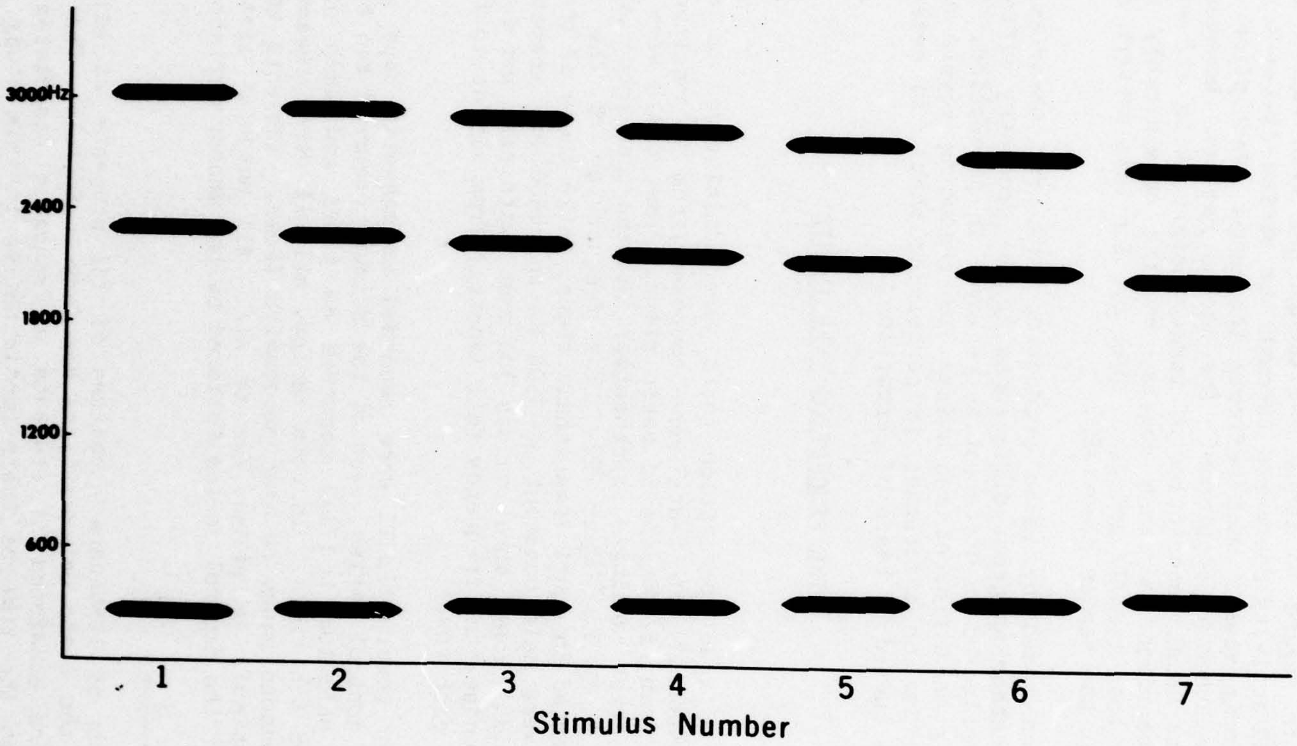


Figure 2: Schematic spectrograms of the seven synthesized vowel stimuli.

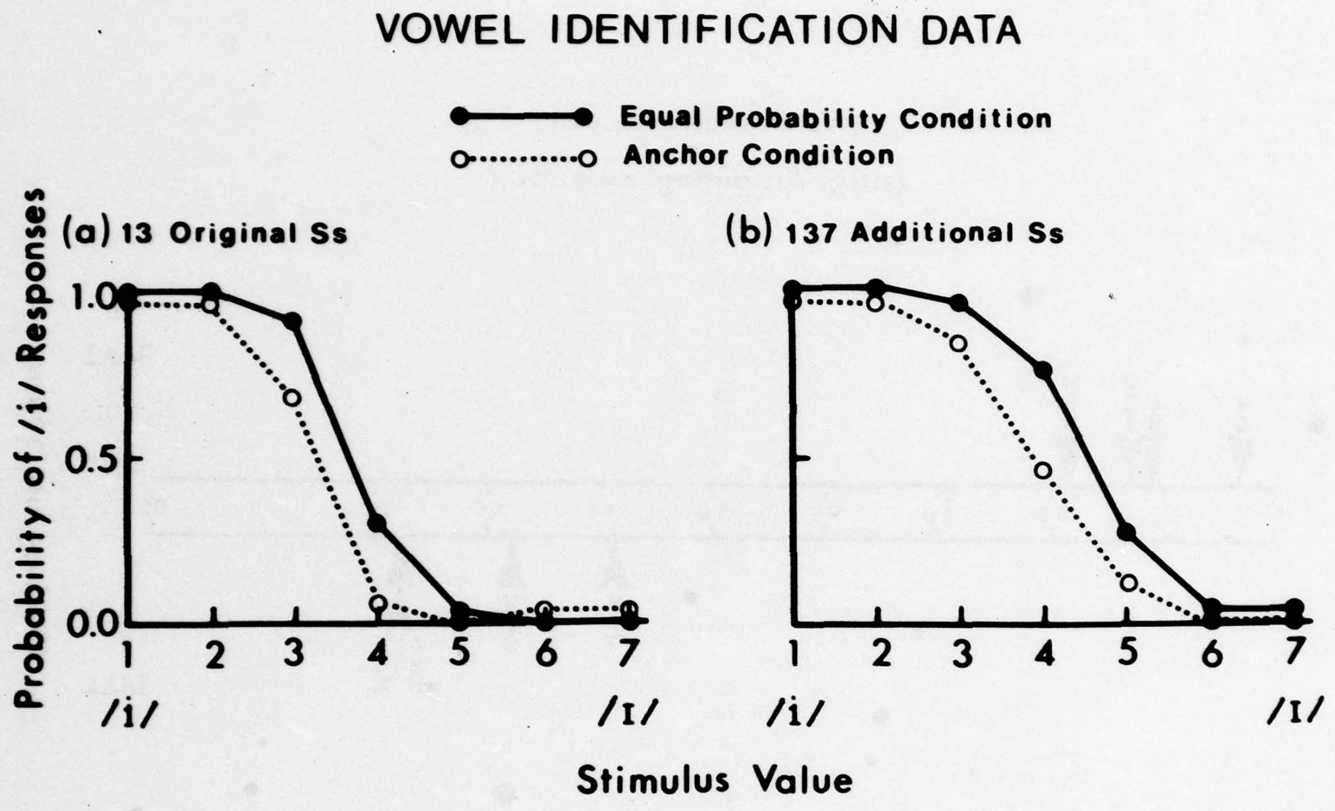
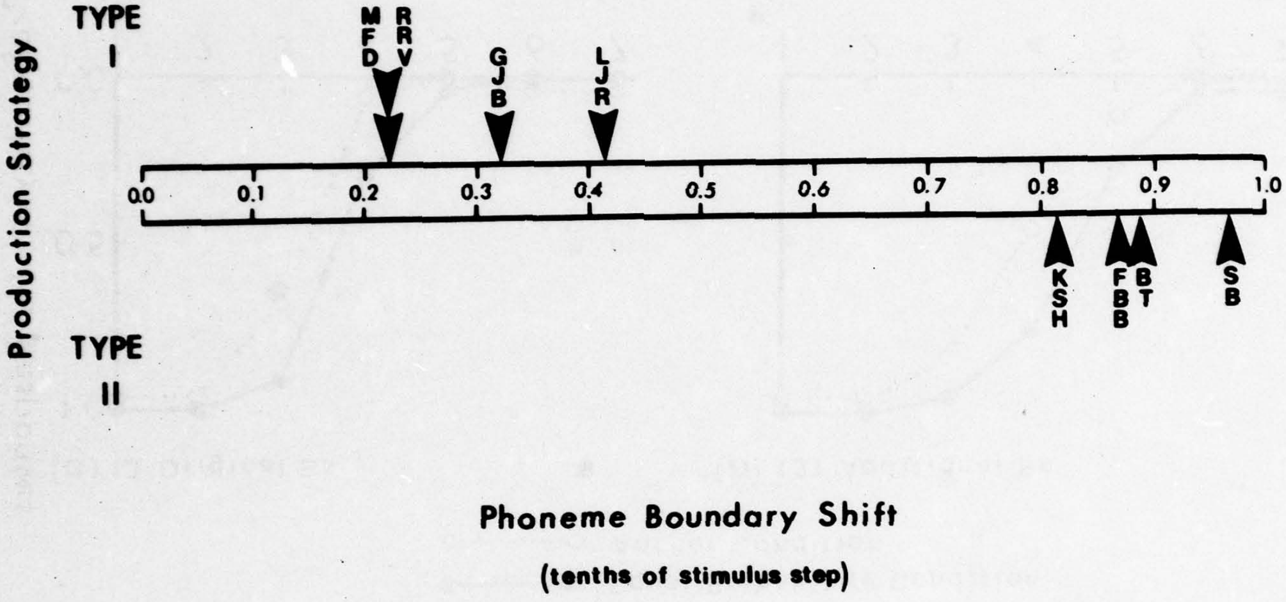


Figure 3: Vowel identification data in equal-probability and anchor-condition tests.
(a) Pooled identification functions from the original 13 subjects.
(b) Pooled identification functions from the additional 137 subjects.

Figure 4: Distribution of phoneme boundary shifts for eight subjects in EMG experiment.



Results

The pooled identification data for the group of 13 subjects, for both control and anchor conditions, are shown in the left panel of Figure 3. The analogous data for the group of 137 subjects are shown in the right panel of Figure 3. Each of the group of 13 subjects showed a shift in the /i-I/ category boundary toward /i/ for the anchor series compared with the control series. The group shift, indicated in the left panel of Figure 3, was highly significant ($t_{(12)}=5.72$, p .001 using a two-tailed, correlated t-test). Of the group of 137 subjects, 133 showed the expected shift in the category boundary toward /i/, a highly significant result ($p=2x10^{-10}$ using a sign test).

COMPARISON OF THE PRODUCTION AND PERCEPTION DATA

Four of the eight EMG subjects displayed a pattern of production in which the traditionally delineated order of tongue height for the front vowels is reflected in peak EMG activity for the genioglossus muscle; these four subjects demonstrated susceptibility to anchoring effects in the unequal-probability condition, as measured by the magnitude of shift in the locus of the phoneme boundary brought about by anchoring. The results for these subjects are shown at the right in Figure 4. The four remaining EMG subjects displayed a pattern of production in which the feature of tongue tension is reflected in the EMG data from the genioglossus muscle; these subjects demonstrated relatively small shifts in the anchoring condition, and are shown at the left of Figure 4.

Figure 5 shows the distribution of magnitudes of category boundary shifts in the anchoring condition for all 150 subjects. This distribution displays prominent peaks in the .2- to .3-unit range and in the .7- to .8-unit range, implying the existence of real differences among listeners in susceptibility to /i/-anchoring. It should also be noted that the average boundary shift for the four EMG subjects who show an /i-e-I-E/ vowel ordering was .29 stimulus units, a value coinciding with the first peak in the distribution shown in Figure 5. The average boundary shift for the four EMG subjects who showed an /i-I-e-E/ vowel ordering was .88 stimulus units, a value slightly larger than the second peak in the group distribution, but reasonably close to the .7- to .8-unit range.

DISCUSSION

Assuming that the results of these experiments have demonstrated the existence of an interaction between individual strategies for speech production and perception, there are several possible explanations for these results worthy of discussion and further study. One explanation is that these results are due to the correspondence between the acoustic patterns of the stimuli in the anchoring experiment and the particular articulatory strategies of the subjects, within each subject's internalized phonetic space.

Larger anchoring effects were found for the subjects displaying a pattern of muscular activity that parallels the ordering of the vowels in the test continuum used in the perceptual experiments; that is, both genioglossus activity, in production, and the changes in formant frequencies, in percep-

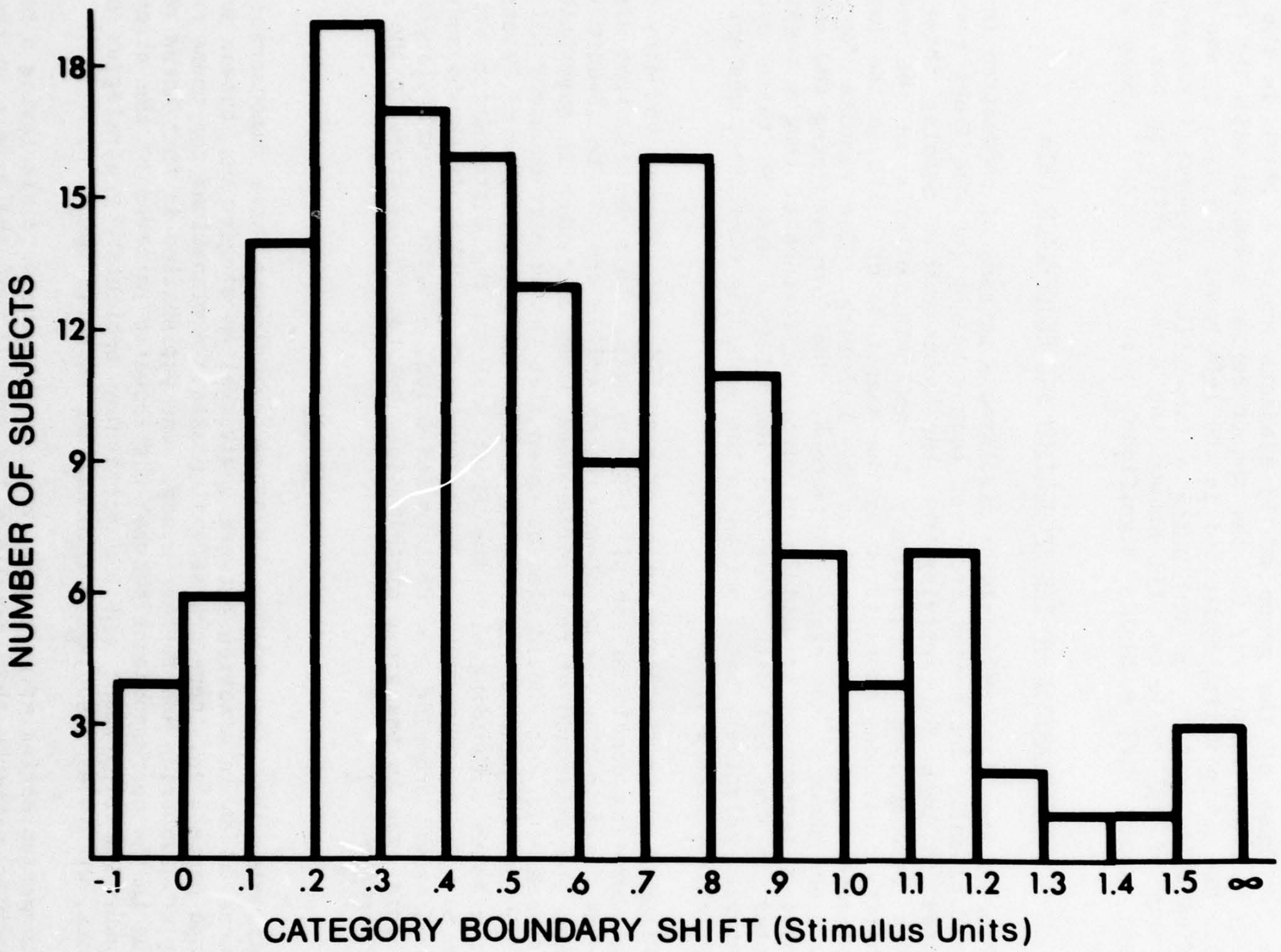


Figure 5: Distribution of shifts in anchoring condition, for 150 subjects.

tion, reflect changes from /i/ to /I/. In contrast, smaller anchoring effects were found for the subjects displaying a pattern of muscular activity that does not correspond to the ordering of the vowels in the test continuum; that is, genioglossus activity decreased from /i/ to /e/ to /I/, whereas the test continuum contained no intervening perceptual category between /i/ and /I/. Thus, susceptibility to anchoring effects in vowel identification may be substantially reduced for these latter subjects because the test stimuli represent vowels that are not contiguous within each subject's articulatory or phonetic space.

Thus, one hypothesis for explaining these data is that differences in perception reflect the different articulatory strategies of tongue height and tongue tension as ways of realizing the vowels in the set /i,I,e,E/. This hypothesis, and alternatives not considered here,[5] should be extended to other perceptual, articulatory and acoustic dimensions, especially temporal dimensions. Although many investigators have argued for the existence of a common mechanism linking speech perception and production, the evidence typically cited in support of these views was, by necessity, often indirect. We believe that the data of the present experiment, although preliminary, provide a more direct and convincing demonstration of the existence of some common mechanism or process that mediates at least some aspects of the production and perception of vowels. The extent to which these initial findings can be replicated and then extended to other phonetic distinctions obviously awaits the results of additional experiments specifically designed to reveal the interaction between dimensions of speech production and perception.

REFERENCES

Cooper, F. S. (1972) How is language conveyed by speech? In Language by Ear and by Eye, ed. by J. F. Kavanagh and I. G. Mattingly. (Cambridge: M. I. T. Press).

Cooper, W. E. (1974) Perceptuomotor adaptation to a speech feature. Perception and Psychophysics 16, 229-234.

Cooper, W. E. and M. R. Lauritsen. (1974) Feature processing in the perception and production of speech. Nature 252, 121-123.

Cooper, W. E. and R. M. Nager. (1975) Perceptuo-motor adaptation to speech: An analysis of bisyllabic utterances and a neural model. Journal of the Acoustical Society of America 58, 256-265.

Eimas, P. D. and J. D. Corbit. (1973) Selective adaptation of linguistic feature detectors. Cognitive Psychology 4, 99-109.

Harris, K. S. (1971) Action of the extrinsic musculature in the control of tongue position: Preliminary report. Haskins Laboratories Status Report on Speech Research SR-25/26, 87-96.

Ladefoged, P., J. DeClerk, M. Lindau and G. Papçun. (1972) An auditory-motor theory of speech production. Univ. Calif., Los Angeles, Working Papers in Phonetics 22, 48-75.

Liberman, A. M., F. S. Cooper, K. S. Harris and P. F. MacNeilage. (1962) A motor theory of speech perception. In Proceedings of the Speech

[5] For example, other accounts of the perceptual differences might rely on durational differences among these vowels.

Communication Seminar, Stockholm, Sept. 1962.

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

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

Raphael, L. J., F. Bell-Berti, R. Collier and T. Baer. (in press) Tongue position in rounded and unrounded front vowel pairs. Language and Speech, 22(1).

Stevens, K. N. and M. Halle. (1967) Remarks on analysis by synthesis and distinctive features. In Models for the Perception of Speech and Visual Form, ed. by W. Wathen-Dunn. (Cambridge: M. I. T. Press).

Stevens, K. N. and A. S. House. (1972) Speech perception. In Foundations of Modern Auditory Theory, ed. by J. Tobias. (New York: Academic Press).

Reflex Activation of Laryngeal Muscles by Sudden Induced Subglottal Pressure Changes

Thomas Baer

ABSTRACT

In measuring the effect of subglottal pressure changes on fundamental frequency of phonation, the effects of changing laryngeal muscle activity must be eliminated. Several investigators have used a strategy in which pulsatile increases of subglottal pressure are induced by pushing on the chest or abdomen of a phonating subject. Fundamental frequency is then correlated with subglottal pressure changes during an interval before laryngeal response is assumed to occur. The present study was undertaken to repeat such an experiment while carefully monitoring electromyographic activity of some laryngeal muscles, since previous investigators may have overestimated the latency of laryngeal response. The results showed a rapid and consistent response to each push. The latency of the response was about 30 msec. However, analyses of fundamental frequency versus subglottal pressure changes during this interval were in general agreement with previously published values. In considering the nature of the electromyographic response, its timing was found to be within the range of latencies appropriate for peripheral feedback, and was also similar to that for an acoustically- or tactually-elicited startle reflex.

INTRODUCTION

The effect of subglottal pressure changes on fundamental frequency of phonation has been a subject of theoretical and research interest for the past twenty years. However, measurement of this effect is difficult because, during normal phonation, changes in subglottal pressure and in the activity of laryngeal muscles are usually correlated (for example, Atkinson, 1978). To measure the effects of subglottal (or transglottal) pressure changes on fundamental frequency, the effects of changing laryngeal muscle activity must be eliminated. A commonly adopted strategy intended to eliminate laryngeal muscle effects on fundamental frequency is to induce changes in subglottal pressure experimentally, and to measure fundamental frequency before laryngeal reaction is presumed to occur. Several studies have been reported in which pulsatile increases of subglottal pressure are produced by pushing suddenly at random intervals on the chest or abdomen of a phonating subject (for example,

Acknowledgment: The author is grateful to Drs. Hajime Hirose and Seiji Niimi, who performed all the physiological procedures reported here. This research was supported by NIH grants DE01774 and NS13870.

van den Berg, 1957; Isshiki, 1959; Ladefoged, 1963; Öhman and Lindqvist, 1966; Fromkin and Ohala, 1968). Van den Berg (1957), for instance, argued that "no known human reflexes can occur faster than 100 msec," so that if fundamental frequency is correlated with subglottal pressure during this interval, laryngeal configuration should be considered constant.

Results from these experiments and others with a somewhat different paradigm (Lieberman, Knudsen and Mead, 1969; Hixon, Klatt and Mead, 1971), have shown general agreement. The sensitivity of fundamental frequency to subglottal pressure is usually found to be in the range of 3 to 7 Hz/cm H20, with the higher values occurring at the higher fundamental frequencies or in falsetto.

Constancy of laryngeal configuration can only be ensured if electromyographic (EMG) signals from the laryngeal muscles are recorded systematically. Although some of the cited investigators apparently monitored EMG signals and eliminated from processing those tokens for which the EMG showed sizable variations, these results were not reported in detail. Furthermore, the experiments reported depend on the unproven assumption that reflex latencies must exceed 100 msec. The assumption is certainly true for voluntary reaction times. For example, Draper, Ladefoged and Whitteridge (1960) report a minimum reaction time of 140 msec in a respiratory control task, and Netsell and Daniels (1974) report a similar latency for EMG signals associated with voluntary lip movements. Voluntary adjustments for initiating phonation also appear to be limited by a similar latency (Izdebski and Shipp, 1976). However, a peripheral reflex might well be much faster. The mechanical eyeblink response to an acoustic startle stimulus begins at about 40 msec (Landis and Hunt, 1939). Reflex activation of respiratory muscles with EMG latencies of 33-80 msec to sudden changes in pressure has been reported by Sears and Newsom Davis (1968). The laryngeal protective-closure reflex is similarly biologically basic, and the laryngeal adductor muscles are among the fastest in the body, with mechanical response times as short as 15 to 20 msec (Sawashima, 1974; Atkinson, 1978).

Because, as the above discussion indicates, reflex adjustments could have interfered with previous measurements, it seemed useful to repeat a chest-pushing experiment while carefully monitoring EMG signals from laryngeal muscles, as well as the subglottal pressure and voice waveforms. A series of experiments with one subject was performed. The results showed a rapid and consistent EMG response to each chest push. Although this response could have affected results of previous investigations, analyses of fundamental frequency versus subglottal pressure during the period before this response could occur were in general agreement with published values. The nature of the EMG response itself is of interest, and is considered further.

Methods

In the initial experiment, the subject sat upright in a dental chair and produced steady phonation at, successively, three different fundamental frequencies (nominally 94, 110 and 220 Hz) in chest voice and one (240 Hz) in falsetto while the experimenter pushed sharply and at random intervals on his chest. Some double pushes were used, but these were eliminated during later stages of processing. Pressure was measured through a catheter inserted into the lower subglottal space of the larynx through the cricothyroid membrane, while the voice signal was recorded through a standard microphone. The EMG signals were recorded with hooked-wire electrodes. Suitable signals were obtained from two laryngeal adductor muscles, the vocalis (VOC) and the interarytenoid (INT). Some variations of this procedure were introduced in two later runs, as discussed below.

The data were processed using the Haskins Laboratories' processing system for physiological signals. Measurements of subglottal pressure and rectified electromyographic activity were obtained by integrating the signals in contiguous, nonoverlapping 5-msec intervals. The voice waveform synchronous with the pressure and EMG signals was sampled and digitized at a 10 kHz rate. From this waveform, an estimate of the fundamental frequency over each 5-msec interval was obtained by an autocorrelation method (Lukatela, 1973). Measures of the amplitude of the audio waveform were also made over 5-msec intervals. Thus, the sampling rate for all these signals was 200/second, and no further smoothing was applied.

Ensemble averages and standard deviations were calculated for each of the four conditions. For each condition, the data were carefully aligned for averaging at the onset of the rise in subglottal pressure. Since the line-up point was the beginning of the push, as measured from the onset of pressure rise, the ensemble averages reflect the average response to a push.

Results

Some results from the first experiment are illustrated in Figure 1. This figure shows results from the lowest F_0 condition (94 Hz). Each column contains waveforms for one variable. The line-up point is shown by the vertical line through each waveform. The top row shows, for each column, the average waveform based on 18 tokens. The four rows underneath contain waveforms from the first four of these 18 tokens. The columns contain, from left to right, waveforms for subglottal pressure, acoustic variables (amplitude envelope and fundamental frequency) and electromyographic signals from the vocalis and interarytenoid muscles.

As this figure shows, the subglottal pressure builds up rapidly at the onset of a push, and then falls more slowly. As could be expected, the amplitude envelope reflects, at least grossly, the variations in subglottal pressure, as do variations in fundamental frequency. The EMG signals for individual tokens are noisy, but clearly contain peaks of activity in an interval after the line-up point. These peaks appear consistently. Although it would be difficult to measure the timing of the peaks from individual pushes, their timing appears grossly constant, so that averaging seems appropriate. Timing of the average waveform is more easily measured than that for individual tokens. This timing will be discussed below.

Figure 2 shows the average waveforms from each of the four conditions in vertical alignment to display the timing relationships clearly. In all four conditions, subglottal pressure, amplitude envelope, and fundamental frequency begin to rise immediately at the line-up point. In all three chest voice conditions, vocalis and interarytenoid activity rise sharply at a latency of about 30 to 40 msec after the line-up point. In the lowest F_0 condition,

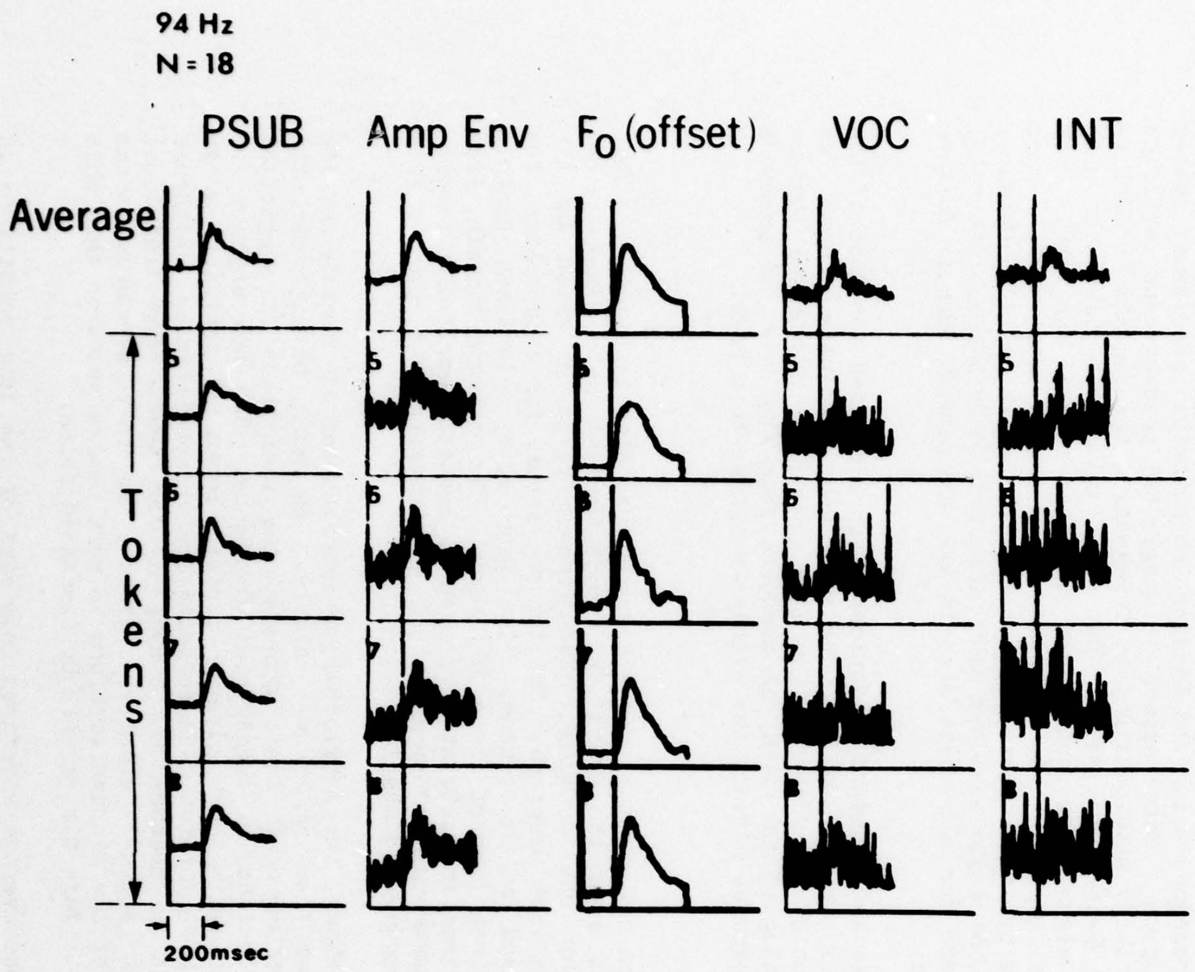


Figure 1: Results from the lowest-fundamental-frequency condition (see text). Ordinate scale values, not shown here, are the same as in the next figure.

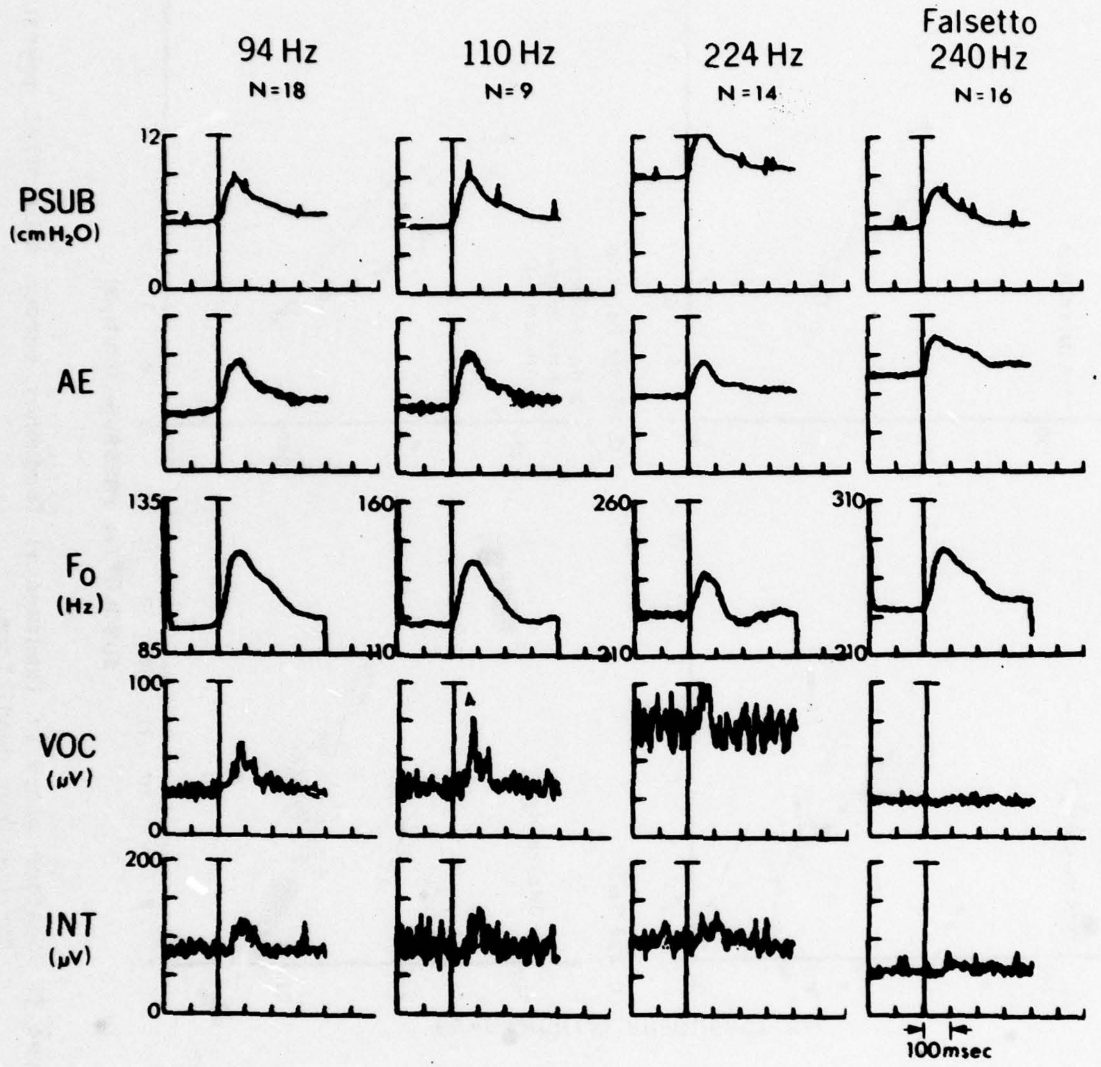


Figure 2: Average waveforms from all four conditions. Note that the scale factor for the F_0 plot in the falsetto condition is different from that for the other 3 conditions.

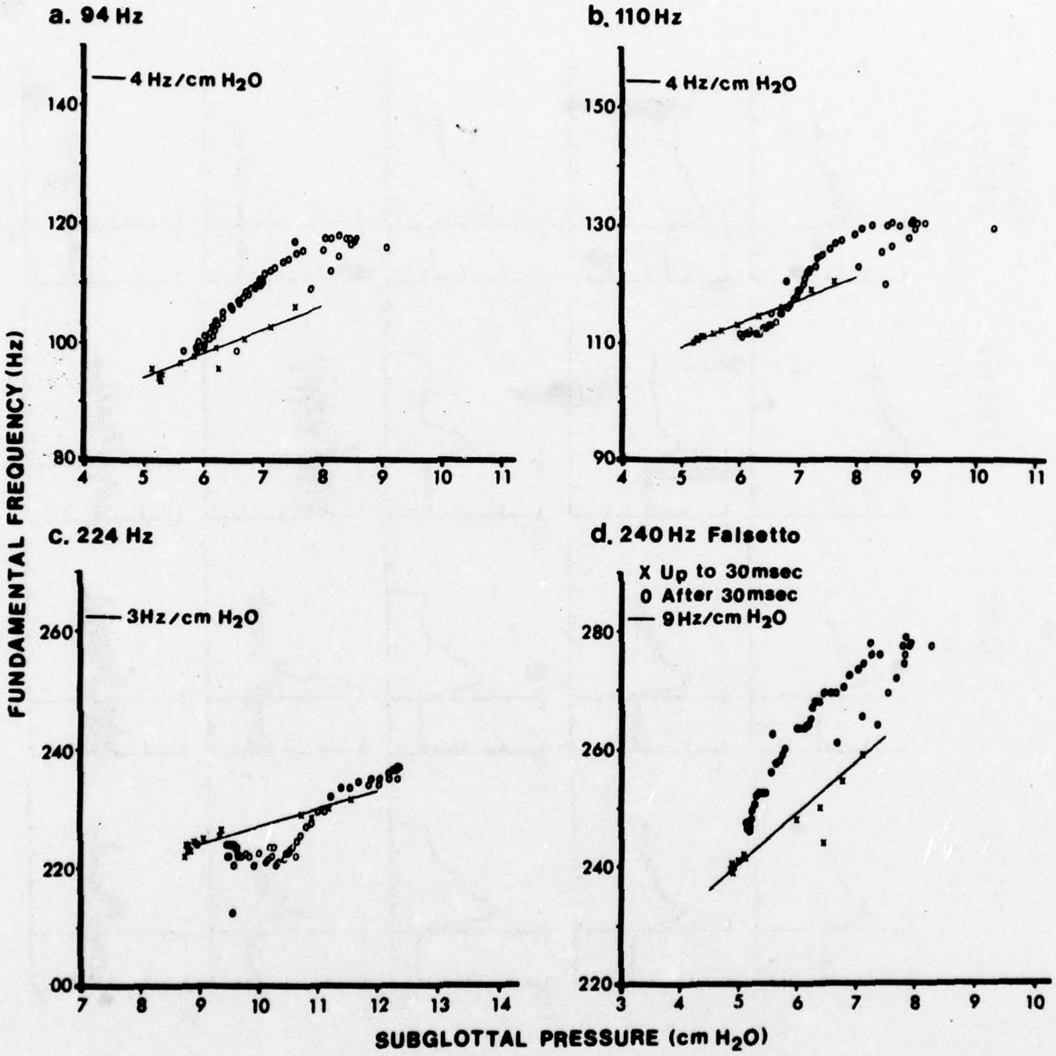


Figure 3: Scatter plots of fundamental frequency versus subglottal pressure for the four conditions.
3a. 94 Hz.
3b. 110 Hz.
3c. 224 Hz.
3d. 240 Hz. (Falsetto)

there is also a gradual rise of activity immediately after the line-up point. This apparent activity may reflect a "microphonic effect." The amplitudes of vocal-fold vibration are very large at such low phonatory frequencies, and the resulting movements probably modulate the interelectrode distance, producing a modulation of the recorded signal at the phonatory frequency. With increases of subglottal pressure, the amplitudes of vibration, and hence the microphonic effect, increase. For the falsetto condition, the activity of the laryngeal adductors was relatively low, and there appeared to be little or no response to the chest push.

Figure 3 shows scatter plots of the average fundamental frequency versus the average subglottal pressure waveforms for the four conditions shown in Figure 2. Points corresponding to times up to 30 msec after the line-up point are plotted with different symbols from those that occur later. The trajectories of the earlier points are reasonably well fit by straight lines. For the three chest-voice conditions (Figures 3a, b, and c), the slopes of these lines are 4, 4, and 3 Hz/cm H20, respectively. For the falsetto condition, shown in Figure 3d, the slope was greater: 9 Hz/cm H20. This aspect of the results agrees qualitatively with the consensus of earlier studies. Thus, although the results confirm the initial hypothesis, that muscle activity might have interfered with previous investigations, the actual results of those investigations are substantiated.

Beyond 30 msec, the trajectories depart from linearity. Perhaps the onset of muscular activity accounts for some of this apparent nonlinearity. However, this departure from linearity cannot be fully explained on the basis of EMG activity whose latency is 30 msec. At least 15 additional msec would be required to account for the mechanical response time of the muscles. Therefore, nonlinearity seems to be inherent in the relationship between fundamental frequency and subglottal pressure.

Response latencies of 30 to 40 msec are too short to be centrally mediated, and therefore must be due to peripheral reflexes; but before it can be concluded that the observed EMG responses were actually due to a peripheral reflex, it must be established that the subject had no cue for the impending chest-push prior to the onset of the pressure rise. In the foregoing results, all latencies were referred to the beginning of the pressure rise. The possibility was considered that the beginning of the push may have significantly preceded the pressure rise, and therefore tactile cues may have been available to the subject.

To establish this latency, a smaller experiment was performed. The chest-pushing procedure was repeated, and a contact switch was fitted on the subject's chest. The experimenter pushed directly on the switch. A signal that suddenly increased with switch closure was recorded in parallel with the voice waveform and with the EMG activity from the vocalis muscle. Subglottal pressure was not recorded, since the foregoing results show that increase of subglottal pressure is immediately reflected in the voice amplitude envelope.

Results from this experiment are shown in Figure 4. The upper plot shows the switch contact signal. Results from 23 pushes have been averaged, and all signals have been aligned with respect to switch closure. The middle plot shows the voice amplitude envelope. This signal begins to increase exactly at

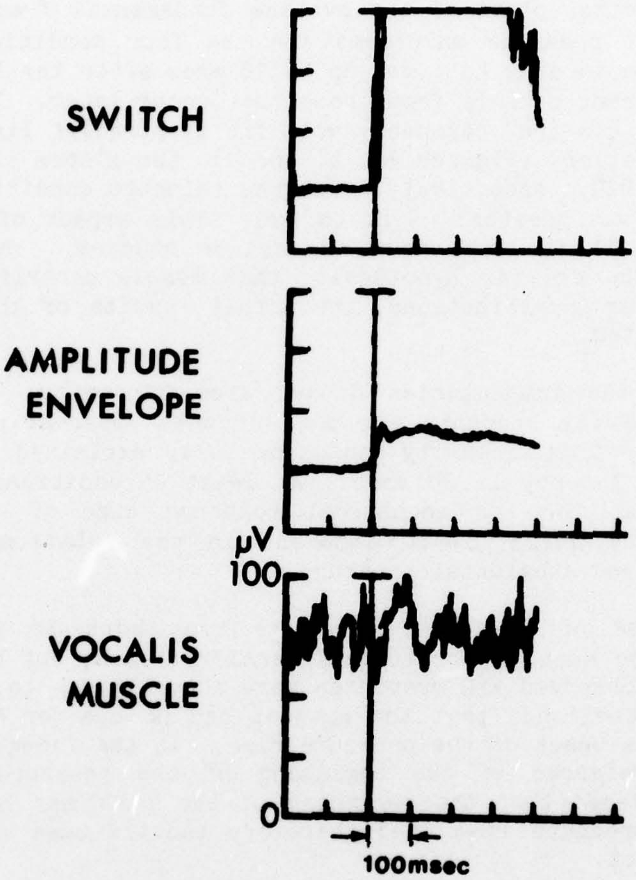


Figure 4: Results from experiment with contact switch. The upper panel shows the average switch contact signal, based on 23 pushes.

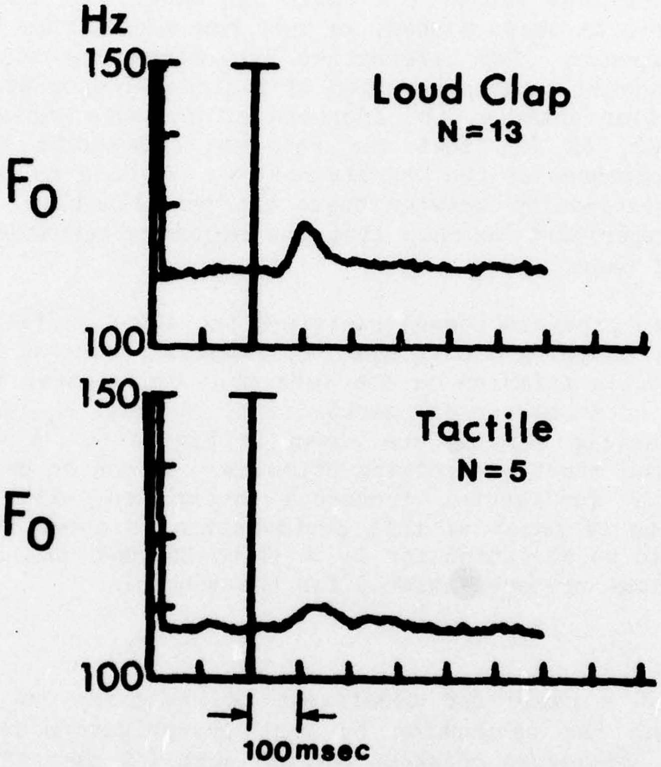


Figure 5: Average perturbations in fundamental frequency caused by startle reflexes elicited by auditory and tactile stimuli.

the line-up point. Hence, the delay between the onset of the physical push and the beginning of the pressure rise must be within the 5 msec sampling interval of the processing system. The bottom plot shows simultaneous activity of the vocalis muscle. Though this is a noisy waveform, it can be seen that the EMG activity sharply increases after a latency of about 30 msec from the line-up point. The figure shows that this latency may be measured from either the onset of pressure rise or from the tactile stimulus of the push. The possibility that other cues, such as visual or auditory correlates of the experimenter's preparation to push, contributed to the subject's time to respond is unlikely.

Having established that there is a rapid and consistent EMG response of the laryngeal adductors to chest pushes, we must now ask what is the nature or function of this response. Two alternative hypotheses are suggested: (1) that the response represents the correction of a closed-loop control system to a perturbing signal (for example, the increase of pressure, which might tend to abduct the folds), or (2) that the response represents the laryngeal protective-closure component of the startle reflex. We have collected no data that allows us to distinguish between these two possibilities. However, we performed a simple experiment to show that the second alternative is consistent with the data at hand.

The subject again produced steady sustained phonation. Startle responses were elicited both by clapping loudly near the subject's ear and by the use of a sudden, painful tactile stimulus on the forearm. Fundamental frequency was tracked with the aid of an electrical glottagraph. Average F_0 curves based on 13 claps and on 5 tactile stimuli are shown in Figure 5. A vertical line marks the onset of the startle-eliciting stimulus. It can be seen that there is a perturbation of fundamental frequency associated with the startle response. The latency of onset of this perturbation is about 50 msec. This combined latency could be accounted for by a 30 to 35 msec EMG latency and a 15 to 20 msec mechanical response latency for the muscle.

Discussion

The existence of a rapid and consistent EMG response to chest pushes appears to contradict the assumption by past investigators that laryngeal configuration can be considered constant for at least 100 msec after the push. Nevertheless, the relationship between fundamental frequency and subglottal pressure, derived by correlating those variables within a 30-msec interval after the push, agrees with the results of past investigations.

The trajectories in Figure 3 exhibit hysteresis in addition to the nonlinearity already noted. One component of the hysteresis is undoubtedly due to muscular forces, both from the adductor muscles whose EMG activity was recorded and from other muscles. The complex nature of some of these trajectories, in which the descending part crosses the ascending part, suggests a pattern of control in which an early F_0 raising response by one set of muscles elicits an F_0 lowering response from another set, which overshoots. However, EMG data must be obtained from a larger set of muscles before any such hypothesis can be tested. In addition, there may also be an inherent component for both the nonlinearity and the hysteresis, as suggested by the results for the falsetto condition, where no muscle response is apparent, and from the results of experiments with excised larynxes (van den Berg and Tan, 1959).

From a consideration of the latency alone, there are two plausible explanations for the EMG response. The results of Sears and Newsom Davis (1968) provide an example of another respiratory muscle with a response at comparable latency to pressure changes introduced at the mouth. These responses were interpreted in terms of peripheral feedback control. However, we have shown that the latency is also consistent with that for a startle response. Measurements from a larger set of muscles, and probably a more elaborate experimental paradigm, are necessary to elucidate the nature of the observed responses. Continuation of these studies is worthwhile for what it may reveal about reflex mechanisms available for the ongoing control of laryngeal function during speech.

REFERENCES

Atkinson, J. E. (1978) Correlation analysis of the physiological features controlling fundamental voice frequency. *Journal of the Acoustical Society of America 63*, 211-222.

Draper, M. H., P. Ladefoged and D. Whitteridge. (1960) Expiratory pressures and airflow during speech. *British Medical Journal 1*, 1837-1843.

Fromkin, V. and J. Ohala. (1968) Laryngeal control and a model of speech production. *Working Papers in Phonetics (UCLA) 10*, 98-110.

Hixon, T. J., D. H. Klatt and J. Mead. (1971) Influence of forced transglottal pressure on fundamental frequency. *Journal of the Acoustical Society of America 49*, 105.

Isshiki, N. (1959) Regulatory mechanism of the pitch and volume of voice. *Oto-Rhino-Laryngology Clinic* (Kyoto) *52*, 1065-1094.

Izdebski, K. and T. Shipp. (1976) Voluntary reaction times for phonatory initiation. *Journal of the Acoustical Society of America 60*, S66.

Ladefoged, P. (1963) Some physiological parameters in speech. *Language and Speech 6*, 109-119.

Landis, C. and W. A. Hunt. (1939) *The Startle Pattern*. (New York: Farrar and Rinehart).

Lieberman, P., R. Knudson and J. Mead. (1969) Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation. *Journal of the Acoustical Society of America 45*, 1537-1543.

Lukatela, G. (1973) Pitch determination by adaptive autocorrelation method. *Haskins Laboratories Status Report on Speech Research SR-33*, 185-194.

Netsell, R. and B. Daniels. (1974) Neural and mechanical response time for speech production. *Journal of Speech and Hearing Research 17*, 608-618.

Öhman, S. and J. Lindqvist. (1966) Analysis-by-synthesis of prosodic pitch contours. *STL-QPSR 1/1966* (Royal Institute of Technology, Stockholm, Sweden), 1-6.

Sawashima, M. (1974) Laryngeal research in experimental phonetics. In *Current Trends in Linguistics*, Vol. 12: *Linguistics and Adjacent Arts and Sciences*, ed. by T. A. Sebeok et al. (The Hague: Mouton), 2303-2348.

Sears, T. A. and J. Newsom Davis. (1968) The control of respiratory muscles during voluntary breathing. *Annals of the New York Academy of Science 155*, (Art. 1), 183-190.

van den Berg, Jw. (1957) Sub-glottal pressure and vibrations of the vocal folds. Folia Phoniatrica 9, 65-71.

van den Berg, Jw. and T. S. Tan. (1959) Results of experiments with excised larynxes. Practica Oto-Rhino-Laryngologica (Basel) 21, 425-450.

Dynamic Aspects of Velopharyngeal Closure[*]

Seiji Niimi[+], Fredericka Bell-Berti[++] and Katherine S. Harris[+++]

ABSTRACT

In a study of normal velopharyngeal closure mechanisms, motion picture films of the velum and lateral pharyngeal wall were taken through a flexible fiberoptic endoscope positioned to provide a view of both the velum and the lateral wall within the nasopharynx. Frame-by-frame measurements were made of velar elevation and medial lateral pharyngeal wall movement (at three levels along its vertical length) during speech articulation. The movement patterns of the points on the lateral wall were strikingly similar, differing primarily in the extent, rather than the time-course, of excursion. This movement of the lateral wall parallels that of the velum, supporting the hypothesis that both movements are caused by the action of a single muscle, the levator palatini, and not by the combined action of the levator palatini and superior constrictor muscles.

INTRODUCTION

A major question in studies of velopharyngeal closure in speech has been whether the closure is achieved by a sphincteric or a trapdoor mechanism, or by a combination of both. Observations of the velum and the lateral pharyngeal walls during speech, in the region of velopharyngeal closure, have revealed superior and posterior movement of the velum and medial movement of the lateral pharyngeal walls in the port-closing gesture. In some sense, then, the closing gesture is indeed sphincteric--that is, both the velum and the lateral pharyngeal walls move in closing the port. However, the questions of what the relationship between the movement patterns of the velum and

[*] A version of this paper was presented at the Third International Congress on Cleft Palate and Craniofacial Anomalies, Toronto, Canada, June 1977.

[+] Also University of Tokyo, Tokyo, Japan.

[++] NINCDS post-doctoral fellow, on leave from Montclair State College, Upper Montclair, New Jersey.

[+++] Also The Graduate School, The City University of New York.

Acknowledgment: This work was supported by NINCDS grants NS-13870, NS-13617, NINCDS post-doctoral fellowship NS-05332, NIDR grant DE-01774, and BRSG grant RR-05596, to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-55/56 (1978)]

lateral walls is and how they are accomplished still remain. More specifically, do the movements result from action of the levator palatini muscle acting alone, an essentially trapdoor mechanism, or from the levator palatini and superior constrictor muscles acting in concert, a more truly sphincteric mechanism?

Several investigators have previously addressed themselves to this question, using a variety of observation techniques, including a) monitoring of articulator position (including cine- and videofluorography, ultrasonic echo recording, and endoscopy), b) recording of electromyographic (EMG) potentials from the muscles of the velar region, and c) analysis of anatomical relationships. These studies have produced conflicting results, but we believe that they can be reconciled. Accordingly, we have begun a series of studies directed toward this goal, and this paper contains the results of our first step.

A limitation of earlier studies is that only one type of data was collected on each occasion--either articulator position and movement data (for example, Skolnick, McCall and Barnes, 1973; Shprintzen, Lencione, McCall and Skolnick, 1974; Zagzebski, 1975), or EMG data from muscles in the velar region (Lubker, 1968; Fritzell, 1969; Bell-Berti, 1973; 1976), or anatomical data (Dickson and Dickson, 1972). Furthermore, even when gathering data of one particular type the scope of the study has been further narrowed to a single source. For example, movements of the velum and lateral pharyngeal walls have frequently been studied independently, so that only one motion has been tracked, or only one time point has been studied (usually the instant of maximum excursion). Still a third limitation is that the phonetic inventory within which the closure mechanism has been studied has also been limited: several investigators have been content to investigate velar behavior in the production of isolated speech sounds, although the importance of phonetic context is well documented (Czermak, 1869; Subtelny, Koepp-Baker and Subtelny, 1961; Moll, 1962; Bzoch, 1968; Lubker, 1968; Fritzell, 1969; Bell-Berti and Hirose, 1975; Bell-Berti, 1973; 1976).

The dynamic articulatory data on velar and lateral pharyngeal wall displacement presented here are part of a larger data set consisting of several other types of observations, including data on electromyographic activity, intraoral air pressure and acoustic measurements. This data set incorporates a fairly large inventory in which phonetic context was systematically manipulated to allow minimal contrasts among groups of utterances. The phonetic inventory used includes two vowels, of markedly different tongue heights, and obstruent (both stop and fricative) as well as nasal consonants.

In a previously reported study, Bell-Berti and Hirose (1975) have shown that the pattern of levator palatini EMG activity is reflected in the velar elevation pattern, although the two are offset in time. In addition, Bell-Berti (1973; 1976) has shown, for three subjects, that the EMG activity patterns of the levator palatini and superior constrictor are not parallel. While the answer to the question of how closure is achieved may only be finally reached by studying the dynamics of port closure from simultaneous measurements of velar and lateral pharyngeal wall motion, and EMG recordings from the levator palatini and superior constrictor muscles, we believe that the directly observed movement data presented here, taken together with Bell-

Berti's (1973; 1976) EMG data from the same speaker provide strong evidence in favor of the conclusion that the levator palatini is solely responsible for velopharyngeal closure, at least for this particular speaker.

METHODS

A native speaker of American English (one of the authors of this paper) served as the subject for this study. An inventory of twenty-four disyllables was used in the experiment, containing nasal-oral and oral-nasal consonant oppositions in utterance medial position, thus using the most extreme contrasts that the system is required to make. The nasal consonant was always /m/, and the oral consonants were /p/,/b/,/f/,/v/,/s/ and /z/. The vowels were /i/ and /a/, with the same vowel occurring in both syllables. Each utterance type began with /f/ and ended with /p/; these phone sequences were designed to avoid lingual coarticulation effects, provide clear oscillographic records of the beginning of the first and the end of the second syllable, and insure initial and final oral articulation. The actual utterance types are listed in Table 1. All utterances were placed in lists in random order, and the lists were read from six to eight times during the recording session.

A thin plastic sheet with grid markings was inserted into one nostril of the subject, and placed onto the nasal floor (nasal surface of the velum) to enhance the contrast between the edge of the supra-velar surface and the posterior pharyngeal wall with a view to easing the task of subsequent frame-by-frame film analysis. A flexible fiberoptic endoscope (Olympus VF Type O) was also inserted into the subject's nostril, and positioned with its objective lens tip at the posterior border of the hard palate, providing a view of the vertical excursion of the velum as well as the side-to-side movement of the lateral pharyngeal wall. The subject was able to articulate under these conditions without any perceptually apparent interference.

A 16mm motion picture film of the nasopharynx was taken through the fiberscope at the rate of 60 frames per second (Figure 1). Synchronization pulses, which were generated frame-by-frame, were recorded on an FM data recorder with other acoustic and physiological data, including electromyographic (EMG) potentials from velopharyngeal muscles, and intraoral air pressure.

The distances between four monitoring points lying in the velopharyngeal region and a fixed reference point in the field of view were measured frame-by-frame, to expose the relationship between the displacement patterns of the velum and lateral pharyngeal wall above the level of the hard palate. These measurement points, shown in Figure 1, included one point on the velum and three points on the lateral nasopharyngeal wall. Velar displacement was measured at the highest visible point on the velum. The lateral wall movements were described at three levels of the lateral pharyngeal wall: at 75 percent, 50 percent and 25 percent of the maximum observed vertical excursion of the velum. The maximum excursion was determined by measuring velar height during blowing.

All four measurements could not be made for every frame. Figures 1a and 1b illustrate the extreme positions, at which certain points were unmeasurable. The left-hand picture was taken when the velum was low, that is, the

TABLE 1: Experimental utterances.

	fVCmVp	fVmCVp
	fapmap	fampap
	fabmap	fambap
V=/a/	fafmap	famfap
	favmap	famvap
	fasmap	famsap
	fazmap	famzap
	fipmip	fimpip
	fibmip	fimbip
V=/i/	fifmip	fimfip
	fivmip	fimvip
	fismip	fimsip
	fizmip	fimzip

TABLE 2: Correlation coeffiecients. Velar elevation vs lateral pharyngeal wall displacement.

fVCmVp		fVmCVp	
	r=		r=
fapmap	0.94	fampap	0.95
fabmap	0.93	fambap	0.97
fafmap	0.89	famfap	0.96
favmap	0.87	famvap	0.92
fasmap	0.89	famsap	0.98
fazmap	0.91	famzap	0.97
fipmip	0.94	fimpip	0.97
fibmip	0.96	fimbip	0.96
fifmip	0.93	fimfip	0.99
fivmip	0.96	fimvip	0.96
fismip	0.86	fimsip	0.95
fizmip	0.96	fimzip	0.96

velopharyngeal port was widely open. In this picture the three measurement points on the lateral pharyngeal wall are visible, but the high point of the velum is below the level of the hard palate and, hence, masked. The right-hand picture represents the tightly closed velopharyngeal port, in which the velum comes up almost to the 50 percent level of its maximum upward deflection. In this case, the lateral wall position is undefined at the 25 percent level, since at levels below the level of the high point of the velum the lateral wall is drawn into the velum.

Data Reduction Procedures

The displacement of each of the points against time was plotted for each repetition of each utterance type with the assistance of a small laboratory computer (Gay, 1977). Thus, the movement patterns of four different points for each trial (repetition of each of 24 utterance types) were obtained. The next step involved forming utterance groups composed of six to eight repetitions of each utterance type. Within each group, movement patterns were aligned with reference to an acoustic event--the boundary between the medial oral and nasal consonants, determined from the acoustic waveform (Figure 2)--and then averaged. Figures 3 and 4 display the component and ensemble-average displacement curves for the four measurement points for two different utterance types, which contrast in vowel, medial obstruent consonant, and order of the medial oral and nasal consonant sequence. These curves are followed by examples of the individual tokens.

It is apparent that the movement data for the individual repetitions and ensemble-averages at each measurement point for each utterance type have essentially the same pattern, with some variability among the tokens, resulting in part from differences in articulatory gestures, and in part from limitations in the resolution of the measuring system.

RESULTS

There is a striking similarity in the movement pattern of the three measurement points on the lateral pharyngeal wall. The differences between the medial displacement of the several levels are primarily differences in the extent, rather than the time-course, of the excursion.

This point is illustrated in Figures 3 and 4, which show four single tokens of the utterances /fimpip/ (3) and /fazmap/ (4). The small irregularities in the points for individual tokens are due to fine-grained measurement error. An inspection of the figures shows that medial movement is very small at the 75 percent point, which is nearest the "hinge" of the lateral wall, and that medial movement is substantially greater at the 50 percent and 25 percent points. The data obtained from the 25 percent point illustrate very well the indeterminate state that occurs when the velum is high. In forming the average curves, we have excluded the regions of the tokens that are undefined, leaving gaps in those curves. Since the movements of all three points are so similar, we have chosen the 50 percent point of the lateral wall movement to compare with velar movement. Several samples are shown in Figure 5.

In utterances having oral-nasal sequences (Figure 5a), both velar elevation and medial displacement of the lateral pharyngeal wall increase for the

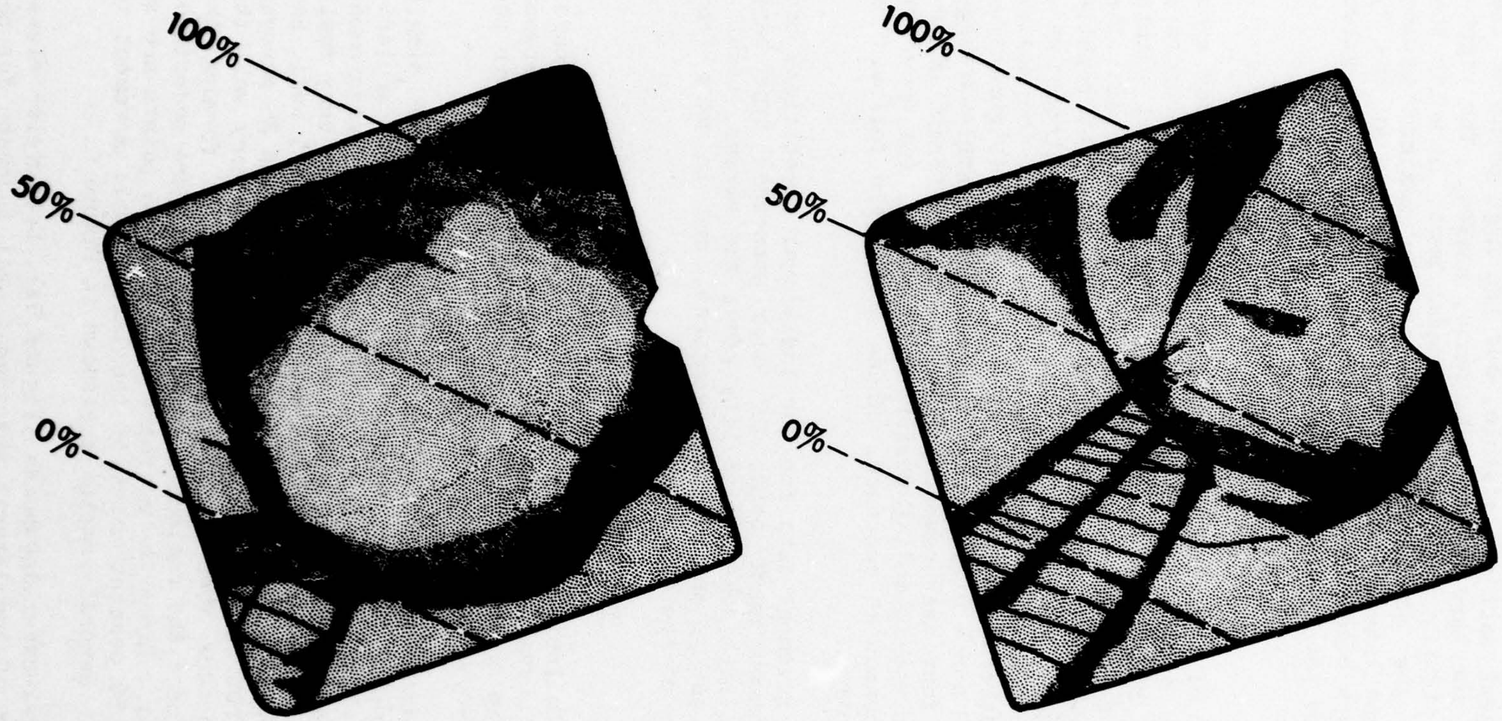


Figure 1a

Figure 1b

Figure 1: Frames reproduced from the experimental film, with measurement reference lines superimposed on each frame. In 1a the velum is low and out of view. In 1b the velum is elevated to nearly 50% of its maximum excursion.

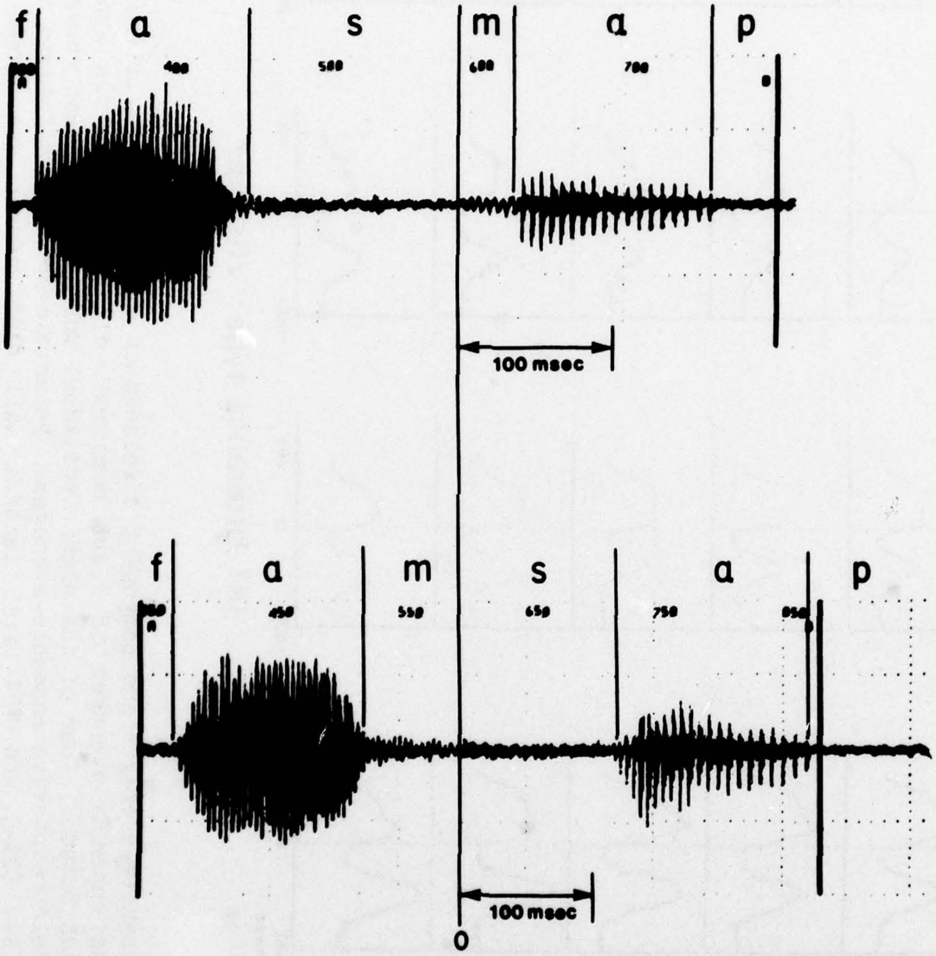
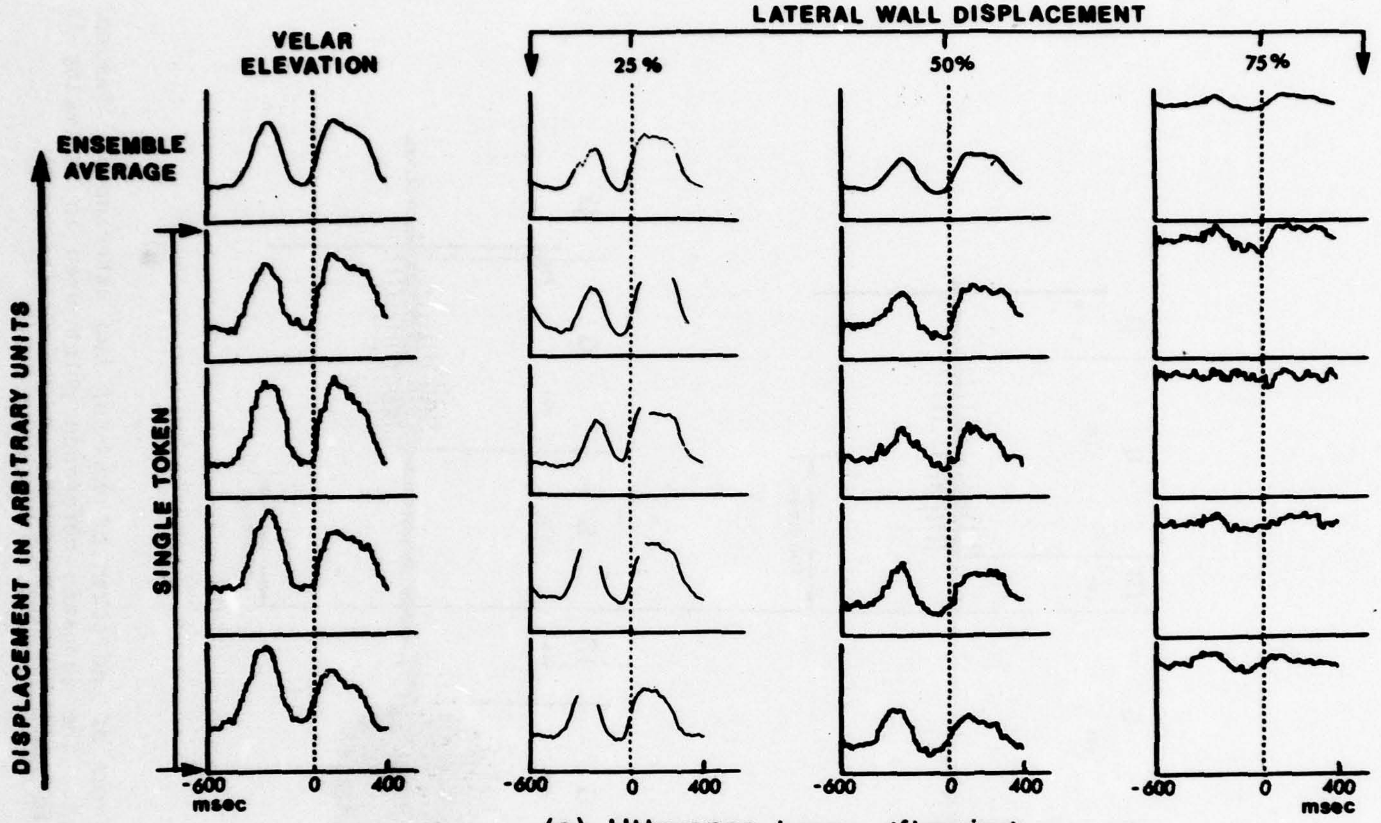
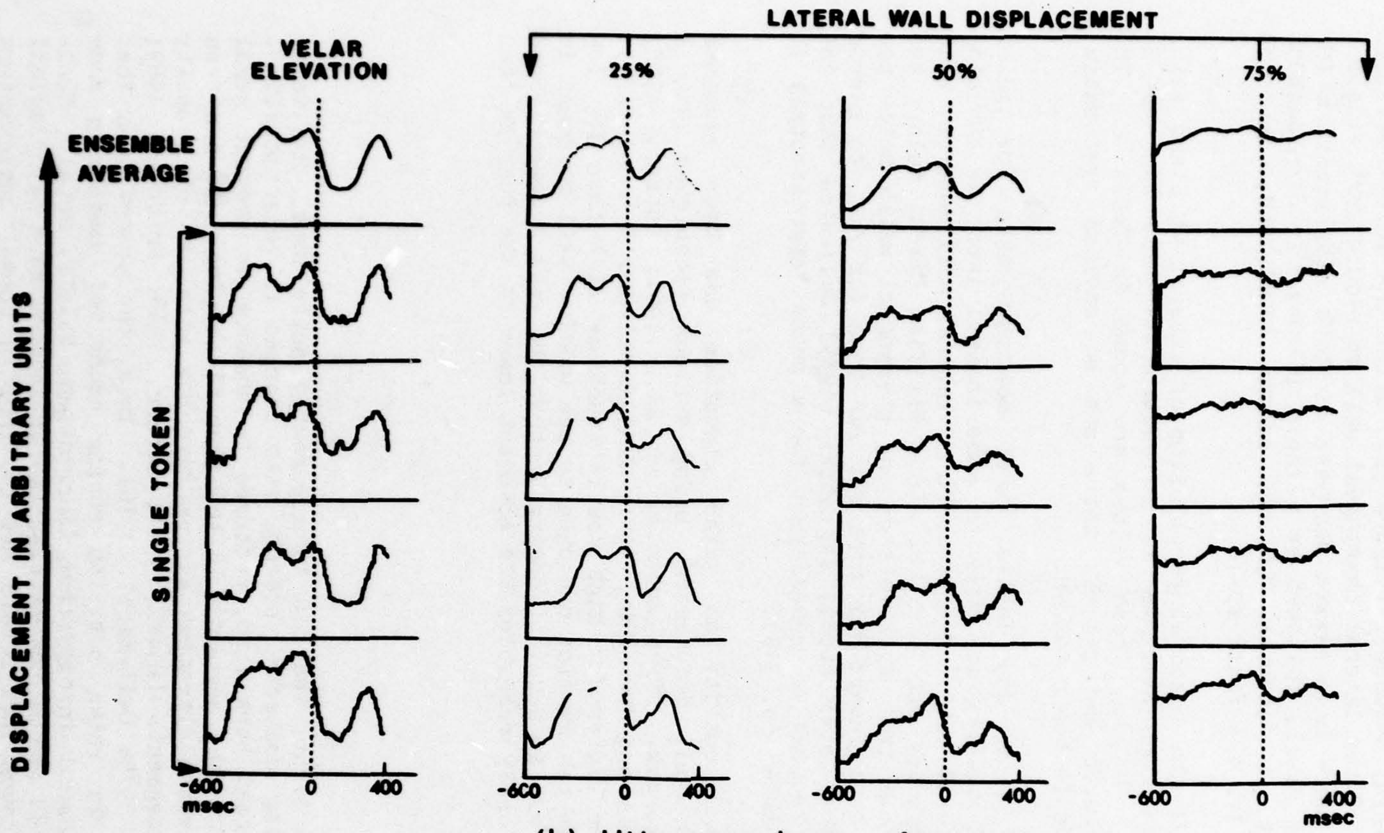


Figure 2: Audio waveforms of one token of each of two utterances, /fasmap/ and /famsap/. The acoustic reference point used in averaging is indicated, for each token, by an arrow.



(a) Utterance type: /fimpip/

Figure 3: Ensemble-averages and component displacement curves for the utterance type, /fimpip/. The ensemble-averages of eight component displacement curves appear at the head of each column. Four of the eight individual token displacement curves are shown beneath their respective ensemble-averages. Velar elevation is displayed in the first column; 25%-, 50%-, and 75%-level lateral wall displacement are shown in the next three columns, respectively.



(b) Utterance type: /fazmap/

Figure 4: Ensemble-averages and component displacement curves for the utterance type, /fazmap/. The ensemble-averages of eight component displacement curves appear at the head of each column. Four of the eight individual token displacement curves are shown beneath their respective ensemble-averages. Velar elevation is displayed in the first column; 25%-, 50%-, and 75%-level lateral wall displacement are shown in the next three columns, respectively.

initial fricative, /f/ (thus decreasing velopharyngeal port size), decrease for the first vowel, increase for the medial oral consonant, decrease markedly for the nasal consonant (to open the port and enhance nasal coupling), and then increase again for the final vowel and stop consonant, /p/. In utterances having nasal-oral sequences (Figure 5b), after the velar elevation and medial displacement of the lateral pharyngeal wall to reduce port size for the initial fricative, there is a marked decrease in both displacements for the first vowel and the nasal, followed by a rapid increase in displacement for the oral consonant, a decrease for the second vowel and, in some utterances, an increase for the final /p/.[1]

In order to quantify the temporal parallelism of lateral wall and velar movement, we have correlated averages as a function of time for the 24 utterances (Kewley-Port, 1973). These values are shown in Table 2. The values range between $\underline{r}$=0.86 and $\underline{r}$=0.99; there are no obvious systematic, phonetically-based trends in the $\underline{r}$ values.

The comparison of velar and lateral wall movement data for pairs differing only in vowel quality may give us some insight into the muscular forces acting on the lateral wall. Measurements of lateral wall movement with ultrasound (Kelsey, Woodhouse and Minifie, 1969; Minifie, Hixon, Kelsey and Woodhouse, 1970) have shown that at levels on the pharyngeal walls below the hard palate, the walls move inward far more for /a/ than for /i/, to narrow the faucal isthmus. In a later study, Zagzebski (1975) suggested that the vowel differences were reduced or nonexistent for a probe "approximately at the level of the hard palate" (p. 315).

The effect of vowel quality on velar elevation has been examined experimentally and been well documented, using various techniques over a hundred-year period (Czermak, 1869; Subtelny et al., 1961; Lubker, 1968; Fritzell, 1969; Sawashima and Ushijima, 1971; Bell-Berti and Hirose, 1975). The velum is found to be higher for high vowels (such as /i/) than for low vowels (such as /a/). We can examine our data to see whether wall movement is greater for /i/ or for /a/, as a way of seeing whether the wall movement above the hard palate more closely resembles the movement lower in the pharynx than the velar elevation.

[1] We might note, in this light, that it is apparently quite usual for velar elevation to vary during connected speech, with changes in velar position, and thus in velopharyngeal port size, produced to enhance or prevent nasal coupling, as needed, for the segments in the phonetic string. It has been shown previously that velar elevation varies directly with the oral cavity constriction of oral segments (Bjork, 1961; Lubker, 1968; Fritzell, 1969; Bell-Berti and Hirose, 1975; Bell-Berti, 1976). Thus, the observation that there are variations in velar position during connected speech in some speakers with velopharyngeal incompetence (Shprintzen, Rakoff, McCall, Skolnick and Zimmerman, 1977) should not necessarily be labeled as a pathological articulatory pattern: such individuals may be using a normal articulatory strategy--but one that fails due to anatomical defect.

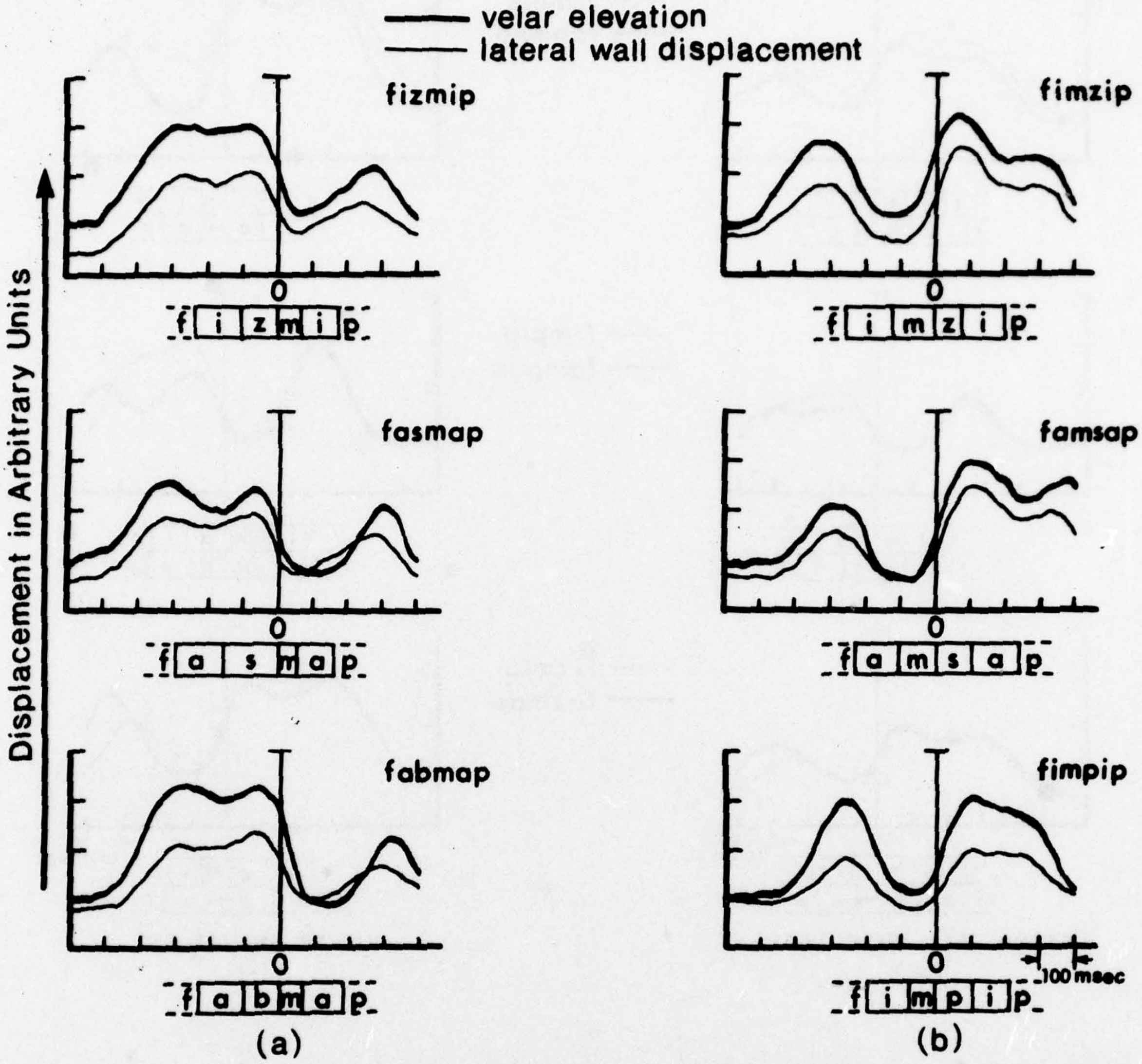


Figure 5: Ensemble-average time-course of velar and lateral wall displacement (50% level) for six utterance types. Velar elevation is represented by the heavy line; lateral wall displacement by the thin line. Zero on the abscissa represents the acoustic reference point for averaging. Displacement is in arbitrary units, with velar elevation and mesial lateral wall displacement increasing in the direction of the arrow at the left of the figure. Average audio segment durations are displayed beneath each graph.

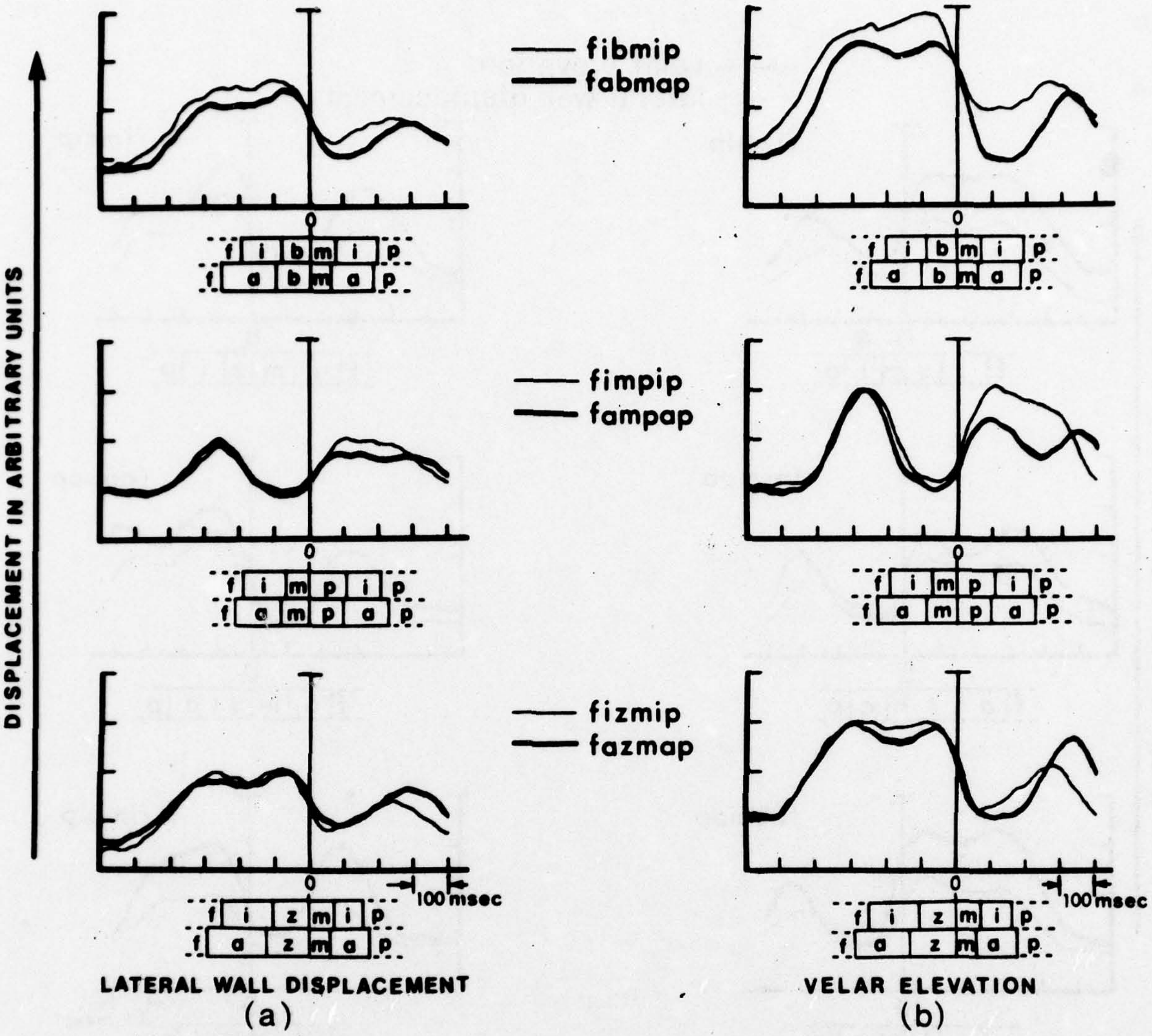


Figure 6: Lateral wall displacement (at the left) and velar elevation (at the right) for three minimal pairs contrasting in vowel quality. Utterances with /i/ are represented by the thin line; utterances with /a/ are represented by the heavy line. Small arrows indicate the point at which the displacement comparisons (of vowels associated with the medial oral consonant) were made (Table 3). Zero on the abscissa represents the acoustic reference point for averaging. Displacement is in arbitrary units, with velar elevation and mesial lateral wall displacement increasing in the direction of the arrow at the left of the figure. Average audio segment durations are displayed beneath each graph.

Results for three utterance types are shown in Figure 6. In general, there is a small tendency for excursions to be greater both in velar elevation and in medial movement for/i/.

In order to make a rough quantitative estimate of the magnitude of the effect for all minimial pairs, we compared the curves at the point of the vowel associated with the medial oral consonant. The comparisons were tabulated as "+" or "-" in Table 3.

TABLE 3: Sign Test. Comparison of velar and lateral pharyngeal wall positions for utterances minimally contrasting in vowel quality. Inspection was made at the time point corresponding to articulation of the vowel associated with the medial oral consonant. "+" indicates greater displacement for the utterance containing /i/; "-" indicates greater displacement for the utterance containing /a/.

	Velum	Lateral Wall		Velum	Lateral Wall
fipmip-fapmap	+	+	fimpip-fampap	+	+
fibmip-fabmap	+	+	fimbip-fambap	+	+
fifmip-fafmap	+	-	fimfip-famfap	+	+
fivmip-favmap	+	+	fimvip-famvap	+	+
fismip-fasmap	+	-*	fimsip-famsap	+	+
fizmip-fazmap	+	+	fimzip-famzap	+	+

*Maximum velar height exceeded the 50% level for some tokens, making impossible measurement of the lateral pharyngeal wall at that point. The average value for the /i/ utterances in these cases is depressed because of the removal from the average of those tokens having the greatest velar elevation.

Viewed in this manner, the results clearly show that, as expected, in all 12 cases velar elevation was higher for /i/ than for /a/; lateral wall movement was greater for /i/ than for /a/ in 10 out of 12 cases. Hence, the data support the hypothesis that both velar elevation (P < .0005) and lateral wall movement (P < .011) are greater for /i/ than for /a/.

DISCUSSION

There is general agreement that the velum is elevated and retracted primarily by the levator palatini muscle (Lubker, 1968; Fritzell, 1969; Lubker, Fritzell and Lindqvist, 1970; Bell-Berti, 1976). The point of controversy revolves around the putative role of other muscles in the velopharyngeal port region in bringing about movement of the lateral pharyngeal walls at various levels relative to the point of velopharyngeal closure. Essentially two points of view have been advanced.

The first of these views is that of Skolnick and his colleagues (Skolnick, 1970; Skolnick, McCall and Barnes, 1973; Shprintzen, McCall, Skolnick and Lencione, 1975; and Skolnick, Zagzebski and Watkin, 1975). On the basis of measurements of frontal and lateral videofluoroscopic films, Skolnick has suggested that the observed lateral pharyngeal wall movement toward the midsaggital line observed during speech cannot be due to the action of the levator palatini, because maximal medial excursion occurs below the level of the levator eminence on the nasal floor.

The second point of view is that of Dickson (Dickson, 1972; 1975; Dickson and Dickson, 1972), whose recent anatomical studies of the velopharyngeal region have shown that the superior margin of the superior constrictor muscle is at the level of the hard palate. Thus, while the superior constrictor might act in velopharyngeal closure, it can do so at, but no higher than, the level of the hard palate. The muscle in the lateral nasopharyngeal walls that might effect the medial displacement of the walls above the hard palate is the levator palatini, which runs infero-antero-medially from its origin on the Eustachian tube to its insertion in the velum (Figure 7). Dickson (1972) has proposed that maximum medial movement of the lateral walls, during speech, is at the level of the torus tubarius, and results from contraction of the levator palatini, which lies laterally to the torus. The Dickson argument excludes the superior constrictor from participation in displacing the lateral pharyngeal walls at the level of closure on two grounds: first, that the superior constrictor does not normally extend above the level of the hard palate, as we mentioned above; and, second, that there is no anteriorly directed movement of the posterior pharyngeal wall at any level during velopharyngeal closure for speech--a movement which would be expected since the constrictor attachment to the medial pharyngeal raphe is a very loose one.

A third possibility does exist, however, although it has not been widely suggested. It is that velopharyngeal closure is accomplished by both sphincteric and trapdoor mechanisms, with the contribution of each differing among speakers.

Electromyographic studies summarized above have shown that velar elevation and retraction are accomplished by contraction of the levator palatini muscle (Lubker, 1968; Fritzell, 1969; Lubker et al., 1970; Bell-Berti, 1973; 1976). However, the contradictory reports of Fritzell (1969) and Bell-Berti (1973; 1976) on the role of the superior constrictor in velopharyngeal closure serve to illustrate the controversy over the nature of the closure mechanism: sphincteric, trapdoor, or, perhaps, a combination of the two.

Electromyographic (EMG) data reported elsewhere (Bell-Berti, 1973; 1976) on the activity patterns of the levator palatini, palatopharyngeus, and superior constrictor indicated that both the levator palatini and the palatopharyngeus are more active for consonant than for vowel articulations, although palatopharyngeus activity is highly influenced by vowel quality, so that the palatopharyngeus is most active in low-vowel environments, helping to narrow the faucal isthmus for such vowels (Bell-Berti, 1973; 1976). That finding amplifies Fritzell's earlier (1969) report that palatopharyngeus activity is most evident in the articulation of /a/. Fritzell (1969) and Bell-Berti (1973; 1976) have reported conflicting data from the superior constrictor, with the former reporting consonant-dependent activity and the latter reporting vowel-dependent activity, with the greatest EMG potentials recorded for the articulation of /a/. Bell-Berti (1973; 1976) also reported that the EMG potentials from the middle constrictor, recorded at the level of the tip of the epiglottis, mirrored those from the superior constrictor, recorded at the estimated level of velopharyngeal closure.

We believe that the levator palatini is the muscle primarily responsible for the medial movement of the lateral pharyngeal wall from the level of velopharyngeal closure (which varies with the type of phonetic segment produced) to the superior limit of that movement. That the interpretation that the levator palatini is responsible for both the lateral wall and velar movements is a valid one is supported by the data presented here, as well as by the data of other investigators, including Harrington (1944), Skolnick (1969), Zagzebski (1975), and Honjo, Harada and Kumazawa (1976), who have also described lateral pharyngeal wall movement, from the level of velopharyngeal closure upward, as paralleling velar movement. We have shown here that the pattern of lateral nasopharyngeal wall movement is the same through its vertical extent, having its lower limit at the level of velopharyngeal closure. The only differences we found in the excursion, at the levels measured, were in the extent of excursion, decreasing at more superior levels. Furthermore, we found that the time courses of lateral wall and velar movement are parallel and that the effects of phonetic environment are the same on lateral wall movement and velar elevation. Since Bell-Berti's (1973; 1976) EMG data for this subject show different effects of phonetic environment on the levator palatini and superior constrictor, and the levator palatini data parallel the velar and lateral wall movement data, we believe that the weight of evidence supports the conclusion that this subject's mechanism is in nature essentially that of a trapdoor.

Additional evidence in support of this conclusion is that the superior constrictor _does not_ extend above the level of the hard palate and that the levator palatini runs obliquely in the lateral wall to its insertion in the velum (Dickson and Dickson, 1972), where its contraction might be expected both to elevate and retract the velum and to draw the lateral walls medially and superiorly. It seems, then, at least as likely that the observed movements are the result of _one_ muscle's contraction as it does that two muscles are responsible for these parallel movements. For the two-muscle description to be correct, one of the muscles, the superior constrictor, would have to function in a radically different manner in its superior (consonant-related action) and inferior (vowel-related action) portions.

Thus, our data appear to support the hypothesis of Dickson and Dickson (1972) and Honjo et al. (1976): that the lateral pharyngeal wall movement component of velopharyngeal closure is caused by contraction of the levator palatini (Figure 7a). These two reports suggest that the greatest medial excursion of the lateral walls occurs at the level of the torus tubarius; but the vertical position on the lateral wall referred to as the torus tubarius

Figure 7a

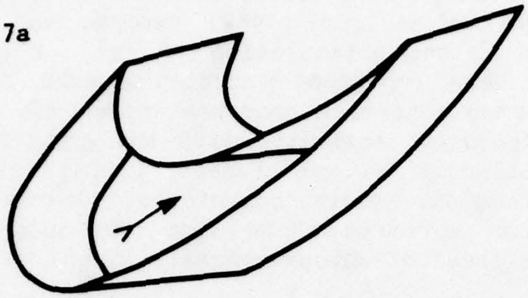


Figure 7: (a) Schematic representation of the unitary action of the velum and lateral nasopharyngeal walls.
(b) Drawing showing the anatomical relationships among the levator palatini, superior constrictor, and hard palate.

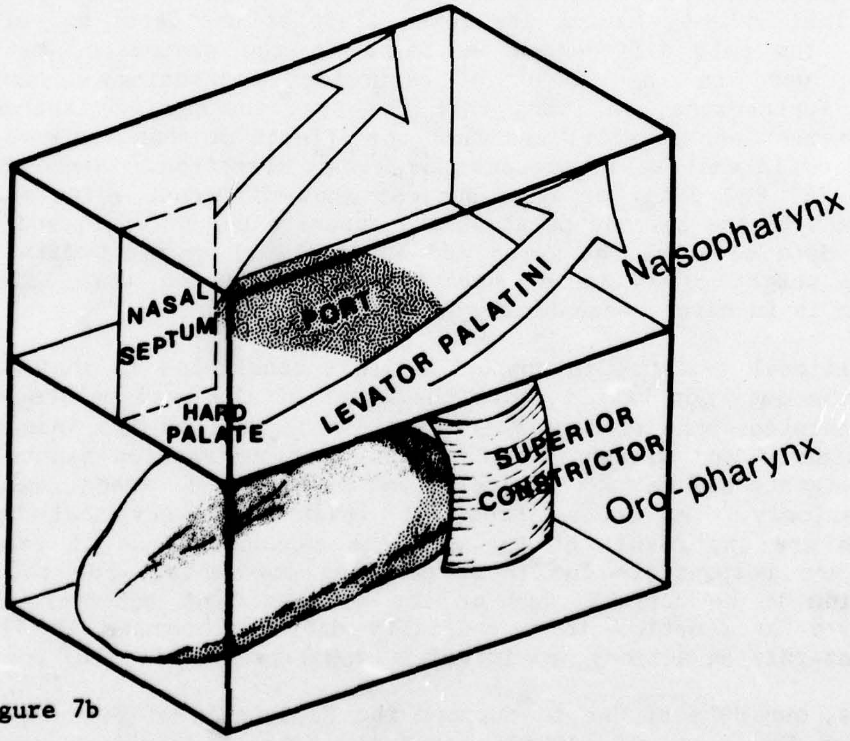


Figure 7b

is not clearly specified in either paper. Our data reveal that maximum medial excursion occurs nearest the velum.

The data of Skolnick et al. (1973) and of Shprintzen et al. (1974), that show the presence of a bulge in the lateral pharyngeal wall above the hard palate and below the levator eminence, might be an artifact of the videofluoroscopic focal plane: a vertical slice through an oblique ridge would look very much like a localized bulge (Figure 7b). This possibility might be explored using multiplane tomography, a technique in which x-rays are taken simultaneously in multiple planes.

Finally, we must also expect some individual differences among normal speakers, and thus it may be that some speakers use the superior constrictor in addition to the levator ʀalatini when they constrict the velopharyngeal port; in such speakers, the vertical dimension of the port should increase directly with velar elevation, since the lower level of closure will not shift its vertical position.

REFERENCES

Bell-Berti, F. (1973) The Velopharyngeal Mechanism: An Electromyographic Study. Unpublished Ph. D. thesis, The Graduate School, The City University of New York.

Bell-Berti, F. (1976) An electromyographic study of velopharyngeal function in speech. Journal of Speech and Hearing Research 19, 225-240.

Bell-Berti, F. and H. Hirose. (1975) Palatal activity in voicing distinctions: A simultaneous fiberoptic and electromyographic study. Journal of Phonetics 3, 69-74.

Bjork, L. (1961) Velopharyngeal function in connected speech. Acta Radiologica, Suppl. 202.

Bzoch, K. R. (1968) Variations in velopharyngeal valving: The factor of vowel changes. Cleft Palate Journal 5, 211-218.

Czermak, J. N. (1869) Wesen und Bildung der Stimm- und Sprachlaute. Czermak's gesammelte Schriften, vol. 2. (Leipzig: Vilhelm Engelman).

Dickson, D. R. (1972) Normal and cleft palate anatomy. Cleft Palate Journal 9, 280-293.

Dickson, D. R. (1975) Anatomy of the normal velopharyngeal mechanism. Clinical Plastic Surgery 2, 235-248.

Dickson, D. R. and W. M. Dickson (1972) Velopharyngeal anatomy. Journal of Speech and Hearing Research 15, 372-381.

Dixit, P. and P. F. MacNeilage. (1972) Coarticulation of nasality: Evidence from Hindi. Journal of the Acoustical Society of America 52, 131(Abs.).

Fritzell, B. (1969) The velopharyngeal muscles in speech. Acta Otolaryngologica, Suppl. 250.

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

Harrington, R. (1944) A study of the mechanics of velopharyngeal closure. Journal of Speech and Hearing Disorders 9, 325-345.

Honjo, I., H. Harada and T. Kumazawa. (1976) Role of the levator veli palatini muscle in movement of the lateral pharyngeal wall. Archives of Oto-Rhino-Laryngology 212, 93-98.

Kelsey, C., R. Woodhouse and F. D. Minifie. (1969) Ultrasonic observation of coarticulation in the pharynx. Journal of the Acoustical Society of

America 46, 1016-1018.

Kewley-Port, D. (1973) Computer processing of EMG signals at Haskins Laboratories. *Haskins Laboratories Status Report on Speech Research SR-33*, 173-183.

Lubker, J. F. (1968) An electromyographic-cineradiographic investigation of velar function during normal speech production. *Cleft Palate Journal 5*, 1-8.

Lubker, J. F., B. Fritzell and J. Lindvist. (1970) Velopharyngeal function: An electromygraphic study. *Quarterly Progress and Status Report*, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden. *STL-QPSR*, *4/1970*, 9-20.

Minifie, F. D., T. J. Hixon, C. Kelsey and R. Woodhouse. (1970) Lateral pharyngeal wall movement during speech production. *Journal of Speech and Hearing Research 13*, 584-594.

Moll, K. L. (1962) Velopharyngeal closure on vowels. *Journal of Speech and Hearing Research 5*, 30-77.

Passavant, G. (1863) *Ueber die Verschliessung des Schlundes beim Sprechen*. (Frankfurt a. M.: J. D. Sauerlander).

Sawashima, M. and T. Ushijima. (1971) Use of fiberoptics in speech research. Research Institute of Logopedics and Phoniatrics, University of Tokyo, *Annual Bulletin*, *5*, 25-35.

Shprintzen, R. J., R. M. Lencione, G. N. McCall and M. L. Skolnick. (1974) A three dimensional cinefluoroscopic analysis of velopharyngeal closure during speech and nonspeech activities in normals. *Cleft Palate Journal 11*, 412-428.

Shprintzen, R. J., G. N. McCall, M. L. Skolnick and R. M. Lencione. (1975) Selective movement of the lateral aspects of the pharyngeal walls during velopharyngeal closure for speech, blowing, and whistling in normals. *Cleft Palate Journal 12*, 51-58.

Shprintzen, R. J., S. J. Rakoff, G. N. McCall, M. L. Skolnick and K. Zimmerman. (1977) The pulsing palate and pharynx: a newly recognized phenomenon of velopharyngeal valving. *Cleft Palate Journal 14*, 350(A).

Skolnick, M. L. (1969) Video velopharyngography in patients with nasal speech, with emphasis on lateral pharyngeal wall motion in velopharyngeal closure. *Radiology 93*, 747-755.

Skolnick, M. L. (1970) Videofluoroscopic examination of the velopharyngeal portal during phonation in lateral and base projections--A new technique for studying the mechanics of closure. *Cleft Palate Journal 7*, 803-816.

Skolnick, M. L., G. N. McCall and M. Barnes. (1973) The sphincteric mechanism of velopharyngeal closure. *Cleft Palate Journal 10*, 286-305.

Skolnick, M. L., J. A. Zagzebski and K. L. Watkin. (1975) Two dimensional ultrasonic demonstration of lateral pharyngeal wall movement in real time --A preliminary report. *Cleft Palate Journal 12*, 299-303.

Subtelny, J. D., H. Koepp-Baker and J. D. Subtelny. (1961) Palate function and cleft palate speech. *Journal of Speech and Hearing Disorders 26*, 213-224.

Zagzebski, J. A. (1975) Ultrasonic measurement of lateral pharyngeal wall motion at two levels in the vocal tract. *Journal of Speech and Hearing Research 18*, 308-318.

Effect of Speaking Rate on the Relative Duration of Stop Closure and Fricative Noise[*]

David Isenberg

ABSTRACT

Words distinguished by a fricative-affricate contrast ("dish" and "ditch") were produced in sentence contexts by native English speakers at five successively increasing speaking rates, from their slowest natural rate to their fastest rate possible without misarticulation. Durations of closure intervals in "ditch" and fricative noises in "dish" and "ditch" were measured from oscillograms. Linear functions relating the duration of these acoustically defined intervals to total sentence duration were fitted to these data. Their slopes were smaller for closure intervals than for fricative noises, indicating that closure intervals are more stable in duration than fricative noises as speaking rate changes. These results offer a rationale for an apparently paradoxical finding in a recent perceptual study of the fricative-affricate distinction (Repp, Liberman, Eccardt and Pesetsky, 1978).

INTRODUCTION

The duration of silent intervals functions as a powerful speech cue in many contexts. Whole phonetic segments may or may not be perceived depending on the amount of silence present in the signal. Silence may cue a stop consonant for which there are no spectral transition cues when it is inserted between a fricative noise and a vocalic region--"slit" can become "split" (Bastian, Eimas and Liberman, 1961), "gray ship" can become "great ship" (Repp, Liberman, Eccardt and Pesetsky, 1978), /si/ can become /ski/ and /su/ can become /spu/ (Bailey and Summerfield, 1978) when silence alone is appropriately inserted into the signal. Conversely, the disyllable /ebde/, for example, goes to /ede/ as the intervocalic silence is shortened to less than about 75 msec, even though all spectral information is preserved (Repp[1]).

The manner feature [continuant] may also be cued by the duration of silence between a vocalic interval and a fricative noise, yielding a fricative

[*] This paper is a revised and expanded version of a talk presented at the 95th Meeting of the Acoustical Society of America, May 1978.

Acknowledgment: The author thanks A. M. Liberman and B. H. Repp for much helpful advice at all stages of work. This research was supported by N.I.H. fellowship NS-05493 and N.I.H. grant HD-01994.

[1] Repp, B. H. (1978) Influence of spectral and temporal properties of vocalic environment on silence as a cue distinguishing single intervocalic stop consonants from stop clusters and geminates. Unpublished manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-55/56 (1978)]

percept when the silence is short and an affricate when the silence is long. Thus, "say shop" becomes "say chop" and "great ship" becomes "great chip" when sufficiently more silence is inserted (Repp et al., 1978). Also, "dish" becomes "ditch" when silence is inserted between the vocalic region and the fricative noise and, conversely, "ditch" becomes "dish" when the naturally produced silence is made sufficiently short or eliminated (Dorman and Raphael, 1977). The work cited above with "slit" and "split" (Bastian et al., 1961) and "si" and "ski" (Bailey and Summerfield, 1978) may also be taken as instances where the duration of a silence may determine the value of the feature [continuant] between a fricative noise and a vocalic interval.

The duration of the fricative noise is another powerful temporal cue known to affect the fricative-affricate distinction. When the fricative noise is long, the tendency is to hear the fricative, and when it is short, the affricate is perceived (Gerstman[2]). Fricative noise duration has been shown to interact perceptually with silence duration in a straightforward and intuitively plausible way--when the fricative noise is lengthened, biasing the percept towards fricative, more silence is needed to hear the stimulus as an affricate, and vice versa (Repp et al., 1978).

Rate of speaking is another variable, cued primarily by temporal parameters, which interacts with the other two temporal cues for the fricative-affricate distinction in a striking and paradoxical fashion. When the same fricative noise is embedded in the carrier sentence "Why don't we say (sh/ch)op again," produced at two different rates of speaking, more silence is needed between "say" and the fricative noise to hear the affricate when the carrier sentence is spoken at the faster rate [Repp et al., (1978); Dorman, Raphael and Liberman, (1976)]. A naive hypothesis (for example, Joos, 1948) would predict the opposite, more intuitively plausible result--that the duration of silence at the shop/chop boundary would decrease in direct proportion to the decrease in the duration of the sentence.

Only a few studies of speech production have examined the effects of changes in rate of speaking. Those that have addressed the problem have shown that rate of speaking does not equally affect the timing of all articulatory gestures (Gay, Ushijima, Hirose and Cooper, 1974) or acoustically defined regions of the speech signal (Kozhevnikov and Chistovich, 1965; Port[3]; Gay, 1978). The data from these studies indicate that consonantal gestures tend to be relatively more stable in duration than gestures associated with vowels as rate of speaking increases.

Assuming that fricative noises behave like vocalic regions regarding duration change across rate of speaking, Repp et al. (1978) reasoned that a listener would tacitly expect noise duration to decrease greatly as speaking rate increased. Hearing no decrease in noise duration in the presence of an increased speaking rate would tend to bias the listener towards hearing a

[2] Gerstman, L. (1957) Cues for distinguishing among fricatives, affricates and stop consonants. Unpublished Ph.D. dissertation, New York University.

[3] Port, R. F. (1977) The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words. Unpublished Ph.D. dissertation, University of Connecticut.

fricative, since the listener would now implicitly assign a relatively longer duration to the fricative noise. Thus correspondingly more silence would be needed to counteract this bias and hear an affricate. The present experiment was designed to determine whether durations of closure intervals would be more stable across different speaking rates than durations of fricative noises. In addition, the duration of a neighboring vocalic interval was also examined as a function of rate of speaking.

METHOD

The present study employed the minimal contrast pair "ditch"/"dish" to examine how speaking rate affects production timing of the closure interval in relation to neighboring acoustically defined intervals. Eight native speakers of American English, four men and four women employed at Haskins Laboratories, served as talkers in this study. None reported any known speaking or hearing impairment nor any neurological problems.

The sentence produced in this study was "I meant to say talk ditch (or dish) fast." Each talker produced six groups of five sentences. Talkers were instructed to produce each group of five so that the first sentence would be produced at his or her slowest natural rate (without hesitations or excessively prolonged continuants) and subsequent sentences would be produced progressively faster, so that the fifth sentence would be produced at the fastest rate possible without misarticulation. Three groups of sentences were produced with "ditch" and three were produced with "dish." Emphatic stress was varied in each of these three groups, with the first group of five articulated with emphasis on "talk," the second group with emphasis on "dish" or "ditch," and the third group with emphasis on "fast."

The sentences were recorded at 7.5 ips on a Crown tape recorder (model SX 822) and digitized at 10 kHz on the Haskins Laboratories PDP-11/45 pulse code modulation (PCM) system. Duration measurements were made using the Haskins Wave Editing and Display (WENDY) software package (Szubowicz[4]).

The durations of several acoustically defined intervals in these sentences were measured. The measure of speaking rate was the duration of the entire sentence measured from the first pitch pulse in "I" to the release burst of the /t/ in "fast." Figure 1 depicts oscillograms of "dish" and "ditch" to illustrate some of the other acoustically defined regions that were measured. The D-burst + vowel interval was defined as the interval from the onset of the D-burst to the last glottal pulse. The closure interval was measured for productions of ditch only.[5] This interval was measured from the end of glottal pulsing to the point where the amplitude of the fricative noise began to increase continuously. The fricative noise was measured beginning at

[4] Szubowicz, L. S. (1977) A tutorial guide to WENDY - the Haskins Wave EditiNg and DisplaY system. Unpublished manuscript.

[5] The amplitude envelope of "dish" usually dipped considerably between the vocalic interval and the fricative noise, but this interval of low energy was never maintained for more than 5 msec and never resembled a closure interval. Furthermore, only one utterance of "ditch" had no closure interval. This utterance was not included in any further analysis.

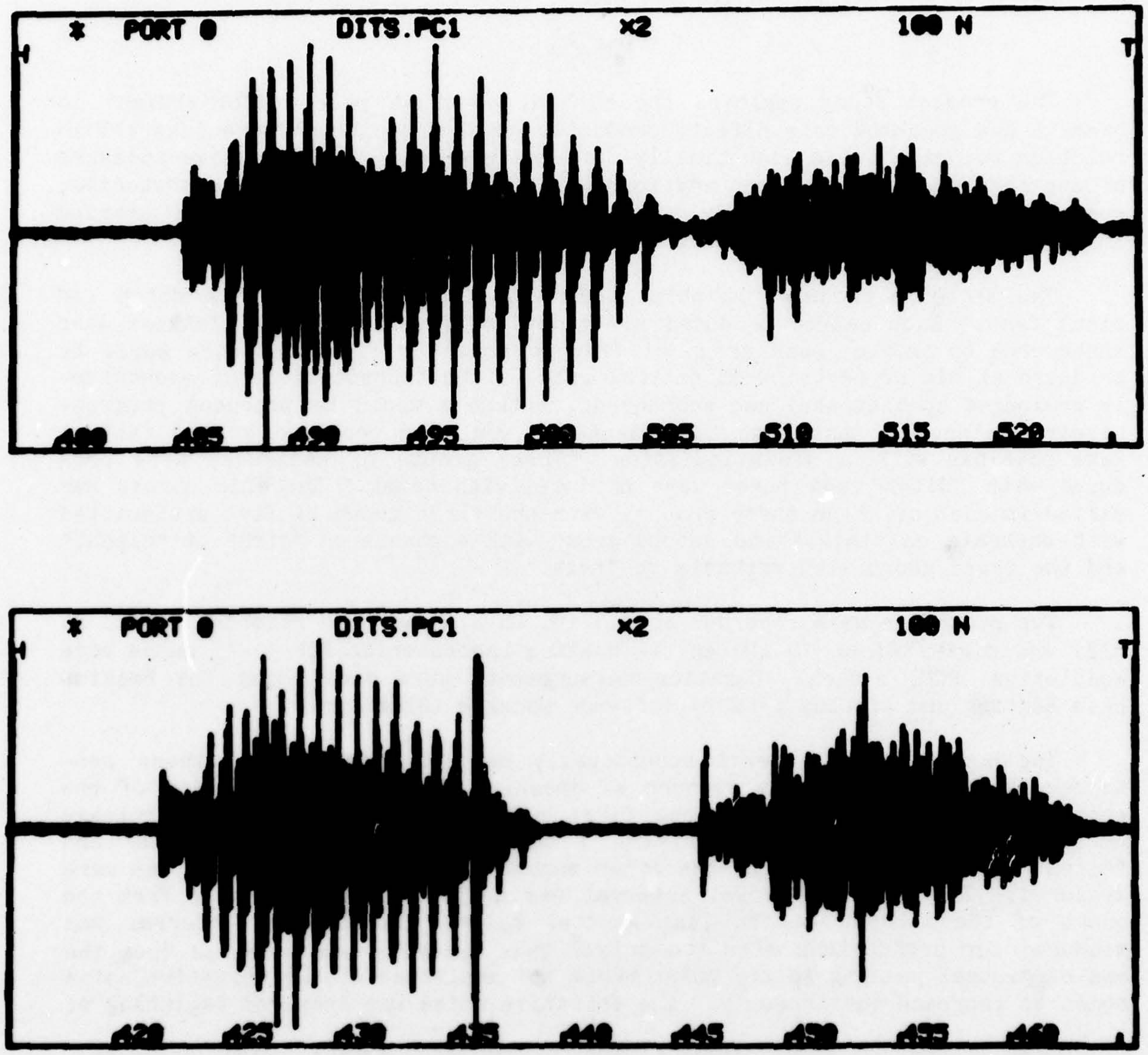


Figure 1: Oscillograms of "dish" (top panel) and "ditch" (bottom panel), spoken in isolation, to illustrate the acoustically defined intervals measured in this study.

the end of glottal pulsing in the case of dish, or the end of the silent interval for ditch, to the point where the amplitude of the noise became minimum, which corresponded to the onset of /f/ in fast.

RESULTS

Sentence Durations.

Analysis of the sentence durations indicate that the talkers in this study were able to follow the instructions about rate of speaking. That is, they were able to systematically increase their rate of speaking in five successive increments. An analysis of variance performed on the total sentence durations, in which word ("dish" or "ditch"), stress (emphasis on "talk," "dish" or "ditch," or "fast") and rate (1 - 5) were within-subjects factors, showed only a main effect of rate [F(4,28) = 208.02, p<.01]. The mean sentence durations at each rate, shown in Table 1, fell in a negatively accelerated fashion from rate 1 to rate 5. The absence of main effects for word or stress (F<1.0 in both cases) or of any higher order interactions indicates that any systematic effects of these manipulations on total sentence duration are minimal.

TABLE 1: Sentence duration (msec) for repetitions 1-5.

Repetition	1	2	3	4	5
Sentence Duration	2265.7	1963.0	1754.9	1606.8	1506.9

Intervals Compared Within Words.

"Dish" and "ditch" contain unequal numbers of acoustically defined intervals. Therefore, two separate analyses of variance were performed on the relative durations of these intervals, expressed as proportions of the sentence in which they occurred. Each analysis had three factors--stress, interval (D-burst+vowel, fricative noise and, for "ditch," silent interval) and rate.

The analysis of variance performed on the data for "dish" showed main effects of rate and interval. The effect of interval indicated that the mean proportion of the sentence subsumed by the fricative noise (.074 = 129.3 msec) was greater than for the D-burst+vowel interval [.058 = 103.5 msec: F(1,7) = 15.81, p<.01]. The effect of rate of speaking indicated that the proportion of the sentence subsumed by a given interval tended to increase from .063 to .068 as the rate of speaking increased [F(1,7) = 3.10, p<.05]. There was no main effect of stress [F(2,14) = 1.26] and there were no higher order interactions. Figure 2 depicts durations for the acoustically defined intervals from "dish" plotted against the corresponding sentence durations. The symbol "1" represents measurements for the D-burst + vowel interval and "3" represents the duration of the fricative noise. The lines depicted in Figure 2 were fitted to group duration data by the method of least squares. Even

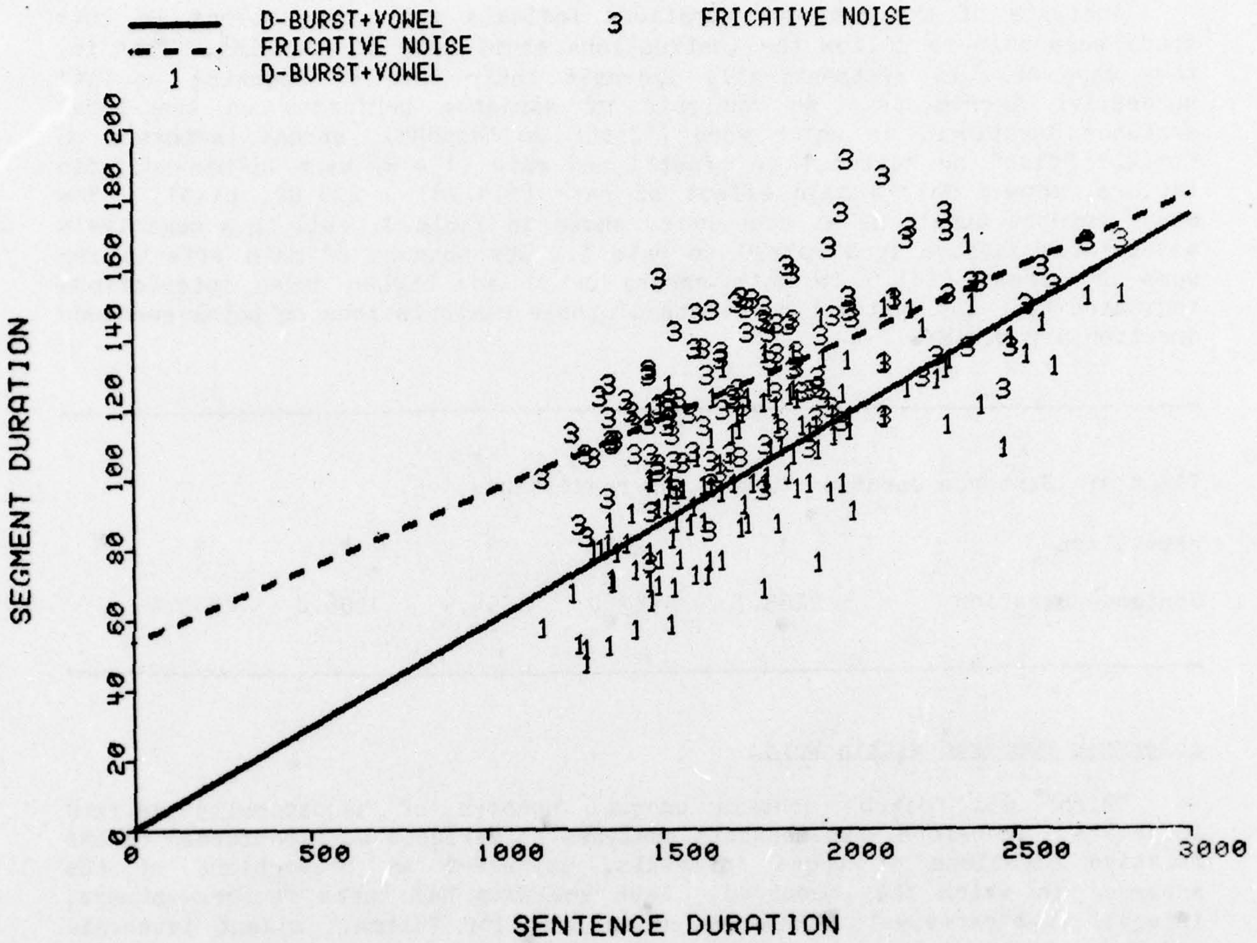


Figure 2: Durations of the acoustically defined intervals of "dish" (msec) plotted against durations of the sentences (msec) in which they occurred.

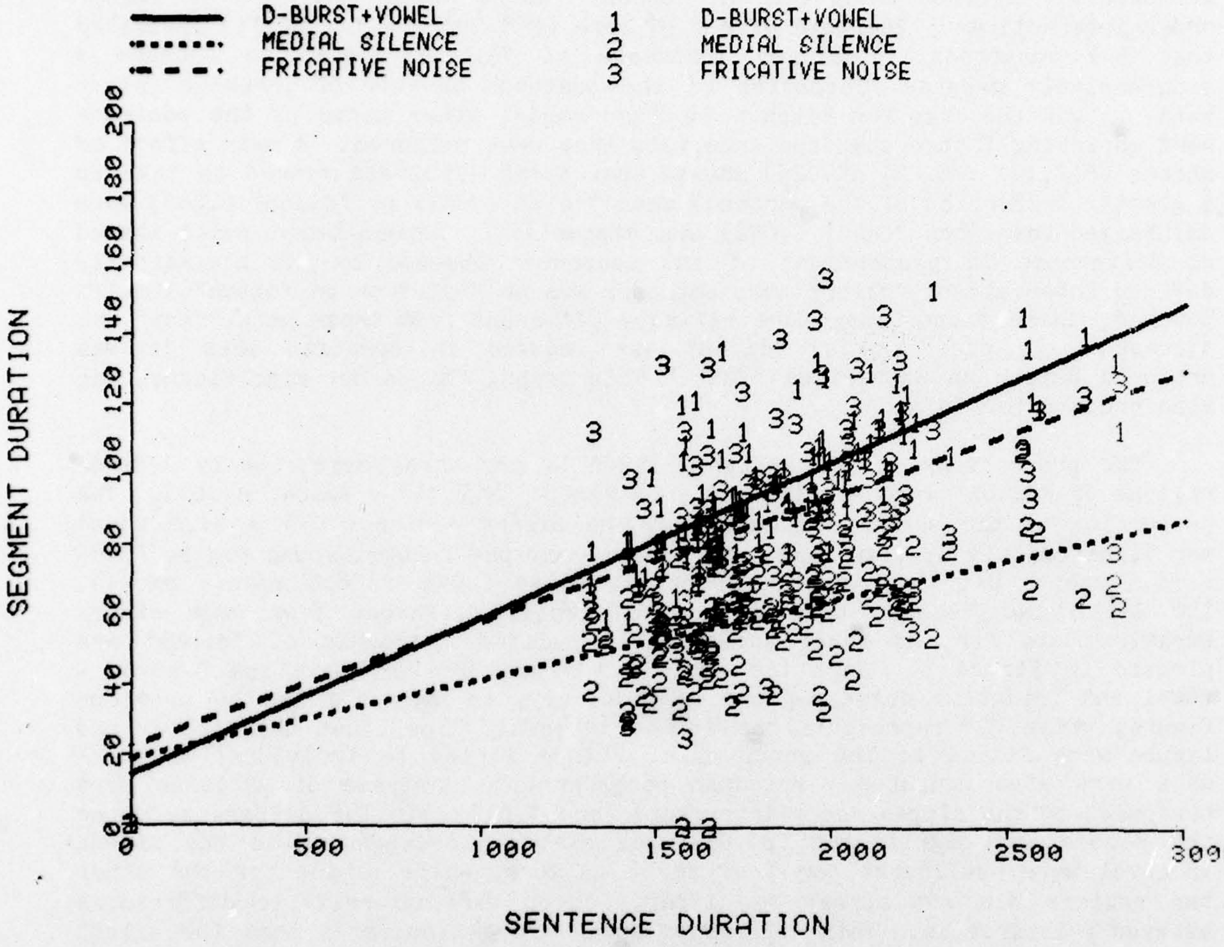


Figure 3: Durations of the acoustically defined intervals of "ditch" (msec) plotted against durations of the sentences (msec) in which they occurred.

though Figure 2 reflects the significant effect of interval, sign-ranks tests on slopes and y-intercepts of the corresponding lines fitted to each talker's data revealed no reliable differences between the two intervals for either parameter (see Table 2).

All three factors yielded significant main effects in the analysis of variance performed on the proportions of the sentence subsumed by the acoustically defined intervals of "ditch." There were, however, no higher order interactions. The main effect of rate [F(4,28) = 4.71, p<.01] indicated that the acoustically defined intervals of "ditch" tended to subsume a progressively greater proportion of the sentence as rate of speaking increased, as was the case for "dish." In other words, other parts of the sentence were shrinking faster than the intervals that were measured. A main effect of stress [F(2,14) = 6.22, p<.025] showed that these intervals tended to take up a greater proportion of the sentence when "talk" (.047) or "ditch" (.046) were emphasized than when "fast" (.042) was emphasized. Newman-Keuls tests showed no difference in proportions of the sentence subsumed by the acoustically defined intervals of "ditch" when emphasis was on "talk" or on "ditch" itself. However, these proportions were reliably different from those when "fast" was stressed. In other words, "ditch" was reduced in duration when it was produced before an emphasized "fast." This trend, though not significant, was also present for "dish."

The proportions of the sentence taken by the three acoustically defined regions of "ditch" were significantly different [F(2,14) = 26.04, p<.01]. The proportion of the sentence subsumed by the silent region (.033 = 61.0 msec) was significantly smaller than that subsumed by the D-burst+vowel region (.053 = 97.1 msec: p<.01) or by the fricative noise (.048 = 88.7 msec: p<.01). The later two regions were not significantly different from each other. Duration data for the three acoustically defined intervals of "ditch" are plotted in Figure 3. Here the symbols "1" and "3" represent the D-burst + vowel and fricative noise regions respectively, as they did in the previous figure, while "2" represents the silent interval. The lines depicted in the figure were fitted to the group data. Lines fitted to individual talkers' data were also computed. Friedman nonparametric analyses of variance were performed on the slopes and y-intercepts (see Table 3). The difference among the slopes was significant (p<.05)--the slopes corresponding to the silent interval were shallowest for 7 of the 8 talkers, while slopes for the other two regions did not appear to differ. There were no reliable differences between y-intercepts. This difference among slopes indicates that the silent interval is less elastic as a function of rate of speaking than the other two intervals. Such a conclusion would have been considerably stronger if a significant interaction between rate of speaking and acoustic interval had been obtained in the analysis of variance.

Intervals Compared Between Words.

The next four figures show the same data plotted so that comparisons may be made directly between various acoustically defined intervals in the two words. Figure 4 shows the durations of the D-burst+vowel interval. The vocalic section for "dish" subsumes a somewhat greater proportion of the sentence (.058) than the same interval for "ditch" [.053: t(7) = 2.87, p<.05]. It can be seen, however, that the intervals from the two words have distributions that overlap a great deal. Figure 5 depicts the durations of the fricative noise from the two words. Notice that the fricative noises of

TABLE 2: Slopes and y-intercepts by subject for the acoustically defined regions of "dish."

Subject	Slopes		Y-intercepts	
	D-burst +vowel	Fricative noise	D-burst +vowel	Fricative noise
1	.030	.046	55.4	28.6
2	.051	.023	26.2	95.9
3	.071	.025	-5.5	85.9
4	.028	.084	42.6	-4.5
5	.040	.071	45.1	14.2
6	.072	.095	-41.5	-23.2
7	.049	.062	-0.3	15.4
8	.047	.039	19.5	55.1

TABLE 3: Slopes and y-intercepts by subject for the acoustically defined regions of "ditch".

Subject	Slopes			Y-intercepts		
	D-burst +vowel	Silence	Fricative noise	D-burst +vowel	Silence	Fricative noise
1	.055	.042	.043	-6.5	-14.8	0.1
2	.028	.040	.030	46.7	-11.3	31.9
3	.069	.017	.040	-16.2	20.4	18.1
4	.038	.028	.091	22.8	-0.9	-51.2
5	.024	.014	.045	65.7	29.8	11.6
6	.026	.001	.084	39.4	68.4	-68.8
7	.056	.015	.025	-13.9	38.8	36.6
8	.045	.009	.017	12.4	49.2	51.1

"dish" are much longer than the fricative noise in "ditch." This is reflected in the fact that the fricative noise takes a substantially greater proportion of the sentence in "dish" (.073) than in "ditch" [.049: t(7) = 12.70, p<.001]. Also, the y-intercepts for the lines fitted to the dish data are higher than the corresponding lines for ditch in 7 of the 8 talkers. Clearly, fricative noise behaves differently with respect to duration in these two cases. Figure 6 shows what happens when the duration of the silent interval in "ditch" is added to its corresponding fricative noise to yield an interval corresponding to the affricate, and is plotted with the fricative noise from "dish," which may be considered to correspond to the fricative segment. The two distributions are much more overlapping when intervals are compared that correspond to phonological segments rather than to acoustically similar regions. Nevertheless, the interval associated with the affricate subsumes a significantly greater proportion of the sentence (.082) than that associated with the fricative [.073: t(7) = 4.09, p<.01]. Figure 7 shows durations for the two words as wholes. Now there is a high degree of overlap and there is no difference between the proportion of the sentence subsumed by "ditch" (.133) and that subsumed by "dish" [.131: t(7) = 0.71]. The longer vocalic region in "dish" has been compensated for by a shorter fricative noise, and conversely, the shorter vocalic section of "ditch" occurs with a longer affricate.

DISCUSSION

The data have confirmed the speculation of Repp et al. (1978) that the durations of silent intervals associated with affricates are relatively more stable than fricative noises across different rates of speaking. In this respect, the silent interval for affricates behaves like that for intervocalic stops relative to vocalic intervals (Gay, 1978).

Inspection of Figure 3 reveals that in production there is more silence at slower rates of speaking--in apparent contrast to the perceptual result. The crucial difference between the perception study of Repp et al. (1978; Dorman et al., 1976) and the present study is that the fricative noise was held constant in the former but varied over a natural range in the latter. Of course, in a production study, it is not possible to hold a given cue absolutely constant. In production, other properties of the fricative noise also vary with rate to sufficiently constrain the segmental identity of the affricate. In the perception study more noise was needed to hear an affricate only in the presence of constant fricative noise duration. We can transform the present data to reflect this situation by mentally rotating the data clockwise with respect to the axes so that the dashed line representing the fricative noise duration is level. This represents a hypothetical situation in which the average timing of the fricative noise in production does not change across sentence duration, to afford a direct comparison of the production data with the perceptual study. When this transformation is performed, the slope of the line fitted to the durations of the closure interval durations is negative, in complete accord with the perceptual result. That is, as one moves right along the abcissa in the direction of an increased rate of speaking, the line fitted to the silent interval gets higher on the ordinate, indicating that more silence is needed as rate of speaking increases.

If a similar transformation is performed upon the data of Figure 3 so that the line fitted to the D-burst+vowel data is now level, it can be argued that the above observations regarding fricative noises may be generalized to

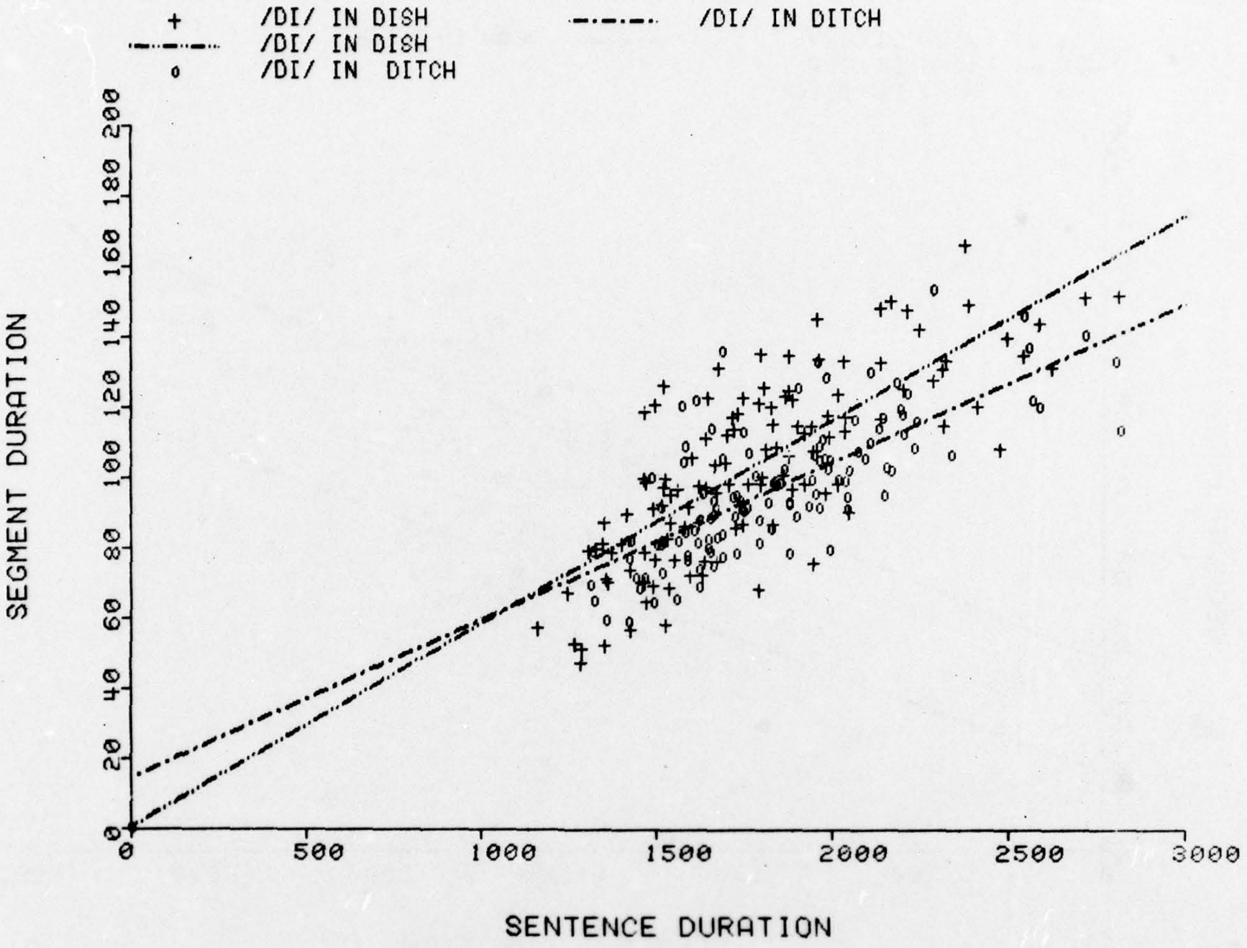


Figure 4: Duraticns of the D-burst+vowel intervals (msec) from "dish" (pluses) and "ditch" (circles) plotted against durations of the sentences (msec) in which they occurred.

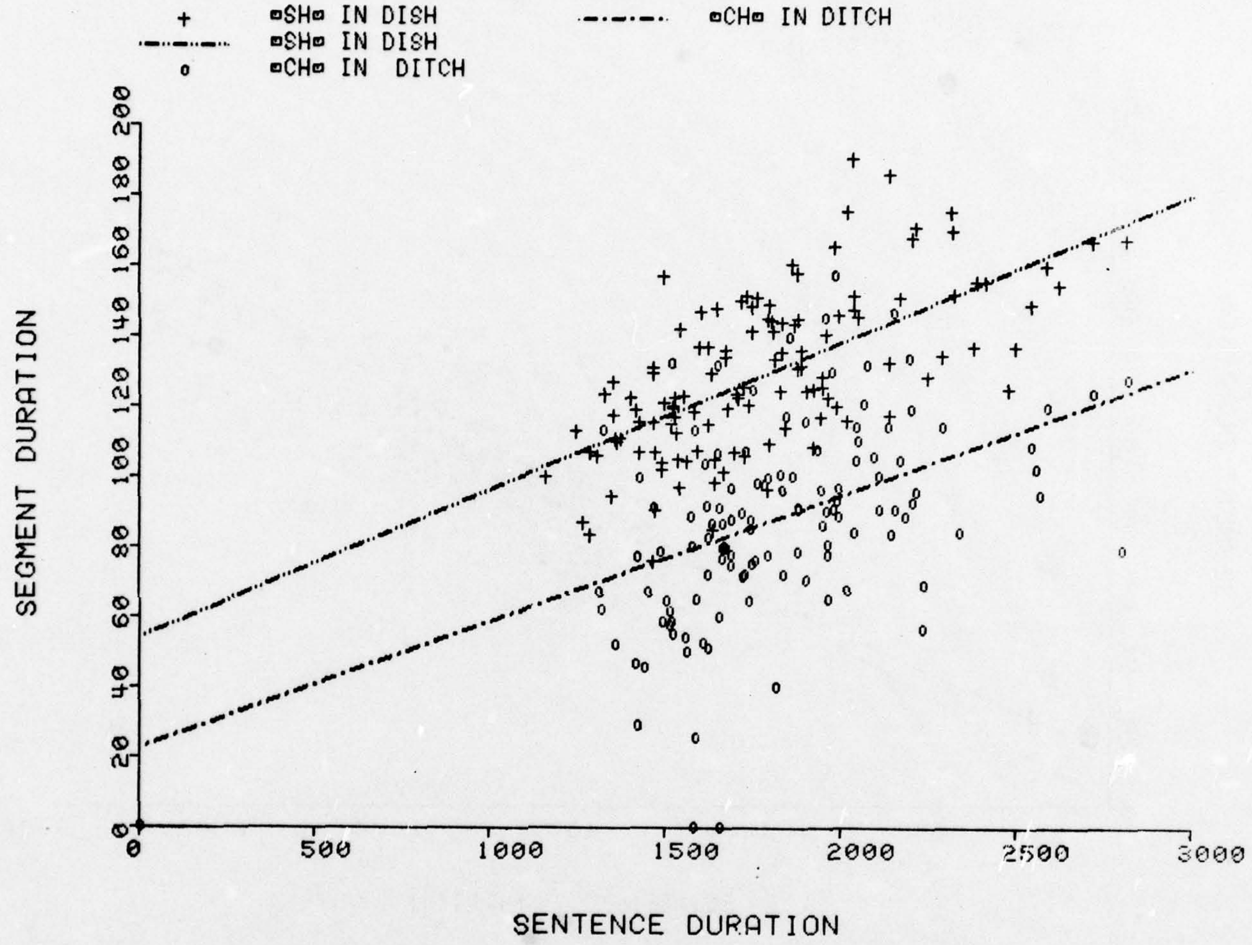


Figure 5: Durations of the fricative noises (msec) from "dish" (pluses) and "ditch" (circles) plotted against durations of the sentences (msec) in which they occurred.

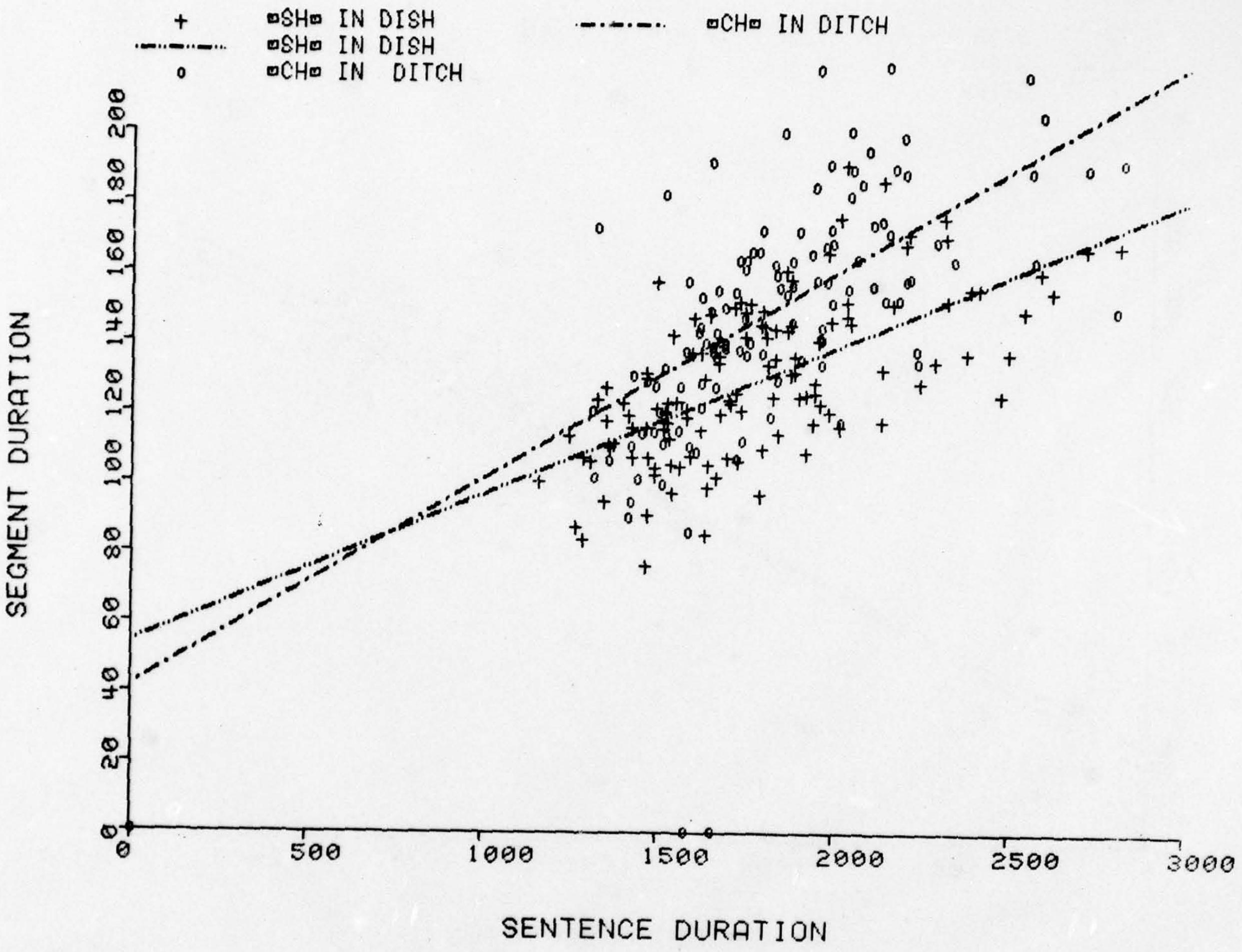


Figure 6: Durations, in msec, of the fricative noise intervals from dish (pluses) and the silence + fricative noise intervals from "ditch" (circles) plotted against the durations of the sentences (msec) in which they occurred.

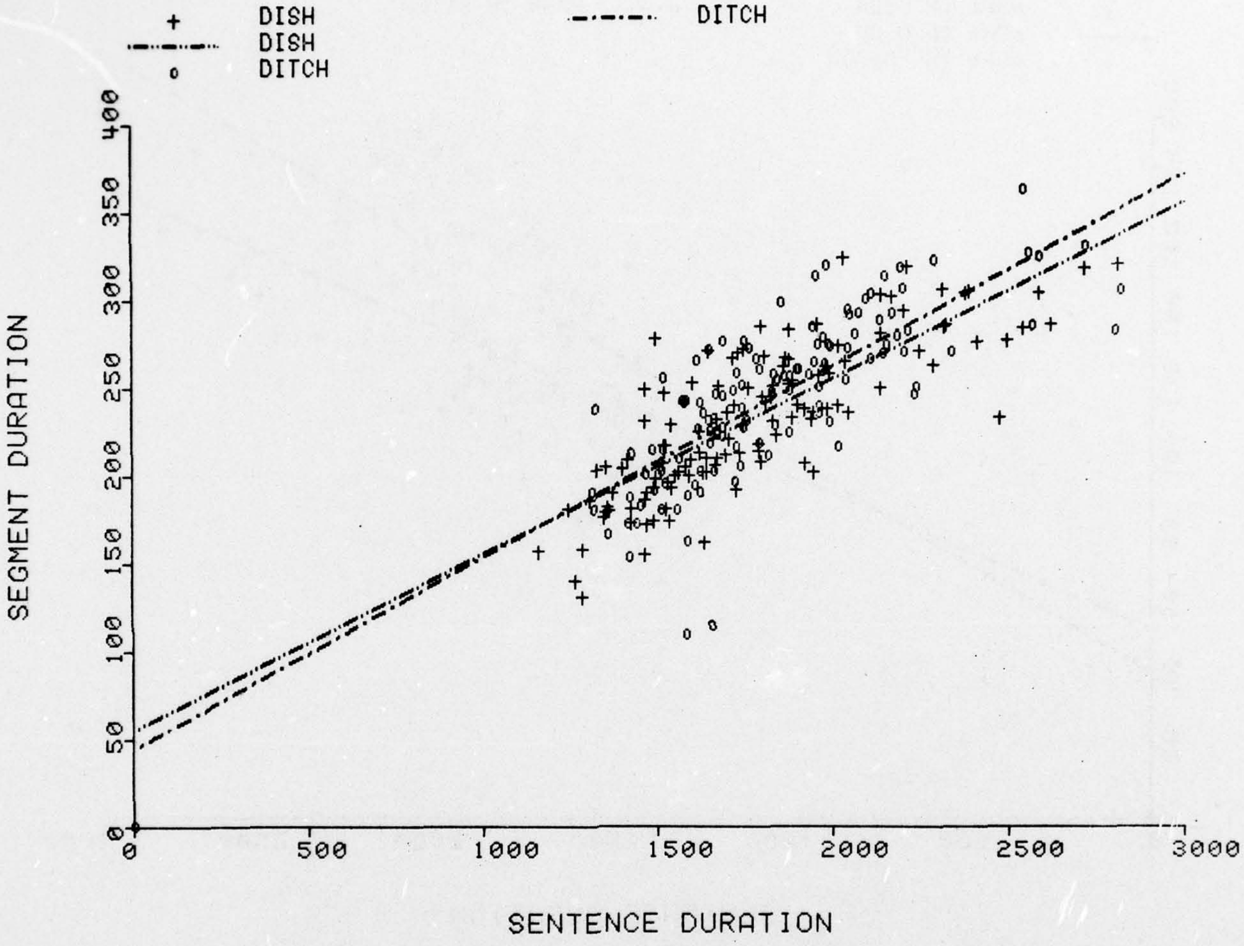


Figure 7: Durations, in msec, of the entire words "dish" and "ditch" plotted against the durations of the sentences (msec) in which they occurred.

vocalic regions as well. First, let us consider that the proportion of the sentence subsumed by the D-burst+vowel interval is greater for "dish" than for "ditch." Bailey and Summerfield (1978) have observed that any acoustic property that differs in the production of two phonemes may have cue value if the other acoustic properties that differentiate the phonemes are sufficiently neutralized. Support for this claim may be found in the present data where strong systematic differences between fricative and affricate in the durations of closure intervals and fricative noises closely parallel well-established perceptual results (Gerstman, see Footnote 2; Dorman et al., 1976; Repp et al., 1978). Thus the prediction may be made from the present data that, all other things being equal, the longer the vocalic section preceding the fricative noise, the more fricative-like will be the resulting percept. This prediction is tantamount to the claim that the absolute duration of a vocalic section is a cue to the fricative-affricate distinction.[6]

Now, let us consider what would result were the relative duration of the vocalic section to be manipulated by holding its absolute duration constant while changing the rate of speaking of its carrier utterance. By the logic of Repp et al. (1978), an increase in rate of speaking would effectively increase the relative duration of the vocalic section, biasing the percept towards a fricative and necessitating more silence to hear an affricate. This interpretation, however, appears to be at odds with their results. That is, increasing the rate of the carrier phrase "Why don't we say..." would presumably shorten the duration of the vocalic section of "say" (cf. Gay, 1978) relative to the silence and fricative noise of "shop." By the above logic this would tend to bias the percept towards affricate and make it possible to hear the affricate with less silence. The effect of rate of speaking found by Repp et al. (1978) goes in the opposite direction--an increase in rate of speaking necessitates more silence to hear the affricate.

There is now good evidence that timing cues associated with the voicing feature (for example, voice-onset time and intervocalic closure duration) are more sensitive to local changes in speaking rate than to global ones (Port, see Footnote 3; Summerfield[7]). If these findings may be generalized to the perception of stop and affricate manner, then the present conflict is exacerbated. The local change in duration of the vocalic section is posited to make the percept more affricate-like as rate of speaking increases, while the effect of overall rate of speaking appears to work in the opposite direction.

One attempt to resolve the conflict between the explanation of the perceptual results presented by Repp et al. (1978) and that generated by extensions of the present data considers the possibility that these durational cues reflect temporal compensation within the syllable. It is known, for

[6] There is now evidence that the vocalic sections of "ditch" and "dish" in natural speech contain cues to the fricative-affricate distinction which are as powerful as any other known cues to this distinction except the closure interval [Dorman, M. F., L. J. Raphael and D. Isenberg. (1978) Acoustic cues for a fricative-affricate contrast in word-final position. Unpublished manuscript].

[7] Summerfield, A. Q. (1978) On articulatory rate and perceptual constancy in phonetic perception. Unpublished manuscript.

example, that a vocalic section preceding a closure associated with a voiceless stop is shorter than that preceding a closure associated with a voiced stop. This difference in duration is compensated for by a silent interval that is longer in the former case, yielding a relatively constant overall duration (Lisker, 1978). The comparison of the acoustic intervals within "dish" and "ditch" indicate that such compensation occurs in these syllables. This evidence indicates that acoustically defined intervals associated with similar phonological segments behave more alike with respect to duration than do intervals associated with similar acoustic segments, but that complete temporal compensation is not achieved until the syllable or word level (cf. Figures 4-7). Therefore, it is possible that any cue value of the duration of a vocalic section is a by-product, so to speak, of this temporal compensation. In other words, a shortened vocalic section would cue an affricate by virtue of the listener's implicit knowledge that a relatively short vocalic section compensates for a longer silence. Given that temporal compensation occurs within syllables, it would be unlikely that the duration of the vocalic section of "say" would interact with the duration of the following closure in the same way that the duration of the vocalic section of "ditch" would interact with its closure.

The rate effect of Repp et al. (1978); (Dorman et al., 1976) appears to conflict with some other known results. In another case where a silence has cue value between a fricative noise and a vocalic region, Marcus (1978) uniformly compressed both non-silent intervals simultaneously and found no effect of this manipulation upon the duration of silence at the perceptual boundary between "slit" and "split." If we assume that the production of the acoustically defined regions of "split" parallels that of "ditch" in that the medial silence decreases less than the other two intervals as rate of speaking increases, there is no way that Marcus' (1978) findings may be reconciled with the Repp et al. (1978) and Dorman et al. (1976) results by an appeal to duration cues as such. Another study by Port (1977) naturally varied the rate of speaking "rabid" or the carrier sentence in which it occurred, to learn how much silence was necessary for the percept to become "rapid" under such conditions. Here is yet another case where vocalic regions in both syllables decrease in duration faster than the medial closure as rate of speaking increases (Gay, 1978). In accord with naive intuition, and in discord with the results under consideration, less silence was needed when the rate of either the word or the carrier sentence was increased.

We do not know whether uniform compression of an isolated syllable is sufficient to cue an increased rate of speaking. We do not know what are the relative contributions of formant movement and overall vocalic section length--or fricative noise duration and spectral shaping of that noise--in the perception of speaking rate. Further studies of these factors, in isolated words and in longer carrier utterances, within and across syllable, word and syntactic boundaries, will reveal more clearly how properties of speech and language at every level are encoded in the speech signal.

REFERENCES

Bailey, P. J. and A. Q. Summerfield. (1978) Some observations on the perception of [s]+stop clusters. Haskins Laboratories Status Report on Speech Research SR-53, vol.2, 25-60.

Bastian, J., P. D. Eimas and A. M. Liberman. (1961) Identification and discrimination of a phonemic contrast induced by a silent interval. Journal of the Acoustical Society of America 33A, 842.

Dorman, M. F. and L. J. Raphael. (1977) On the cues for the fricative-affricate distinction in syllable final position. Paper presented at the International Congress of Acoustics, Madrid.

Dorman, M. F., L. J. Raphael and A. M. Liberman. (1976) Further observations on the role of silence in the perception of stop consonants. Haskins Laboratories Status Report on Speech Research SR-48, 199-207.

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

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

Joos, M. (1948) Acoustic Phonetics. Linguistic Society of America Language Monograph No. 23. (Baltimore: Waverly Press).

Kozhevnikov, V. A. and L. A. Chistovich. (1965) Speech, Articulation, and Perception. [U.S. Department of Commerce: National Technical Information Service (JPRS-305430)].

Lisker, L. (1978) Segment duration, voicing and the syllable. Haskins Laboratories Status Report on Speech Research SR-54, 175-189.

Marcus, S. M. (1978) Distinguishing "slit" and "split" - an invariant timing cue in speech perception. Perception & Psychophysics 23, 58-60.

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

Voicing in Intervocalic Stops and Fricatives in Dutch

René Collier,[+] Leigh Lisker,[++] Hajime Hirose[+++] and Tatsujiro Ushijima[+++]

ABSTRACT

This study represents an addition to the literature describing the role of the larynx in the production of voiced and voiceless stop and fricative consonants. Electromyographic (EMG) recordings from the intrinsic laryngeal musculature and measurements of sub- and supraglottal air pressures were obtained from a speaker of Standard Dutch, who produced nonsense forms preceded by a short carrier Dutch phrase. The forms included intervocalic voiced and voiceless stops and fricatives, as well as certain combinations of these consonants. The data obtained are in general conformity with previous studies of larynx management in consonant voicing, indicating systematic differences for voiced vs. voiceless and for stop vs. fricative categories. The different EMG patterns suggest that the voicing dimension primarily involves the varying adjustment of static glottal width by means of the adductor-abductor muscles, while the stop-fricative difference involves both this variable and a feature of longitudinal vocal fold tensing. The evidence is negative so far as providing support for a view that this latter feature plays a significant role in the voicing distinction.

INTRODUCTION

This study describes some systematic differences and correspondences in the activity of certain intrinsic laryngeal muscles and in the variation of subglottal and intraoral air pressure that characterize the realization of the "voiced/voiceless" and "stop/fricative" distinctions in consonants.

We have been encouraged in the pursuit of this aim by the growing evidence, offered in previous studies, that the analysis of these variables

[+] On leave from the University of Antwerp (UFSIA), Belgium.

[++] Also University of Pennsylvania.

[+++] University of Tokyo, Japan.

Acknowledgment: This research was supported in part by the National Institute of Neurological and Communicative Disorders and Stroke, Grant NS-13870.

[HASKINS LABORATORIES: Status Report on Speech Research SR-55/56 (1978)]

[redacted]

can reveal important aspects of the speech production process in general, and of the physiological implementation of phonologically relevant distinctions in particular.

It has been our intention to duplicate certain electromyographic (EMG) and air pressure observations of the past, based now on another language (Dutch), while extending the range of such observations by investigating both single consonants and consonant clusters and combining EMG and air pressure data for the same subject.

EXPERIMENTAL SET-UP

Data Collection and Processing

In a first experiment we have attempted to record the EMG signals in the following muscles: the interarytenoid (INT), the vocalis (VOC), the lateral cricoarytenoid (LCA), the posterior cricoarytenoid (PCA) and the cricothyroid (CT). The preparation of the hooked-wire electrodes and the techniques of their insertion have been explained by Hirose (1971a). and Hirose and Gay (1972).

In a second experiment the air pressure data were recorded. Subglottal air pressure (P_{sg}) was measured immediately below the glottis by means of a flexible plastic tube inserted through the cricothyroid membrane. Intraoral air pressure (P_{io}) was measured in the pharyngeal cavity by means of a catheter inserted through the nose. The two pressure recording tubes were coupled to two pressure transducers (Setra Systems, model 236L). In the second experiment we have again measured the EMG activity in the VOC and CT muscles. This partial repetition of the EMG recordings was done in order to compare the pattern of muscle activity and its timing across the two experiments, in order to decide whether the EMG data of the first experiment could indeed be combined with the pressure data of the second.

The physiological signals, the audio signal and timing pulses were recorded on a 14-channel instrumentation recorder (Consolidated Electrodynamics VR-3300). The visual editing of the raw data and their computer processing were performed on the Haskins Laboratories' EMG data processing system. Details of the successive procedures have been explained by Port (1971) and Kewley-Port (1973, 1974). The EMG and pressure signals have been integrated with a time constant of 50 and 25 msec, respectively. The comparison of the VOC and CT data of the two experiments revealed very similar patterns of activity in the respective muscles; the timing of the activity patterns was nearly identical. The data of the two experiments can therefore be combined in the presentation of the results below. Reliable data could be obtained for all the variables under investigation, except for the PCA muscle.

Speech Materials and Subject

Dutch has the following stop and fricative phonemes: /p,b,t,d,k,f,v,s,z,x,γ,H/. There is no phoneme /g/ in Dutch, but [g] can occur as an allophone of /k/. Also lacking in the phoneme inventory is any voiceless glottal fricative in contrast with /H/. To this set of consonants we have added the glottal stop [ʔ], which can occur as the nondistinctive "hard attack" of a word-initial vowel or between adjacent vowel sounds.

In order to study the same consonants as elements in clusters of two consonants, we have included the following intervocalic combinations: [pt, tp, kp, fp, sp, xp, bd, d̂b, gb, vb, zb, γb]. It should be noted that in these clusters the two consonants are the same with regard to the feature "voice." Indeed, Dutch phonology has a rule to the effect that in combinations of two consonants (stops or fricatives), both segments become voiceless, except where the second element is either /b/ or /d/, in which case both segments become voiced. For example, /pv/ is realized as [pf], /zp/ as [sp], /tb/ as [db], /kd/ as [gd], and so on. The consonants under study were embedded in nonsense words of the form /'baC(C)at/. The test words were preceded by the carrier phrase "Waar ligt ______?" ([wa:rlɪxt_____]), meaning "Where is _____ located?"

The complete list of test words is given in Table 1 below. It should be noted that the second column in Table 1 does not contain all possible combinations of two stops or fricatives in Dutch. We have limited the list to those C_1C_2-combinations in which C_2 is a stop.

TABLE 1: List of test words containing one and two intervocalic consonants as used in the experiment.

['bapat
'batat
'bakat
'bafat
'basat
'baxat
'baʔat
'babat
'badat
'bagat
'bavat
'bazat
'baγat
'baHat

'babdat
'badbat
'bagbat
'bavbat
'bazbat
'baγbat
'baptat
'batpat
'bakpat
'bafpat
'baspat
'baxpat]

The order of the test sentences was randomized. The speech materials were read by one subject, the first author, who is a native speaker of the variety of Standard Dutch spoken in the northern part of Belgium.

RESULTS

In presenting the results we will show the data for selected voiceless/voiced and stop/fricative oppositions with single consonants and clusters of two. The figures show the EMG activity in the INT, LCA, and VOC muscles, as well as the P_{io} and P_{sg} variations. The data on PCA activity deteriorated in the course of the experimental session. The activity of CT appeared not to be relevant to the consonant distinctions under investigation, at least not in our subject. Therefore the PCA and CT data will not appear in the figures below.

It can be seen in those figures that, during the production of vowels, P_{io} remains at approximately 1 cm aq above atmospheric pressure, rather than being equal to it, as one would expect. We have attempted to simulate, post factum, the conditions under which this phenomenon may occur. We have found that it may be caused by a small degree of clogging in the catheter that picks up the pressure in the pharyngeal cavity. It appeared that if the clogging is not too severe, the peak P_{io} values are not significantly affected. The reliability of these peak values remains questionable, but since our P_{io} data are in good agreement with those that have been published by other researchers, we have decided to present and discuss them, be it with some caution.

Single Stops and Fricatives

The Voiceless/Voiced Contrast in Stops. Figure 1 exemplifies the voiceless/voiced opposition as it is found in the /t-d/ contrast. Table 2, A and B, presents the exact numerical values for the physiological variables at selected points in time. In Table 2 it can also be checked to what extent the /t-d/ opposition is typical of the voicing contrast between Dutch stops in general.

INT: At about 100 msec before the line-up point, INT activity starts to decrease for the following /t/, but not for the following /d/. The lowest level of activity around the line-up point is 97 microvolts for /t/ and 121 microvolts for /d/. There is a high peak of INT activity after /t/, but not after /d/.

LCA: During the last 50 msec before the line-up point there is a small decrease in LCA activity for the following /t/, but not for /d/. At the line-up point the level of LCA activity is 27 microvolts for /t/ and 30 for /d/. Following /t/ there is a small peak in LCA activity; after /d/ there is no such peak, but rather a decrease (which, however, is not to be found after /b/ or /g/).

VOC: At about 100 msec before the line-up point VOC activity starts to decrease for both the following /t/ and /d/. At the line-up point the level of VOC activity is 72 microvolts for /t/ and 70 for /d/. After /t/ there is strong increase of VOC activity, but not after /d/.

P_{io}: Intraoral air pressure rises to a maximum of 5.83 cm aq during the closure of /t/, and 3.15 cm aq during that of /d/. The pressure curves also reflect the durational difference between the voiceless and the voiced stop.

P_{sg}: Measured at the moment of maximum P_{io}, the level of P_{sg} is 6.74 cm aq for /t/ and 6.84 cm aq for /d/. The pressure drop across the glottis, ΔP (namely, $P_{sg} - P_{io}$), measured at the moment of maximum P_{io}, is 0.91 cm aq for /t/ and 3.69 cm aq for /d/.

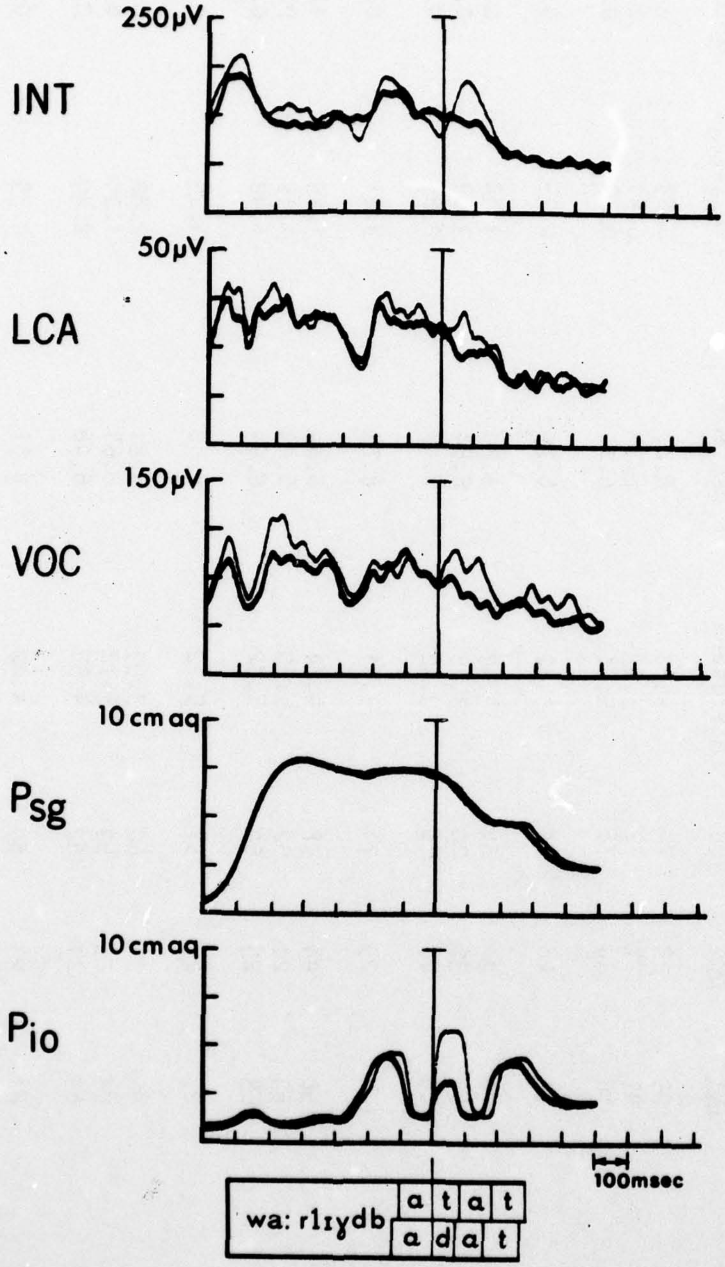


Figure 1: Averaged EMG and pressure data for Dutch /t/ (thin line) and /d/ (thick line).

TABLE 2: Numerical values at selected points in time for EMG and air pressure signals in single stops and fricatives.

	ADDUCTOR MUSCLES Minimum values (in microvolts) within 25 msec distance from line-up point.			INTRAORAL AND SUBGLOTTAL AIR PRESSURE (in cm aq)			
	INT	LCA	VOC	Maximum P_{io} During Stricture	P_{sg} at moment of maximum P_{io}	$\Delta P(=P_{sg}-P_{io})$	P_{sg} Drop/Rise During Stricture
A) p	95	28	74	5.81	6.80	0.99	0.18
t	97	27	72	5.83	6.74	0.91	0.08
k	86	30	76	5.81	6.70	0.89	0.22
Average	93	28	74	5.82	6.75	0.93	0.12
B) b	122	29	64	3.47	6.85	3.38	0.11
d	121	30	70	3.15	6.84	3.69	0.14
g	104	27	78	3.92	6.47	2.55	0.11
Average	116	29	71	3.51	6.72	3.21	0.12
C) f	50	24	58	4.84	6.09	1.25	1.03/ -
s	69	21	52	5.30	6.49	1.19	0.82/0.25
x	53	25	60	5.22	6.64	1.42	0.96/0.07
Average	57	23	57	5.12	6.40	1.29	0.93/0.10
D) v	80	26	60	3.99	6.81	2.82	0.21
z	93	23	59	4.72	6.67	1.95	0.32
γ	69	21	62	4.56	6.96	2.40	0.38
Average	81	23	60	4.42	6.81	2.39	0.30
E) 6	69	21	62				

As far as the carrier phrase itself is concerned, it can be observed that the adductor muscles show a momentary burst of activity before the onset of phonation. At about 200 msec before the line-up point there is a drop in the activity of INT, VOC and LCA that can be associated with the cluster /γdb/. This cluster results in a first increase of P_{io}. The second increase of P_{io} is associated with the consonants under study, while the third corresponds to the final consonant /t/ of the test word. The utterances were read with a rising-falling intonation on the first word, [wa:r], and on the test word ['bɑCɑt]. This variation of F_0 over time is somewhat reflected in the course of the P_{sg} curve. In particular, the falling F_0 on the second syllable of the test word is clearly reflected in the falling P_{sg} during the first 200 msec after the line-up point. In other words, the falling P_{sg} following the line-up point is to be associated with prosodic properties of the utterances, not with the consonants proper.

It can be concluded from the data in Table 2, A and B, that the same tendencies also hold for the comparisons /p - b/ and /k - g/.

The Voiceless/Voiced Contrast in Fricatives

Figure 2 illustrates the opposition between a voiceless and a voiced fricative, as it is found in the /f - v/ contrast. Precise numerical data on these and the other fricatives are to be found in Table 2, C and D.

INT: At about 200 msec before the line-up point INT activity begins to decrease for the following consonant. The relaxation in this muscle is greater for /f/ than for /v/: the minimum level of activity is 50 and 80 microvolts, respectively. There is a higher peak of INT activity after /f/ than after /v/.

LCA: At about 75 msec before the line-up point LCA activity decreases for both /f/ and /v/. The lowest level of LCA activity is 24 microvolts for /f/ and 26 microvolts for /v/. There is no difference in LCA activity after the two fricatives.

VOC: At about 75 msec before the line-up point VOC activity decreases to the same extent for the following /f/ and /v/. The lowest level of activity in VOC is 58 microvolts for /f/ and 60 microvolts for /v/. The peak in VOC activity is slightly higher after /f/.

P_{io}: Intraoral pressure rises to a maximum of 4.84 cm aq for /f/ and 3.99 cm aq for /v/. The pressure curve also reflects the difference in stricture duration between the two fricatives. This timing difference is also to be found in a slightly later increase of INT and LCA activity after the longer fricative /f/.

P_{sg}: Measured at the moment of maximum P_{io}, the level of P_{sg} is 6.09 cm aq for /f/ and 6.81 cm aq for /v/. During the stricture of /f/, P_{sg} drops by 1.03 cm aq; during /v/ this pressure drop is only 0.21 cm aq. The ΔP value, measured at the moment of maximum P_{io}, is 1.25 cm aq for /f/ and 2.82 cm aq for /v/.

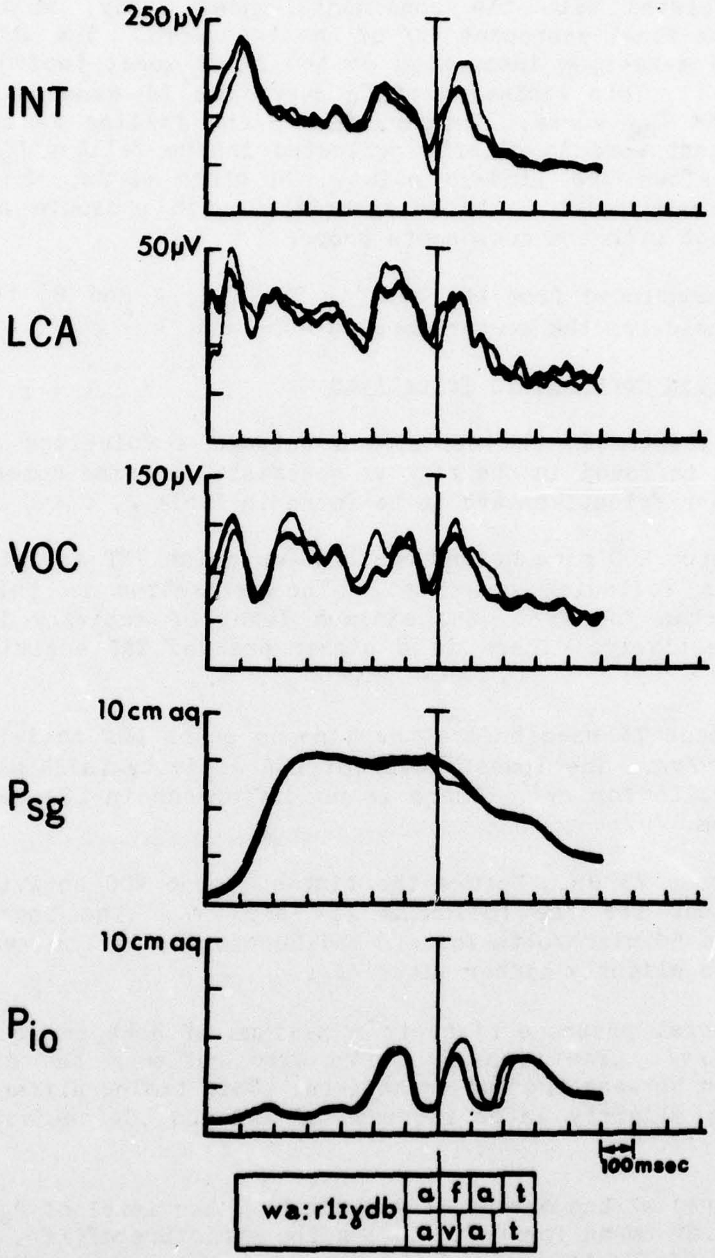


Figure 2: Averaged EMG and pressure data for Dutch /f/ (thin line) and /v/ (thick line).

It can be seen in Table 2, C and D, that the same tendencies apply to the comparisons /s - z/ and /x - γ/. Notice that the degree of INT relaxation is smaller for /s/ and /z/ than for the other fricatives. These two fricatives also have higher P_{io} values than the others. The P_{sg} drop at the beginning of /s/ and /x/ is followed by a small increase.

Some General Comparisons

Based on a comparison of the data in Figures 1 and 2, and of the averages in Table 2 we may conclude that:

1. LCA shows almost no reduction of activity for a stop, but it relaxes somewhat for a fricative. There are no major differences in the pattern of (reduced) LCA activity that correspond to the voiceless/voiced distinction in stops and fricatives.

2. VOC shows some reduction of activity for a stop and a much stronger relaxation for a fricative. The degree of relaxation is nearly the same for a voiceless as for a voiced consonant, but the level of VOC activity tends to be somewhat higher after a voiceless consonant.

3. INT shows no significant decrease of activity for a voiced stop. Its activity is clearly reduced for a voiceless stop and even more so for a voiced fricative. INT relaxes most for a voiceless fricative. The degree of INT activity at resumption after a consonant is proportional to the degree of relaxation for that consonant.

4. P_{io} is higher in voiceless stops and fricatives than in their voiced counterparts.

5. P_{sg} decreases momentarily at the beginning of fricatives, especially voiceless ones. In voiceless fricatives the P_{sg} drop may be followed by a slight rise.

6. The pressure drop across the glottis, measured at the moment of maximum P_{io}, is greater in voiced consonants than in voiceless ones. It is also greater in voiced stops than in voiced fricatives.

Discussion

The difference between voiceless and voiced stops is most clearly reflected in the pattern of INT activity (and, presumably, in the converse pattern of its antagonist, PCA): voiceless stops are characterized by partially suppressed INT activity, whereas for voiced stops there is practically no INT suppression. The voiced/voiceless distinction is only indirectly reflected in a difference in VOC activity: this muscle tends to be somewhat more active after a voiceless stop than after a voiced one. There is only a small difference between voiceless and voiced stops as far as P_{sg} is concerned: the former may have a small drop in P_{sg} at the beginning of occlusion. P_{io} is higher in voiceless than in voiced stops.

The difference between voiceless and voiced fricatives is also most readily accounted for in terms of differences in INT activity: this muscle shows more relaxation in the voiceless than in the voiced case. VOC activity is related only indirectly to the voicing distinction in fricatives, in that the level of VOC activity tends to be higher after a voiceless than after a voice. fricative. P_{io} is higher in voiceless fricatives than in voiced ones and the momentary P_{sg} drop is more pronounced with the former than with the latter.

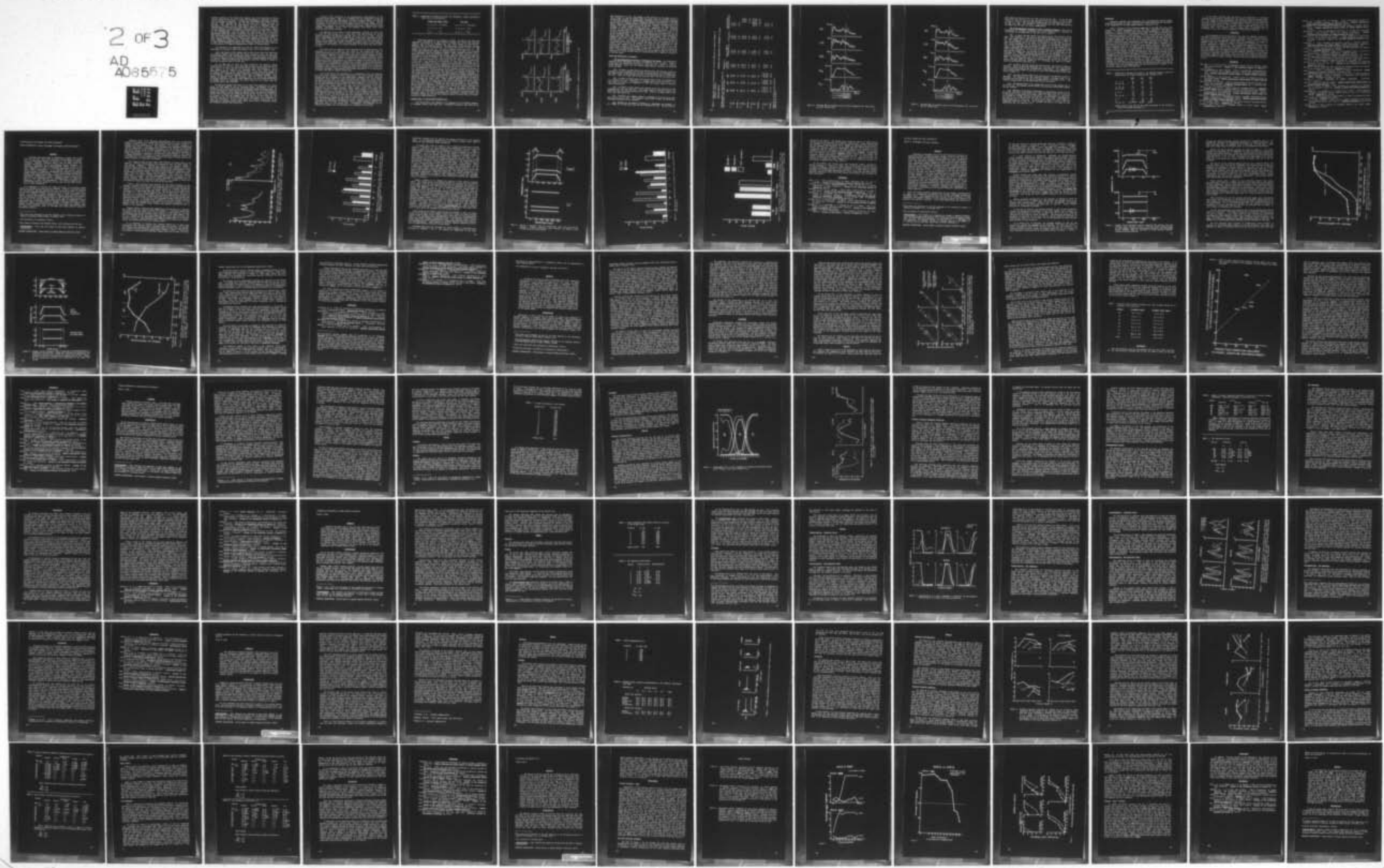
An overall comparison of stops and fricatives indicates that, on the whole, there is less adductor muscle activity in fricatives than in stops. It may also be observed that voiceless stops and voiced fricatives have a fairly similar pattern of INT activity, but that they differ more strongly in terms of LCA and VOC activity. From this we may infer that voiceless stops and voiced fricatives differ in glottal width and/or glottal shape. The unaspirated voiceless stop may be produced with a glottis that is generally closed, but with a small degree of opening at the posterior end, a configuration effected by relaxation of INT, with LCA (and VOC) contracted. When both INT and LCA relax for a voiced fricative, there should be a slightly larger opening that involves also a separation at the level of the vocal processes (assuming that the degree of PCA activity is not widely different in both cases). Voiced stops show practically no relaxation in the adductor muscles and may have the same degree of glottal width as vowels. Finally, voiceless fricatives show the strongest degree of adductor muscle relaxation and, presumably, have the largest degree of glottal opening.

Stops show less VOC relaxation than fricatives, suggesting that the vocal folds are slacker in the latter case. Possibly the slackening of the vocal folds in fricatives also contributes to their abduction. In the case of voiced fricatives the slackening might be said to facilitate the maintenance of vocal fold vibration during the constriction. Unfortunately for this view, since VOC relaxation also characterizes voiceless fricatives, we should have to admit that slackening of the folds is not incompatible with voicelessness.

Let us now turn to a comparison of the findings described above and some available data on the articulation of consonants other than Dutch. Hirose, Lisker and Abramson (1972) have examined laryngeal muscle activity in the five types of bilabial stops that are common in Sindhi (as produced by a phonetician who was not a native speaker). They observed no INT or LCA relaxation for [b], but some for [p]. Our observations are in good agreement with their findings. Hirose and Gay (1972) mention that for the articulation of both voiceless and voiced English stops (in intervocalic, poststressed position) there is a slight decrease of VOC activity. In our data we observe the same tendency. On the other hand, in the data of these authors there is the same degree of VOC relaxation in fricatives as in stops, whereas our results show more VOC relaxation in the case of fricatives. The conclusion then would be that at least in the present data, VOC activity differentiates very little between voiced and voiceless consonants; however, it clearly correlates with the stop/fricative distinction. This would be in agreement with the results of an EMG study involving Danish consonants (Fischer-Jørgensen and Hirose, 1974a). These authors have found that in their data the

AD-A065 575   HASKINS LABS INC NEW HAVEN CONN   F/G 5/7
SPEECH RESEARCH. A REPORT ON THE STATUS AND PROGRESS OF STUDIES--ETC(U)
DEC 78 A M LIBERMAN   PHS-HD-01994
UNCLASSIFIED   SR-55/56(1978)   NL

2 OF 3
AD
A065575



activity pattern of VOC seems more complex than that described by Hirose (1971b), where VOC was active in vowels and suppressed in consonants irrespective of the type of consonant. Our data are in agreement with those of Hirose and Gay (1972), showing that LCA and VOC have very similar patterns of activity. INT (and PCA for that matter) is also in agreement with the observations of Hirose and Gay (1972) and of Hirose et al. (1972), showing an activity pattern of fine adjustment rather than an all-or-none type. Indeed, our data indicate that in the activity of INT, several (at least three) levels can be distinguished: one for vowels and voiced stops, one for voiceless unaspirated stops and voiced fricatives, and one for voiceless fricatives. This three-level distinction in the pattern of INT (and PCA) activity suggests that the corresponding three classes of Dutch speech sounds also differ in their degree of glottal width. Fiberoptic high-speed cinefilms of, among others, Sawashima, Abramson, Cooper and Lisker (1970) indicate that there are at least three, possibly as many as five, distinct degrees of glottal aperture in the production of English consonants and vowels in running speech.

The outcome of our experiment can also be related and compared to the hypotheses put forward by Halle and Stevens (1967, 1971) and Stevens (1975).

Halle and Stevens (1967) assume that in both voiced stops and fricatives the glottis remains open during the entire vibratory cycle, and that this overt adjustment in vocal fold position toward a more open state is necessary in order to maintain vocal fold vibration with a reduced pressure drop across the glottis. They also assume that the degree of glottal opening is larger in voiceless consonants than in voiced ones.

In our EMG data we find no clear indication of adductor muscle relaxation for /b/ and /d/. We do find some for /g/, in which stop the ΔP for /g/ is still 2.5 cm aq, and this value is well above the minimum of 1 cm aq that appears to be required for continued vocal fold vibration during obstruents (Lindqvist, 1972). Voiceless stops, on the other hand, appear to be produced with reduced INT muscle activity, indicating some separation of the arytenoids. Halle and Stevens (1971) make a different proposal. They assume, as against their 1967 proposal, that for voiced and unaspirated voiceless stops there is no vocal fold separation, and the voicing distinction is brought about by slackening the vocal folds in the voiced case and stiffening them in the voiceless.

Our EMG data suggest that there is no vocal fold separation for voiced stops, but that there is one for voiceless stops. Furthermore, we find no evidence in VOC activity for vocal fold stiffening during voiceless stops. As far as fricatives are concerned, Halle and Stevens (1971) do not make explicit claims with regard to their degree of glottal width, but they do hypothesize "stiff" vocal folds for the voiceless fricatives and "slack" vocal folds for the voiced. Our EMG data suggest that the glottis is more open for the voiceless than for the voiced fricatives, but that there is no difference in vocal fold stiffness. In fact, both types of fricative would have to be considered as implying "slack" vocal folds, since there is evidently strong VOC relaxation during their production.

Stevens (1975) elaborates on the physiological characteristics of the different larynx modes. Again it is assumed that the unaspirated voiced and voiceless stops have the same degree of glottal opening, namely, the neutral position of the arytenoid cartilages that is also typical of vowels. Now the "stiffness" of the vocal folds in voiceless stops is no longer sought in their longitudinal tensing, but in their vertical stretching, resulting from larynx raising. Conversely, voiced stops are produced with slack vocal cords resulting from larynx lowering.

Our EMG data do not speak to these hypotheses. However, we would like to point out that Hirose et al. (1972) have found EMG evidence for active larynx lowering in the implosive stop [ɓ] only, not in [b] or [bh]. The EMG data on pharyngeal cavity size expansion for voiced stops, presented by Bell-Berti (1975), indicate that not all speakers actively lower their larynx during the articulation of these consonants. Finally, the x-ray analysis of Perkell (1969) shows that "there is little observable effect of the different consonants on the behavior of vertical movement of the hyoid bone and larynx" (p. 42).

We believe that the general picture emerging from the physiological data available so far is, that--at the level of the larynx--the type of voiced/voiceless distinction discussed in this paper correlates more strongly with different degrees of glottal width than with different degrees of vocal fold stiffness. The four classes of (unaspirated) consonants labeled "voiced stop," "voiceless stop," "voiced fricative" and "voiceless fricative" can be separated by unique combinations of a specific degree of glottal width and a specific degree of supraglottal constriction. A model of the voiced/voiceless distinction in consonants may do without an additional parameter of vocal fold stiffness, since, even in the opinion of Halle and Stevens (1971), increased stiffness is only effective in inhibiting vocal fold vibration if there is a sufficient decrease in the pressure drop across the glottis, or if the glottis is either wide open or tightly constricted.

Let us now direct the discussion to a comparison of our air pressure data with those reported by other researchers. Our data indicate that the average peak P_{io} is highest for voiceless stops, somewhat lower for voiceless fricatives, still lower for voiced fricatives and lowest for voiced stops. The same rank order is to be found in the P_{io} data of Prosek and House (1975) for the corresponding classes of speech sounds in English. In particular, our data confirm the earlier observation that voiced fricatives have higher peak P_{io} values than voiced stops. Our data are also in agreement with those presented by, among others, Netsell (1969), Lisker (1970), Slis (1970) and Löfqvist (1974b). Prosek and House (1975) have reported "a tendency for consonants produced in the back of the mouth to have greater peak pressures than consonants produced anteriorly (p. 140)." Table 3 compares the Prosek and House data to ours. It can be seen that both sets agree in the rank order of P_{io} values as a function of the place of articulation. It should be noted that the agreement between the two sets of data can only be found if, in our data, <u>all</u> the homorganic consonants are pooled. However, in any subgroup of consonants the rank order is different, and no single subgroup exhibits the same rank order as the pooled data. Furthermore, the magnitude of the differences between, say, two consonants in a subgroup varies widely.

TABLE 3: Comparison of pooled P_{io} values for consonants, ranked according to their place of articulation.

Prosek and House (1975)		Our Data	
p, b, f, v:	4.5 cm aq	p, b, f, v:	4.52 cm aq
t, d, s, z:	4.9	t, d, s, z:	4.75
k, g :	5.0	k, g, x :	4.87

Concerning subglottal air pressure in stops, our data do not show any significant differences in peak pressure between the voiced and voiceless cognates when these are averaged over the three places of articulation. However, when we compare stops of the same place of articulation, the P_{sg} difference may well be significant. Also note that while /b/ and /d/ show higher P_{sg} than /p/ and /t/, /g/ has a markedly lower P_{sg} than /k/. The absence of a significant difference between the /p, t, k/ and /b, d, g/ groups is in agreement with the findings of Netsell (1969), McGlone and Shipp (1971), Ohala and Ohala (1972) and Löfqvist (1974a). The very small drop in P_{sg} that we observed at the beginning of the closure phase of voiceless stops was also observed by Löfqvist (1974a), but he showed that the difference in this respect between voiceless and voiced stops was not statistically significant in his data. P_{sg} variation in fricatives is less well documented in the literature. Our data show that in voiced fricatives there is a slight drop in P_{sg} at the very beginning of the supraglottal constriction gesture; this drop is more pronounced with voiceless fricatives and can be of the order of 1 cm aq. The momentary decrease in P_{sg} is in fact already initiated at the end of the preceding vowel, indicating that the glottis is already relatively wide open before the oral constriction (and hence P_{io}) has reached its maximum. As the constriction becomes narrower, P_{sg} may slightly rise again, especially in voiceless fricatives. These characteristic P_{io} and P_{sg} variations are in good agreement with the airflow variations in voiceless fricatives as studied by Klatt, Stevens and Mead (1968). These authors observed that the airflow traces of (English) voiceless fricatives show a characteristic "double peak," which they explain as a consequence of the relative timing of the laryngeal and articulatory gestures. During the vowel that precedes the fricative, the glottis begins to open while the vocal cords continue to vibrate, resulting in an increase in airflow and a lowering of P_{sg}. As the upper teeth begin to make contact with the lower lip, there is a decrease in airflow, a rise in P_{io} and a rise (or stabilization) of P_{sg}. Then, as the lower lip moves away from the upper teeth, air flow increases and P_{io} decreases. Finally, vocal cord vibration resumes and airflow and pressure return to values that are characteristic of vowel production.

Glottal Stop [?] and Glottal Fricative /H/

The intervocalic stop appears to be produced by the following sequence of laryngeal events (see Figure 3A): Preceding the glottal stop there is no

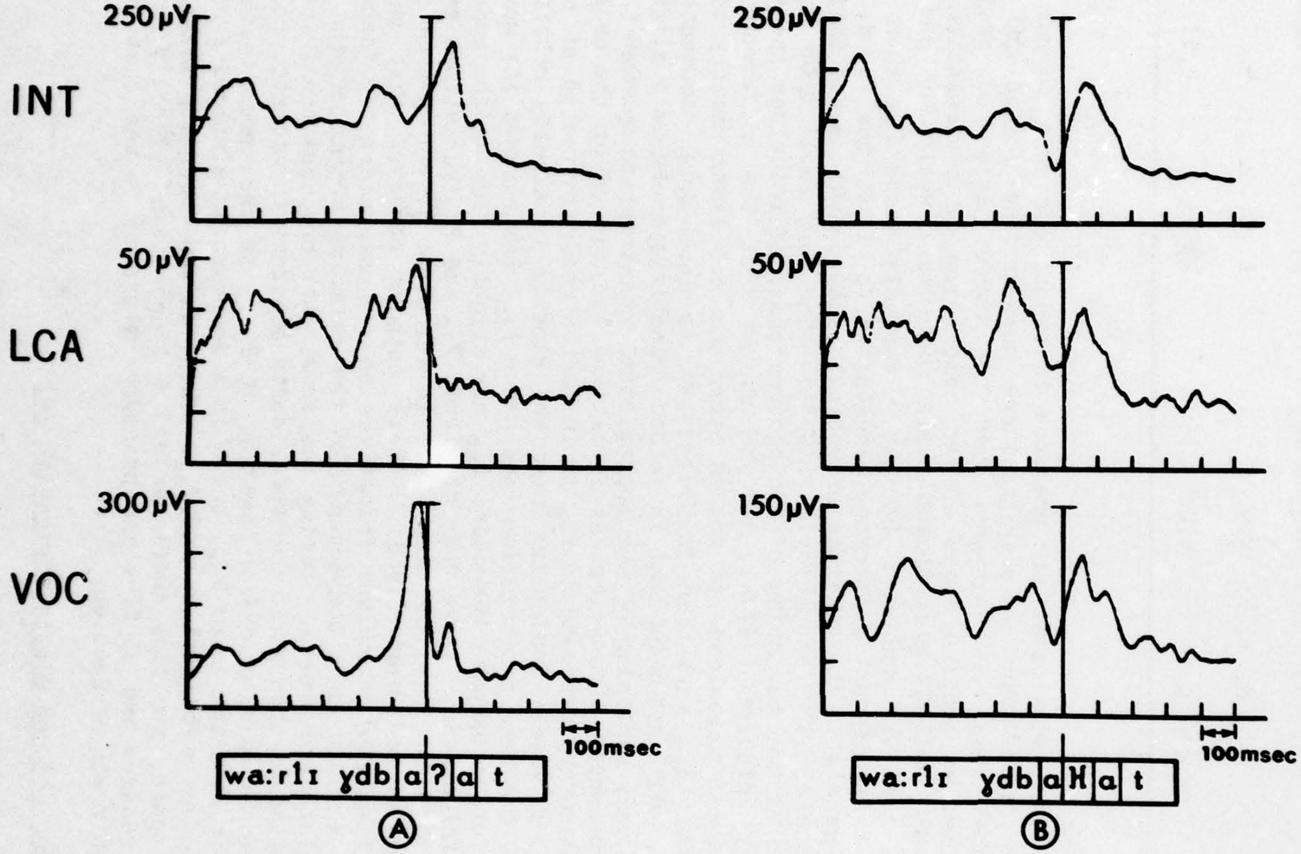


Figure 3: Averaged EMG data: comparison of [?] (A) and Dutch /H/ (B).

INT relaxation; in fact, the degree of INT activity is the same as for a voiced stop, so that we assume that the vocal folds are loosely adducted. Then follows a moment of strong medial compression of the folds by an increase of INT and LCA activity and by unusually strong VOC contraction. The next moment there is a sudden release of the glottal occlusion, brought about by an abrupt, large scale relaxation of VOC and LCA. Finally, the vocal folds are adjusted to their normal voicing position, apparently by the continued increase in INT activity and a momentary burst of VOC contraction. As far as the air pressure data are concerned, (not shown in Figure 3), there is of course no P_{io} increase for [ʔ]; a small decrease of P_{sg} is observed at the end of the preceding vowel. Our EMG data are in agreement with those reported by Hirose and Gay (1973) who studied "hard vocal attack," and by Fischer-Jørgensen and Hirose (1974b) dealing with Danish "stød." In these two studies, however, there is no indication of strong INT activity after the release of the glottal stop. The pattern of EMG activity for the production of the voiced glottal fricative /H/ (Figure 3B) shows partial suppression of INT activity and strongly reduced VOC and LCA activity. Since there is no oral constriction, there is no increase in P_{io}, and the separation of the vocal folds leads to a marked drop in P_{sg} of more than 1 cm aq. As can be seen in Table 2, the pattern of EMG activity for /H/ is the same as that for the velar fricative /γ/, which accords with its classification as a "glottal fricative" in the IPA chart.

Stops and Fricatives in Clusters

The Voiceless/Voiced Contrast in Clusters of Two Stops. In intervocalic position Dutch consonants can occur in clusters of two or more. Figure 4 illustrates the intervocalic contrast of /tp/ and /db/. Table 4, A and B, gives numerical values for the various physiological variables at selected points in time.

INT: At about 150 msec before the line-up point INT activity starts to decrease for the following /tp/ cluster, but much less so for following /db/. The lowest level of INT activity for /tp/ is 86 microvolts and 117 microvolts for /db/. There is a high peak of INT activity after /tp/ but not after /db/.

LCA: At about 175 msec before the line-up point there is a decrease in LCA activity for /tp/, resulting in a minimum value of 22 microvolts. For /db/ there is little or no relaxation in LCA, the lowest level of activity being 30 microvolts.

VOC: At about 50 msec before the line-up point the activity of VOC decreases for both /tp/ and /db/. The lowest level of activity of VOC is 59 microvolts for /pt/ and 64 microvolts for /db/. There is a higher peak of VOC after /pt/ than after /db/.

P_{io}: Intraoral air pressure rises to a maximum of 6.56 cm aq in /tp/ and 4.03 cm aq in /db/. The pressure curves also reflect the durational differences between the two clusters.

P_{sg}: Measured at the moment of maximum P_{io}, subglottal air pressure is at a level of 7.18 cm aq in /tp/ and at 7.39 cm aq in /db/. The ΔP value at

TABLE 4: Numerical values at selected point in time for EMG and air pressure signals in consonant clusters.

		ADDUCTOR MUSCLES			INTRAORAL AND SUBGLOTTAL AIR PRESSURE (in cm aq)			
		Minimum values (in microvolts) within 25 msec distance from line-up point.			Maximum P_{io} During Stricture	P_{sg} at moment of maximum P_{io}	$\Delta P(=P_{sg}-P_{io})$	P_{sg} Drop/Rise During Stricture
		INT	LCA	VOC				
A)	pt	82	20	60	6.70	7.05	0.35	0.07
	tp	86	22	59	6.56	7.18	0.62	0.03
	kp	84	23	66	6.50	7.03	0.53	0.14
Average		84	22	62	6.58	7.08	0.50	0.08
B)	bd	120	32	70	4.19	7.46	3.27	/0.01
	db	117	29	63	4.03	7.39	3.36	/0.08
	gb	112	29	73	4.15	6.79	2.63	/0.04
Average		116	30	69	4.12	7.21	3.08	/0.04
C)	fp	52	24	53	6.27	6.77	0.50	0.95/0.29
	sp	68	21	58	6.03	6.30	0.27	0.84/0.23
	xp	56	25	64	5.72	6.18	0.46	1.49/0.28
Average		59	23	58	6.00	6.41	0.41	1.09/0.26
		* +	* +	* +				
D)	vb	88/128	27/30	64/84	5.08	7.19	2.11	0.01
	zb	101/116	24/26	69/77	4.96	6.99	2.03	0.10
	Yb	91/113	25/29	69/83	4.74	6.83	2.09	0.09
Average		93/119	25/29	69/83	4.92	7.00	2.07	0.06

*measured at 40 msec before line-up
+measured at line-up

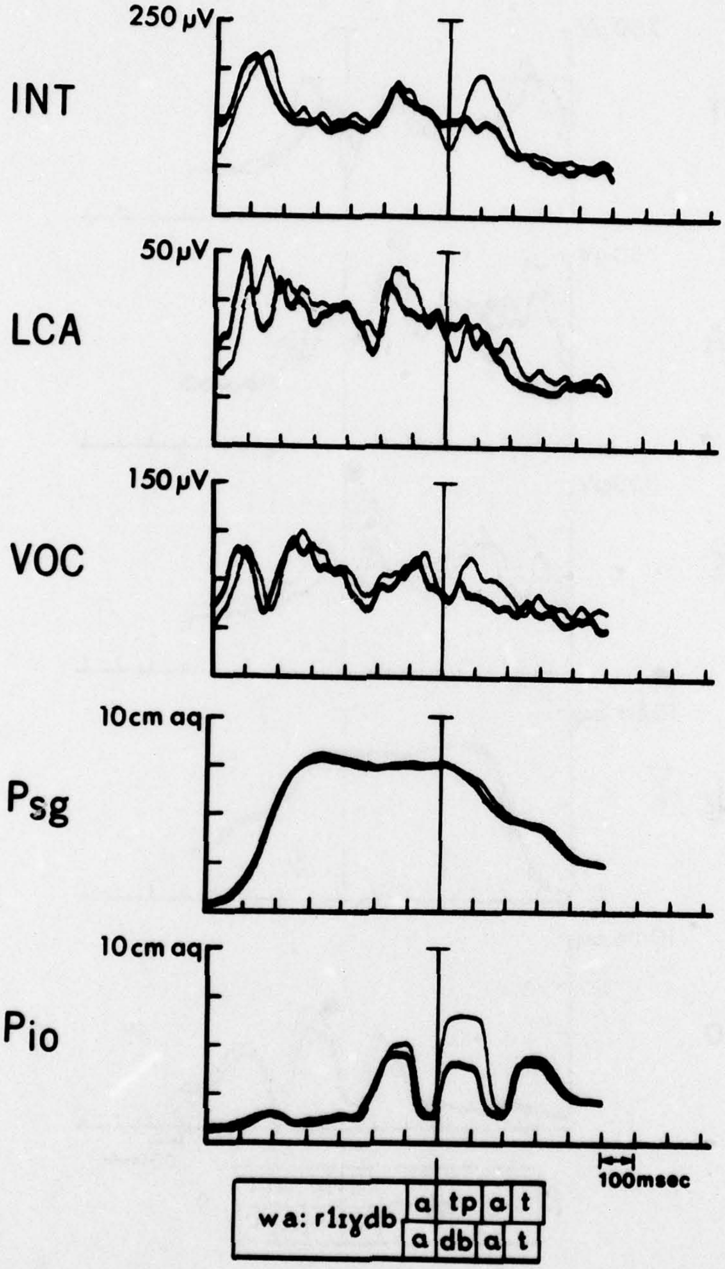


Figure 4: Averaged EMG and pressure data for Dutch sequences /tp/ (thin line) and /db/ (thick line).

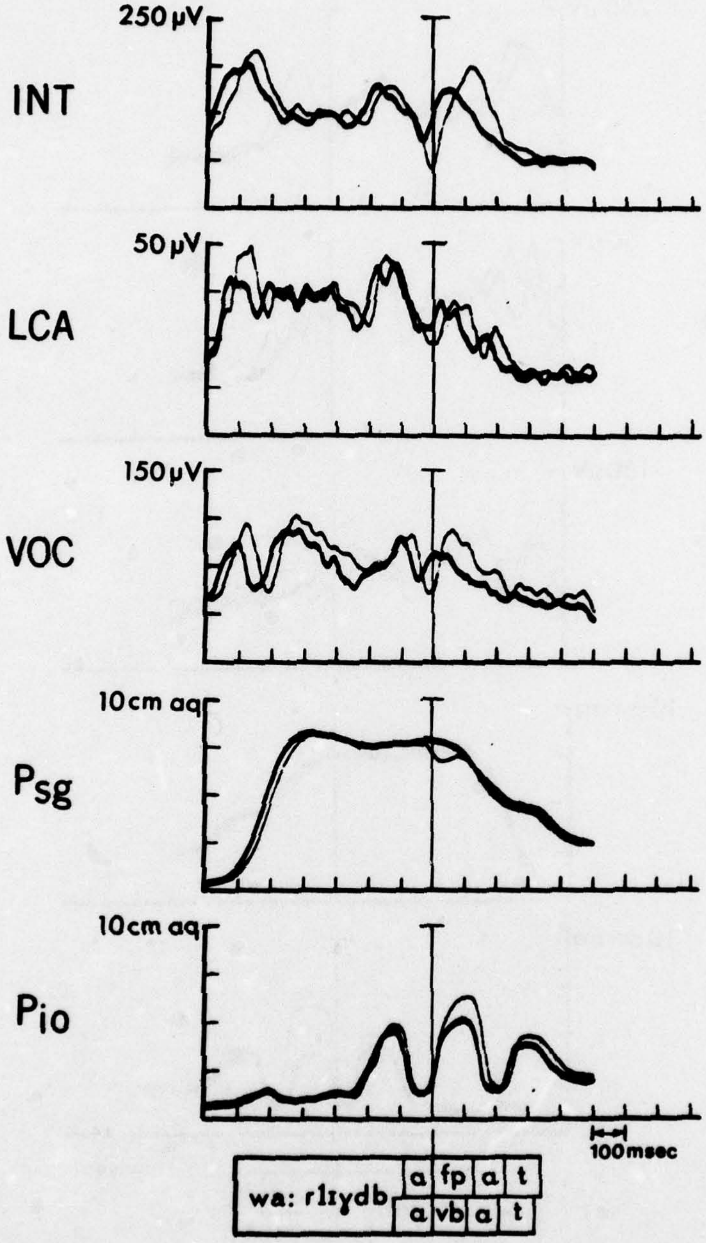


Figure 5: Averaged EMG and pressure data for Dutch sequences /fp/ (thin line) and /v/ (thick line).

that moment is 0.62 cm aq for /tp/ and 3.36 cm aq for /db/. In /tp/ ΔP gets down to a value of 0.24 cm aq at the end of the /p/-closure. It can be seen in Table 4, A and B, that the same tendencies hold for the comparison of /pt-bd/ and /kp - gb/. The minimum ΔP value during /pt/ and /kp/ (not given in the Table) is 0.24 and 0.12 cm aq, respectively.

The Voiceless/Voiced Contrast in Stop + Fricative Clusters. The voicing distinction in clusters consisting of a stop followed by a fricative is illustrated by the comparison of /fp/ and /vb/ in Figure 5.

INT: At about 150 msec before the line-up point, INT activity decreases. This reduction of activity is much more pronounced for /fp/ than for /vb/, the lowest level of INT activity being 52 and 88 microvolts, respectively. The resumption of INT activity starts earlier in the case of /vb/, so that at the line-up point the level of activity is already 128 microvolts. This timing difference is not merely the result of the shorter duration of /vb/; it also indicates that the glottis is opened for /v/ and closed for /b/. Indeed, comparing INT activity in the sequence /vb/ and /vɑ/, we have found very little difference. In the case of /fp/ the later resumption of INT activity may suggest that the glottis is opened for /f/ and kept open for the following /p/. However, comparing INT activity in the sequences /pɑ/, /fɑ/ and /fp/, we have found that the resumption of INT activity starts at the same moment in the three cases, but that it builds up more slowly in the case of /fp/. This fact might be taken as an indication that in the /fp/ cluster the glottis is being partially closed for /p/. Following /fp/ there is a higher level of INT activity than after /vb/.

LCA: At about 125 msec before the line-up point, LCA activity starts to decrease. There is more LCA relaxation for /fp/ than for /vb/, the minimum level of activity being 24 and 27 microvolts, respectively. The resumption of LCA activity starts later in the case of /fp/, and develops more slowly than after a single /f/.

VOC: At about 100 msec before the line-up point, VOC activity starts to decrease. The reduction of VOC activity reaches a minimum level of 53 microvolts in the case of /fp/ and of 64 microvolts in the case of /vb/. VOC activity resumes later in the former case than in the latter.

P_{io}: The maximum value of P_{io} during /fp/ is 6.27 cm aq; during /vb/ it is 5.08 cm aq. The pressure curves reflect the durational differences between the two clusters.

P_{sg}: Measured at the moment of maximum P_{io}, P_{sg} attains a value of 6.77 cm aq for /fp/ and 7.19 cm aq for /vb/. With /fp/ there is a drop of 0.95 cm aq from about 50 msec before the line-up point to 30 msec after it. This drop is followed by a slight increase (0.29 cm aq). At the moment of maximum P_{io}, the value of ΔP is 0.50 cm aq in /fp/ and 2.11 cm aq in /vb/. The lowest ΔP value during /fp/ is 0.38 cm aq. It can be seen in Table 4, C and D, that the tendencies observed in the /fp - vb/ contrast are also to be found in the oppositions /sp - zb/ and /xp - γb/.

Discussion

Generally speaking, the differences and correspondences between single stops and fricatives in their voiceless and voiced conditions are reproduced in the comparison of the same speech sounds in clusters of two.

Table 5 presents a more detailed comparison of the EMG data that were sampled for single consonants and for the same consonants occurring in clusters. It can be observed that there is more relaxation in the adductor muscles for clusters of two voiceless stops than for a single voiceless stop in intervocalic position. Specifically, there is evidently more LCA relaxation in the former case than in the latter. On the other hand, the comparison of single voiced stops and the same stops in clusters of two does not reveal any major differences. As far as the voiceless fricatives are concerned, the degree of adductor muscle relaxation is the same when they occur singly and when they are first members of a fricative + stop combination. However, if a voiced fricative occurs in a fricative + stop cluster, there is less adductor muscle relaxation than for the same fricative occurring singly. The level of adductor muscle activity is approximately 15 percent higher in the voiced fricative in the cluster. We have pointed out in the first part of this paper that, in the case of single stops and fricatives, the voicing contrast is not systematically reflected in the degree of LCA and VOC activity, which mainly correlates with the stop/fricative distinction. In a stop + stop cluster, however, the pattern of LCA activity correlates systematically with the voicing distinction, and in a fricative + stop combination the voicing contrast is reflected in both LCA and VOC activity. Therefore it may be concluded that all three adductor muscles, and specifically INT, simultaneously reflect both the voiceless/voiced and the stop/fricative contrast in the particular clusters under investigation.

TABLE 5: Comparison of minima of activity in the adductor muscles around the line-up point for single consonants and clusters of two.

	INT	LCA	VOC
p, t, k	93	28	74
pt, tp, kp	84	22	62
b, d, g	116	29	71
bd, db, gb	116	30	69
f, s, x	57	23	57
p	95	28	74
fp, sp, xp	59	23	58
v, z	81	23	60
b	121	29	64
vb, zb, γb	93/119(*)	25/29	69/85

(*)The values to the left of the slash correspond to the fricative, those to the right represent /b/.

It is worth noticing that the data on the production of consonant clusters are even more at variance with the claims of Halle and Stevens (1967, 1971) than those on single consonants. For one thing, there is no indication of overt changes in intrinsic laryngeal muscle activity to facilitate the continuation of glottal pulsing in voiced stop + stop clusters, even with increased closure duration. For another, there is more VOC relaxation in the voiceless stop + stop clusters than in single stops, indicating that no stiffening of the vocal folds is required to inhibit their vibration.

CONCLUSION

Our experiment has confirmed that at the level of laryngeal adjustment systematic differences can be found that correlate with the voiced/voiceless and the stop/fricative contrasts among consonants. Four groups of Dutch consonants can be distinguished along these two dimensions. In order to separate these four classes in terms of articulatory differences, it is of primary importance to specify for each its particular degree of glottal width and of supraglottal constriction. The degree of longitudinal vocal fold stiffness appears not to be a crucial factor in the voicing distinction. Our data also suggest that the degree of glottal aperture as well as the shape of the glottis may vary as a function of the combined difference in voicing and in manner of consonant production. This is to say that static glottal width and glottal shape not only correlate with the positive or negative specification of the phonetic feature [voice], but with that of the feature [sonorant] as well.

REFERENCES

Bell-Berti, F. (1975) Control of pharyngeal cavity size for English voiced and voiceless stops. Journal of the Acoustical Society of America 57, 456-467.

Fischer-Jørgensen, E. and H. Hirose. (1974a) A preliminary electromyographic study of labial and laryngeal muscles in Danish stop consonant production. Haskins Laboratories Status Report on Speech Research SR-39/40, 231-254.

Fischer-Jørgensen, E. and H. Hirose. (1974b) A note on laryngeal activity in the Danish "stød." Haskins Laboratories Status Report on Speech Research SR-39/40, 255-259.

Halle, M. and K. N. Stevens. (1967) On the mechanism of glottal vibration for vowels and consonants. Quarterly Progress Report (M.I.T. Research Laboratory of Electronics) 85, 267-271.

Halle, M. and K. N. Stevens. (1971) A note on laryngeal features. Quarterly Progress Report (M.I.T. Research Laboratory of Electronics) 101, 198-213.

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

Hirose, H. (1971b) An electromyographic study of laryngeal adjustments during speech articulation: A preliminary report. Haskins Laboratories Status Report on Speech Research SR-25/26, 107-116.

Hirose, H. and T. Gay. (1972) The activity of the intrinsic laryngeal muscles in voicing control. Phonetica 25, 104-164.

Hirose, H. and T. Gay. (1973) Laryngeal control in vocal attack. Folia Phoniatrica 25, 203-213.

Hirose, H., L. Lisker and A. S. Abramson. (1972) Physiological aspects of certain laryngeal features in stop production. *Haskins Laboratories Status Report on Speech Research SR-31/32*, 183-191.

Kewley-Port, D. (1973) Computer processing of EMG signals at Haskins Laboratories. *Haskins Laboratories Status Report on Speech Research SR-33*, 173-183.

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

Klatt, D. H., K. N. Stevens and J. Mead. (1968) Studies of articulatory activity and airflow during speech. In *Sound Production in Man*, ed. by M. Krauss. (New York: Annals of the New York Academy of Sciences) *155*, 42-55.

Lindqvist, J. (1972) Laryngeal articulation studied on Swedish subjects. *Quarterly Progress and Status Report* (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm) *STL-QPSR 2-3/1972*, 10-27.

Lisker, L. (1970) Supraglottal air pressure in the production of English stops. *Language and Speech 13*, 215-230.

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

Löfqvist, A. (1974a) Subglottal pressure during stop production. 4th Phonetics Symposium, University of Essex, Colchester, 4-6 January 1974.

Löfqvist, A. (1974b) Variations in subglottal pressure during stop production. Speech Communication Seminar, Stockholm, 1-3 August 1974.

McGlone, R. E. and T. Shipp. (1971) Comparison of subglottal air pressures associated with /p/ and /b/. *Journal of the Acoustical Society of America 51*, 664-665.

Netsell, R. (1969) Subglottal and intraoral air pressures during the intervocalic contrast of /t/ and /d/. *Phonetica 20*, 68-73.

Ohala, J. J. (1974) A mathematical model of speech aerodynamics. *Speech Communication Seminar*, Stockholm, 1-3 August, 1974.

Ohala, M. and J. Ohala. (1972) The problem of aspiration in Hindi phonetics. *Annual Bulletin*, Research Institute of Logopedics and Phoniatrics. (Tokyo: University of Tokyo) *6*, 39-46.

Perkell, J. (1969) *Physiology of Speech Production*. (Cambridge: M.I.T. Press).

Port, D. (1971) The EMG data system. *Haskins Laboratories Status Report on Speech Research SR-25/26*, 67-72.

Prosek, R. A. and A. S. House. (1975) Intraoral air pressure as a feedback cue in consonant production. *Journal of Speech and Hearing Research 18*, 133-147.

Sawashima, M., A. S. Abramson, F. S. Cooper and L. Lisker. (1970) Observing laryngeal adjustments during running speech by use of a fiberoptics system. *Phonetica 22*, 193-201.

Slis, I. H. (1970) Articulatory measurements on voiced, voiceless and nasal consonants. *Phonetica 21*, 193-210.

Slis, I. H. and A. Cohen. (1969) On the complex regulating the voiced-voiceless distinction, part I and II. *Language and Speech 12*, 80-102 and 137-155.

Stevens, K. N. (1977) Physics of laryngeal behavior and larynx modes. *Phonetica 34*, 264-279.

Insufficiency of the Target for Vowel Perception[*]

Donald Shankweiler,[+] Robert Verbrugge[+] and Michael Studdert-Kennedy[++]

ABSTRACT

Listening tests were made by excerpting a single pitch pulse from each of nine vowels and iterating it sufficiently to produce sets of pseudovowels matched in duration to the parent syllables. Listeners' judgments contained nearly twice as many errors for iterated pseudovowels as for the unedited natural versions. An additional experiment with OVE-synthesized vowels yielded error rates comparable to those for iterated pseudovowels, but not to their natural counterparts. It is apparent that natural vowels contain sources of linguistic information not captured by sustained target formant frequency values. A final study created stylized CVC syllables by adding linear formant transitions to the set of OVE-synthesized steady-state vowels. Listeners showed a significant gain in vowel identification for OVE-synthesized CVC vowels in comparison to their #V# counterparts. The results suggest that time-varying spectral properties of more than one kind may contribute to vowel recognition.

It is customary to describe the vowels as points in an acoustic space defined by frequencies of the first two formants. Indeed, it has been assumed that the essential physical specification of a vowel can be stated in terms of its acoustic spectrum measured over as brief an interval as a single pitch period. To attempt to obtain perceptual data on such unnaturally brief stimuli presents psychophysical problems of signal generation and reception that are irrelevant to the central issue of the cues for vowel quality. A fairer test of the acoustic target idea can be made by excerpting a single pitch pulse from a naturally produced sustained vowel and iterating it for a sufficient number of repetitions to produce a stimulus of natural vowel length. Such a procedure would eliminate all cues other than spectral cues for vowel identification, and, at the same time, would yield a signal of appropriate duration.

[*] This paper was presented at the 95th meeting of the Acoustical Society of America, Providence, Rhode Island, 16-19 May, 1978.

[+] Also University of Connecticut, Storrs.

[++] Also Queens College and the Graduate Center, City University of New York.

Acknowledgment: This work was aided by NIH Grant HD01994 to Haskins Laboratorfes.

Recently, Bond (1976) obtained identification data for pseudovowels created by iteration of a central pulse excerpted from a token of each of a speaker's vowels produced in an hVd context. Her subjects displayed an average error rate of 47 percent. Such strikingly poor perceptual performance requires explanation. Bond examined the spectra of her stimuli and concluded that the first and second formant frequencies were in the expected range, an unsettling finding if we are to maintain that the stationary spectrum at target frequencies is sufficient for accurate identification of the vowels. It seemed worthwhile, therefore, to make a deliberate comparison between the intelligibility of truly steady-state vowels and vowels in which some of the natural sources of variability are present.

Five tokens of each of nine vowels were produced in random order by two speakers. Iterated pseudovowels were constructed from the vowel tokens of each speaker. Speaker 1 produced the vowels in pVp context; Speaker 2 produced them in isolation. The recorded utterances were sampled at a rate of 8 kHz and digitized on the Haskins Laboratories' pulse code modulation (PCM) system. The iterated stimuli were made by excerpting and repeating a pitch period from the syllable center. Care was taken to select a pulse for which a single cycle fit neatly between zero crossings. Each pulse was iterated sufficiently to produce a stimulus approximately matched in duration to the parent syllable token. The control tests consisted of randomized sets of natural isolated vowels produced by each speaker. For Speaker 2 these were the identical tokens from which the iterated set was made.

Figure 1 shows spectral cross sections for two vowels produced by this speaker. Spectral cross sections of the parent token and the iterated version are shown for each of two vowels. The solid contour is a section through the midpoint of the natural vowel; the dotted contour is a section made after iteration of a center pulse from that token. The sections on the left are for the vowel /I/. As is readily apparent, the fit is very close. Sections of the vowel /u/ are shown on the right. Here the fit between iterated vowel and parent vowel is less close. These tokens were chosen for display because they represent instances of the best and worst agreement between iterated stimulus and parent vowel. In neither case, however, do we see evidence that the formant values were substantially altered by the iteration procedure. It is apparent that the iterated vowels contain the spectral information that is present in the originals.

Figure 2 shows a vowel-by-vowel comparison of errors made by 21 subjects when identifying the stimuli derived from Speaker 2. The figure shows that for seven of the nine vowels more errors occurred in identifying the iterated pseudovowels than natural isolated vowels produced by the same speaker. The pair of histograms on the far right shows the mean error rates for each type of item. Errors for the pseudovowels averaged 46 percent and for the natural vowels, 27 percent. These data show that iterated vowel stimuli are substantially less identifiable even than isolated vowels.

It is important to discover why iterated vowels are so much less intelligible than isolated vowels from which they were derived. The purpose of creating the pseudovowels was to produce stimuli in which the only cues for vowel quality are the sustained steady-state formants. Figure 1 showed that the process of excerpting and iterating a single cycle from a vowel to form a

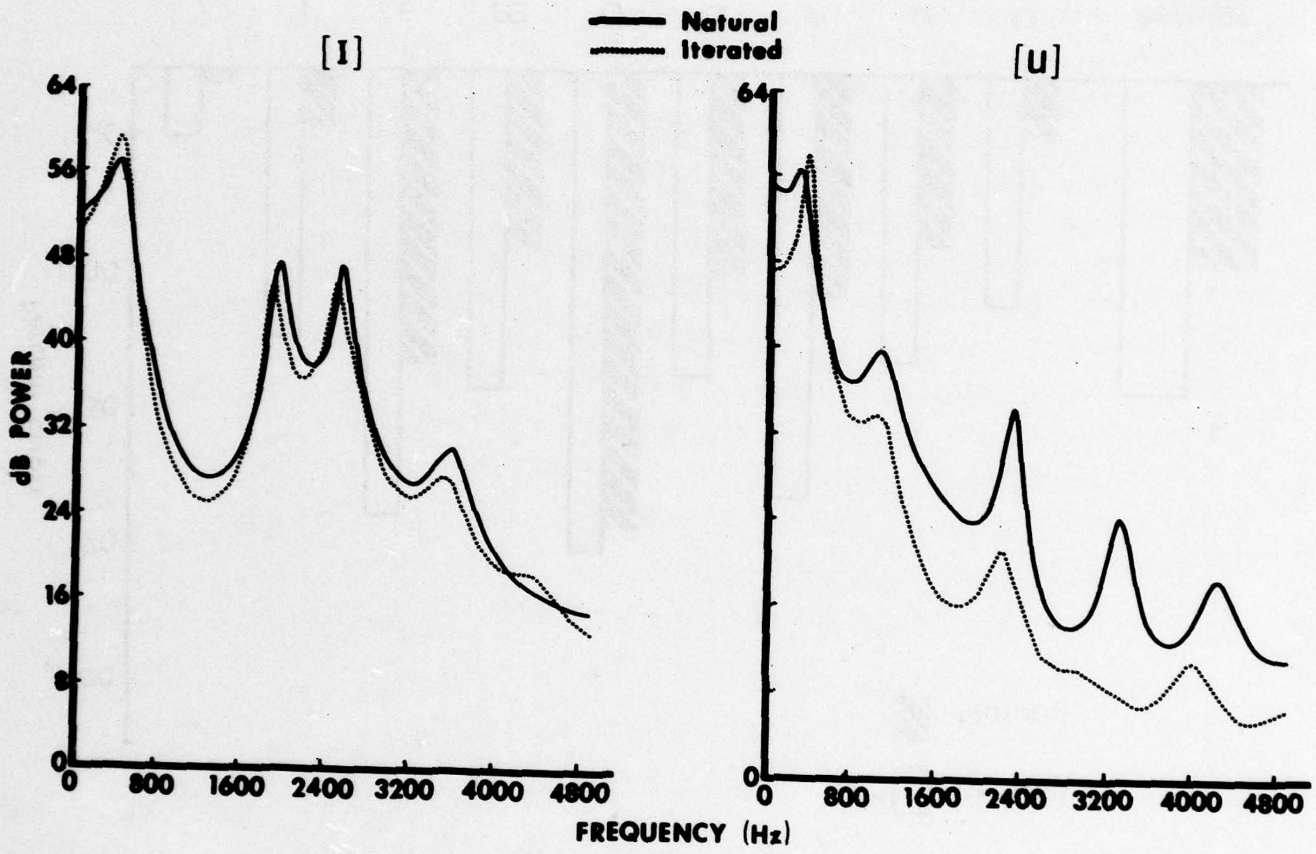


Figure 1: Spectral cross section of natural (parent) vowel superimposed on section taken from iterated pseudovowel. Instance of best fit (left figure) and poorest fit (right figure). Stimuli by Speaker 2.

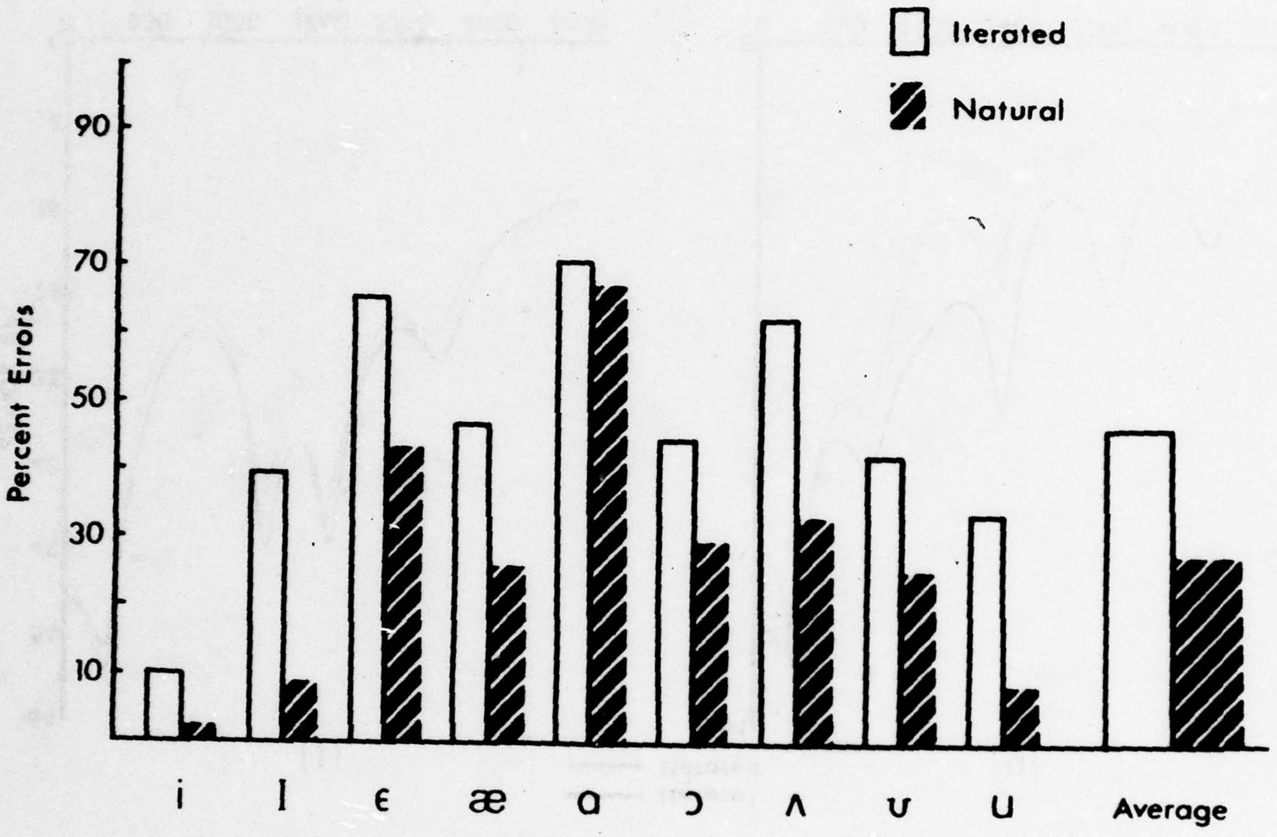


Figure 2: Mean errors in identification of nine natural isolated vowels by Speaker 2 and pseudovowels created from iteration of pulse excerpted from each of these.

continuous stimulus does not destroy the formant structure of the original vowel. It is nonetheless possible that such stimuli contain other types of distortion that adversely affect perception.

To bypass that issue we created a set of synthetic steady-state vowels, similar in structure to the iterated pulse pseudovowels but free from the kinds of distortion that might accompany the iteration process. Figure 3 shows schematic spectrograms of three-formant patterns produced on the OVE III synthesizer. In the lower portion of the figure a steady-state vowel is displayed in which the formant frequencies correspond to average values for adult males in the Peterson and Barney (1952) sample for the vowel /ɑ/. The upper portion of the figure is a highly-stylized CVC syllable, [bɑb], created by appending symmetrical linear 30-msec transitions to the steady-state vowel shown below, to create a syllable of the same length. Five tokens of each of the nine vowels were made in this fashion and placed in random order. Figure 4 shows a vowel-by-vowel comparison of errors of identification for the steady-state isolated OVE vowels and the schematic OVE bVb syllables. For six of the vowels, error rates are higher for isolated vowels than for those same vowel nuclei flanked by simplified transitions. The average difference of 19 percent was significant. These data agree with our earlier findings (Strange, Verbrugge, Shankweiler and Edman, 1976; Shankweiler, Strange and Verbrugge, 1977) that vowels in CVC context are more accurately perceived than comparable vowels in isolation.

Figure 5 permits us to compare the results of the various listening tests. The middle group of bar graphs depicts the results for true steady-state vowels: that is, the iterated vowels and the synthetic OVE vowels. It is striking that the OVE vowels were as poorly identified as the iterated vowels from Speakers 1 and 2. All of these steady-state vowels showed an average of 45-50 percent errors. It is clear from this that the low identifiability of the iterated-pulse vowels cannot be attributed to artifacts introduced by the process of iteration, since no artifacts are present in the OVE vowels. Both types of steady-state vowels represent a strong test of the theory that a vowel's linguistic identity is determined primarily by cross-sectional formant values. It is certainly true that the steady-state vowels, whether synthetic or natural in origin, are unrepresentative in many respects: in onset and offset characteristics, in having flat pitch and amplitude contours. According to canonical vowel theory, however, these kinds of acoustic variations should not affect intelligibility as long as formant values are well defined. It is embarrassing for this theory that identification is worst where the formants are defined most clearly.

Error rates on the natural isolated vowels of Speakers 1 and 2 are shown in the left portion of the figure. While error rates for these isolated vowels are still high, they are substantially lower than for the steady-state vowels. This indicates that intelligibility is affected by characteristics other than target formant frequencies alone. The isolated vowels contain natural onset and offset characteristics, natural amplitude and pitch contours and diphthongization. These aspects of syllable dynamics apparently play a role in identification, and merit further study.

Listening data were also obtained for vowels spoken in stop-vowel-stop syllables by Speakers 1 and 2. The results are summarized in the right

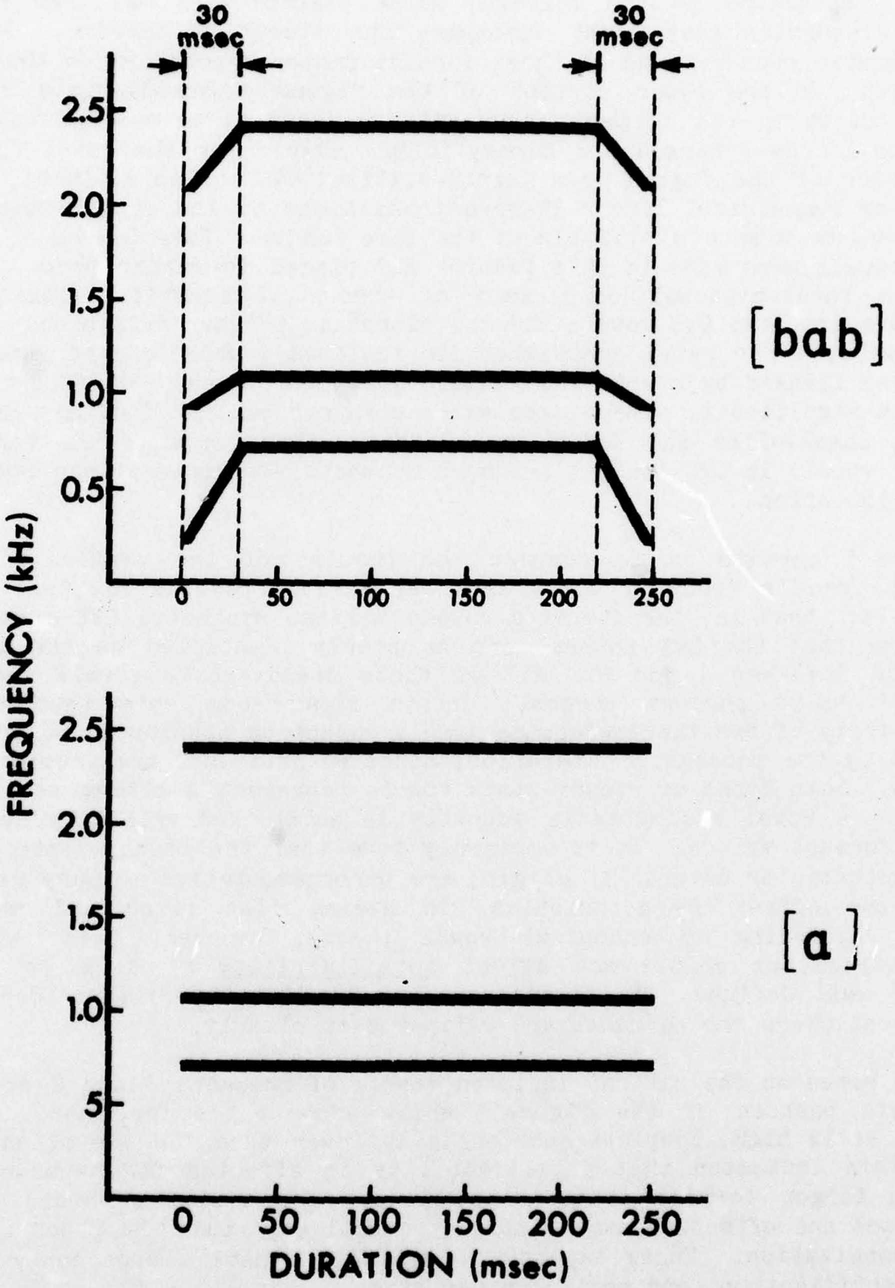


Figure 3: Spectra of synthetic OVE III steady-state vowel and stylized CVC syllable. Formant frequencies from Peterson and Barney (1952) average values for adult males.

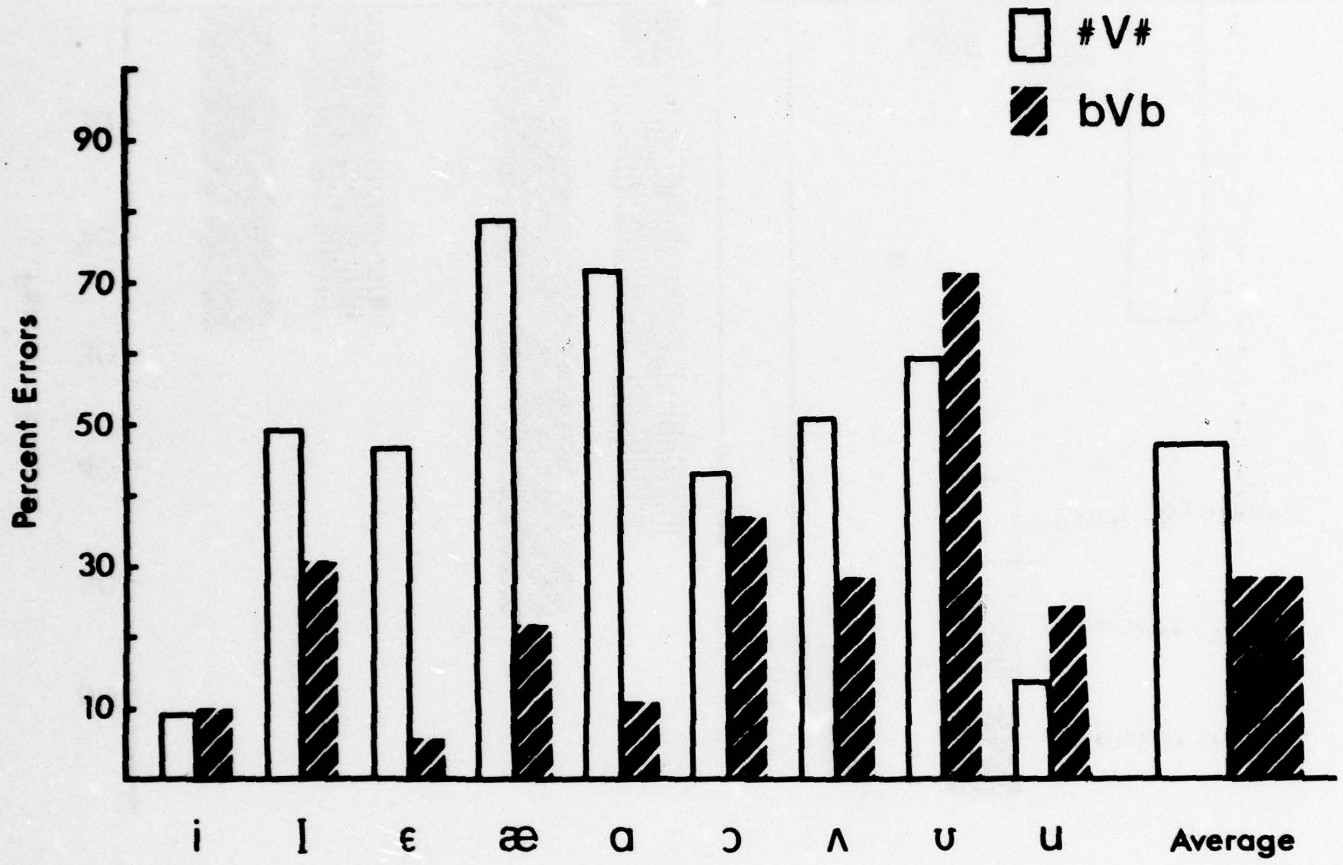


Figure 4: Mean errors in identification of synthetic isolated vowels (#V#) and bVb syllables.

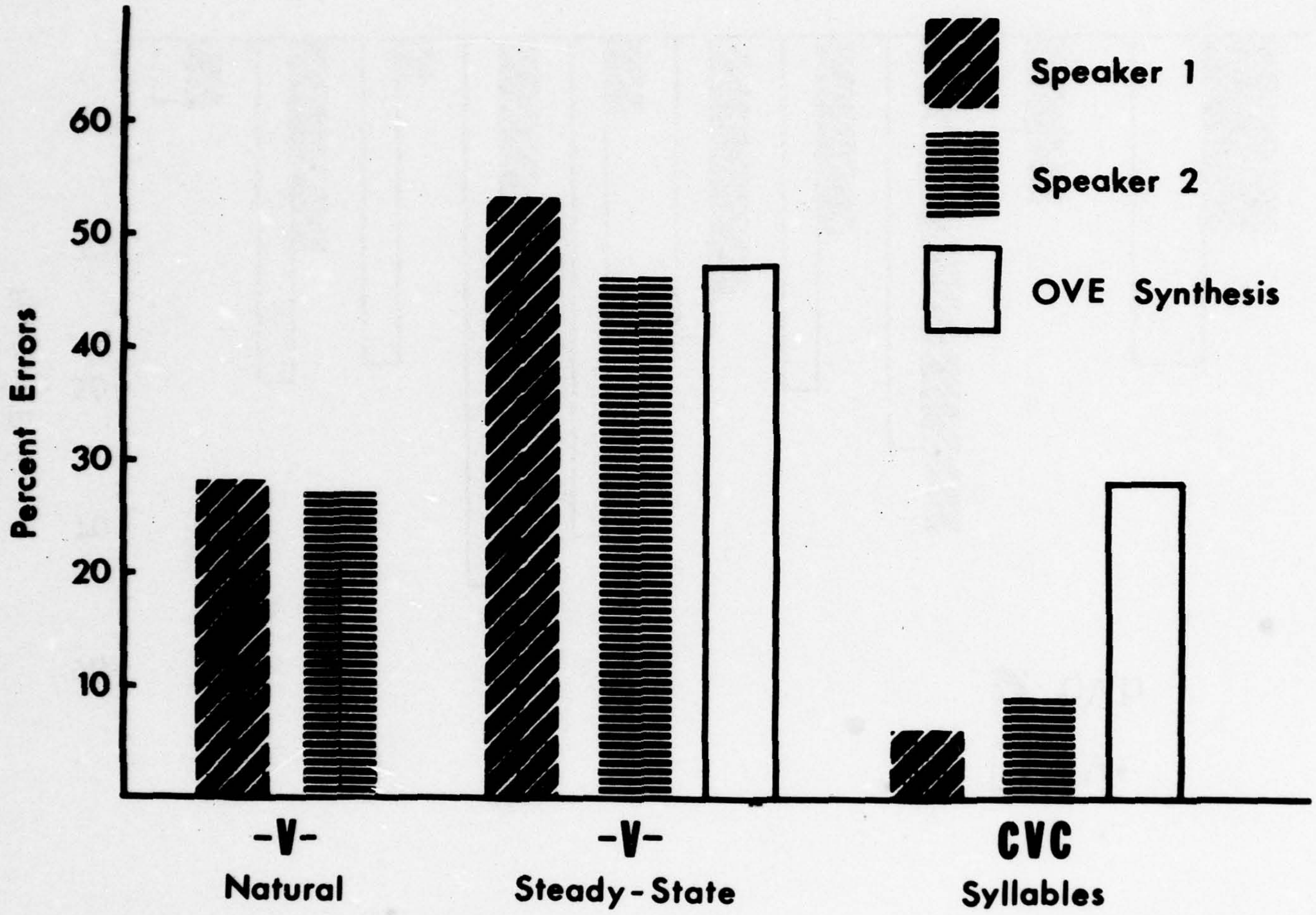


Figure 5: Errors averaged over nine vowels for isolated natural vowels (by Speakers 1 and 2), steady-state pseudovowels and synthetic steady-state vowels, and bVb syllables produced by Speaker 1, Speaker 2 and the OVE III synthesizer.

portion of the figure. The average error rates for vowels in these natural CVC syllables were 6 percent and 9 percent, respectively. These findings also are contrary to canonical vowel theory. *First*, in comparison with the results for the synthetic CVC syllables, it is clear that the dynamic structure of a natural CVC syllable, where no steady-state typically is present, is more informative than the structure of our OVE syllables, which were built around steady-state nuclei with highly stylized transitions. *Second*, in comparison with the results for natural *isolated* vowels, vowels in the natural CVC syllables showed significantly higher identifiability. It may be that the formant trajectories accompanying consonants do not simply facilitate the extraction of underlying vowel targets, but are in themselves carriers of distinctive information for vowel identity.

At all events, the results show that sustained formant frequencies, whether synthetic or natural in origin, are not sufficient to specify the set of vowels of English. The results thus complement a growing body of evidence (Lindblom and Studdert-Kennedy, 1967; Strange et al., 1976; Strange, Jenkins and Edman, 1977; Verbrugge and Shankweiler, 1977) indicating that time-varying spectral information, both within and beyond the syllable, may contribute to vowel perception.

REFERENCES

Bond, Z. S. (1976) Identification of vowels excerpted from /l/ and /r/ contexts. *Journal of the Acoustical Society of America 60*, 906-910.

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

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

Shankweiler, D., W. Strange and R. R. Verbrugge. (1977) Speech and the problem of perceptual constancy. In *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, ed. by R. Shaw and J. Bransford. (Hillsdale, N.J.: Lawrence Erlbaum Associates).

Strange, W., J. J. Jenkins and T. R. Edman. (1977) Identification of vowels in "vowel-less" syllables. *Journal of the Acoustical Society of America 61* U4(A).

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

Verbrugge, R. R. and D. Shankweiler. (1977) Prosodic information for vowel identity. *Journal of the Acoustical Society of America 61*, U3(A).

Syllable Timing and Vowel Perception*

Robert R. Verbrugge+ and David Isenberg

ABSTRACT

Consonantal environment may aid in specifying vowel identity by supplying critical information about timing. Several vowel pairs in American English are distinguished by temporal as well as spectral variables, and these temporal differentiae vary with articulatory rate. Two studies were designed to explore the following paradox: when consonantal formant transitions are introduced into a steady-state vowel, holding syllable duration constant, a response shift is observed toward longer vowel alternatives, even though steady-state duration has been reduced. The first study verified this finding for the vowel pair /ɛ/-/æ/ in comparisons of #V# and bVb continua. Pairs of continua were defined separately by F_1 variation and by duration variation, and both continuum types evidenced the paradox to some degree. A second study varied the rate of symmetric consonantal transitions in F_1-varying CVC continua (V = /ɛ,æ/, C = /#, b, w/) in order to test whether transition rate might specify an articulatory rate that effectively scales vowel duration. Vowel responses did not vary monotonically with either transition rate or steady-state duration, but interacted with the perceived identity of the initial consonant. Listeners' judgments may demonstrate a sensitivity to constraints on the relative timing of consonantal and vocalic gestures.

There are several possible explanations for the greater identifiability of vowels in a consonantal environment (see Strange, Verbrugge, Shankweiler and Edman, 1976). One possibility, which is not often considered, is that consonantal environments supply critical information about the timing of the gestures that comprise a syllable. It has been demonstrated that variation in

*This paper was presented at the 95th meeting of the Acoustical Society of America, Providence, R.I., 15-19 May 1978.

+Also University of Connecticut, Storrs.

Acknowledgment: The research was supported by a grant to Haskins Laboratories from the National Institute of Child Health and Human Development (NICHD Grant HD-01994; BRSG Grant RR-05596), by a fellowship to R. R. Verbrugge from the University of Michigan Society of Fellows, and by a fellowship to D. Isenberg from the National Institutes of Health (NS05493-01).

[HASKINS LABORATORIES: Status Report on Speech Research SR-55/56 (1978)]

the rate and rhythm of a carrier phrase can alter the identity of a vowel in an embedded syllable [Ainsworth, (1974); Nooteboom, (1974); Verbrugge, Strange, Shankweiler and Edman, (1976); Verbrugge and Shankweiler, (1977)]. It is possible that single syllables, in themselves, contain information about the rate and rhythm of articulation, and that this information (like that carried over longer contexts) can alter the identity of the vowel perceived.

In a study reported by Shankweiler, Verbrugge and Studdert-Kennedy (1978), synthetic steady-state vowels were compared with synthetic b-vowel-b syllables of the same acoustic duration. The major finding of that study was an overall decrease in perceptual errors for the medial vowels in contrast to the steady-state vowels. A second and more subtle finding of that study was a systematic exception to the overall reduction in errors: short vowels were misidentified as long vowels more often in the bVb syllables than in the steady-state vowels. This was a surprising result: the introduction of consonantal transitions shortened the duration of the steady-state, but lengthened the perceived vowel.

Several studies have shown that some fraction of consonantal formant transitions or frication contributes to the perceived duration of a vowel in the same syllable--that is, vowel duration is not to be identified solely with duration of a steady-state [for example, Raphael, Dorman and Liberman, (1975); Mermelstein, Liberman and Fowler, (1977)]. Estimates of the fraction vary widely, showing dependencies on a number of factors, including the manner class, place and syllable position of the consonants employed. However, if we seek a comparable interpretation of the results of Shankweiler, et al. (1978), we are forced to an unusual and troubling conclusion: a fraction greater than 100 percent of the transition is treated as vowel duration. The pattern with transitions is heard as containing a longer vowel than the pattern without transitions, even though the total acoustic durations of the two patterns are the same.

Before lavishing explanations on this paradox, we decided it would be best to verify it in a design that would allow the response shift to be assessed more parametrically. Using an OVE III synthesizer, we designed continua to study the short-long vowel pair /ɛ/ and /æ/. Typical formant patterns are illustrated in Figure 1.

In one pair of continua, the contrast between /ɛ/ and /æ/ was achieved by varying the steady-state frequency of the first formant (F_1) through nine steps of approximately 15 Hz, over a range from 626 to 744 Hz. This is indicated by the vertical arrows in Figure 1. One continuum consisted of bVb syllables (illustrated on the top) and the second continuum consisted of steady-state vowels (at the bottom). All of the patterns in these continua were 235 msec in total duration.

In a second pair of continua, the contrast between /ɛ/ and /æ/ was achieved by varying the duration of the syllables. For these two series, a fixed, intermediate value of F_1 (702 Hz) was used for all patterns. The duration-varying patterns ranged in 20-msec steps from 175 to 315 msec in total duration, as indicated by the horizontal arrows in Figure 1. The F_1-

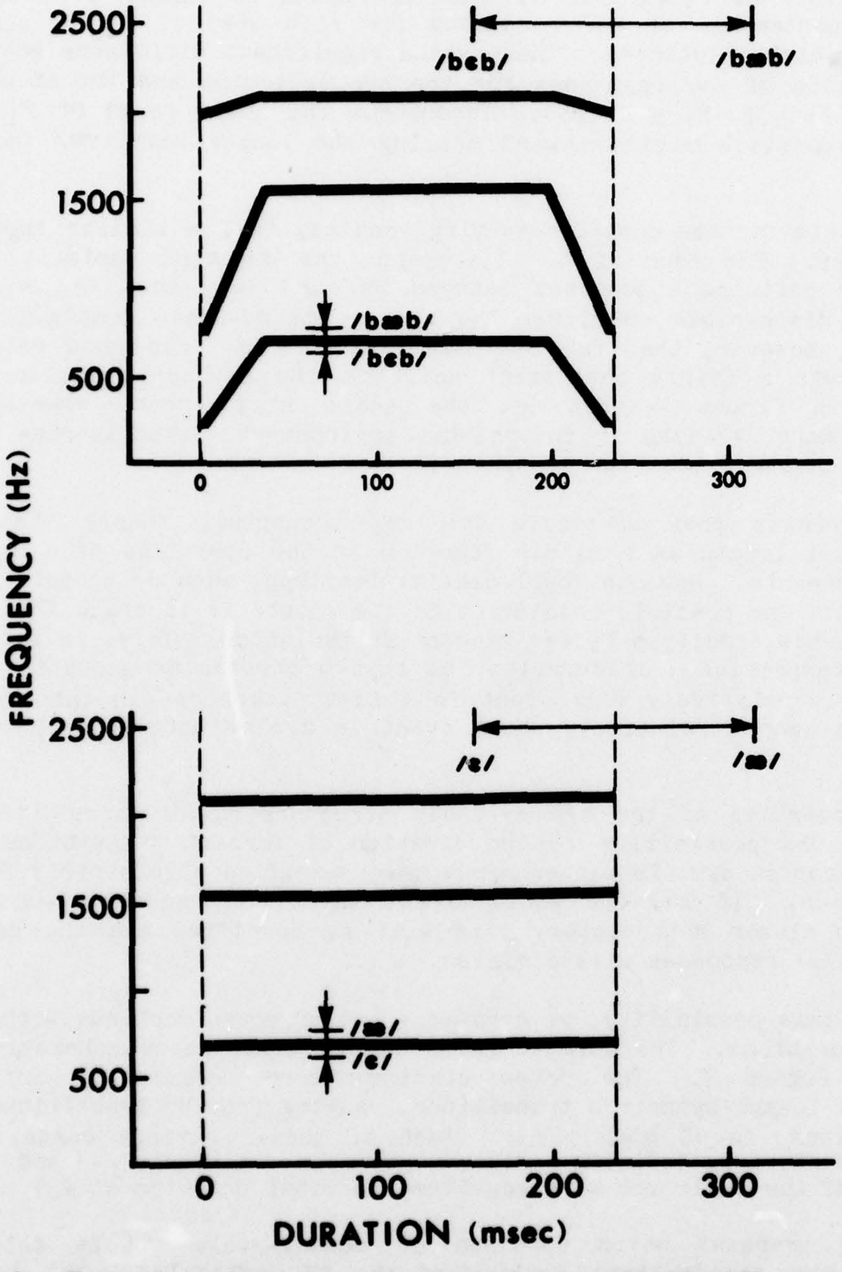


Figure 1: Formant control parameters used in synthetic bVb continua (top) and isolated vowel continua (bottom). Vertical arrows indicate range of first-formant variation (F_1-varying continua), and horizontal arrows indicate range of variation in total syllable duration (duration-varying continua).

varying and duration-varying continua allowed two independent tests of the paradoxical influence of consonantal environment on vowel perception. For each continuum type, the paradox would be verified if there were a shift toward the longer vowel alternative (/æ/) in the consonantal environment.

Results for the F_1-varying continua are shown in Figure 2. The curves present the percentage of /æ/ responses for each step on the F_1 continuum, averaged over eight listeners. There was a significant difference between the total proportion of /æ/ responses for the bVb continuum and the steady-state continuum ($\underline{t}(7) = 2.18$, $\underline{p} < .05$). Throughout the lower range of F_1 values, there was a consistent bias toward hearing the longer vowel /æ/ in the bVb syllables.

The results for the duration-varying continua tell a similar though less coherent story. For four of our listeners, the duration variation was not successful in defining a contrast between /ɛ/ and /æ/, that is, no phoneme boundary was discernible on either the bVb or steady-state continuum. It is worth noting, however, that for the four listeners who did show categorization, there was a fairly consistent shift in the boundary similar to that illustrated in Figure 2--that is, the short steady-state vowels became consistently more /æ/-like in consonantal environment. (The average boundary shift was 52 msec for the four listeners.)

These results pose a puzzle for any perceptual theory that treats perceived vowel length as a simple function of the durations of a syllable's acoustic components. How can vowel quality lengthen, when no acoustic measure is lengthened? One possible resolution to the puzzle is to argue that the bVb syllables somehow specify a faster rate of articulation. This, in turn, could scale the interpretation of duration, so that a particular acoustic duration would specify a relatively long event in a fast utterance (in this case, the bVb syllables) and a relatively short event in a slow utterance (the steady-state vowels).

What properties of the bVb syllable structure might carry information about rate? One possibility is the duration of formant transitions--the 35 msec transitions in our bVb patterns may have specified a relatively fast rate of articulation. If so, one would expect that, as transition duration is lengthened, a slower articulatory rate will be specified and the degree of shift toward /æ/ responses will diminish.

To test this possibility, we created a set of seven continua with varying transition durations. The formant patterns are illustrated schematically at the top of Figure 3. The seven continua were defined by introducing progressively longer symmetric transitions, ranging from no transitions to 60-msec transitions, in 10 msec steps. Each of these continua contained nine steps of F_1 variation, defining a contrast between the vowels /ɛ/ and /æ/. As before, all of the syllables were equalized in total duration at 235 msec.

We also prepared seven continua of steady-state vowels matched in duration to the steady-state regions of the CVC syllables. The lower two patterns in Figure 3 illustrate one of the CVC syllables (in this case with

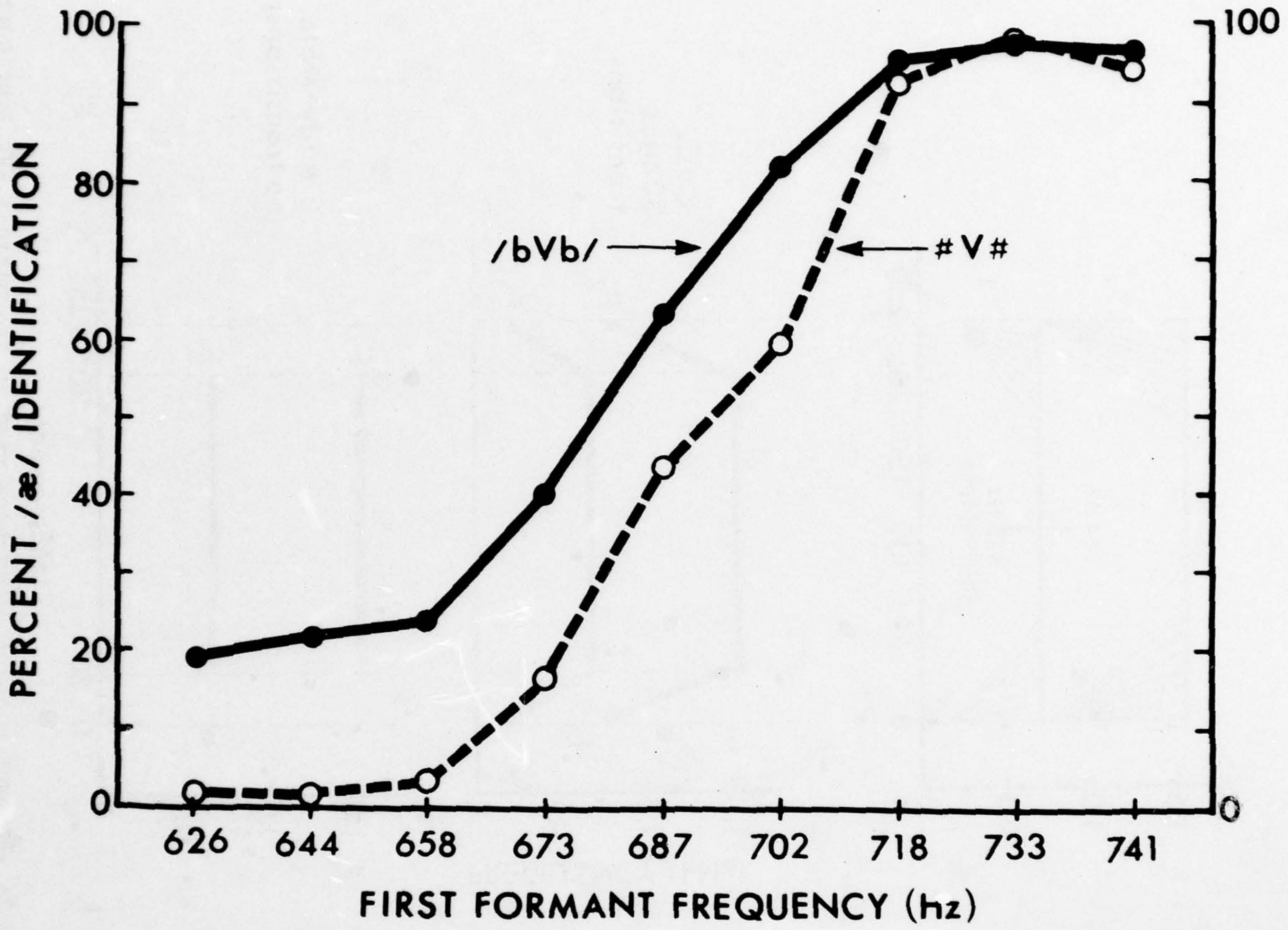


Figure 2: Average percent /æ/ identification for the F_1-varying continua. Each point represents an average for eight listeners, 20 judgments per listener.

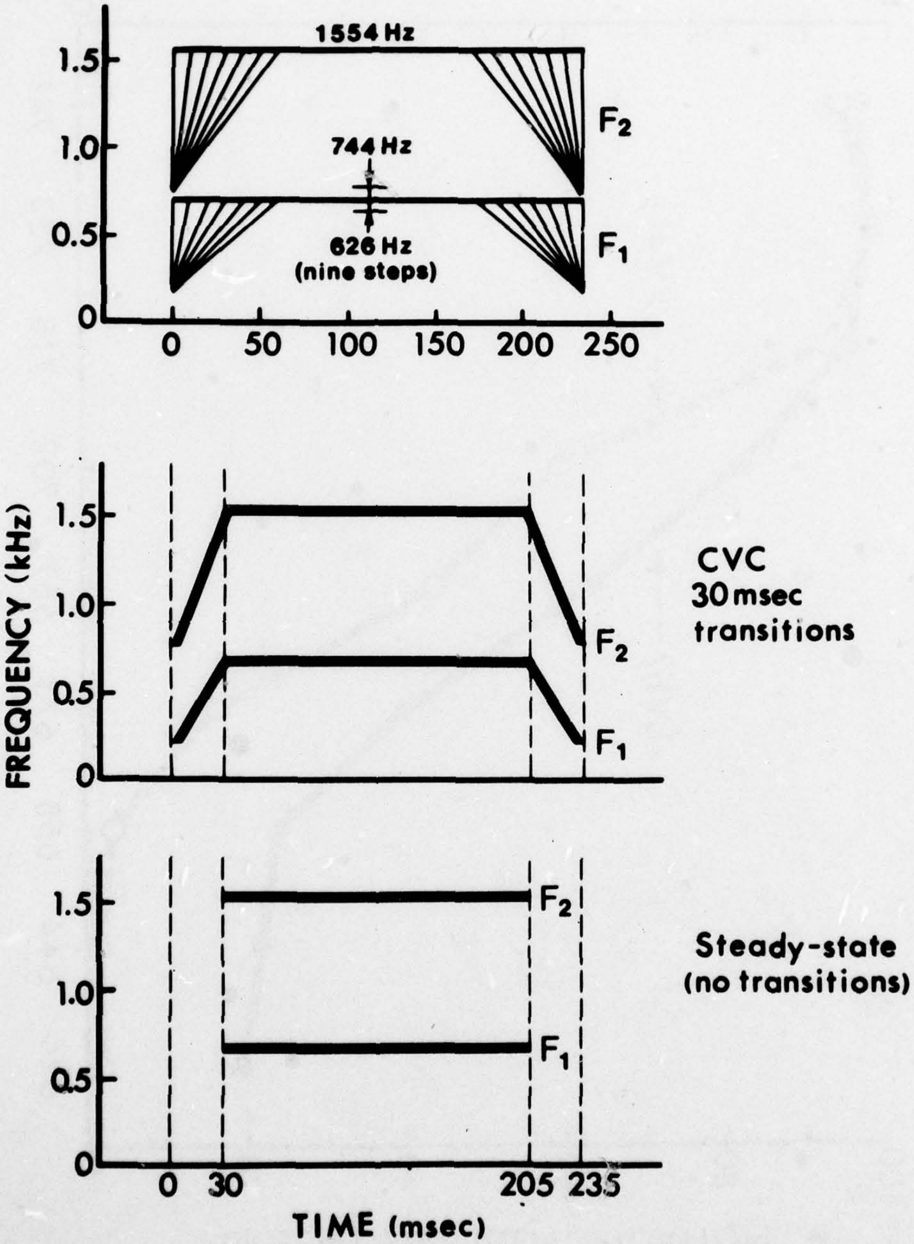


Figure 3: Formant control parameters used in synthetic F_1-varying continua in transition duration test. Seven such continua were defined by the progressive introduction of symmetric transitions ranging from 0 to 60 msec in duration (indicated by fan-shaped lines at top). Sample CVC and steady-state patterns are illustrated at the center and bottom.

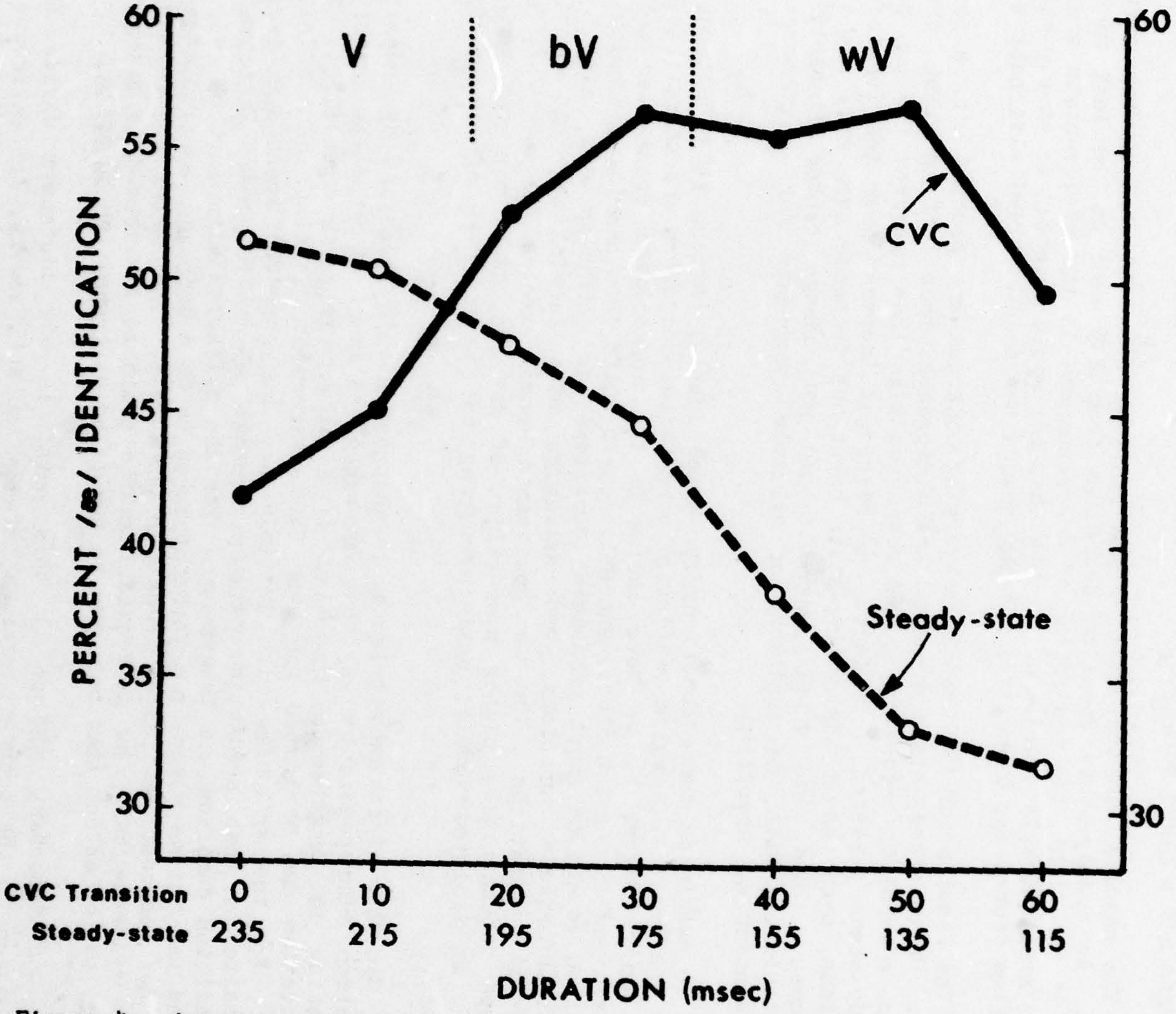


Figure 4: Average percent /æ/ identification for the CVC continua and steady-state continua. Each point represents data pooled over a continuum of nine steps, and pooled over eight listeners, 20 judgments per step per listener.

30-msec transitions) and its corresponding steady-state control.

In any one block of trials, listeners heard a randomized series of either the transition-varying syllables or the steady-state controls. Results for eight listeners are presented in Figure 4. Each point plots the total percentage of /æ/ responses for a particular spectral continuum (for which all patterns had a fixed duration).

The results for the steady-state continua are connected by the dotted line. As steady-state duration decreases from 235 msec at the left to 115 msec at the right, the proportion of /æ/ responses in the steady-state vowels drops monotonically. This is exactly what one would expect: as steady-state duration decreases, there is a biasing toward the shorter vowel alternative.

The results for the transition-varying continua are plotted with a solid line. While steady-state duration again decreases from left to right, there is not a monotonic decrease in /æ/ responses like that observed for the pure steady-state patterns. In fact, /æ/ responses increase over the first four continua, even though the steady-state durations decrease over this range by 60 msec. Thus, the vowel judgments did not change either linearly or monotonically with the durations of syllable components when a consonantal environment was specified.

In examining these functions for vowel identification, it is helpful to know the regions in which particular consonantal neighbors are heard. Across the top of Figure 4, we have indicated the approximate cross-over points (determined in a pilot study) where listeners shift from hearing predominantly isolated vowels at short transition durations, to syllables with initial /b/ at intermediate durations, and syllables with initial /w/ at the long transition durations. The two consonantal environments, /b/ and /w/, both showed substantially higher proportions of /æ/ responses than the patterns heard as isolated vowels, again verifying the paradoxical effect we found before.

Can this effect be attributed to changes in rate specified by transition duration? Our prediction was that /æ/ responses should decrease as transition duration is lengthened, that is, as a slower rate is specified. This prediction can be tested for each CVC environment, initial /b/ and initial /w/. For the syllables with initial /b/, the observed trend was exactly opposite to that predicted: the percentage of /æ/ responses increased as transition duration was lengthened. For the syllables with initial /w/, the trend was not as clear, but there appeared to be a decrease, as predicted, in the /æ/ responses at longer transition durations. Thus there was no consistent evidence that the paradoxical shifts could be attributed to changes in rate, to the extent that rate may be specified by transition durations.

One important unknown at this point is how listeners divide their responses in the boundary regions between isolated vowels, /b/ environments, and /w/ environments. It is possible, for example, that the rising slope in the bV region of Figure 4 is an artifact of the averaging process. The underlying proportions in all three phoneme regions might be stable or falling as a function of transition duration. We are currently studying contingencies between consonant and vowel responses to assess these possibilities.[1]

Another alternative that we are currently exploring is to view the vowel as an event that extends beyond what is acoustically realized. The duration of a stop-vowel-stop syllable extends throughout the silent closure intervals that precede initial release and follow final closure. Thus, the vowel's duration is not limited to the period of time that the vocal tract has a nonzero output amplitude. As an articulatory event, its duration may encompass part of the silent closure intervals as well. From this perspective, it would not be at all paradoxical for a bVb syllable to specify a longer vowel than is specified by a steady-state pattern of equal _acoustic_ duration. In effect, consonantal transitions might lengthen the perceived vowel, not by specifying a _faster_ rate of articulation, but by specifying a _slower_ one.

Our simple paradox has become absorbingly complex. In our efforts to explain it, we have come to realize how little is known about the timing relationships between consonants and vowels, how these relationships are changed or preserved with changes in rate, and how the acoustics of a syllable specify them to a listener. Answers to these questions may be basic to understanding why consonantal environments facilitate vowel identification. The presence of consonants may make the temporal as well as spectral identity of vowels more determinate.

REFERENCES

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

Mermelstein, P., A. M. Liberman and A. Fowler. (1977) Perception of vowel duration in consonantal context and its application to vowel identification. _Journal of the Acoustical Society of America 62_, S101. (The full text of the paper appears elsewhere in _Haskins Laboratories Status Report on Speech Research SR-55/56_.)

Nooteboom, S. G. (1974) Some context effects on phonemic categorization of vowel duration. _IPO Annual Progress Report 9_. (Eindhoven: Institute for Perception Research), 47-55.

Raphael, L. J., M. F. Dorman and A. M. Liberman. (1975) The perception of vowel duration in VC and CVC syllables. _Haskins Laboratories Status_

[1] Our preliminary analyses suggest that the proportions of /æ/ responses _decrease_ for all three phoneme environments (initial /b/, initial /w/ and no initial consonant). These results are consistent with the view that longer transitions specify a slower rate and, in turn, a shorter vowel, but they are also consistent with the simpler view that shorter steady-state duration (which is correlated with the longer transitions) specifies a shorter vowel. However, the latter approach is not able to explain the differences _between_ the three environments--in particular, why syllables with initial /b/ or /w/ are perceived to contain a longer vowel than those heard as steady-state vowels.

Report on Speech Research SR-42/43, 277-284.

Shankweiler, D., R. R. Verbrugge and M. Studdert-Kennedy. (1978) Insufficiency of the target for vowel perception. This paper appears elsewhere in Haskins Laboratories Status Report on Speech Research SR-55/56.

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

Verbrugge, R. R. and D. Shankweiler. (1977) Prosodic information for vowel identity. Haskins Laboratories Status Report on Speech Research SR-51/52, 27-35.

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

Perception of Vowel Duration in Consonantal Context and its Application to Vowel Identification*

Paul Mermelstein+, Alvin M. Liberman++ and Anne E. Fowler+++

ABSTRACT

To assess the extent to which a consonant-vowel (CV) transition contributes to the perceived duration of the vowel, we asked listeners to judge whether isolated vowels of specified duration were shorter or longer than the same vowels in CV context. To test the extent to which such a transition contributes to the identification of the vowel, we had the same listeners identify the vowel as /ɛ/ or /æ/, both when it was presented in isolation and when it was embedded in CV context. On the average, about half of the CV transition was included in the duration of the vowel as perceived. For five of the seven subjects that same perceived duration was almost exactly equal to the duration that determined whether they identified the vowel as /ɛ/ or /æ/. Thus, for those subjects the duration of the CV that was relevant to the linguistic judgment was the perceived duration of the vowel. (The data for the remaining two subjects departed from that rule in opposite directions.) The results underline the need to distinguish between vowel duration as perceived and vowel duration as measured from acoustic events noted on the speech signal.

INTRODUCTION

Variations in vowel duration are known to affect the perception of phonetic segments, for example, vowel identity [Stevens, (1959); Mermelstein, (1978)] and voicing of syllable-final stops [Denes, (1955); Raphael, (1972)]. How is the relevant duration to be defined in the acoustic signal, and what is its relation to the perceived duration of the vowel? Neither answer is obvious because, as is well known, the processes of articulation smear the phonetic information. Thus, information about more than one phone is often signaled simultaneously by the same parameters of the acoustic signal; and,

*This paper was presented in part at the 94th Meeting of the Acoustical Society of America, Miami Beach, Florida, 1977.

+Also Bell-Northern Research and Institut national de la recherche scientifique-Télécommunications, Verdun, Quebec, Canada.

++Also Yale University and University of Connecticut, Storrs.

+++Also Yale University; presently at University of Pennsylvania.

[HASKINS LABORATORIES: Status Report on Speech Research SR-55/56 (1978)]

conversely, several distinct acoustic segments often carry information about a single phone (Liberman, 1970).

One result of this complex relation between acoustic segment and phonetic unit is that it has not been possible to use direct acoustic criteria for the purpose of dividing speech into its constituent phones. Although some workers have arrived at operational definitions of phone durations for the purposes of their own research [for example, Peterson and Lehiste, (1960); Umeda, (1977)], no generally applicable segmentation procedure has been found. Indeed, the lack of effective segmentation techniques has been a major impediment to the achievement of automatic speech recognition (Hyde, 1972). The problem manifests itself not only in analysis but also in synthesis. Recall, for example, that Harris (1958) long ago found it impossible to produce acceptable speech by concatenating phone-size segments that had been excised from stretches of speech in which the segments had appeared in other phonetic contexts. In contrast, investigators have had better success in synthesizing speech by concatenating larger segments such as diphones (Estes, Kerby, Maxey and Walker, 1964) or demisyllables (Fujimura, 1976).

Evidence bearing more directly on the duration of the acoustic signal that is actually used in the perception of vowels comes from several experiments. In one of the earlier of these, Lindblom and Studdert-Kennedy (1967) demonstrated that the phonetic identity of the syllabic vowel can depend on the consonant-vowel and vowel-consonant transitions. More recently, Strange, Jenkins and Edman (1977) showed that vowels can still be quite reliably identified when the medial 'vowel' segment is excised from a consonant-vowel-consonant (CVC) syllable and the listener is presented with only the initial and final transitions separated by silence equal in duration to the excised segment. Of course, it is only in synthetic speech that one can readily distinguish transitional and 'vowel' (steady-state) segments. In natural speech, steady-state segments (that is, stationary acoustic patterns) rarely extend over significant intervals, so the marking of transition and steady-state intervals becomes an ad hoc procedure. Given the lack of generally applicable acoustic criteria for the segmentation of vowels and consonants, we should doubt the suggestion of Myers, Zhukova, Chistovich and Mushnikov (1975) that the perceiver employs a process of auditory segmentation to delimit consonant and vowel segments before making decisions about their phonetic identity.

Our concern in this paper is with the relationship between perceived vowel duration as determined in a direct psychophysical test and the vowel duration that is used as a cue for a phonetic decision. Directly relevant to that concern is a study by Raphael, Dorman and Liberman (1975), who took as their point of departure the fact that perceived voicing of stops in syllable-final position is cued by changes in the duration of the preceding vowel and undertook then to find out what acoustic interval constituted the relevant duration. To that end, they measured the extent to which the voicing boundary of the syllable-final stop was affected by the addition of an initial consonant-vowel transition. By this process, an initial /ɛd/ vs. /ɛt/ contrast was converted to a contrast between /dɛd/ vs. /dɛt/. The result was that the duration boundary at which listeners reported voiced and voiceless final stops with equal frequency shifted so as to indicate that the relevant stimulus duration included a large part of the syllable-initial transition.

The experiment just described revealed that voicing of a syllable-final stop depended, not just on the duration of the signal that conveyed information about the vowel, but also on earlier-occurring parts that included cues for a syllable-initial stop consonant; it was not designed to discover how the listeners assigned the relevant duration to the several phonetic units they perceived. One possibility, of course, is that they attributed part to the duration of the syllable-initial consonant and part to the vowel, but made the linguistic judgment about the voicing of the syllable-final stop in terms of a rule that takes the relevant duration to be the sum of the perceived durations of initial consonant and medial vowel. It seems more likely, however, that the linguistic judgment was controlled by the perceived duration of the vowel and that the syllable-initial transition contributed, in whole or in part, to that perception. As for the perceived duration of the initial stop consonant, it might well have been based on a stretch of acoustic signal that was also included in the perceived duration of the vowel, if indeed the perceived duration of a syllable-initial stop consonant can be judged with sufficient reliability. At all events, we suppose that the perceived duration of the vowel is determined from the duration of all, or almost all, of the acoustic syllable and then also used as a basis for such linguistic judgments--voicing of syllable-final stop or phonetic identity of the vowel--as the variable of duration may be relevant to.

The purpose of the experiment to be reported here is to test that hypothesis, given a setting in which phonetic identity of a vowel is the linguistic judgment to which duration is relevant. To that end, we will first determine how listeners perceive the duration of vowels in CV (stop consonant-vowel) context by comparison with spectrally identical V's (vowels in isolation). Then we will find in both the CV and V conditions how duration controls the phonetic identification of the vowels as /ε/ or /æ/. Appropriate comparisons should reveal the extent to which the phonetic identification of the vowel segments depends on the same processes that underlie their perceived durations.

PROCEDURE

Preliminary experiments revealed that synthesized vowels with formant frequencies at F_1 = 650 Hz, F_2 = 1800 Hz, and F_3 = 2500 Hz could be identified by most listeners as either /ε/ or /æ/, the difference depending solely on the duration of the stimuli. The boundary for equal probability of identifying either vowel was in most cases located at a duration value near 100 msec. To study the effects of context, we selected the simplest possible linguistic context into which the vowel could be embedded. For this purpose, we created a CV syllable, heard as /bV/, by preceding the above steady-state vowel with a 48-msec linear transition from initial formant values of F_1 = 100 Hz, F_2 = 1000 Hz and F_3 = 2000 Hz.

All stimuli were generated with the aid of a programmed (software) synthesizer modeled after the one presented by Rabiner (1968). The first three formants were adjustable under program control, the fourth and fifth formants were fixed at 3500 and 4500 Hz respectively. Bandwidth values were fixed at 50, 80, 100, 175, and 281 Hz respectively. The stimuli were entirely voiced and generated with a fundamental frequency of 125 Hz at a sampling frequency of 10,000 Hz.

Tape-recorded stimuli were prepared that divided the subjects' tasks into three parts. The first task was a control test of duration judgment. For this test only isolated vowels of the above spectral specification were used. Subjects were asked to judge probe vowels of 72, 80, 88, 96, 104, 112 and 120 msec, respectively, as "longer" or "shorter" than a 96-msec standard. Responses were obtained for four presentations of each stimulus-pair, two in the order standard followed by probe, and two in the reversed order. Any one presentation included two repetitions of the pair, standard plus probe or probe plus standard, separated by 2 seconds of silence. One second was allowed between standard and probe. Immediately preceding this and each of the other tasks, the subject heard a practice run of 10 presentations so that he could accustom himself to the equipment, the synthesized speech and the task at hand. Subjects responded by noting on paper whether the second stimulus was shorter, equal to or longer than the first. The perceived vowel duration was estimated as the 50 percent intercept on the psychometric function fitted to the total number of "longer" plus one-half of "equal" responses plotted as functions of probe vowel duration.

In the second part of the experiment we asked listeners to rate an isolated probe vowel of variable duration relative to the same vowel in CV context and having durations of 40, 96, and 152 msec, respectively. In each case the probe vowel durations used were 0, 8..., 40, and 48 msec longer than the standard in CV context. The 8-msec duration increment was selected to equal the pitch period and thereby avoid duration artifacts arising from presentation of incomplete pitch periods. Each listener responded to each of the 21 comparisons four times, as in the control situation described above. To eliminate time order effects, two presentations followed the V-CV order, and two others the reverse order.

In the last part of the experiment, subjects identified the vowel alone or in CV context as /æ/--as in "bat"--or as /ε/--as in "bet." Here each subject was provided representative pairs, heard a practice run, and was allowed to ask questions before performing the task. In each case presentations included steady-state vowel segments of 40, 80, 96, 112 and 152 msec duration. Any one identification response was based on two repetitions of each stimulus separated by a 1-sec silent interval. V and CV stimuli were presented in one randomized list and identification responses were analyzed by subject and by stimulus. The duration value corresponding to the identification boundary for each context was estimated from the 50 percent identification value of the psychometric function fitted to the curve of /æ/ vowel responses.

The subjects were all students at Yale University and were paid $2.00 per hour for their services. They ranged in age from 18 to 28 and all had at least an introductory knowledge of phonetics. Three subjects completed the vowel identification before they did the duration ratings; five completed the rating task before they attempted to identify stimuli.

RESULTS

Figure 1 shows the results of the experiment as three sets of data points that represent the judgments of vowel duration and vowel identity for vowels in isolation and vowels in CV context. Before saying what those data show, we

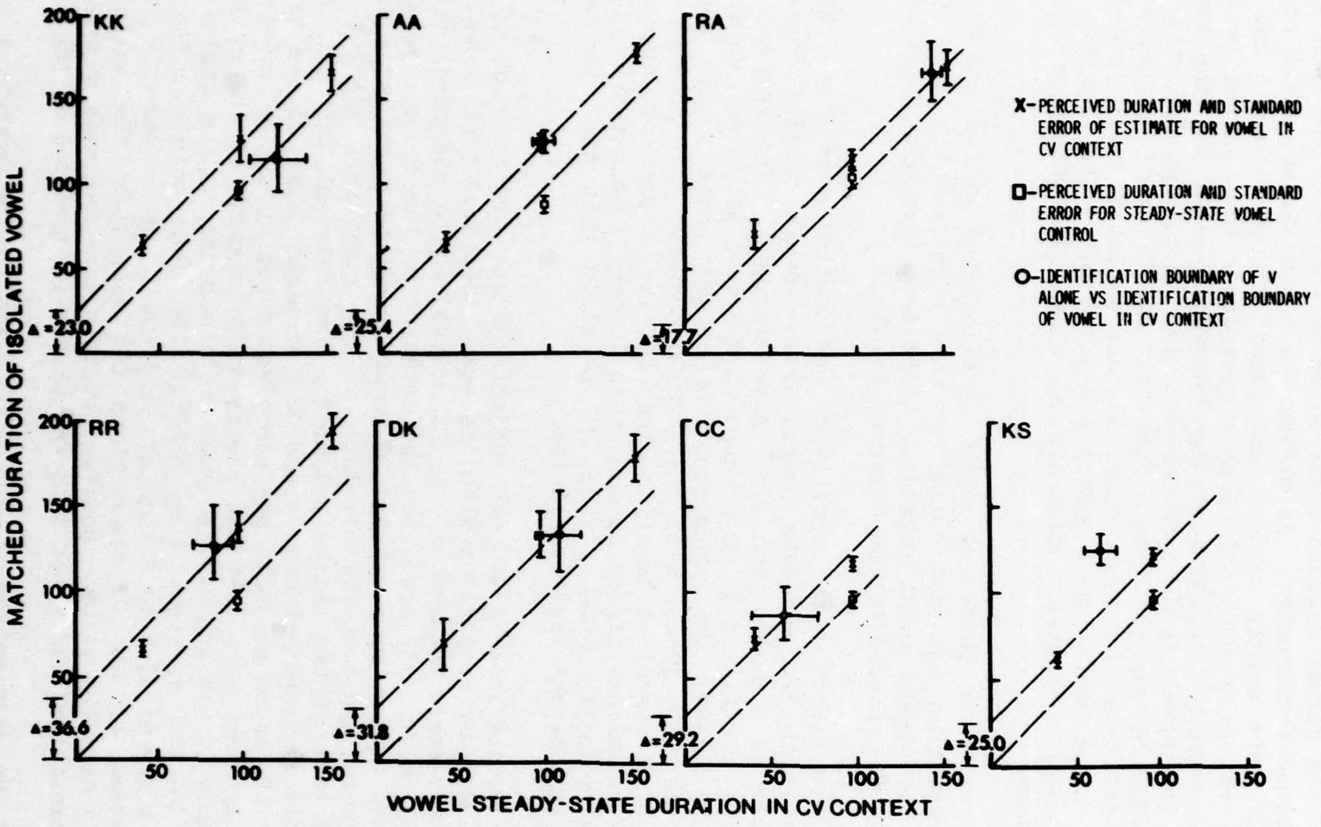


Figure 1: Duration of the isolated vowel matched to a vowel heard alone and in CV context. An estimate of the relationship between the identification boundary for the vowel alone and in CV context is also shown for each subject.

should explain how the values shown in the figure were computed.

All data points, and their corresponding standard errors, were calculated with the use of a probit analysis program (Finney, 1964) that determines the parameters of the normal sigmoid curve so as to arrive at the best fit to the observed response frequencies. In making the calculations, duration responses judged "equal" were evenly divided and added to the "longer" and "shorter" responses to obtain a curve of "longer" responses that varied between 0 and 100 percent. When the duration of the "standard"--that is, the steady-state duration of the vowel in CV context--was 152 msec, two subjects failed to provide a sufficient number of both "shorter" and "longer" responses to permit an accurate estimate of the duration of that isolated vowel judged equal in duration; in those cases, data points are not shown. One subject failed in general to yield consistent judgments of vowel identification; his results were eliminated.

The increment in perceived vowel duration due to the addition of the consonant transition was not found to vary significantly with the actual duration of the steady vowel segment. Therefore, straight lines with slopes of 45 degrees were fitted to the observed points.

Let us look first at the results of the control condition--that is, the condition in which the subjects judged probe vowels ranging from 72 to 120 msec as longer or shorter than a 96 msec standard. The relevant data are represented in the figure as the squares. Looking at these, we see a close correspondence between the judgments made on the probe vowels and, on the other hand, the true duration of the standard vowel: for no subject did the computed value of perceived duration--that is, the 50 percent intercept on the psychometric functions--differ by more than 10 msec from the true duration (96 msec) of the standard vowel. Thus, it is plain that the task we set our subjects was not beyond their abilities.

Next we should look at how our subjects judged the duration of the vowel in a CV syllable by comparison with the same (isolated) probe vowel. In this case the steady-state segment of the CV syllable (the standard) had duration values of 40, 96 and 152 msec, and the duration of the probe (isolated) vowel ranged between values equal to and 48 msec longer than the standard. Thus, the probe varied between the duration of the stationary segment of the CV syllable and that of the entire syllable. The relevant parts of the figure are the points represented as x's and the straight lines fitted to them. We see that for all seven subjects the perceived duration of the vowel in the CV context was greater than just the steady-state segment of the syllable. The amount by which it was greater is shown by the point on the ordinate that is intercepted by the straight line through the data points. The increment ranges from a low of 17.7 to a high of 36.6 msec. In no case is it as great as the duration of the CV transitions (48 msec); on the average, it is approximately half the value. However, it is clear that some significant part of the CV transitions did enter into the perceived duration of the vowel.

Finally, we examine the effect of duration on the perceived identity of the vowel as /ɛ/ or /æ/ to see what the controlling duration is for the vowel in CV context. The relevant data points are those marked by the circles, since these show the vowel identification boundaries for the isolated probe vowels *vs* the corresponding boundaries for the vowels in CV context. It is evident that for five of the seven subjects these points fall squarely on the straight-line function, described in the preceding paragraph, that represents the perceived duration of the vowel in CV context. Thus, for those subjects the duration in the CV context that is relevant to the linguistic judgment--that is, the identification of the vowel as /ε/ or /æ/--is, in fact, the perceived duration of the vowel in that same context. For the remaining two subjects that is not the case. For one of them (KK), the duration that controls the linguistic judgment is less than the perceived duration--indeed, it appears to be approximately equal to the duration of the steady-state segment. For the other subject (KS), it is greater than the perceived duration of the vowel, greater, in fact, than the overall duration of the CV syllable.

A more nearly exact comparison between perceived duration of the vowel and the duration that controlled the linguistic judgment can be made from the data presented in Table 1. For each subject the increment in perceived vowel duration produced by the CV transitions is shown, and also the effect of those same transitions on the phonetic identification boundary. To make it easier to see the fit between the two sets of data points, those data have been put into the scatter plot of Figure 2.

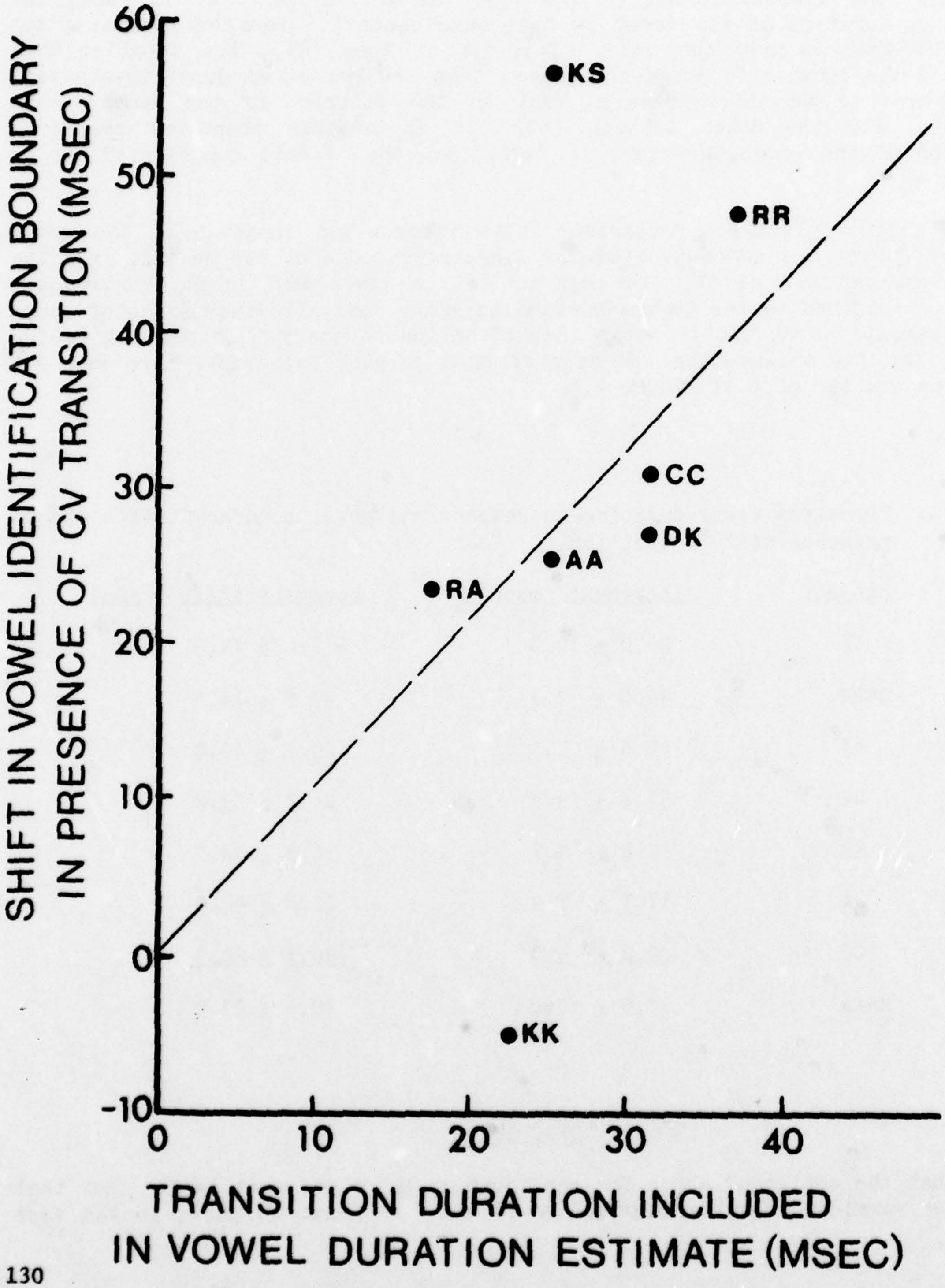
TABLE 1: Perceived vowel duration increments and vowel boundary shifts due to presence of CV transition.

Subject	Increment (msec)	Boundary shift (msec)
KK	23.0 ± 10.3	- 5.6 ± 19.3
RR	36.6 ± 7.1	47.6 ± 24.4
AA	25.4 ± 5.1	25.9 ± 10.0
DK	31.8 ± 14.5	26.7 ± 19.2
KS	25.0 ± 4.7	56.5 ± 12.5
RA	17.7 ± 7.9	22.8 ± 48.0
CC	29.2 ± 4.9	30.7 ± 24.3
Mean	27.0 ± 10.2	29.2 ± 31.2

DISCUSSION

That the estimated error for the boundary shift was much larger than that for the vowel-duration increment may be due, at least in part, to the fact

Figure 2: Shift in vowel identification boundary plotted against the transition duration included in the estimate of vowel duration for seven listeners.



that the judgments used to determine the boundary shift were obtained from stimuli that were spaced more widely (five stimuli at intervals of 16 msec) than those used to determine the perceived durations (seven stimuli at intervals of 8 msec). The difference in error may also reflect a difference between the tasks: judging the identity of the vowels, which is the basis on which the boundary shift is determined, may be more difficult than judging their duration.

Consider now why the perceived duration of the vowel and the controlling duration for the linguistic judgment should, for most of the subjects, have included just half of the CV transition. In fact, we are quite uncertain, the more so because other studies have found that more of the CV transition is sometimes included. Thus, in the study by Raphael, Dorman and Liberman (1975) that was referred to in the Introduction, generally larger fractions of initial CV transitions were included in the duration that controlled the perception of a syllable-final stop as voiced or voiceless. Also relevant is a study by Verbrugge and Isenberg (1978) on the role of duration in controlling the perceived identity of a vowel in a CVC context. In that study, the CV transitions contributed to the duration that controlled the linguistic judgment by an amount greater than their own durations.

It is difficult to know what to make of the differences between these experiments and ours in the duration of consonant-vowel transition that contributed to the linguistic judgment. One suspects that aspects of the different stimulus patterns may have implied different rates of articulation, hence different corrections of syllable duration that listeners may have made in order to take account of rate; or, perhaps, there were psychoacoustic factors--including, for example, differences in rise-time--that produced differences in the perceptually effective durations of the transitions. It seems that it will be possible, with properly designed experiments, to throw light on this question, but for the moment speculation is idle. At all events, such factors as those experiments might uncover can hardly affect the conclusion we wish to draw from ours, since that conclusion is based on a comparison of the results of two conditions--judgments of vowel duration and a linguistic judgment about vowel identity--in which the stimuli were identical. We note again, therefore, the major results of our experiment: the perceived duration of a vowel in CV context included a significant part of the initial formant transitions, and, for most of our subjects, the duration that controlled a judgment about the phonetic identity of the vowel included a part of the initial transitions almost exactly equal to the part that contributed to the perceived duration of the vowel. Thus, the duration that controlled the linguistic judgment was the duration of the vowel as perceived. Given a variety of contexts, the factors that affect the one might affect the other equally, in which case our findings would exemplify a general rule. That, however, remains to be seen.

The results underline the distinction that must be made between perceived duration of vowels and vowel-duration as measured in the acoustic signal. Clearly no time-localized acoustic events mark the mid-point of the CV transition heard as the start of the vowel in the CV context. Thus, vowel duration measurements from acoustic data (Allen, 1978) do not generally reflect the perceptually relevant intervals. Further systematic experiments are required to allow the formulation of an integrated theory of vowel perception in speech context.

REFERENCES

Allen, G. D. (1978) Vowel duration measurement: A reliability study. *Journal of the Acoustical Society of America 63*, 1176-1185.

Denes, P. (1955) Effects of duration on the perception of voicing. *Journal of the Acoustical Society of America 27*, 761-768.

Estes, S. E, H. R. Kerby, H. D. Maxey and R. J. Walker. (1964) Speech synthesis from stored data. *IBM Journal of Research and Development 8*, 2-13.

Finney, D. J. (1964) *Probit Analysis - A Statistical Treatment of the Sigmoid Response Curve*. (Cambridge: Cambridge University Press).

Fujimura, O. (1976) Syllables as concatenated demisyllables and affixes. *Journal of the Acoustical Society of America 59*, S55(A).

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

Hyde, S. R. (1972) Automatic speech recognition: A critical survey of the literature. In *Human Communication, a Unified View*, ed. by E. E. David and P. B. Denes. (New York: McGraw-Hill), 399-438.

Liberman, A. M. (1970) The grammars of speech and language. *Cognitive Psychology 1*, 301-323.

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

Mermelstein, P. (1978) On the relationship between vowel and consonant identification when cued by the same acoustic information. *Perception & Psychophysics 23*, 231-236.

Myers, T. F., M. G. Zhukova, L. A. Chistovich and V. N. Mushnikov. (1975) Auditory segmentation and the method of dichotic simulation. In *Auditory Analysis and Perception of Speech*, ed. by G. Fant and M. A. A. Tatham. (New York: Academic Press), 243-274.

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

Rabiner, L. R. (1968) Digital-formant synthesizer for speech-synthesis studies. *Journal of the Acoustical Society of America 43*, 822-828.

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

Raphael, L. J., M. F. Dorman and A. M. Liberman. (1975) The perception of vowel duration in VC and CVC syllables. *Haskins Laboratories Status Report on Speech Research SR-42/43*, 277-284.

Stevens, K. N. (1959) Effect of duration upon vowel identification. *Journal of the Acoustical Society of America 31*, 109(A).

Strange, W., J. J. Jenkins and T. R. Edman. (1977) Identification of vowels in "vowel-less" syllables. *Journal of the Acoustical Society of America 61*, S39(A).

Umeda, N. (1977) Consonant duration in American English. *Journal of the Acoustical Society of America 61*, 846-858.

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

Stimulus Dominance in Fused Dichotic Syllables

Bruno H. Repp

ABSTRACT

Fifteen synthetic syllables from a /bæ/-/dæ/-/gæ/ continuum were dichotically fused with three selected stimuli from the same continuum and presented to listeners for identification. It was found that stimuli in the vicinity of phonetic category boundaries make weaker dichotic competitors than stimuli from well within a phonetic category. This result is in agreement with the hypothesis that dichotic stimulus dominance relationships reflect the relative category goodness of the competing stimuli, that is, their relative perceptual distances from the listener's category prototypes.

INTRODUCTION

In the large majority of dichotic studies using speech sounds, the focus of interest has been the direction and magnitude of the ear dominance effect (or ear advantage). However, there is a second factor that plays an important role in dichotic perception. This factor, called "stimulus dominance" (Repp, 1976), is the tendency of one stimulus in a specific dichotic pair to receive more correct responses than the other stimulus, regardless of the ear in which it occurs. Ear dominance and stimulus dominance are independent factors that jointly determine the listener's responses to dichotic stimulus pairs.

Most dichotic experiments in the past have used stimuli that did not fuse and thus were heard as more or less separate events. Stimulus dominance effects have occasionally been noted (for example, Berlin, Lowe-Bell, Cullen, Thompson and Loovis, 1973), but they did not appear to be theoretically interesting. More recent work using fused dichotic syllables has changed this state of affairs (Repp, 1976, 1977a, 1977b, 1978a). Stimulus dominance plays an important role in assessing ear dominance with fused stimuli, comparable to the role of response bias in signal detection tasks (Repp, 1977b). In addition, stimulus dominance relationships may reveal some interesting facts about the nature of dichotic stimulus interaction. By identifying the properties that make one stimulus dominate another, important information may be obtained about the level at which perceptual competition between fused dichotic stimuli takes place.

Acknowledgment: This research was supported by NIH Grant DE00202 to the University of Connecticut Health Center and by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. I would like to thank Virginia Mann for her comments on an earlier draft of this paper, and Franklin Cooper and David Zeichner for technical advice.

There are at least three different (not mutually exclusive) levels at which dichotic competition between speech sounds may occur. One is the phonetic level (Studdert-Kennedy, Shankweiler and Pisoni, 1972). If dichotic competition occurred solely between categorical phonetic representations (syllables, phonemes, or features), acoustic stimulus variations within a phonetic category should have no influence on the degree of dominance that a stimulus from that category exerts over stimuli from other categories. However, there is now ample evidence that this is not true: within-category changes in the voice onset times (VOTs) or the formant transitions of competing syllable-initial stop consonants do significantly change stimulus dominance relationships (Miller, 1977; Repp, 1976, 1977a, 1978a). Thus, dichotic competition is certainly not exclusively phonetic in nature, although there may be a phonetic component to the process.

Perhaps the most obvious level at which dichotic competition between fused syllables might occur is that of auditory processing. The phenomenon of dichotic fusion itself is determined by rather low-level auditory properties of the signals; thus, fusion can easily be prevented by introducing slight discrepancies in fundamental frequency or temporal alignment (Halwes[1]; Cutting, 1976). However, even though fusion occurs at a relatively low level in the auditory system, it seems unlikely that the individual auditory properties of the fused stimuli are lost at this early level by direct auditory interactions, such as masking or integration. Rather, they are probably retained in some superimposed or mixed form, and the dominance of one stimulus property over another comes about because it is, in some sense, more salient. This "auditory salience" hypothesis predicts that stimulus dominance relationships will change (if they change at all) as a relatively smooth and continuous function of changes in acoustic stimulus parameters, as long as these changes do not introduce discontinuities in auditory perception.

In experiments on dichotic voicing contrasts (syllable-initial stop consonants contrasting only in the voicing feature, such as /ba/-/pa/), this prediction has generally been supported: the voiceless stimulus in a pair becomes more (less) dominant as its VOT is increased (decreased); and similar, although smaller, effects are obtained by manipulating the VOT of the voiced stimulus in a pair (Miller, 1977; Repp, 1977a, 1978a). However, in a study of dichotic place contrasts (syllable-initial stop consonants contrasting only in place of articulation, such as /ba/-/da/), Repp (1976) found an interesting irregularity in the effects of acoustic parameters on stimulus dominance, which led to the present study.

In this earlier experiment, seven different consonant-vowel syllables were presented in all possible dichotic pairings. The stimuli were distinguished only by the initial transitions of the second and third formants (F_2 and F_3) whose onset frequencies were varied to form a continuum from /bæ/ to /dæ/ to /gæ/. When presented dichotically, these stimuli were perfectly fused and sounded like single syllables presented binaurally. The results showed a clear effect of variations in F_2 and F_3 transitions on stimulus dominance,

[1] Halwes, T. G. (1969) Effects of dichotic fusion on the perception of speech. Unpublished doctoral dissertation, University of Minnesota.

even when these variations occurred within a phonetic category. However, this effect was not monotonic with the changes in F_2 and F_3. In particular, when stimulus 1 (/bæ/) was paired with stimuli 3-7 (/dæ/, /gæ/), the percentage of B responses showed a local maximum in the pairing with stimulus 5, which happened to be ambigous between /dæ/ and /gæ/. This finding suggested that yet another property of the stimuli may be important in dichotic competition: the "category goodness" of the stimuli. Stimulus 5 was ambiguous and hence not a good example of any category. This may have been the reason why it was a weaker dichotic competitor for /bæ/ than its neighbors on the stimulus continuum.

This interpretation suggests a determinant of dichotic stimulus dominance intermediate between the auditory and phonetic levels. In the context of recent models of phoneme recognition, this intermediate level has been termed the "multicategorical" (Repp, 1976, 1977a) or "prototype matching" stage (Oden, in press; Oden and Massaro, 1978). At this level, the perceptual system is assumed to determine how well a stimulus matches any of several category prototypes. When two competing dichotic stimuli enter the system, a stimulus that is close to a prototype will tend to dominate over a stimulus that is far from any prototype. This prediction is made by any of several possible models of dichotic interactions, such as a "race for the nearest prototype" in auditory space.

Since changes in acoustic parameters change not only the perceptual distance of stimuli from the category prototypes but possibly also the salience of auditory cues, the predictions of the "prototype matching" hypothesis for changes in stimulus dominance are often difficult to distinguish from those of the auditory salience hypothesis. Nor are the two hypotheses mutually exclusive. The evidence in favor of the prototype matching hypothesis, as applied to dichotic listening, rests primarily on the demonstration that ambiguous stimuli are weak in dichotic competition with unambiguous stimuli. Repp's (1976) results were merely suggestive in that regard. It was the purpose of the present study to investigate this question in more detail.

In order to obtain more precise data than in the earlier study (which had used only seven different stimuli), a 15-member /bæ/-/dæ/-/gæ/ continuum was created by varying the onset frequency of the F_2 transition. (The F_3 transition was held constant, as explained below.) All stimuli were dichotically paired with themselves, and with stimuli 1, 8, and 15, which were representative of the three phonetic categories. Consider the results that might be obtained for stimulus 1 (/bæ/) when paired with all others and presented for identification as B, D or G. When the percentage of B responses is plotted as a function of the number (that is, location on the continuum) of the stimulus competing with stimulus 1, a "stimulus dominance function" is obtained that describes changes in the relative dominance of stimulus 1 as the formant transitions of the competing stimulus are changed. According to the auditory salience hypothesis, this function should be smooth and continuous; its precise shape will depend on whether or not the other stimuli tend to dominate stimulus 1. (We assume, for the time being, that auditory salience is a smooth function of acoustic changes in formant transitions.) The prototype matching hypothesis, on the other hand, predicts a significant local increase in B responses in the region of the /dæ/-/gæ/ category boundary, where the competing stimulus is ambiguous (and therefore expected to be weaker in dichotic competition). A similar stimulus dominance function may be obtained for stimulus 15 (/gæ/) paired with all others, with the percentage of G responses as the dependent variable. Here, the prototype matching hypothesis predicts a local peak in the function around the /bæ/-/dæ/ boundary.

In order to make sure that these local peaks, if obtained, are really related to the category boundaries -- and thus to the relative ambiguity of the competing stimuli -- two different stimulus sets were used in the present experiment. They were distinguished by the presence or absence of a rising F_3 transition which remained constant over the whole stimulus continuum. The F_3 transition is known to affect the perception of place of articulation, particularly the size of the alveolar category on continua such as used here (Harris, Hoffman, Liberman, Delattre and Cooper, 1958; Hoffman, 1958). Thus, the two different stimulus series were expected to have their category boundaries in different locations. Accordingly, the local peaks in the stimulus dominance functions predicted by the prototype matching hypothesis should appear at corresponding different locations in the two stimulus series.

The prototype matching hypothesis also predicts that stimuli from within a given category should dominate stimuli ambiguous between this same category and a neighboring category. For example, a good /bæ/ (stimulus 1) should dominate stimuli ambiguous between /bæ/ and /dæ/. In the stimulus dominance function for stimulus 1, this should be reflected in a high level of B responses extending from the /bæ/ category beyond the /bæ/-/dæ/ boundary. Confirmation of all these predictions would constitute strong evidence in favor of the prototype matching hypothesis.

METHOD

Subjects

Six subjects participated. One of them was the author; the others were paid volunteers (four Yale undergraduates and one high-school student). All but one had participated in earlier experiments and had been selected because of their accurate performance with synthetic speech stimuli. One subject had no experience with synthetic speech but did just as well as the others.

Stimuli

The stimuli were two sets of 15 synthetic syllables produced on the Haskins Laboratories parallel resonance synthesizer (frame rate = 200/sec) and ranging perceptually from /bæ/ to /dæ/ to /gæ/. All syllables were 280 msec long, had a constant fundamental frequency of 114 Hz, a VOT of -15 msec (that is, prevoicing), 45-msec stepwise-linear formant transitions, and no burst but an abrupt onset of energy following the prevoicing. (Very similar stimuli had been used by Pisoni[2] in his studies on categorical perception.) Within each

[2] Pisoni, D. B. (1971) On the nature of categorical perception of speech sounds. Unpublished Ph.D. dissertation, University of Michigan, 1971.

set, the stimuli differed only in the onset frequencies of F_2, which are shown in Table 1. One stimulus set had a constant F_3 transition that rose from 2180 Hz to 2862 Hz ("rising F_3"). The other set had no F_3 transition; F_3 remained constant at 2862 Hz in all stimuli ("flat F_3"). The presence of a rising F_3 transition was expected to bias perception against D, and more toward B and G.

TABLE 1: F_2 onset frequencies of the stimuli.

Stimulus No.	F_2 onset (Hz)
1	1155
2	1232
3	1312
4	1386
5	1465
6	1541
7	1620
8	1695
9	1772
10	1845
11	1920
12	1996
13	2078
14	2156
15	2234
Steady state	1620

All stimuli were digitized at 8 kHz using the pulse code modulation (PCM) system at Haskins Laboratories. Dichotic pairs were created from the digitized waveforms with a special-purpose computer program. At the time the experimental tape was recorded, the digital sampling procedure led to a random error in dichotic stimulus alignment of up to one sampling period (0.125 msec), which was considered insignificant. The tape contained three sequences of 228 stimulus pairs each. Each sequence contained the following combinations within each stimulus set: all 15 stimuli paired with themselves twice (30 binaural pairs); stimulus 1 paired with all others, in both channel assignments (28 dichotic pairs); stimulus 8 paired with all others (28 pairs); and stimulus 15 paired with all others (28 pairs). (Pairs 1+8, 1+15, and 8+15 were unnecessarily duplicated and occurred twice as often as the other pairs.) Stimuli from different sets were never paired with each other. All in all, there were 30 + 3 x 28 = 114 pairs from each stimulus set, which were randomized together. The interpair interval was 2 sec. The three sequences of 228 pairs were separated by longer intervals.

Procedure

The experiment required three one-hour sessions, held on different days. The subjects were instructed to identify the initial consonants in each (fused) stimulus as B, D, or G. In each session, the stimulus tape was played twice; thus, at the end of the experiment, each subject had given a total of 36 responses to each stimulus pair (18 to each channel assignment of a dichotic pair), except for the three duplicated pairs, which received 72 responses each. The subjects knew that there were both binaural and dichotic stimuli in the sequence; but, as Repp (1976) has shown, fused dichotic syllables contrasting only in their initial formant transitions are practically indistinguishable from binaural syllables, and the present stimulus sequences indeed sounded like homogeneous lists of single syllables originating in the center of the listener's head.

The tape was played back on an Ampex AG-500 tape recorder. The subjects listened over Telephonics TDH-39 earphones. Playback amplitude was adjusted and monitored on a Hewlett-Packard voltmeter, and special care was taken to equalize the amplitudes of the two channels at about 85 dB SPL (peak deflections on a voltmeter for a single stimulus). The tape recorder channels were reversed electronically halfway through each session, in order to counterbalance any possible quality differences between tape tracks.

RESULTS

Binaural Identification

The average labeling functions for the two sets of 15 stimuli when presented binaurally (randomized together with the dichotic pairs) are shown in Figure 1. The solid functions represent the set with a rising F_3. It can be seen that the D category was fairly narrow, and D responses did not exceed 90 percent to any stimulus. This was expected, since alveolar stops normally require a falling F_3 transition. Clearly, however, an F_2 transition in the range appropriate for alveolars perceptually overrode the conflicting F_3 transition. In the stimulus set with no F_3 transition (dotted functions), the D category was more prominent and occupied a larger region on the stimulus continuum. As expected, both category boundaries shifted outward as the onset frequency of F_3 was raised (rising vs. flat transition), thus reducing the frequency of B and G responses, which normally require a low F_3 onset. Again, however, the F_2 transition remained the dominant cue for the perceived place of articulation.

Category boundaries were estimated by fitting normal ogives to the labeling functions of individual subjects, using the probit algorithm (Finney, 1971), and by determining the 50-percent intercept of the fitted functions. These boundary values were submitted to a 2 x 2 analysis of variance with the factors boundaries (B-D and D-G) and stimulus sets (rising F_3 and flat F_3). The outward boundary shifts caused by the change in F_3 were reflected in a significant interaction between the two factors ($F_{1,5} = 19.2$, $p < .01$). Figure 1 indicates that the B-D boundary shifted more than the D-G boundary. This difference did not quite reach significance ($F_{1,5} = 5.4$, $p < .10$), since it was shown by only five of the six subjects. Identification of stimuli within the B category also seemed to suffer more from the elimination of the

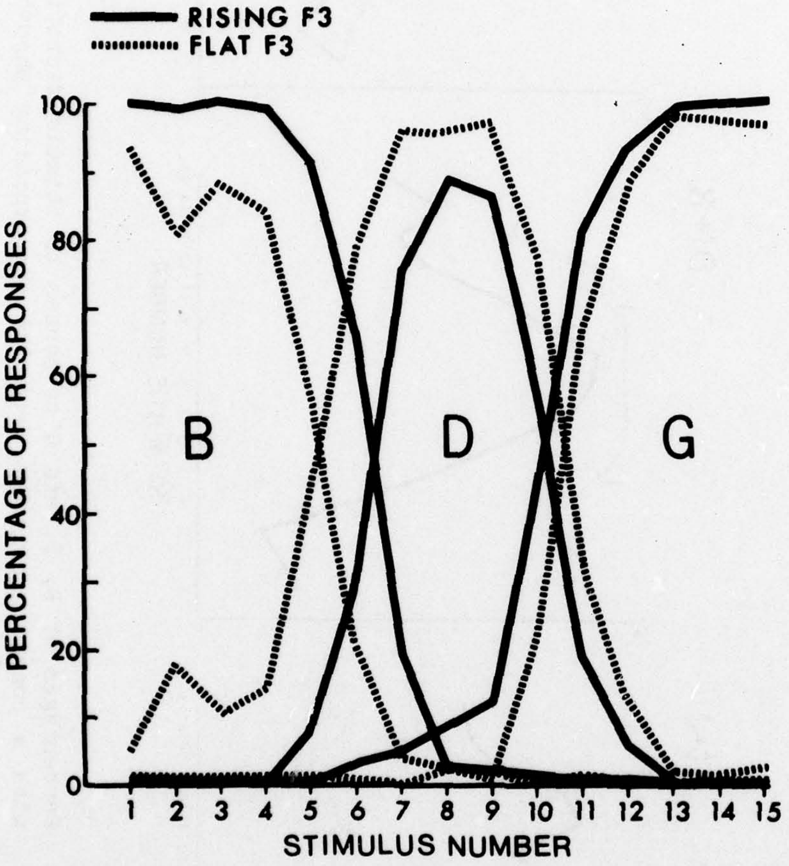


Figure 1: Percentages of B, D, and G responses to binaurally presented stimuli with a rising or flat third formant (F_3).

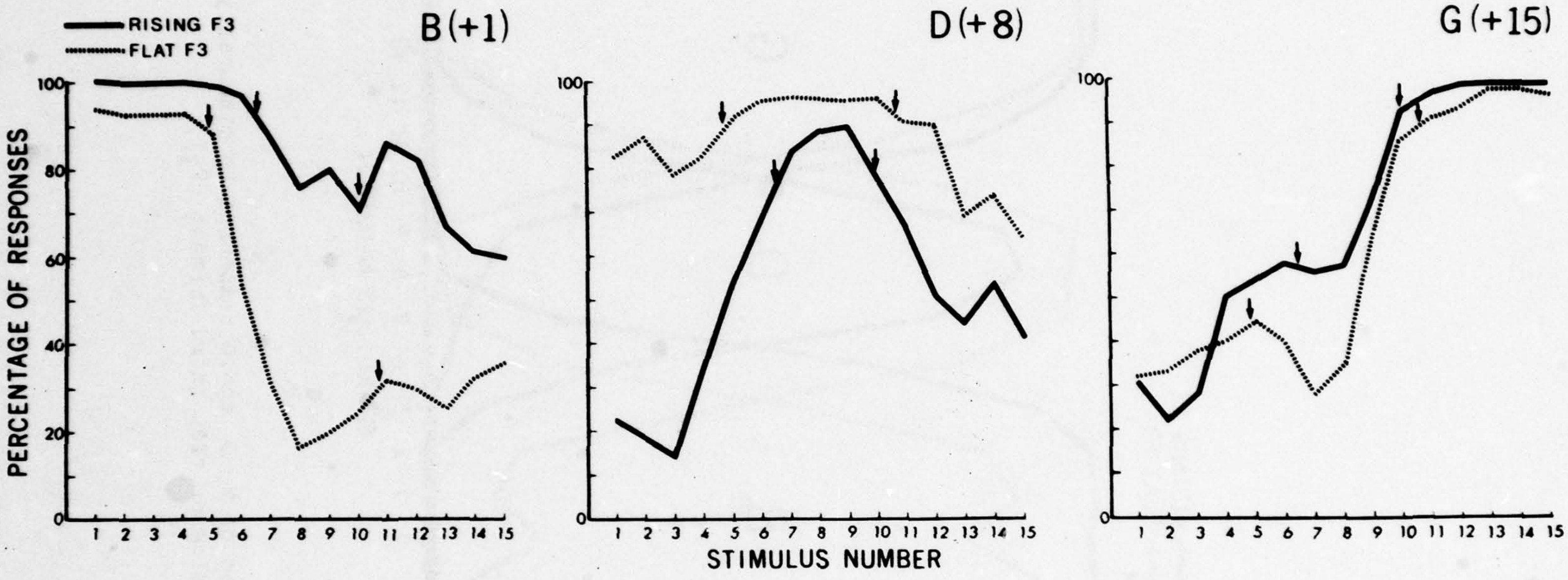


Figure 2: Percentages of B, D, and G responses to stimuli dichotically fused with a constant stimulus from the corresponding phonetic category (+1, +8, +15).

rising F_3 transition than stimuli in the G category. Individual differences in the sensitivity to the change in F_3 were quite substantial: the shifts of the B-D boundary ranged from 0.64 to 2.78 stimulus units, and those of the D-G boundary ranged from -0.04 to 1.92 stimulus units.

Although the basic effect of the F_3 transition on the perceived place of articulation confirmed the earlier results of Harris et al. (1958) and Hoffman (1958), it was much smaller than in these earlier studies (though adequate for the purpose of the present experiment). Harris et al. and Hoffman obtained hardly any D responses to stimuli with a rising F_3 transition less steep than the present one. The reason for this discrepancy presumably lies in the relative amplitudes of F_3 and F_2. Harris et al. and Hoffman, who constructed their stimuli on the Pattern Playback, did not report any formant amplitudes, but a cautious estimate suggests that F_3 was 6-10 dB below F_2 during the transitional portion. In the present stimuli, on the other hand, F_3 was at least 12 dB below F_2, which may explain the difference in results. Clearly, the relative amplitude of a formant determines its salience as a perceptual cue.

Dichotic Stimulus Dominance Functions

Figure 2 shows the stimulus dominance functions separately for the two stimulus sets (solid vs. dashed lines). The results for the six subjects have been averaged in this figure. The left-hand panel shows the percentage of B responses for the combinations of stimulus 1 (/bæ/) with stimuli 1-15 (+1 pairs); the center panel shows the percentage of D responses for the combinations of stimulus 8 (/dæ/) with stimuli 1-15 (+8 pairs); and the right-hand panel shows the percentage of G responses for the combinations of stimulus 15 (/gæ/) with stimuli 1-15 (+15 pairs). The locations of the category boundaries for the binaural stimuli (Figure 1) are indicated by arrows in Figure 2. Ear of presentation was ignored in this analysis, and all data were collapsed over this factor.

Consider first the panel labeled B(+1). The solid line describes the extent to which stimulus 1 (/bæ/) dominated the other stimuli on the rising-F_3 continuum, as reflected in the percentage of B responses. In combinations with stimuli 1-5, which themselves were identified as B in binaural presentation, there were, of course, almost 100 percent B responses. This portion of the function is of little interest. From stimulus 6 on, B responses fell below 100 percent, indicating that the dichotic stimuli were phonetically conflicting. However, for all stimulus combinations on the rising-F_3 continuum, the percentage of B responses remained substantially above 50 percent, the level of perceptual equilibrium. Thus, stimulus 1 was perceptually dominant in all these dichotic pairs.

The prototype matching model predicted that the stimulus dominance function for stimulus 1 would exhibit a peak at the location of the D-G category boundary. There was indeed a clear peak in the dominance function (solid line); however, it occurred at stimulus locations 11 and 12, and thus fell somewhat to the right of the D-G boundary (located at 10.02). Stimulus 10, which was ambiguous between /dæ/ and /gæ/, was less dominated by stimulus 1 than stimuli 11 and 12, which received predominantly G responses in binaural presentation. Thus, while the existence of a peak in the function would seem to support the prototype model, its precise location does not agree with the predictions.

In the dominance function under consideration, a peak at the B-D boundary cannot be discerned because of the high response ceiling to the left of the boundary. Nevertheless, the prototype matching hypothesis does predict that stimuli ambiguous between /bæ/ and /dæ/ should be strongly dominated by a good /bæ/. Indeed, the pairing of stimulus 1 with the ambiguous stimulus 6 (66.7 percent B responses binaurally) received 96.8 percent B responses, which indicates that stimulus 6 was almost completely dominated by the less ambiguous stimulus 1. This is in agreement with the prototype matching hypothesis.

Turning now to the second function in the B(+1) panel of Figure 2 -- that for the stimuli with a flat F_3 (dotted line) -- we note first the drastic reversal in the relative dominance of stimulus 1. In all combinations with stimuli 7-15, stimulus 1 was the less dominant component, and B responses constituted a minority. The large effect of the change in F_3 on the dichotic response frequencies is surprising in view of the fact that it reduced the binaural identifiability of stimulus 1 as B by only 6.5 percent (from 100 to 93.5 percent). The same change in F_3 reduced the percentage of B responses to dichotic pairings of stimulus 1 with phonetically conflicting stimuli by as much as 60 percent.

As in the rising-F_3 function, a peak appeared at stimulus locations 11 and 12 in the flat-F_3 function also. This peak was rather small but, because of the shift in the D-G boundary to the right (now located at 10.69), it was somewhat closer to the predicted location. The most ambiguous stimulus at the B-D boundary, stimulus 5 (56.9 B responses binaurally), was strongly dominated by stimulus 1 (88.0 percent B responses, as compared with 93.5 percent for stimulus 1 binaurally). This lends support to the prototype matching hypothesis.

Consider now the center panel in Figure 2, which shows the dominance functions (percentage of D responses) for stimulus 8 (/dæ/) paired with all others. In the rising-F_3 set, stimulus 8 was dominated by stimuli 1-4 (/bæ/) but in approximate perceptual equilibrium with stimuli 12-15 (/gæ/). When F_3 was flat (dotted line), on the other hand, stimulus 8 dominated all other stimuli on the continuum, despite the fact that a natural /dæ/ typically has a falling F_3 transition. The change in dominance was especially pronounced in competition with stimuli 1-4, in agreement with the results shown in the B(+1) panel.

Since both category boundaries are adjacent to the critical response category, D, we cannot look for distinct peaks in the D(+8) functions, but stimulus 8 nevertheless should have strongly dominated ambiguous stimuli in the boundary regions. Taking the clearest example, D responses to stimulus 10 with a rising F_3 changed from 57.4 percent binaurally to 78.2 percent when fused with stimulus 8 (as compared with 88.4 percent to stimulus 8 binaurally). This indicates only moderate dominance of stimulus 8 over the ambiguous stimulus 10, so that no strong support for the prototype matching model can be derived from this portion of the data.

Finally, consider the G(+15) dominance functions in the right-hand panel of Figure 2. Stimulus 15 (/gæ/) with a rising F_3, tended to be dominated by /bæ/ stimuli but was in approximate equilibrium with /dæ/ stimuli (solid line). Elimination of the F_3 transition made /gæ/ weaker in competition with /dæ/ stimuli but slightly stronger in competition with /bæ/ stimuli (dotted line). Both dominance functions were decidedly nonmonotonic in the region of the B-D boundary, as predicted by the prototype matching hypothesis: they exhibited broad peaks whose centers coincided approximately with the boundary locations. (The solid function did not show a peak but a plateau; however, this may be interpreted as a peak superimposed on a steep gradient.) Ambiguous stimuli in the D-G boundary region were strongly dominated by stimulus 15. For example, stimulus 10 with a rising F_3 received 41.2 percent G responses binaurally, but 92.1 G responses when fused with stimulus 15.

It would be good to know whether the peaks in the dominance functions were significant, or whether they were simply due to variability in the data. Unfortunately, there is no easy significance test, since the overall shapes of the stimulus dominance functions cannot be predicted; moreover, they varied from subject to subject. Instead of a numerical test, a qualitative summary of the individual data is presented in Table 2. This summary indicates that the peaks were reasonably consistent across subjects.

In summary, the pattern of the data supports the predictions of the prototype matching model fairly well. The constant stimulus in each panel of Figure 2 strongly dominated most ambiguous stimuli adjacent to the critical category, and there were peaks roughly in the region of the distant category boundaries in all four functions where such peaks had been predicted. Only the precise location of these peaks was not always as predicted. While this fact requires explanation, we note for the time being that the non-monotonic shapes of the dominance functions are not easily explained by reference to the purely auditory properties of the stimuli. Further discussion of this issue will be postponed until some secondary results have been described.

Psychoacoustic Fusion

One apparent contradiction in the data may be noted by comparing the results for the stimulus pair 1+15 in the B(+1) panel of Figure 2 with those for the same pair in the G(+15) panel. There was a large effect of the change in F_3 on B responses, but no effect at all on G responses. The solution lies in the differential occurrence of D responses. D responses to dichotic combinations of stimuli heard as B and G, respectively, in isolation, have been ascribed to "psychoacoustic fusion" by Cutting (1976). These responses were infrequent when F_3 was rising, but quite frequent when F_3 was flat. Of course, this is in agreement with the general increase in D responses to stimuli with a flat F_3. Thus, the stimulus pair 1+15 received 9.7 percent D responses when F_3 was rising, but 32.2 D responses when F_3 was flat. By comparison, the percentages of D responses to stimuli 1 and 15 in isolation were 0 and 0.5, respectively, when F_3 was rising, and 5.1 and 2.8, respectively, when F_3 was flat. Thus, the effect of F_3 on the frequency of psychoacoustic fusion responses, like its effect on stimulus dominance, was much larger than its effect on binaural identification scores. This reflects the increased perceptual weight of F_3 in the presence of conflicting F_2 transitions.

TABLE 2: Peaks in stimulus dominance functions in vicinity of distant category boundaries: qualitative summary of individual data.

Subject	B(+1)		G(+15)	
	Rising F_3	Flat F_3	Rising F_3	Flat F_3
BHR	large (11-12)	small (10-11)	small (6)	small (5)
KH	large (11-12)	floor	none	small (4)
NK	ceiling	small (11-12)	large (4-5)	large (3-5)
SM	ceiling	small (11)	hump (6-7)	large (4-6)
WT	large (11-15)	small (12)	hump (4-7)	small (4-7)
JK	hump (11-12)	none	large (3-5)	small (5-6)

Note: Numbers in parentheses are stimulus locations at which peaks occurred. A "large" peak generally is an elevation of 20 percent or more. A "hump" is a peak superimposed on a steep gradient (as in the rising-F_3, G(+15) function in Figure 2). Cases where no peak could be distinguished are indicated by "ceiling," "floor," or "none," depending on the overall level of the function in the critical region; when a function was at the ceiling, a peak may have existed but could not be observed.

TABLE 3: Ear dominance indices.

Subject	Rising F_3		Flat F_3	
	e'	e	e'	e
BHR	+0.04	+0.01	-0.01	+0.03
KH	+0.05	0.00	-0.11	-0.08
NK	+0.28	+0.25***	+0.30	+0.29***
SM	+0.13	+0.15*	+0.17	+0.17***
WT[a]	+0.22	+0.21***	+0.20	+0.17***
JK	-0.08	-0.06	-0.07	-0.06
Average	+0.11	+0.08	+0.08	+0.09

[a] Left-handed.

*$p < .05$

***$p < .001$

Ear Dominance

Ear dominance effects were of secondary interest in the present study, but they deserve a brief discussion, especially since fused syllables offer certain methodological advantages in measuring lateral asymmetries (Repp, 1977b). It seemed important to establish that the present stimuli yield significant (right-)ear dominance effects, that these effects are not sensitive to changes in F_3, and that they do not vary too much among individual stimulus pairs. Positive evidence on all three counts is a prerequisite for fused dichotic syllables to be methodologically useful in assessing hemispheric dominance for speech perception. Significant right-ear dominance effects have been obtained by Repp (1976) with very similar stimuli.

The individual ear dominance effects are shown in Table 3, separately for the two stimulus sets. Two ear dominance indices are reported, e' and e. (For a detailed discussion of these, see Repp, 1977b.) Their values are generally very similar. While e' is preferable on theoretical grounds, only the e index can easily be tested for significance; therefore, both are reported here. The significance test is based on a weighted standard error derived from the variability of ear asymmetries across individual stimulus pairs. Both indices take stimulus dominance effects into account; the e' index, especially, represents an unbiased estimate of ear dominance. Its value, like that of the e index, ranges from -1 (perfect left-ear dominance) to +1 (perfect right-ear dominance) and represents the linearly scaled intercept of a bilinear ROC (isolaterality) function with the negative diagonal of the unit square. (See Repp, 1977b.)

Table 3 shows that three subjects (including one left-hander) were significantly right-ear dominant, while the other three subjects showed no significant ear asymmetry. The ear dominance coefficients for the two stimulus sets differing in F_3 were generally very similar, the correlation being +0.91 for e' and +0.96 for e. Thus, the change in F_3 did not affect the degree of ear dominance. The variability among individual stimulus pairs was considerable, however. For example, subject NK's right-ear advantage (the largest of all) was based on 32 stimulus pairs, 11 of which showed no ear dominance or left-ear dominance. (Many other stimulus pairs received only a single type of response and therefore did not enter into the estimate of e'.) Of the 11 stimulus pairs that contributed most to the estimate of NK's ear dominance (because of their small stimulus dominance effects), two still showed left-ear dominance. It cannot be determined from the present data whether this was merely due to large variability, or whether it represented a genuine difference in ear dominance among individual stimulus pairs.

The conclusion from these results is that fused dichotic syllables (place contrasts) do tend to show a right-ear dominance effect, although these effects are smaller than those observed with dichotic voicing contrasts (Repp, 1977a, 1978a). Ear dominance seems to be highly variable, so that a large number of trials is needed to obtain a precise estimate. Changes in F_3 do not affect the ear asymmetry.

DISCUSSION

The present experiment has shown that stimulus dominance functions obtained by pairing a constant stimulus with all other stimuli along a place-of-articulation continuum exhibit local peaks in the vicinity of category boundaries. Such peaks were predicted by the prototype matching model outlined in the Introduction. However, the peaks were not located exactly at the boundaries but typically somewhat more toward the ends of the continuum. Some preliminary explorations with a formalized version of the prototype matching model suggest that its quantitative fit to the data is far from satisfactory. Of course, the quantification of such a complex model requires numerous assumptions and decisions whose consequences still need to be explored. Until a more refined quantitative analysis is available, the present results merely point toward, but cannot be taken as direct support for, the prototype matching model.

The possibility that the peaks in the dominance functions were determined by purely auditory factors must be given serious consideration. However, if there is an auditory explanation, it will not be a simple one. It has been hypothesized that phonetic category boundaries generally coincide with points of natural psychoacoustic discontinuity (for example, Kuhl and Miller, 1978). However, in the case of place-of-articulation distinctions, there is as yet no clear evidence in support of this hypothesis (see Bailey, Summerfield and Dorman, 1977). If the hypothesis were true, it might be argued that auditory cues are less salient in the vicinity of phonetic boundaries because they are less salient at psychoacoustic boundaries. However, there is no direct support even for this latter contention.

An explanation in auditory terms is complicated by the fact that the present boundaries, as well as the peaks in the dominance functions, were not symmetrically located with respect to the stimulus with a flat F_2 transition, stimulus 7 on the continuum (see Figures 1 and 2): the B-D boundary was closer to stimulus 7 than the D-G boundary. The psychoacoustic boundaries for detecting the presence and direction of rising and falling F_2 transitions would be expected to be roughly equidistant from the stimulus with no transition at all. Of course, it is possible, even likely, that these boundaries critically depend on the frequencies and trajectories of other formants that are simultaneously present. For example, the rising F_1 transition in the present stimuli may have made rising F_2 transitions easier to detect than falling F_2 transitions, although this seems somewhat counter-intuitive. (I am not aware of any psychoacoustic experiments demonstrating auditory interactions of this sort.) On the other hand, perceptual boundaries for phonetic distinctions invariably agree with acoustic relationships obtained in natural speech production. Therefore, the hypothesis that identification and dichotic competition of speech sounds are somehow mediated by relationships to internal prototypes retains its plausibility.

A complete model of dichotic competition needs to account not only for local features of the stimulus dominance functions, such as peaks and plateaus, but for their entire shape. Simple auditory hypotheses, such as "rising transitions dominate falling transitions (or vice versa)" or "steep transitions dominate flat transitions (or vice versa)" find little support in the present data. A change in the (constant) F_3 transition changed the whole shape of the dominance function. For example, when F_3 was rising, /dæ/ (stimulus 8) dominated /bæ/ stimuli more than /gæ/ stimuli, but the opposite was true when F_3 was flat. Again, one might appeal to complex auditory interactions in this case. However, further complications arise from the considerable individual differences in the shapes of dominance functions, and from the finding that superficially rather similar stimuli may show radically different dominance relationships (Repp, 1978b). The prototype model seems inherently more flexible to handle these effects than an auditory model: the simplified synthetic tokens of a /bæ/, /dæ/ and /gæ/ may differ in the degree to which they approach the natural prototypes; for example, the absence of an initial burst may be more critical in the case of /dæ/ than in the case of /bæ/. There may be individual differences in the nature of perceptual prototypes, perhaps reflecting the perceiver's own articulatory habits, that could account for the large individual differences in dominance functions. Clearly, however, the determinants of the overall shape of the dominance functions are not well understood at present.

A study related to the present one has been reported by Whitaker and Porter (1976). They paired a constant "target" syllable (drawn from a /ba/-/da/-/ga/ continuum) with a number of stimuli varying in their F_2 transitions. In contrast to the present study, these dichotic competitors were not full syllables but consisted of F_2 in isolation -- stimulus patterns known as "bleats" in the literature. When presented simultaneously with a full syllable, bleats provide quite effective perceptual competition. In their preliminary report, Whitaker and Porter report only a small fraction of their data, so that it cannot be determined to which degree their results support the present findings. Apparently, they did not find significant peaks in the dominance functions. This result does not necessarily contradict the prototype matching model, one version of which assumes that the stimuli in each ear, despite their phonemonal fusion, are separately related to the relevant prototypes before the dichotic combination of information. Since bleats are sufficiently removed from the relevant phonetic prototypes to show little variation in their relative category goodness (or rather, category poorness) as a result of acoustic changes, they would show no loss in competitive power at points corresponding to category boundaries for full syllables. On the other hand, if the F_2 transition of a bleat were fused and combined with that of a target syllable before the prototype matching stage, results similar to those of the present study would be expected. Thus, a comparison of full syllables and bleats as dichotic competitors may provide information about the precise form that a prototype matching model should take.

REFERENCES

Bailey, P. J., A. Q. Summerfield and M. F. Dorman. (1977) On the identification of sine-wave analogues of certain speech sounds. Haskins Laboratories Status Report on Speech Research SR-51/52, 1-27.

Berlin, C. I., S. S. Lowe-Bell, J. K. Cullen, Jr., C. L. Thompson and C. F. Loovis. (1973) Dichotic speech perception: An interpretation of right-ear advantage and temporal offset effects. Journal of the Acoustical Society of America 53, 699-709.

Cutting, J. E. (1976) Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. Psychological Review 83, 114-140.

Finney, D. J. (1971) *Probit Analysis.* 3rd ed. (Cambridge: University Press).

Harris, K. S., H. S. Hoffman, A. M. Liberman, P. C. Delattre and F. S. Cooper. (1958) Effect of third-formant transitions on the perception of voiced stop consonants. *Journal of the Acoustical Society of America 30*, 122-126.

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

Kuhl, P. K. and J. D. Miller. (1978) Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America 63*, 905-917.

Miller, J. L. (1977) Properties of feature detectors for VOT: The voiceless channel of analysis. *Journal of the Acoustical Society of America 62*, 641-648.

Oden, G. C. (in press) Integration of place and voicing information in the identification of synthetic stop consonants. *Journal of Phonetics*.

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

Repp, B. H. (1976) Identification of dichotic fusions. *Journal of the Acoustical Society of America 60*, 456-469.

Repp, B. H. (1977a) Dichotic competition of speech sounds: The role of acoustic stimulus structure. *Journal of Experimental Psychology (Human Perception and Performance) 3*, 53-70.

Repp, B. H. (1977b) Measuring laterality effects in dichotic listening. *Journal of the Acoustical Society of America 62*, 720-737.

Repp, B. H. (1978a) Stimulus dominance and ear dominance in the perception of dichotic voicing contrasts. *Brain and Language 5*, 310-330.

Repp, B. H. (1978b) Categorical perception of fused dichotic syllables. *Haskins Laboratories Status Report on Speech Research SR-55/56*.

Studdert-Kennedy, M., D. Shankweiler and D. B. Pisoni. (1972) Auditory and phonetic processes in speech perception: Evidence from a dichotic study. *Cognitive Psychology 3*, 455-466.

Whitaker, R. G. and R. J. Porter, Jr. (1976) Dichotic and monotic masking of CV's by second formants with different transition starting values. *Journal of the Acoustical Society of America 60* (Supplement No. 1), S119(A).

Categorical Perception of Fused Dichotic Syllables

Bruno H. Repp

ABSTRACT

Synthetic syllables from a /ba/-/da/-/ga/ continuum were paired with either of the two endpoint stimuli (/ba/,/ga/) in dichotic or mixed presentation. The resulting fused hybrid stimuli were presented in identification and AXB discrimination tests. The hybrids were perceived quite categorically, and there was little difference between dichotic and mixed modes of presentation. These results do not replicate earlier data (Repp, 1976b) that had suggested that discrimination of dichotic hybrid stimuli might be based on a level of representation preceding phonetic categorization. The possibility remains that such an earlier level is accessed only when the stimuli are highly ambiguous; they were less so in the present experiment than in Repp's earlier study.

INTRODUCTION

One of the most reliable findings in speech perception research is the categorical perception of stop-consonant-vowel syllables varying in the formant transitions that cue place of articulation. Syllables from such a "place continuum" invariably exhibit sharp category boundaries in a labeling task, and their discrimination is little better than predicted under the assumption that all a listener retains of a stimulus is its phonetic label (Liberman, Harris, Hoffman and Griffith, 1957; Pisoni[1]).

However, Repp (1976b) recently reported a curious result that does not fit the customary pattern. His stimuli were hybrid syllables composed of two different inputs to the two ears. The syllables were taken from a place continuum and thus contrasted only in their initial formant transitions. Dichotic pairs of such syllables fuse into a single stimulus perceived as originating in the center of the head, and the listener has no indication that two different inputs have occurred (Repp, 1976a). Repp (1976b) presented these fused stimuli in a discrimination task and compared the obtained performance with predictions obtained from earlier identification data for the

[1] Pisoni, D. B. (1971) On the nature of the categorical perception of speech sounds. Unpublished Ph.D. dissertation, University of Michigan.

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to the Haskins Laboratories. I would like to thank Georgann Witte for her help in analyzing the data.

same stimuli (Repp, 1976a). In the discrimination task, the stimulus in one ear was held constant, so that the difference to be detected occurred only in one component of the fused stimuli. Performance was predicted to be very poor, since fusion with a constant stimulus greatly reduced the phonetic distinctiveness of the stimuli. However, the obtained discrimination scores were much better than predicted and, moreover, exhibited characteristic peaks that coincided with those obtained in a control condition in which the stimuli were presented binaurally (not fused with a constant stimulus). This seemed to provide an instance of categorical perception without clearly defined categories--a paradoxical result.

This finding was especially interesting, since it could be explained by the author's theory of dichotic competition (Repp, 1976a, 1977a, 1978). This theory assumes that the integration of information from the two ears takes place at a level intermediate between auditory and phonetic representations. At this intermediate level, termed "multicategorical," a stimulus is represented as a vector whose elements are the perceptual distances of the stimulus from the several relevant category prototypes in auditory space. Dichotic fusion of two stimuli is assumed to result in the averaging of their multicategorical vectors. If it is further assumed that the discrimination of two stimuli is based on the distance between their vectorial representations, the paradoxical results of Repp (1976b) can be explained. Obviously, the relative distances among the members of a stimulus set remain unchanged when all stimuli are fused with the same constant stimulus; only the absolute magnitude of the distances decreases. This corresponds to a decrease in overall discriminability, without any change in the relative discriminability of different stimulus pairs, as reflected in the peaks and troughs of the discrimination function. This is precisely what Repp (1976b) obtained.

Unfortunately, there were some procedural problems in these earlier experiments. Repp (1976b) realized that the temporal alignment of the dichotic stimuli contained a random error that may have led to artifacts in the discrimination task. He replicated the experiment, taking great pains to align the stimuli on the two tape tracks as precisely as possible, and obtained essentially the same results. However, it has since transpired that, unbeknownst to him, the specific procedure he used in the replication experiment (two-channel output of stimuli digitized at a 20-kHz sampling rate) did not function properly at the time and may have resulted in intensity and quality differences on the two tape channels. It would be difficult to explain how the quite different technical problems in the two experiments could have led to similar patterns of results that, moreover, resembled those obtained in a binaural control condition. Nevertheless, it seemed advisable to conduct another replication experiment that was free from all previous procedural problems. That was the purpose of the present study.

The present experiment included an important change in the stimulus set used. Although superficially similar to the old set (/ba/, /da/, /ga/ instead of /bæ/, /dæ/, /gæ/) and chosen primarily for its more natural qualities, the new stimulus set led to a profound change in the response pattern to dichotic stimuli. As will become clear below, this change resulted in the failure to replicate one of the basic features of the previous study--the poorly defined category boundaries for fused stimuli. Thus, the present experiment was not an exact replication. Nevertheless, it provided a valid and procedurally clean test of the hypothesis suggested by the earlier data.

The present experiment went beyond the previous ones by including a condition in which the syllables were not dichotically fused but electronically mixed and presented binaurally. A comparison of dichotic and mixed conditions can provide important information about the level at which dichotic stimuli interact. Halwes[2], who was the first to compare dichotic and mixed stimulus pairs, obtained quite different patterns of results in the two conditions. However, when only syllables contrasting in place of articulation are considered, the differences between dichotic and mixed conditions seem less striking (Repp, 1976a). They consist in moderate shifts in the relative weights of individual stimuli in perceptual competition. The present study is the first to compare dichotic and mixed stimuli in a discrimination task.

METHOD

Subjects

The subjects were eight paid volunteers recruited from Yale University. All had participated in at least one earlier experiment using synthetic speech and had proven to be good listeners.

Stimuli

The stimuli were eight syllables from a "place continuum" ranging from /ba/ to /da/ to /ga/, created on the OVEIIIc serial resonance synthesizer at Haskins Laboratories. All syllables were 295 msec in duration and had a constant fundamental frequency of 94 Hz. They had no release bursts and differed only in the transitions of the second and third formants (F_2 and F_3) that occupied the first 40 msec. The transition onset frequencies and steady-state frequencies of F_2 and F_3 are shown in Table 1. In addition, each stimulus had a transition in F_1 that rose from 285 Hz to a steady state of 771 Hz; F_4 and F_5 were hardware-fixed. All transitions were stepwise linear in 5-msec time segments.

The stimuli were digitized at 8 kHz using the Haskins Laboratories Pulse Code Modulation (PCM) system. The onset of the first sampling period was time-locked to stimulus onset, as was the occurrence of the first pitch pulse in synthesis. Two dichotic tapes were prepared: one for identification, the other for discrimination.

The identification tape contained 5 blocks of 42 stimuli each. These 42 stimuli were a random sequence of 16 identical pairs (two presentations of each stimulus paired with itself) and 26 non-identical pairs (all pairings of stimulus 1 with stimuli 2-8, and of stimulus 8 with stimuli 2-7, in both channel assignments). The interstimulus interval (ISI) was 3 sec and blocks were separated by 6 sec.

[2] Halwes, T. G. (1969) Effects of dichotic fusion on the perception of speech. Unpublished Ph.D. dissertation, University of Minnesota.

TABLE 1: Onset frequencies and steady states of F_2 and F_3 in the stimuli used.

Stimulus	F_2 (Hz)	F_3 (Hz)
1	859	1795
2	1037	2150
3	1224	2502
4	1404	2998
5	1588	2998
6	1770	2502
7	1770	2197
8	1770	1902
Steady states	1233	2520

TABLE 2: Ear dominance coefficients.

Subject	Identification		Discrimination
	e'	e	
1	-0.38	-0.32^{*}	-0.40
2	0.00	+0.02	-0.15
3	+0.67	$+0.39^{*}$	+0.14
4	-0.10	0.00	0.00
5	+0.59	$+0.57^{***}$	-0.01
6	-0.05	-0.14	+0.16
7	+0.63	$+0.40^{**}$	+0.13
8	+0.23	$+0.28^{***}$	-0.03

$^{*}p < .05$

$^{**}p < .01$

$^{***}p < .001$

In the discrimination task, an AXB paradigm was used. This procedure differs from the more commonly used ABX paradigm in that the first and the third stimulus are always different from each other, while the second stimulus is identical with either the first or the third.[3]

The discrimination tape contained four blocks of AXB triads. Blocks 1 and 4 contained only identical pairs of stimuli, that is, each individual stimulus in a triad consisted of the same stimulus recorded on both channels. There were the same 50 AXB triads in each of these blocks, including all seven one-step (1 vs. 2, 2 vs. 3, etc.) and all six two-step (1 vs. 3, 2 vs. 4, etc.) stimulus comparisons in all four AXB arrangements (AAB, ABB, BAA, BBA). By mistake, two AXB triads (1-3-3 and 7-5-5) were omitted, reducing the number of triads per block from (7 + 6) x 4 = 52 to 50. Blocks 2 and 3 constituted a single series of 208 AXB triads, divided by a pause in the middle. These 208 triads resulted from the following design: the same 52 stimulus triads as in Blocks 1 and 4 (that is, all one-step and two-step comparisons in all AXB arrangements, with no omissions) occurred on one channel, while a constant stimulus occurred on the other channel. The constant stimulus was either stimulus 1 or stimulus 8, and it could occur on either of the two channels, so that, all in all, there were four times as many triads as in Blocks 1 or 4 (4 x 52 = 208). The ISIs were 500 msec within triads and 3 sec between triads.

Procedure

Each subject was tested in two 2-hour sessions. In one of these sessions (the dichotic condition), the two channels of the tapes were directed to different ears. In the other session (the mixed condition), the two channels were electronically mixed and presented binaurally. The sequence of the dichotic and mixed conditions was counterbalanced across subjects. Since the stimuli on the two channels were exactly simultaneous and in phase, all stimulus pairs were perceived as single syllables originating in the center of the listener's head, both in dichotic and in mixed presentation. In each session, the identification tape was presented first and repeated once, after the tape recorder channels had been electronically reversed. The discrimination tape was also repeated once, again with channels being reversed prior to repetition.

The subjects were fully informed about the nature of the stimuli. They were instructed to respond with B, D, or G in the identification task, guessing if necessary. In the discrimination task, the response was to be A if the second stimulus was identical to the first and B if the second stimulus

[3] The AXB paradigm was chosen to prevent the listener strategy of attempting to compare the first with the third stimulus--a strategy that is likely to be ineffective because the auditory traces of the first stimulus may be lost by the time the third arrives. Thus, the AXB paradigm avoids the memory limitations that may reduce performance in the ABX paradigm (Pisoni and Lazarus, 1974). Although there is, at present, no empirical evidence that the AXB paradigm is superior to the ABX paradigm, there is no reason to believe that it would be inferior. In fact, the AXB procedure may be considered a condensed version of the 4IAX paradigm, which has been shown to lead to higher discrimination performance than the ABX paradigm (Pisoni, 1973; Pisoni and Lazarus, 1974).

was identical to the third; again, guessing was required in the case of uncertainty.

The tapes were played back on an Ampex AG-500 tape recorder, and the subjects listened over Telephonics TDH-39 headsets. The electronic mixer was built at Haskins Laboratories. The amplitudes of the stimuli in the two channels were carefully equalized at a comfortable listening level, using a Hewlett-Packard voltmeter. In the mixed condition, the output was attenuated by 10 dB after mixing; this made the amplitudes approximately equal in the dichotic and mixed conditions.

RESULTS

Identification: Identical Pairs

The responses to pairs of identical stimuli constituted the baseline identification data for the eight syllables. The labeling probabilities were expected to be unaffected by the mode of presentation--binaural or mixed. Figure 1 shows that this was true. In this figure, the top panels are for the dichotic condition, and the bottom panels are for the mixed condition; the three panels in each row represent B, D, and G responses, respectively, as a function of stimulus number. The data for identical pairs are represented by the solid lines. It can be seen that the dichotic and mixed results were practically identical, and that the two category boundaries--between B and D and between D and G--were unusually sharp. There was a complete switch from B to D percepts between stimuli 3 and 4, and the change from D to G between stimuli 6 and 7 was almost as abrupt. In view of the fact that these represent the average results of 8 subjects, the consistency of the labeling responses is quite remarkable; they reflect the high quality of the synthetic stimuli.

Identification: Non-identical Pairs

The labeling results for non-identical pairs are shown as the broken lines in Figure 1. The dashed line represents pairings of the stimuli on the abscissa with the constant stimulus 1 (+1 pairs); the dotted line represents pairings with the constant stimulus 8 (+8 pairs).

Stimulus 1 was perceived as B in isolation. Therefore, it was expected that the labeling probabilities for +1 stimulus pairs would be biased toward B, relative to the labeling probabilities for identical pairs. The extent of the bias reflects the degree to which stimulus 1 perceptually dominated phonetically conflicting stimuli. Figure 1 shows that stimulus 1 was a weak dichotic competitor: B responses to its pairings with stimuli 5-8 reached only 20 percent, on the average, indicating that B was strongly dominated by both D and G. Only in the pairing with stimulus 4, approximate equilibrium of responses was reached (40 percent B responses, 60 percent D responses). Mode of presentation affected the pattern of responses: in the mixed condition, B was completely dominated by D and G, so that there were practically no B responses to pairings of stimulus 1 with stimuli 4-8.

Turning now to the +8 pairs, the most relevant information is contained in the right-hand panels of Figure 1, which show G responses. Since stimulus

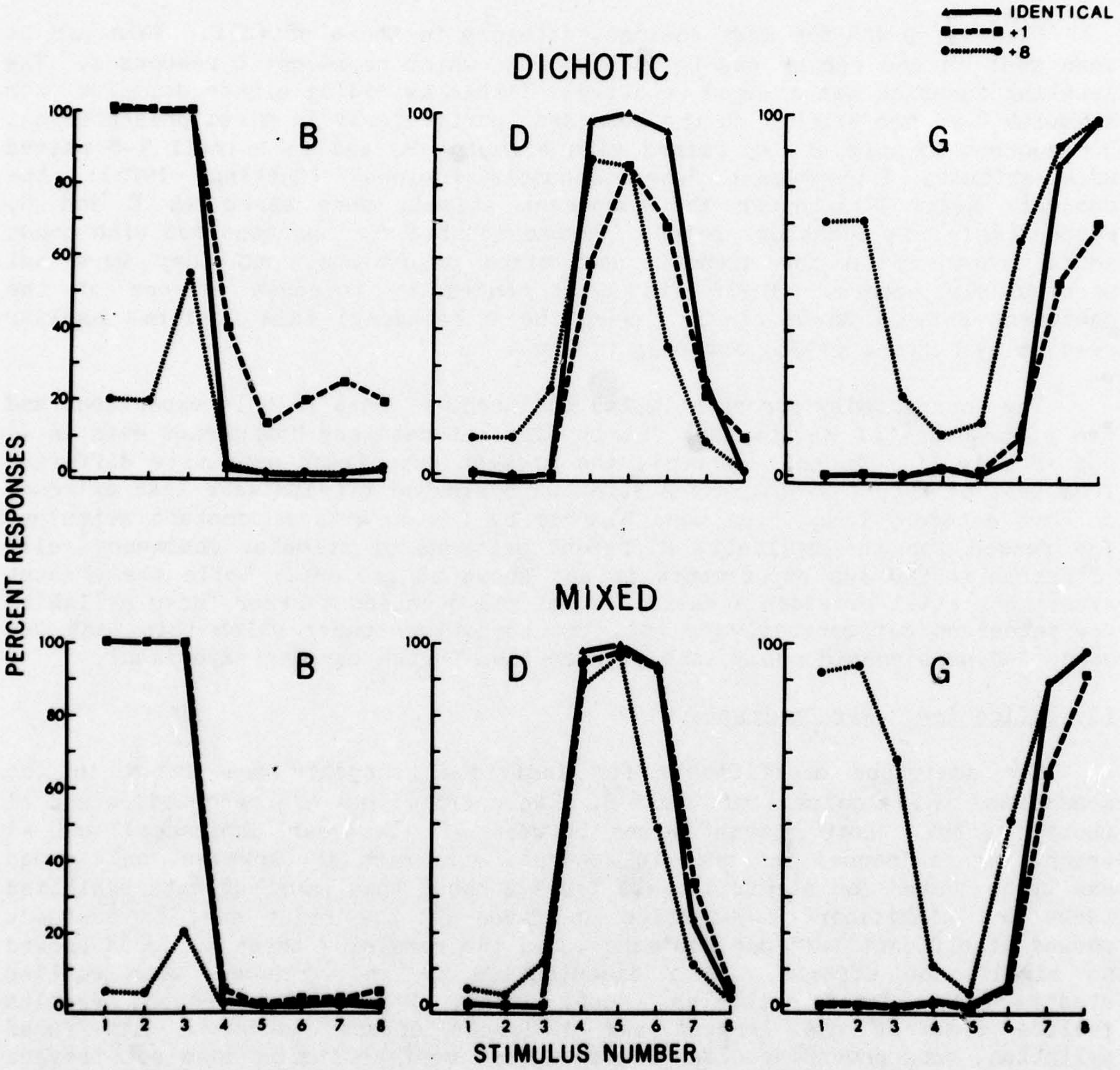


Figure 1: Percentages of B, D, and G responses to identical and non-identical stimulus pairs in dichotic and mixed presentation.

8 was heard as G, it was expected to increase the number of G responses when paired with other stimuli, relative to the labeling function for these other stimuli in identical pairs. It can be seen that this was the case, but the extent of the increase in G responses varied widely depending on the nature of the competing stimulus. More specifically, G (stimulus 8) dominated B (stimuli 1 and 2), but it was dominated by D (stimuli 4 and 5). This resulted in a U-shaped function for G responses. In mixed presentation, this trend was even more pronounced than in dichotic presentation.

Clearly, D was the most dominant category in these stimuli. This can be seen best in the center panels of Figure 1, which represent D responses. The labeling function was changed relatively little by adding either stimulus 1 or stimulus 8 to the stimuli on the abscissa, particularly in mixed presentation. D responses to stimuli 1-3 paired with stimulus 8, and to stimuli 7-8 paired with stimulus 1 represent "psychoacoustic fusions" (Cutting, 1976): the subjects heard D although the component stimuli were heard as B and G, respectively, in identical pairs. Psychoacoustic fusions occurred with about equal frequency in the dichotic and mixed conditions, and they were not particularly common. Their frequency tended to increase as one of the component stimuli moved closer toward the D category; this confirms earlier results by Cutting (1976) and Repp (1976a).

The unexpectedly strong stimulus dominance effects in this experiment had the consequence of maintaining fairly distinct category boundaries even in +1 and +8 stimuli. In this respect, the present experiment was quite different from that of Repp (1976b), where stimulus dominance effects were less extreme, so that category boundaries were blurred by fusion with a constant stimulus. The reason for the radically different patterns of stimulus dominance relationships in the two experiments is not known at present. While the present experiment still provides a valid test of the question whether fused syllables are perceived categorically or not, the conditions under which this test was conducted were considerably less extreme than in the earlier experiment.

Identification: Ear Dominance

Ear dominance coefficients for individual subjects are shown in the second and third columns of Table 2. Two coefficients are reported, e and e' (Repp, 1977b). Both assume values between -1 (left-ear dominance) and +1 (right-ear dominance) and are, in general, very similar; however, only e can easily be tested for significance. Table 2 shows that four subjects exhibited large and significant asymmetries in favor of the right ear, one subject showed significant left-ear dominance, and the remaining three subjects showed no significant effects. This distribution is in agreement with earlier studies using fused syllables (Repp, 1976a, 1978). The present results include some of the largest ear dominance effects observed with fused syllables, thus providing clear evidence that perfect fusion does not prevent lateral asymmetries. To which extent these asymmetries actually reflect hemispheric dominance for speech perception is a question that the present experiment cannot answer, but the higher incidence of right-ear dominance suggests the involvement of speech-specific mechanisms.

Discrimination: Identical Pairs

The discrimination of identical pairs was expected to follow the familiar pattern of categorical perception: high performance across category boundaries and low performance within categories. Because of the unusually sharp category boundaries, this pattern was expected to be especially pronounced in the present data. Predictions were derived from the labeling probabilities using the standard formula given in Pollack and Pisoni (1971). This formula assumes that all the information the listener has available are the category labels of the stimuli, and that the stimuli in an AXB triad are categorized independently. The predictions were computed separately for each subject and then averaged.

The left-hand panels in Figure 2 show obtained and predicted discrimination scores in the dichotic and mixed conditions, respectively. In each panel, there are four functions: obtained one-step discrimination scores (triangles, solid line), obtained two-step scores (circles, solid line), predicted one-step scores (triangles, dotted line), and predicted two-step scores (circles, dotted line). It can be seen that the predicted extreme peaks and valleys in the discrimination functions were indeed obtained. The match between predicted and obtained functions was generally good, with obtained performance being somewhat higher than predicted, which is a common finding. Performance in the dichotic and mixed conditions was extremely similar, as predicted. These results indicate that the stimuli were perceived highly categorically.

Discrimination: Non-identical Pairs

The remaining panels of Figure 2 show the predicted and obtained discrimination results for +1 and +8 pairs in the dichotic and mixed conditions. The functions are exactly analogous to those for identical pairs, except that the stimuli to be discriminated were fused with a constant stimulus. Consider first the +1 condition. Since, as we have seen, stimulus 1 was perceptually dominated by all other stimuli and thus had only a very small effect on labeling probabilities in +1 pairs, discrimination scores in +1 pairs were also expected to be fairly similar to those for identical pairs. The data confirmed these expectations. While discrimination scores were generally somewhat lower than for identical pairs, particularly in the dichotic condition, the pattern of results was quite similar, and the match between predicted and obtained +1 discrimination functions was quite good, particularly in the dichotic condition. Again, obtained scores exceeded predicted scores by some margin, probably reflecting an auditory memory component.

Because of the marginal effect that fusion with stimulus 1 had on performance, the +1 condition could hardly serve to test the model of dichotic competition outlined in the Introduction. However, the +8 condition provided a better opportunity to do so. As can be seen in the right-hand panels of Figure 2, predicted and obtained scores for +8 pairs deviated from those for identical pairs, due to the larger effect that fusion with stimulus 8 had on perception. We note that, again, the fit between obtained and predicted functions was reasonable, although somewhat less convincing than with the other stimulus pairs.

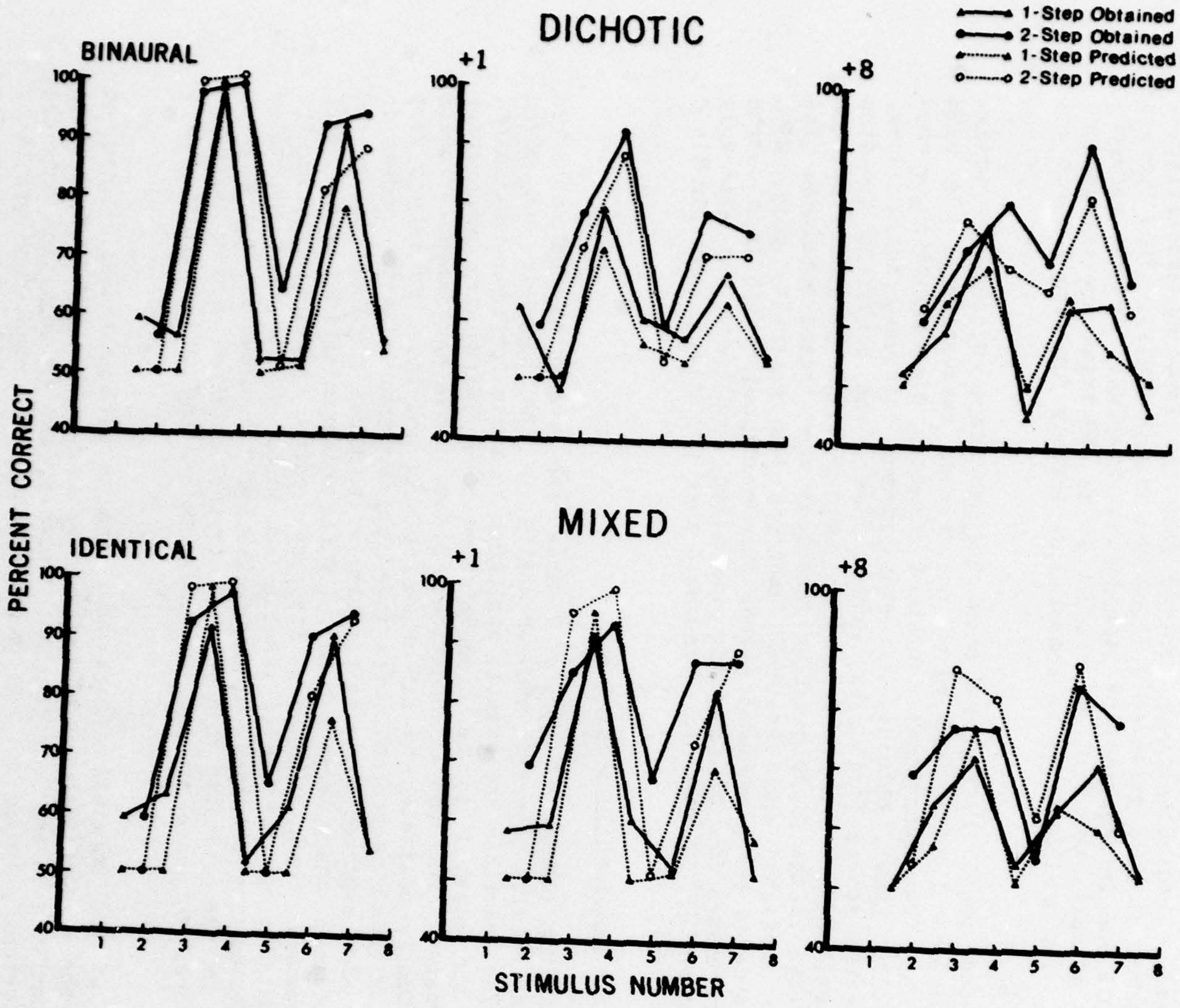


Figure 2: Predicted and obtained 1-step and 2-step percent-correct discrimination scores for identical (binaural) and non-identical stimulus pairs in dichotic and mixed presentation. Data points are plotted midway between the stimulus components to be discriminated.

According to the hypothesis that fused stimuli are perceived categorically, there should be no systematic deviations between predicted and obtained discrimination functions, except perhaps for a slight difference in overall level, due to auditory memory. On the other hand, the hypothesis, outlined in the Introduction, that discrimination is based on "multicategorical" stimulus representations predicts that the obtained +8 discrimination functions should match the pattern for identical pairs (shifted to a lower level of accuracy) more closely than the pattern predicted for +8 pairs. In order to decide this issue, we need to search for instances where the predicted functions for identical and +8 pairs have different trends. For example, scores for the 2-4 comparison were predicted to be higher than for the 3-5 comparison in the +8 condition, while equal performance, or perhaps even higher performance for 3-5, was predicted in identical pairs. The results in both the dichotic and mixed +8 conditions were closer to the latter than to the former. Another, especially clear instance is the relationship between comparisons 5-7 and 6-8, for which opposite directions were predicted in identical and in +8 pairs. As can easily be seen in Figure 2, in this case the results in the +8 condition are closer to the +8 predictions. There are three relationships between one-step comparisons that may be similarly examined. For two of these (1-2 vs. 2-3, and 4-5 vs. 5-6), the +8 results go with the +8 predictions. For the third (5-6 vs. 6-7), another very clear difference between identical and +8 pairs, the +8 results deviate from the +8 predictions in the direction of the results for identical pairs, but they do not nearly approach the extreme difference observed there.

In summary, the data fail to provide consistent evidence that obtained +8 scores are closer to those for identical pairs than to the predicted +8 scores. Although there are some trends in this direction, the overall evidence is too weak to reject the categorical perception hypothesis. It must be concluded that the fused stimuli in the present experiment were, in essence, perceived categorically in both dichotic and mixed modes of presentation.

Discrimination: Ear Dominance

Ear dominance coefficients for the discrimination task are reported in column 4 of Table 2, where they may be compared with the coefficients obtained in the identification task.[4] The coefficients in column 4 correlated +0.66 ($p < .05$) with e' and +0.55 ($p < .10$) with e, indicating some consistency in individual ear asymmetries across different tasks. What is more striking,

[4] The coefficients computed from the discrimination scores are not strictly equivalent with e' and e in the identification task, although they also range from -1 to +1. The coefficient used is $(R^*-L^*)/(R^*+L^*)$, where $R^* = 2(R-50)$ and $L^*=2(L-50)$, and R and L are the percent-correct scores for the two ears. Thus, it incorporates a very crude correction for guessing, but, unlike e' and e, it does not properly take into account the variability between individual stimulus pairs and therefore may underestimate the size of the ear dominance effects (cf. Repp, 1977b). A better index of ear dominance in a discrimination task remains to be worked out.

however, is that there was no longer any overall tendency toward right-ear dominance in the discrimination task. Two of the subjects who had been strongly right-ear dominant in identification lost this asymmetry completely in discrimination. This indicates that even ear asymmetries for fused syllables may be sensitive to task characteristics (cf. Haggard, 1976).

DISCUSSION

The main conclusion of the present experiment is that the fused syllables were categorically perceived. Whatever deviations occurred from the predicted performance pattern were not systematic enough to warrant interpretation. Thus, the listeners in the present experiment apparently relied on phonetic category labels in making their discrimination responses, regardless of whether the syllables were binaural singles or dichotic hybrids.

The contrasting results of Repp (1976b) had suggested that an earlier, multicategorical representation of fused dichotic stimuli can be accessed in a discrimination task. The possibility remains that such an earlier level can be utilized only when the stimuli are highly ambiguous at the categorical phonetic level. As pointed out above, this prerequisite was not sufficiently met in the present experiment. There is increasing evidence in the literature that categorical perception can be transcended by leading subjects to ignore phonetic categories and focus on auditory differences, either by long practice or by presenting only stimuli from a single phonetic category [for example, Carney, Widin and Viemeister, (1977); Ganong[5]]. Thus, in order to assess the nature of earlier levels of stimulus representation, phonetic categorization must somehow be prevented or de-emphasized. This hypothesis might be further tested by presenting fused hybrid syllables in some of the experimental paradigms designed to reduce the role of phonetic categorization.

By indicating that the subjects did not make use of an earlier, multicategorical stimulus code, the present results do not imply that such a level does not exist. Therefore, the failure to replicate the findings of Repp (1976b) cannot be taken as evidence against the hypothesis that information from the two ears is combined at such a multicategorical stage (Repp, 1976a, 1977a, 1978). One detail of the present data, however, is relevant to that hypothesis; and, unfortunately, it is not favorable. The hypothesis predicts that stimuli close to a phonetic category boundary should be weaker in dichotic competition than stimuli farther removed from a boundary (cf. Repp, 1976a, 1978). Figure 1 shows, however, that stimulus 3 was stronger in competition with stimulus 8 than stimuli 1 and 2 (the peak in the dotted functions in the left-hand panels of Figure 1). While such a detailed result may not be sufficient to reject the multicategorical hypothesis, it does add to the increasing evidence that the underlying model does not fit the detailed structure of dichotic data very well (Repp, 1978). A good alternative model is needed.

[5] Ganong, W. F. III. (1977) Selective adaptation and speech perception. Unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

REFERENCES

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

Cutting, J. E. (1976) Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. Psychological Review 83, 114-140.

Haggard, M. P. (1976) Dichotic listening. Speech Perception (Series 2) 5, 40-55. (Belfast: Department of Psychology, The Queen's University of Belfast).

Liberman, A. M., K. S. Harris, H. S. Hoffman and B. C. Griffith. (1957) The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology 54, 358-368.

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

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

Pollack, I. and D. B. Pisoni. (1971) On the comparison between identification and discrimination tests in speech perception. Psychonomic Science 24, 299-300.

Repp, B. H. (1976a) Identification of dichotic fusions. Journal of the Acoustical Society of America 60, 456-469.

Repp, B. H. (1976b) Discrimination of dichotic fusions. Haskins Laboratories Status Report on Speech Research SR-45/46, 123-139.

Repp, B. H. (1977a) Dichotic competition of speech sounds: The role of acoustic stimulus structure. Journal of Experimental Psychology (Human Perception and Performance) 3, 53-70.

Repp, B. H. (1977b) Measuring laterality effects in dichotic listening. Journal of the Acoustical Society of America 62, 720-737.

Repp, B. H. (1978) Stimulus dominance in fused dichotic syllables. Haskins Laboratories Status Report on Speech Research SR-55/56.

Stimulus Dominance and Ear Dominance in Fused Dichotic Speech and Nonspeech Stimuli

Bruno H. Repp

ABSTRACT

The patterns of stimulus dominance (the tendency of one stimulus to perceptually dominate another, competing stimulus) and ear dominance in fused dichotic stimuli were compared among five stimulus sets of decreasing speechlikeness: two-formant CV syllables, bleats (F_2 only), transitions (F_1 and F_2 transitions only), chirps (F_2 transition only), and timbres (brief steady-state F_2 resonances). The patterns of stimulus dominance relationships showed significant similarities across the five stimulus sets, the highest correlation being obtained between CV syllables and timbres. Ear dominance varied widely between stimulus sets and between individuals, but followed no systematic pattern. Thus, these data provide no evidence that either ear dominance or stimulus dominance in fused dichotic stimuli is a function of the degree to which the stimuli resemble speech.

INTRODUCTION

Research on speech perception has led to several findings that, for some time, appeared specific to the perception of speech sounds. These include, among others, categorical perception and the dichotic right-ear advantage (Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967). More recently, however, it has become apparent that certain nonspeech stimuli lead to very similar results; for example, tones differing in rise-time are perceived categorically (Cutting and Rosner, 1974), and although they do not exhibit a right-ear advantage when presented dichotically (Cutting, Rosner and Foard, 1975), they do so if a monaural reaction time procedure is used (Blechner, 1976). Findings such as these have reopened the question about the determinants of several phenomena previously believed to be specific to speech.

Two such phenomena have been selected for study in the present experiment. Both are reliably observed in dichotic listening to speech sounds. One is the ear advantage or ear dominance effect; it is usually--that is, in the

Acknowledgment: This research was supported by NIH Grant DE00202 to the University of Connecticut Health Center, and by NICHD Grant HD01994 and BRS Grant RR05596 to the Haskins Laboratories. I would like to thank Virginia Mann and Dick Pastore for helpful comments on the manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-55/56 (1978)]

[redacted]

majority of subjects--in favor of the right ear. A right-ear advantage (REA) has been taken to indicate that the stimuli under investigation are treated preferentially by the left hemisphere. While the effect is obtained for most speech sounds, there are some, such as isolated steady-state vowels, that yield little or no ear dominance (Godfrey, 1974; Shankweiler and Studdert-Kennedy, 1967), as well as certain nonspeech sounds that do lead to right-ear dominance (Halperin, Nachshon and Carmon, 1973; Cutting, 1974; Blechner, 1976). It is possible, therefore, that it is not the speechlikeness of a stimulus per se that creates the right-ear advantage, but certain auditory characteristics (for example, rapid frequency change) that, while typical for speech, are also shared by certain nonspeech stimuli.

The other phenomenon--which, in fact, constitutes the primary focus of the present study--is dichotic stimulus dominance. This is the finding that, when two specific stimuli are presented in dichotic competition, one of them often tends to dominate the other, regardless of the ear to which it is presented. Such stimulus dominance effects seem to be especially pronounced when the dichotic stimuli are fused, so that the listener hears only a single stimulus that sounds more like the dominant component than like the nondominant component (Repp, 1976). A priori, there is no reason why stimulus dominance should be specific to speech, although, at present, little is known about the extent of such effects for nonspeech sounds. However, Repp (1976, 1978) studied stimulus dominance effects in detail, using a set of stop-consonant-vowel syllables differing in the cues for place of articulation, and observed a pattern of stimulus dominance relationships that seemed difficult to explain on purely auditory grounds. Rather, he suggested, what seems to determine the relative dominance of a stimulus is its "category goodness," as defined by its perceptual distance from a number of relevant category prototypes. According to this hypothesis, the pattern of stimulus dominance effects within a set of speech stimuli is, at least in part, speech-specific; nonspeech stimuli would not be perceived in relation to phonetic categories and therefore would show a different (probably simpler) pattern determined by auditory factors.

The approach taken in the present study follows that taken in several studies of categorical perception: the perception of full CV syllables is compared with the perception of nonspeech stimuli derived from these CV syllables by deleting some constant (nondistinctive) acoustic component. For example, Mattingly, Liberman, Syrdal and Halwes (1971) compared CV syllables varying in the second-formant (F_2) transition with F_2 transitions presented in isolation (nonspeech sounds commonly referred to as "chirps"). In the present study, five different kinds of stimuli were used: two-formant CV syllables varying in the F_2 transition, and four kinds of nonspeech sounds derived from these syllables (see Figure 1). These four types of stimuli were: "bleats" (F_2 only), "transitions" (F_1 and F_2 transitions only), "chirps" (F_2 transitions only), and "timbres" (brief steady-state resonances). These stimuli were successively more unlike the original CV syllables, and hence more and more nonspeechlike. The purpose of the experiment was to investigate whether any systematic changes in stimulus dominance relationships or in ear dominance accompany the progression from speech to nonspeech.

There is little previous research on the dichotic competition of stimuli such as those just described, except for several unsuccessful attempts to determine ear advantages for chirps[1]. While earlier attempts apparently failed because of the difficulty listeners had in consistently assigning labels to different chirps, the most recent experiment[2] failed because of extreme stimulus dominance effects that made ear advantages for both CVs and chirps extremely unreliable. Both these problems are avoided in the present experiment. It uses an AXB paradigm, that does not require labeling of the stimuli, and the CV stimuli, at least, have been used in previous experiments and are known not to lead to complete dominance of one stimulus over another (Repp, 1976, 1978).

Two possible criticisms of the present research shall be met head-on before describing the method in more detail. First, experiments using chirps and similar "nonspeech controls" for CV syllables have been criticized for neglecting the fact that the psychoacoustic characteristics of the distinctive cues are changed by removing some component of the signal. For example, the finding that CV syllables are perceived categorically while chirps are not (Mattingly et al., 1971) may have a purely psychoacoustic reason: perceptual discontinuities in transition perception may arise only when a constant reference signal (a steady-state resonance) follows[3]. Naturally, this argument also applies to the stimuli used here, but it seems less critical to the purpose of the investigation. The present experiment is concerned with the relative strengths of two cues (different F_2 transitions) in dichotic competition, given different acoustic environments; and to the degree that both cues are similarly affected by a change in environment, no effect on dichotic stimulus dominance would be expected. Of course, highly complex psychoacoustic interactions can never be ruled out a priori.

Second, not all the "nonspeech" stimuli described above are nonspeech in the strictest sense. (Nor, for that matter, are the rather primitive, two-formant CV syllables speech in the strictest sense.) Rather, the stimuli form a conceptual continuum from speech to nonspeech; that is, it becomes increasingly more difficult to apply phonetic labels to them. The CV syllables can be classified by most listeners as /bæ/, /dæ/ and /gæ/. Bleats, while perceived as nonspeech by most listeners--and frequently described plainly as nonspeech stimuli in the literature--definitely resemble syllables beginning with the homorganic nasal stops, /mæ/, /næ/ and /ŋæ/. Even the transitions in isolation can, with some effort, be related to the plosive noises of a /b/, /d/ or /g/. Only chirps and timbres sound definitely nonspeechlike. The use of phonetic labels for CVs, bleats and transitions was not rigorously controlled in the present study, but it was generally encouraged since it facilitated the listeners' rather difficult task.

[1] Liberman, A. M.: personal communication.

[2] Hagenow, Vanessa. (1976) Senior essay, Yale University.

[3] Pastore, R.: personal communication.

METHOD

Subjects

Five subjects participated in the full experiment. These were the author (an experienced listener), a colleague who had relatively little experience as a listener but was highly motivated, and three paid volunteers who had proven to be careful listeners in earlier similar experiments. The author did the experiment twice, so that one complete replication of his data was available. (The results of his two runs were combined before computing group averages.) Five additional paid volunteers who were less experienced in listening to synthetic speech participated only in the two easiest conditions, involving CV syllables and timbres.

Stimuli

All stimuli were created on the Haskins Laboratories parallel resonance synthesizer. The CV syllables were similar to those used by Repp (1976, 1978) in earlier studies of dichotic fusion and competition, except that they had only two formants and a more gradual rise in amplitude at onset than the earlier stimuli. There were seven syllables ranging from /bæ/ to /dæ/ to /gæ/, distinguished only by their F_2 onset frequencies and transitions. The F_2 onset frequencies are listed in Table 1. All F_2 transitions were approximately linear in 5-msec steps and reached the steady state of 1620 Hz after 35 msec. All stimuli had a constant F_1 transition that rose from 181 Hz to 743 Hz in 45 msec, a constant fundamental frequency (114 Hz), a constant amplitude contour with initial and final ramps, and a constant duration of 260 msec.

Each of the other four stimulus sets consisted of seven stimuli derived from the CV syllables. They are illustrated schematically in Figure 1. The seven *bleats* were created by completely removing F_1 from the CV syllables, so that F_2 alone remained. For the *transitions*, F_1 was retained, but all stimuli were cut off after the eighth time frame, resulting in a duration of only 40 msec--just enough to include the F_2 transition. *Chirps* were created by eliminating F_1 from the transitions. Thus, chirps stand in the same relation to transitions as bleats to CV syllables (F_1 absent or present), and chirps stand in the same relation to bleats as transitions to CV syllables (short vs. long duration, or absence vs. presence of steady state). Finally, the *timbres* were steady-state F_2 segments of 40 msec duration, corresponding to the seven F_2 onset frequencies of the chirps (see Table 1 and Figure 1).

All stimuli were digitized at 8 kHz using the Haskins Laboratories pulse code modulation system (PCM). For each stimulus type, a tape containing exactly the same randomized stimulus sequences was recorded. These sequences consisted of a series of examples, a binaural discrimination sequence, and three dichotic "similarity judgment" sequences. The initial examples consisted of the seven stimuli of a given series played three times in ascending order, with an interstimulus interval (ISI) of 3 sec. The binaural discrimination sequence contained 60 AXB triads of binaural stimuli in a random arrangement. The 60 triads resulted from the five 2-step comparisons (stimuli

TABLE 1: Onset frequencies of F_2.

Stimulus	F_2 onset (Hz)
1	1232
2	1386
3	1541
4	1695
5	1845
6	1996
7	2156

TABLE 2: Average percent correct discrimination in the binaural discrimination task.

Stimulus set	Stimulus pairs					
	1-3	2-4	3-5	4-6	5-7	Mean
First five subjects						
Timbres	100.0	100.0	95.0	98.3	100.0	99.0
Chirps	77.5	83.3	88.3	85.0	69.2	80.7
Transitions	75.8	84.2	86.7	82.5	72.5	80.3
Bleats	70.0	74.2	83.3	81.7	79.2	77.7
CV syllables	64.2	82.5	91.7	95.8	85.8	84.0
Second five subjects						
Timbres	91.7	98.3	100.0	93.3	93.3	95.3
CV syllables	65.0	75.0	90.0	85.0	68.3	76.7

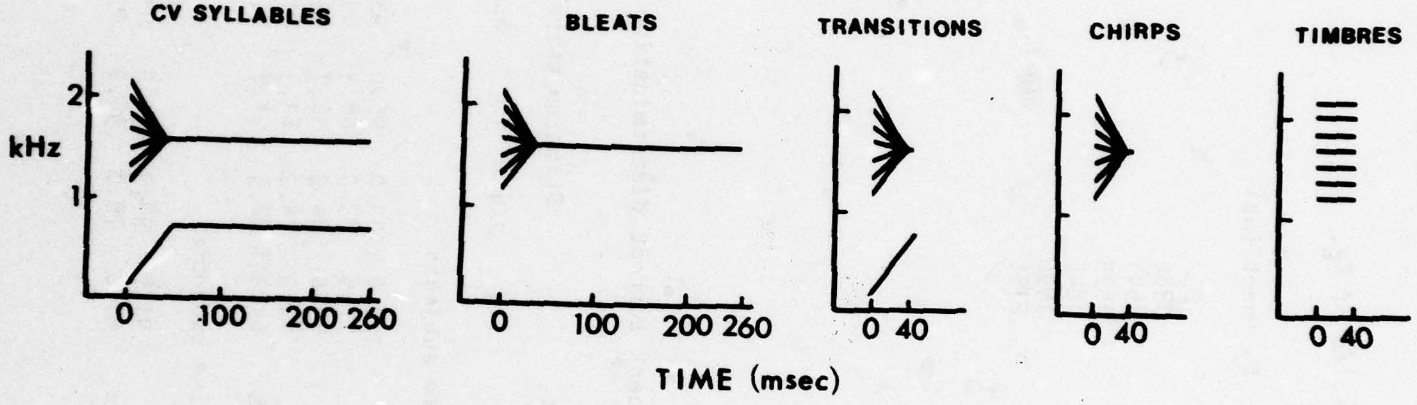


Figure 1: Schematic spectrograms of the five sets of stimuli used in the experiment.

1-3, 2-4, 3-5, 4-6, 5-7) presented three times in each of the four AXB configurations (AAB, ABB, BAA, BBA). The ISI was 1 sec within triads and 3 sec between.

Each of the three dichotic sequences contained a different randomization of 60 AXB triads. In each triad, stimuli A and B were presented binaurally, while stimulus X consisted of A and B presented simultaneously to the two ears. The 60 triads resulted from all combinations of stimuli at least two steps apart--15 in all--presented in each of four possible AXB configurations (A[AB]B, A[BA]B, B[AB]A, B[BA]A, where the order of the stimuli within brackets refers to the two different earphone channels). The ISIs were the same as in the binaural sequence. Because of perfect dichotic fusion, the middle stimulus in dichotic triads did not sound qualitatively different from the other two (binaural) stimuli; this was true for all five stimulus types.

Procedure

The subjects were tested in a quiet room. The tapes were played back over an Ampex AG-500 tape recorder, and the subjects listened over calibrated Telephonics TDH-39 earphones. At the beginning of each session, the amplitudes of the stimuli on the two tape recorder channels were carefully equalized, using a series of prerecorded calibration signals (prolonged / / vowels). All tapes were played back at the same level; the effective loudness of the different types of stimuli varied according to their temporal and spectral characteristics.

Each experimental condition required a one-hour session. Thus, the full experiment consisted of five sessions spread out over a number of days. All subjects listened to the conditions in the order of increasing speechlikeness, starting with timbres and ending up with CV syllables. In a given session, the subject first listened to the sample continuum, after having been informed about the nature of the stimuli. Then the binaural sequence was presented, and the subject was told to write down "A" whenever X equalled A, "B" whenever X equalled B, and to guess when uncertain. For the dichotic AXB sequences, the five subjects who took the reduced version of the experiment (timbres and CV syllables only) were told to write down "A" whenever X was more similar to A than to B, and "B" whenever X was more similar to B than to A. The five subjects who participated in all five conditions were asked to use a 6-point rating scale instead. The six numbers stood for: "very similar to A," "similar to A," "slightly more similar to A than to B," "slightly more similar to B than to A," "similar to B," "very similar to B." It was expected that the ratings might reveal more categorical perception (that is, frequent use of extreme ratings) for the more speechlike stimuli, but the data were not clear on that point. Therefore, the numerical ratings were collapsed into two response categories (1-3, 4-6) equivalent to those used by the other five subjects in the experiment.

In each session, the three dichotic sequences were repeated after a pause during which the tape recorder channels were reversed electronically. Thus, each subject gave six responses to each individual AXB triad, which yielded a total of 24 responses for each stimulus combination.

RESULTS

Binaural Discrimination

The results of the binaural discrimination task are shown in Table 2, separately for the two groups of subjects. Although the task was primarily intended to familiarize the listeners with the stimuli and did not yield enough observations to warrant statistical analysis, the scores do contain some relevant information. Clearly, the timbres were easiest to discriminate; only a few errors were committed. The other four types of stimuli were considerably more difficult, although all individual stimulus combinations were discriminated well above chance (50 percent correct), on the average. There were no large differences in difficulty among these four types of stimuli, and the order of difficulty of the individual stimulus pairs was similar within each condition: the 1-3 and 5-7 discriminations were usually the most difficult, and the 3-5 and 4-6 discriminations were the easiest. These observations suggest that discrimination was easier when the F_2 transitions were relatively flat (in the middle of a continuum) than when they were steep (at the ends).

The listeners' ability to discriminate the stimuli binaurally set a limit to their accuracy in the dichotic task. However, the dichotic sequences included not only two-step comparisons (as in the binaural task), but also combinations of stimuli separated by three steps or more, whose discriminability was expected to be higher than that of the two-step pairs. Also, the subjects' sensitivity to stimulus differences probably increased in the course of a session, thus increasing their accuracy in the dichotic task. Nevertheless, it is important to keep in mind that low discriminability of the A and B stimuli in a dichotic AXB triad leads to random guesses and consequent reductions in ear dominance and stimulus dominance effects for such triads.

Dichotic Stimulus Dominance

The patterns of stimulus dominance relationships in the five stimulus sets are shown in Figures 2 and 3. Figure 2 shows the results for timbres and CVs, separately for the two groups of five subjects. Figure 3 shows the results for the remaining three conditions participated in by the first group of five subjects. In each panel, the dependent variable is the percentage of trials on which the dichotic stimulus was judged to be more similar to the component stimulus with the lower number on the continuum (that is, with the lower F_2 onset frequency). Thus, data points above the 50-percent "equilibrium" line indicate low-frequency dominance, while data points below the equilibrium line indicate high-frequency dominance. In each panel, the data points are arranged into five groups whose members are connected by lines. Each group contains all combinations of a constant stimulus with stimuli two or more steps higher on the continuum. Thus, group 1 (squares) contains all pairings of stimulus 1 with stimuli 3-7, group 2 (circles) contains all pairings of stimulus 2 with stimuli 4-7, and so on.

Consider first the results for timbres, shown in the upper left-hand panel of Figure 2. Since timbres were easiest to discriminate (Table 2), these results were expected to be the most reliable. It can be seen that the data points fell into an orderly pattern. First of all, there was a strong

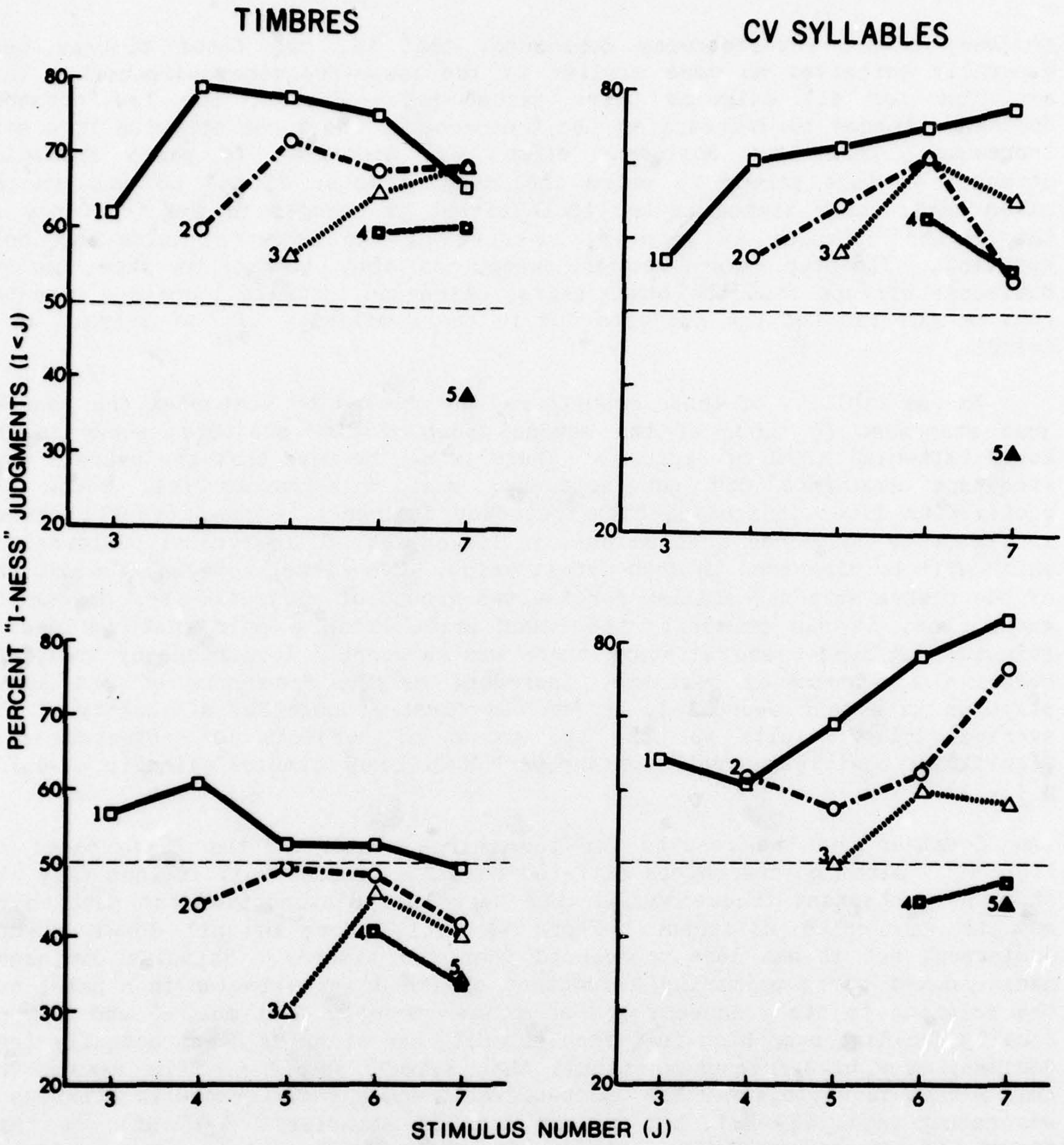


Figure 2: Stimulus dominance patterns for timbres and CV syllables, shown separately for two groups of five subjects (upper vs. lower panels). Each data point represents a dichotic pair of two stimuli, I and J (I < J), where I is indicated in the graph and J on the abscissa. On the ordinate is the percentage of "I-ness" judgments, i.e., the percentage of trials on which the dichotic stimulus was judged to be more similar to the lower component stimulus (I).

tendency toward low-frequency dominance, that is, two fused timbres were generally perceived as more similar to the lower-frequency component. This was true for all stimulus pairs except 5-7. Second, the low-frequency dominance tended to decrease as the frequency of the lower stimulus in a pair increased. Thus, the dominance effect was strongest in pairs including stimulus 1, less strong in pairs including stimulus 2, and so on. On the other hand, there seemed to be little effect of changes in the frequency of the higher stimulus in a pair, as long as the lower stimulus was held constant. Two-step pairs are an exception; they tended to show smaller dominance effects than the other pairs, which may indicate increased uncertainty on the part of the subjects due to the similarity of the stimuli in a triad.

The reliability of these results may be checked by comparing the results just discussed to those of the second group of five subjects, shown in the lower left-hand panel of Figure 2. There it can be seen that the overall low-frequency dominance did not hold up; most data points fell below the equilibrium line, indicating high-frequency dominance. Thus, overall trends in frequency dominance must be due, in large part, to individual preferences, which will be discussed in more detail below. Otherwise, however, the pattern of the timbre data was similar for the two groups of subjects. For the second group, too, it was primarily the lower stimulus in a pair that influenced stimulus dominance, and although there was no overall low-frequency bias, the relative low-frequency dominance increased as the frequency of the lower stimulus in a pair decreased, as for the first group. The similarity of the average timbre results for the two groups of subjects is reflected in a significant positive correlation across the fifteen stimulus pairs ($r = +0.70$, $\underline{p} < .01$).

Consider now the results for the chirps, shown in the first panel of Figure 3. Although the chirps differed from the timbres only in that they all ended at a constant frequency and thus were less distinctive, the pattern of results was quite different. There was still some overall low-frequency dominance, but it was less pronounced than for timbres. Stimulus dominance again seemed to be primarily a function of the lower stimulus in a pair, but the relation to its frequency was no longer orderly. Stimuli 1 and 2 were equally dominant over high-frequency stimuli, and stimulus 3 was actually less dominant over high-frequency stimuli than stimuli 4 and 5. The reason for this result is not clear. The onset-offset frequency difference in stimulus 3 was rather small (79 Hz), but so was that in stimulus 4 (-75 Hz), so that amount of frequency change per se does not seem to have been a factor influencing stimulus dominance. The correlation between the chirp and timbre results for the first group of subjects was +0.43 ($\underline{p} > .05$).

The results for the transitions are shown in the middle panel of Figure 3. These stimuli were, subjectively at least, the most difficult to judge, and the data therefore contained a considerable amount of noise. That granted, some similarity to the earlier two conditions can be detected, although overall low-frequency dominance was practically absent here. The transition results correlated moderately with the results for chirps ($r = +0.43$, $\underline{p} > .05$) and timbres ($r = +0.51$, $\underline{p} < .05$).

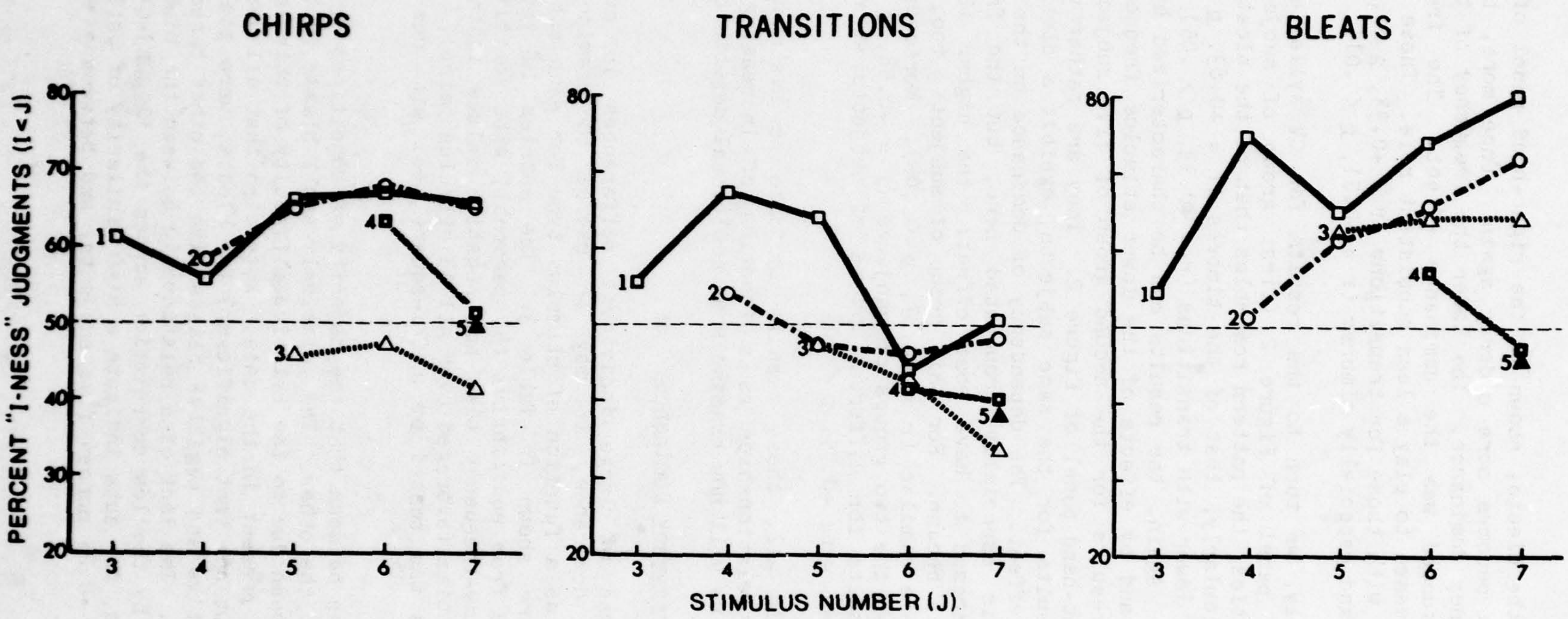


Figure 3: Stimulus dominance patterns for chirps, transitions, and bleats. First group of five subjects only.

With the bleats, shown in the right-hand panel of Figure 3, the pattern of results becomes more orderly again. Once more, there is strong overall low-frequency dominance. The lower the frequency of the lower stimulus, the more pronounced was the dominance effect. The frequency of the higher stimulus seemed to play a less consistent role. These results were positively correlated with those for transitions (r = +0.33, $\underline{p} > .05$), chirps (r = +0.37, $\underline{p} > .05$), and especially timbres (r = +0.71, $\underline{p} < .01$).

Finally, we turn to the results for CV syllables, shown in the upper right-hand panel of Figure 2 (first group of subjects). Apart from some irregularities, the pattern resembles that of the bleats (r = +0.75, $\underline{p} < .001$) and, particularly, that of the timbres (r = +0.83, $\underline{p} < .001$). The correlations were lower with transitions (r = +0.33, $\underline{p} > .05$) and chirps (r = +0.38, $\underline{p} < .05$). Again, the results can be characterized by overall low-frequency dominance and by effects of the lower stimulus frequency on dominance. The parallel results for the second group of five subjects may be found in the lower right-hand panel of Figure 2. They are rather orderly and, unlike the timbre results for the same subjects, exhibit a clear overall low-frequency dominance effect. The dependency of dominance on the frequency of the lower stimulus is especially pronounced here, but the frequency of the higher stimulus seemed to have some effect: the higher its frequency, the less dominant it became. For this group of subjects, too, the results of CVs and timbres were similar (r = +0.57, $\underline{p} < .05$). Moreover, the CV results were similar for the two groups of subjects (r = +0.58, $\underline{p} < .05$), and the timbre and CV results for different groups of subjects were also correlated (both correlations were +0.59, $\underline{p} < .05$).

All in all, these results indicate no systematic change in stimulus dominance relationships as a function of increasing speechlikeness of the stimuli. We will now examine some additional aspects of the data.

Overall Frequency Dominance

Because of large individual differences in overall low- vs. high-frequency dominance, it may be useful to examine changes in frequency dominance as a function of stimulus type for each subject separately. These results are shown in Table 3. The entries in the table are percentage deviations from equilibrium (50 percent), with positive values indicating an overall low-frequency bias and negative values indicating an overall high-frequency bias (averaged over all 15 stimulus pairs). The significance of the deviations was tested by a Chi-square test with one degree of freedom (N = 360).

It can be seen that the majority of conditions showed a significant bias one way or the other. The relatively small biases in the transition condition may have been due to the relative difficulty of this task, so that much random noise was present in the data. Note also that all dominance effects for CVs, and all but one (not significant) for bleats, were positive, while there were both positive and negative biases with the other three, less speechlike types of sounds. The lack of a relationship between the biases for timbres and CVs is shown by the low correlation across the 10 subjects: r = -0.20. On the other hand, the data indicate a high similarity of patterns between bleats and CVs (r = +0.85 across five subjects) and between timbres and chirps (except

TABLE 3: Overall frequency dominance (deviations from equilibrium in percent).

Subject		Stimulus set			
	Timbres	Chirps	Trans	Bleats	CVs
BHR (I)	4.4	-0.3	-6.9**	8.1**	18.6***
(II)	-5.0	1.9	2.8	12.2***	13.6***
DS	-17.5***	-11.7***	6.9**	18.9***	20.3***
DE	27.8***	23.1***	-3.6	9.2***	5.3*
CW	32.5***	20.0***	-1.1	26.7***	20.3***
JK	34.4***	-19.2***	-6.1*	-3.6	5.0
BW	16.1***	----	----	----	4.2
ER	-28.3***	----	----	----	6.1*
AL	-38.6***	----	----	----	20.0***
TM	8.3**	----	----	----	0.6
RG	18.9***	----	----	----	25.8***

Note: Positive values indicate low-frequency dominance.

$^{*}p < .05$
$^{**}p < .01$
$^{***}p < .001$

TABLE 4: Order effects (deviations from equilibrium in percent).

Subject		Stimulus set			
	Timbres	Chirps	Trans	Bleats	CVs
BHR (I)	-3.9	-0.8	9.2***	1.4	9.2***
(II)	1.7	2.5	7.8**	0.0	13.6***
DS	-5.8*	-2.2	3.1	2.2	10.8***
DE	-8.3**	-13.6***	-11.9***	-13.6***	-21.9***
CW	-5.8*	-2.8	5.0	7.8**	6.9**
JK	-2.2	-11.4***	-10.0***	4.2	-1.1
BW	2.2	----	----	----	5.3*
ER	-4.4	----	----	----	-2.2
AL	-8.1**	----	----	----	-9.4***
TM	-18.3***	----	----	----	2.8
RG	-1.1	----	----	----	0.8

Note: Positive values indicate a bias to judge the critical stimulus as being more similar to the following stimulus than to the preceding one.

$^{*}p < .05$
$^{**}p < .01$
$^{***}p < .001$

for subject JK). Thus, there is some evidence that overall frequency dominance changed with stimulus type, but in different ways for different listeners.

Order Effect

Another effect that deserves a brief examination is that of temporal order: a tendency to perceive the dichotic stimulus as more similar to either the first or the third stimulus in AXB triads, regardless of the identity of the stimuli. The results are shown in Table 4. As in Table 3, biases in one or the other direction are shown as a percentage deviation from equilibrium (50 percent); negative values represent a bias toward the preceding (first) stimulus, whereas positive values represent a bias toward the following (third) stimulus.

Again, a large number of significant individual biases and a wide variety of individual patterns can be seen. The biases for timbres and chirps were mostly negative, that is, these stimuli tended to be perceived as more similar to the preceding than to the following stimulus. Bleats and CVs, on the other hand, showed both positive and negative biases, depending on the individual listeners. As with frequency dominance, the temporal order bias for timbres and CVs was practically unrelated ($r = +0.26$), while those for adjacent stimulus types showed more similarity. Thus, order biases, too, seemed to have some relation to stimulus type, although the large individual differences make this relation difficult to describe.

Ear Dominance

Table 5 displays the individual *ear dominance* effects. Two values are reported for each condition: the deviation from equilibrium in percent (Table 5a), tested for significance by a Chi-square test, and the e' coefficient of ear dominance (Table 5b). The e' coefficient (Repp, 1977) ranges from -1 to +1 and corrects for constraints due to stimulus dominance (but not for those due to temporal order effects). The significance levels for the e' coefficient are estimates based on the variability across individual stimulus pairs; strictly speaking, they are the estimated significance levels of the related e coefficient, which usually assumes very similar values (Repp, 1977). For both indices, a negative value indicates left-ear dominance and a positive value right-ear dominance.

The pattern of results is bewildering. First, it must be recognized that, although fused CV syllables have tended to yield an overall right-ear advantage (REA) in the past (Repp, 1976, 1978), this trend is very weak here. Of the ten subjects, six showed a REA (three significant), whereas four showed a LEA (one significant). With all stimuli, significant ear dominance effects were the rule rather than the exception. This was true especially for timbres, where some of the largest asymmetries occurred. Of the ten subjects, three showed a REA for timbres (two significant), whereas seven showed a LEA (four significant). Thus, there may be a small trend to shift from LEAs to REAs with increasing speechlikeness of the stimuli, but, clearly, the effect does not approach statistical significance. In contrast to the frequency dominance and order effects discussed earlier, there was a sizeable correlation between ear asymmetries for timbres and CVs ($r = +0.54$, $\underline{p} < .10$);

TABLE 5a: Ear dominance effects (deviations from equilibrium in percent).

Subject		Stimulus set			
	Timbres	Chirps	Trans	Bleats	CVs
BHR (I)	15.0***	19.7***	0.8	7.5**	2.5
(II)	10.6***	14.2***	-2.8	1.7	3.1
DS	7.5**	0.0	3.1	8.3**	8.6**
DE	-8.9***	-5.8*	-3.1	-4.2	-5.3*
CW	-0.8	2.8	8.9***	2.8	-2.5
JK	-5.0	-7.5**	-11.7***	-4.7	-5.6*
BW[a]	-18.3***	----	----	----	0.3
ER	1.1	----	----	----	8.3**
AL	-1.9	----	----	----	2.8
TM	-15.0***	----	----	----	-8.3**
RG	-5.0	----	----	----	8.6**

[a] Left-handed.

Note: Positive values indicate right-ear dominance.

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

TABLE 5b: Ear dominance coefficients (e').

Subject		Stimulus set			
	Timbres	Chirps	Trans	Bleats	CVs
BHR (I)	+0.43***	+0.52***	+0.01	+0.21***	+0.11
(II)	+0.32**	+0.36***	-0.08	+0.02	+0.08
DS	+0.23**	-0.01	+0.07	+0.29***	+0.33***
DE	-0.42***	-0.25**	-0.10	-0.12*	-0.21
CW	0.00	+0.08	+0.23**	+0.15*	-0.13
JK	-0.45***	-0.19**	-0.40***	-0.12*	-0.16
BW[a]	-0.69***	----	----	----	-0.04
ER	+0.01	----	----	----	+0.22***
AL	-0.12	----	----	----	+0.16
TM	-0.38***	----	----	----	-0.21***
RG	-0.19*	----	----	----	+0.33***

[a] Left-handed.

Note: Positive values indicate right-ear dominance.

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

however, as far as one can tell from the results of five subjects, there was an even closer similarity between timbres and chirps and between bleats and CVs. Of the five subjects in the first group, four tended to show ear asymmetries in the same direction for all types of stimuli; two of these subjects were left-ear dominant and two were right-ear dominant.

It is interesting to note in passing that there was strong evidence from the results of several subjects with large ear asymmetries for timbres and chirps, that the magnitude of the ear dominance varied between stimulus pairs. Specifically, the ear asymmetry increased with the frequency of the higher stimulus in a pair. Thus, stimulus pairs containing stimulus 7 showed the largest ear dominance effects. No such tendency was observed with the more speechlike stimuli.

DISCUSSION

The present experiment provides little evidence for systematic changes in either stimulus dominance or ear dominance as a function of the degree to which the stimuli resemble speech. Obviously, the study cannot be considered definitive in view of the small number of subjects employed. However, the negative results are not due to excessive noise in the data, as attested by the numerous significant effects at the individual level. Rather, it is the enormous variability between individuals that prevents any general conclusions and that calls for more detailed study.

There was a significant similarity of stimulus dominance patterns across different sets of stimuli, particularly between CV syllables and timbres. This suggests that stimulus dominance effects arise at an auditory level and are primarily a function of frequency relationships. Such a conclusion would be contrary to the "prototype matching" hypothesis of stimulus dominance in speech sounds (Repp, 1976, 1978). However, the generality of the present results needs to be tested using a different set of CV syllables. Informal observations have indicated that syllables from a /ba/-/da/-/ga/ continuum show quite different stimulus dominance relationships than syllables from a /bæ/-/dæ/-/gæ/ continuum. If this is true, the high correlation between the results for CV syllables and timbres obtained here must have been accidental.

A comparison of ear dominance effects for different kinds of speech and nonspeech sounds has recently been reported by Divenyi and Efron (1978). Their results suggest that ear dominance does not change direction as long as the dichotic stimulus difference remains a spectral one; only when a temporal difference is introduced (such as in two syllables contrasting in voice-onset-time) is there a tendency to shift toward right-ear dominance. Their analysis, which is based on a similarly small number of subjects, would suggest that ear dominance for spectral differences, even when they are carried by speech stimuli, originates at a subcortical auditory level. The present data are not in contradiction with this view. Clearly, however, the issue requires further investigation. A partial replication of the present experiment using different stimuli is now in progress.

REFERENCES

Blechner, M. J. (1976) Right-ear advantage for musical stimuli differing in rise time. Haskins Laboratories Status Report on Speech Research SR-47, 63-69.

Cutting, J. E. (1974) Two left-hemisphere mechanisms in speech perception. Perception and Psychophysics 16, 601-612.

Cutting, J. E. and B. S. Rosner (1975) Categories and boundaries in speech and music. Perception and Psychophysics 16, 564-570.

Cutting, J. E., B. S. Rosner and C. F. Foard. (1975) Rise time in nonlinguistic sounds and models of speech perception. Haskins Laboratories Status Report on Speech Research SR-41, 71-93.

Divenyi, P. L. and R. Efron. (1978) Right-ear advantage for voicing in subjects left-ear dominant for pure tones. Journal of the Acoustical Society of America 63 (Supplement No. 1), S19-20 (A).

Godfrey, J. J. (1974) Perceptual difficulty and the right ear advantage for vowels. Brain and Language 1, 323-335.

Halperin, Y., I. Nachshon and A. Carmon. (1973) Shift of ear superiority in dichotic listening to temporally patterned nonverbal stimuli. Journal of the Acoustical Society of America 53, 46-50.

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

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

Repp, B. H. (1976) Identification of dichotic fusions. Journal of the Acoustical Society of America 60, 456-469.

Repp, B. H. (1977) Measuring laterality effects in dichotic listening. Journal of the Acoustical Society of America 62, 720-737.

Repp, B. H. (1978) Stimulus dominance in fused dichotic syllables. Haskins Laboratories Status Report on Speech Research SR-55/56.

Shankweiler, D. and M. Studdert-Kennedy. (1967) Identification of consonants and vowels presented to left and right ears. Quarterly Journal of Experimental Psychology 19, 59-69.

On Buzzing the English /b/*

Leigh Lisker+

ABSTRACT

The status of closure voicing as a necessary and/or sufficient property of the /bdg/ phonemes of American English has not yet been conclusively determined. Initially it is well established that glottal pulsing before release is not an essential property of the class; in fact it is quite normal for /bdg/ in this context to be voiceless stops, in the technical sense. Medially, and finally too, the role of glottal pulsing during closure is not entirely clear, perhaps because discussion has more often centered on the role of closure duration and the duration of a preceding vowel as determinants of stop labeling behavior. Evidence from experiments in the perception of edited natural speech indicates that the presence of closure buzz is a strong cue to /bdg/ in medial position, but that its absence does not invariably trigger "ptk" responses. For a word token in which presence/absence of closure buzz produced a shift in phoneme labeling, the effect of varying the intensity and within-closure duration of the buzz was determined. Results suggest that closure buzz must be attenuated more than 10 dB for it to be no longer a decisive cue to /bdg/, but that at a naturally produced intensity it may fill as much as one-half the duration of a long closure (140 msec) without eliciting predominantly "bdg" responses.

INTRODUCTION

The phonetic innocent trying to find out from the literature just what the basis is for partitioning the English stops into the /bdg/ and /ptk/ sets will discover that they are usually called "voiced" and "voiceless" respectively, and that these terms refer to the presence vs. absence of laryngeally produced signal during the interval of oral closure. At the same time, the perceptual importance of this feature of closure buzz is often played down, sometimes to the point of being dismissed as irrelevant to the distinction (Jakobson, Fant and Halle, 1952). The basis for this view is that both initial and final /bdg/ are often produced without closure buzz, whereas

*This paper was presented at the 95th Meeting of the Acoustical Society of America, Providence, R.I., 16-19 May, 1978.

+Also University of Pennsylvania.

Acknowledgment: This research was supported by NIH Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-55/56 (1978)]

medial position experiments in tape editing and in synthesis have shown that the duration of the closure, in association with that of the preceding vowel and in the absence of buzz, can be a sufficient cue to the category contrast [Lisker, (1957); Port, (1978)]. However, in medial position /bdg/ closures are rarely without buzz (Lisker, Abramson, Cooper and Schvey, 1969), and it has not yet been convincingly demonstrated that closure buzz, and its absence as well, play no significant role in signaling /bdg/ as distinct from /ptk/. Despite claims in the older literature that /ptk/ can be produced with voicing (Stetson, 1951), it seems more likely that any stop in whose production the larynx participates as a source of uninterrupted quasiperiodic signal will be heard as /bdg/. It is the absence of closure buzz whose status as a stop category cue is still ambiguous.

EXPERIMENTS

Closure Duration + Buzz

Experiments in which natural tokens of words of the type rapid and rabid were edited so as to vary the duration of a silent interval between the medial closing and opening transitions have been reported, showing that an original rapid is heard as rabid when the interval is reduced in duration, while rabid with silenced and lengthened closure suffers the reverse transformation. Figure 1 shows the labeling responses of six subjects to stimuli derived by editing the waveforms of one token each of rabid and rapid. The curves representing labelings of stimuli without closure buzz demonstrate the earlier-reported effect of varying silent closure duration on medial stop identifications. Such data have been understood to constitute evidence that closure duration serves as a cue to the listener in identifying words in normal speech. It should be noted, however, that according to this particular set of data, the /p/ closure had to be reduced to 30 msec for rabid to be the preferred response, but that the closure durations observed in productions of rabid spoken in isolation by the talker who provided the test stimuli ranged from 100 msec to not less than 85 msec. Two other talkers had minimum closures for /b/ that were shorter, but no closure shorter than about 50 msec was recorded (see also Suen and Beddoes, 1974). This would suggest that the closure duration of a normally produced /b/ may not in itself be a very reliable index of /b/ as against /p/. On the other hand, these data show that rabid went to rapid when the closure duration was no greater than 90 msec, a value that for two of the three talkers recorded was in the /b/ rather than the /p/ range for their natural productions. Replacing the buzz by silence had then, in this experiment, some effect even for durations not incompatible with /b/ as normally pronounced. In other words, normally produced /b/ may sometimes have a closure duration that is less than optimal for the perception of /b/, particularly if closure buzz is discounted as a cue. The particular token of rabid that served as one of the stimulus sources used in this experiment was produced, in fact, with a buzzed closure of 95 msec duration.

The Buzz Intensity Boundary

The data of Figure 1 do not support the view that closure buzz is linguistically irrelevant to the /b/-/p/ distinction in intervocalic position: /b/ may go to /p/, upon deletion of buzz, when the original closure duration is maintained, despite preservation of all other features associated with

FIGURE CAPTIONS

Figure 1: The curves represent percentage "rapid" judgments reported by six native speakers of American English. Stimuli were derived by waveform editing of one naturally produced token of rabid (upper panel) and one of rapid (lower panel). Closure durations were varied in 15-msec steps from 30 to 150 msec; each tested duration was either entirely silent (-buzz) or entirely filled with buzz (+buzz) derived from the original rabid token.

Figure 2: Percentage "rabid" responses of 19 listeners, all native speakers of American English. A naturally produced token of rabid was altered by waveform editing to have a closure duration of 140 msec. This interval was entirely filled by buzz derived from the originally recorded source word, and the intensity level of this buzz was varied, in 1 dB steps, from between -3 to -13 dB below the level of the original recording. The mean buzz level of this original recording was estimated to be about -25 dB relative to the mean level of the stressed /æ/.

Figure 3: Percentage "rabid" judgments of three native speakers of American English. Stimuli were generated by waveform editing from naturally produced tokens of rabid and rapid, which were varied in closure duration from 60 to 140 msec (20-msec steps). Closures were also varied in the extent to which they were filled with buzz, so that, for example, 0 on the abscissa of the upper left display represents stimuli with 60 msec of silent closure, while 60 on the same axis represents 60 msec of closure that is entirely filled by buzz.

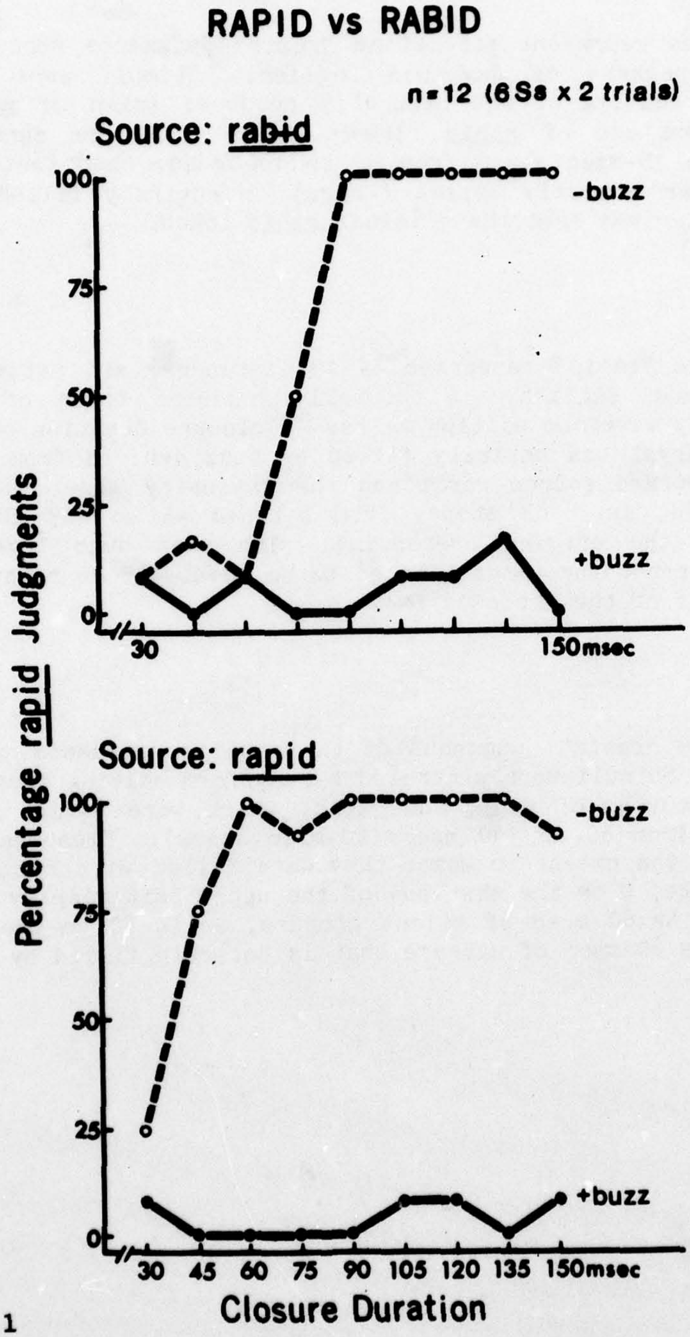

RAPID vs RABID
n=12 (6Ss x 2 trials)
Source: rabid
100
75
50
25
0
-buzz
+buzz
30
150msec
Percentage rapid Judgments
Source: rapid
100
75
50
25
0
-buzz
+buzz
30
45
60
75
90
105
120
135
150msec
Closure Duration

Figure 1

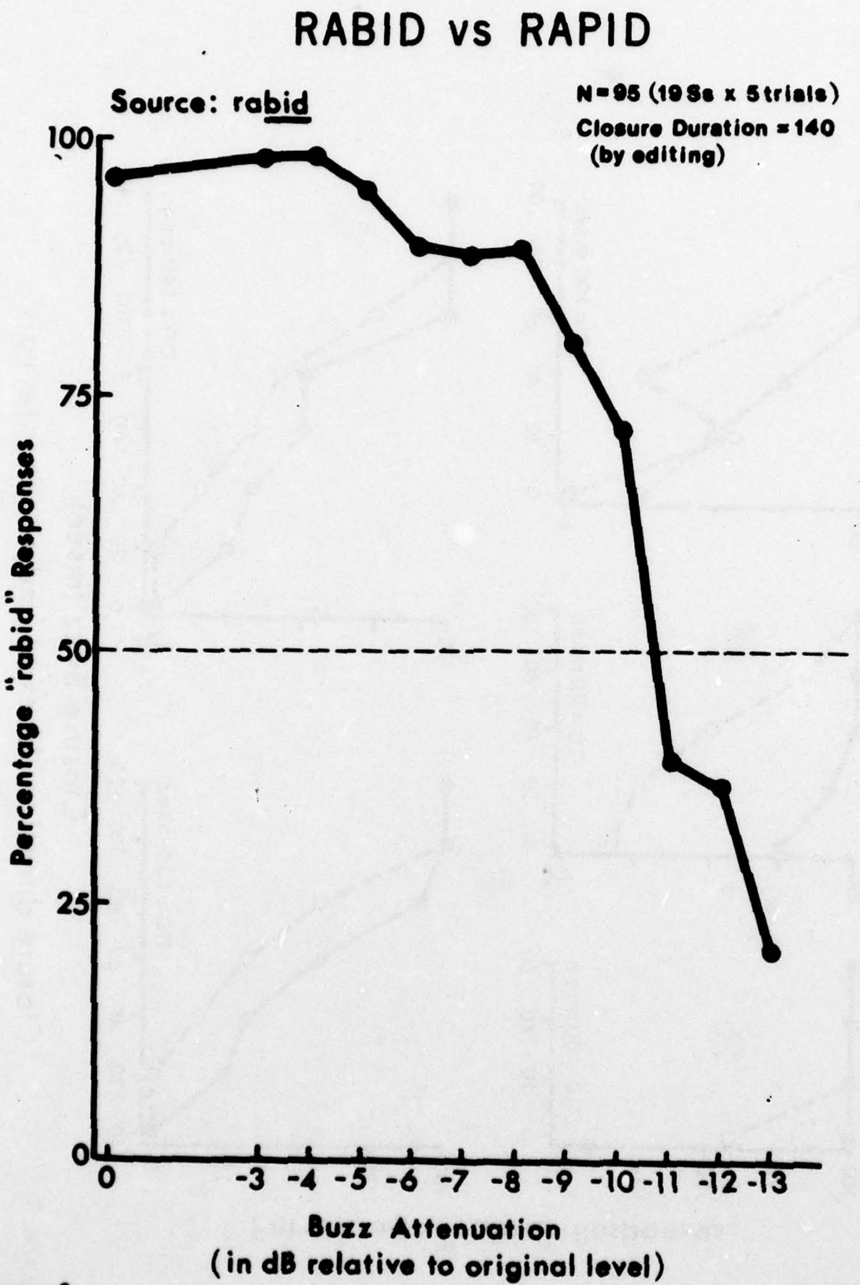

RABID vs RAPID
Source: rabid
N=95 (19 Ss x 5 trials)
Closure Duration = 140
(by editing)
Percentage "rabid" Responses
100
75
50
25
0
0
-3
-4
-5
-6
-7
-8
-9
-10
-11
-12
-13
Buzz Attenuation
(in dB relative to original level)

Figure 2

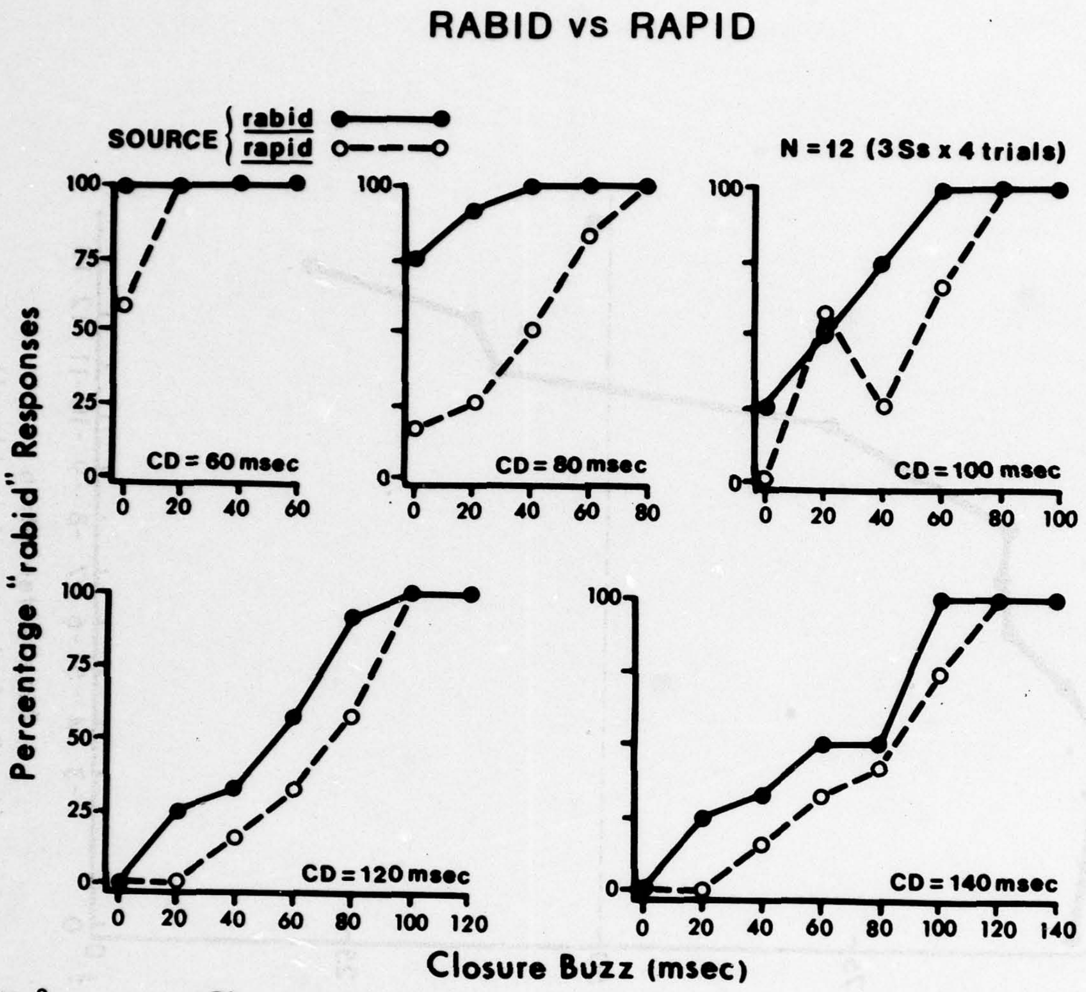


Figure 3

Closure duration: buzzed and silent intervals medial /b/. On the other hand, all extra-closure features of /p/ are insufficient to inhibit /b/ responses when buzz is supplied by editing.

If we take it that the presence of closure buzz is a cue for /b/, and that the absence of buzz may under certain conditions be a cue for /p/, then it is appropriate to consider the following questions: 1) Where along the intensity dimension is the perceptual-phonetic boundary between effective presence and absence of the buzz? 2) For a closure duration greater than one eliciting only /b/ labelings, how much of that interval must be buzz-filled for listeners to report /b/, or, alternatively, how much must be silent for /p/ to be heard?

Figure 2 shows the pattern of labeling responses to a set of stimuli derived from a token of _rabid_ whose /b/ closure was extended to a duration of 140 msec, with this interval entirely filled with buzz. Stimuli were presented at a comfortable listening level. As estimated by eye from the waveform, the mean buzz level of the originally recorded _rabid_ was about 25 dB below that of the preceding vowel. When buzz level was attenuated by more than an additional 10 dB, /p/ responses exceeded /b/. Similar tests using stimuli derived from other tokens of the same word gave slightly different crossover values, and there was some intersubject variation, but the results represented in Figure 2 held generally true. They suggest that at an attenuation of about 12 ± 3 dB below the normal buzz level, judgments divide evenly between /b/ and /p/. It is still to be determined whether the threshold is relative to the level of the immediately adjacent signal; work on this question is now in progress.

Closure: _Buzz_ + _Silence_

To answer the second question, that is, to determine the effect of varying the allocation of the closure interval between buzz and silence, stimuli were prepared from a token each of _rabid_ and _rapid_, varying closure duration from 60 to 140 msec and varying also the amounts of buzz and silence within each closure. The step size used was 20 msec, and buzz was at the level of the source token, that is, -25 dB relative to the stressed vowel. The data obtained (Figure 3) are compatible with earlier findings on closure duration, in that for the shortest durations, /b/ is most often reported. For 80 msec closure the _rapid_-derivatives are divided evenly between /b/ and /p/ when buzz occupies half the closure. With increasing closure duration the amount of buzz required for /b/ increases, up to a duration of about 85 msec for the longest closure. The _rabid_-derivatives elicited, as we would expect, more /b/ responses overall, but for the longest closure to be heard as /b/ more than 80 msec of buzz was needed. Putting the question the other way, we find that the silence duration at the category boundary increases somewhat as the closure is lengthened, but differences are small, and, given the small number of responses so far obtained, are of doubtful significance. Perhaps they warrant this statement: _rapid_-derivatives require more than 40 msec of silence to be heard as _rapid_, while _rabid_-derivatives need somewhat more than 80 msec of silence to be reported as _rapid_.

CONCLUSION

In summary: the main finding of the experiments just reported is that closure buzz is a non-negligible feature of stops in intervocalic position before unstressed vowels; not only is its presence a decisive cue for /b/, but its absence can sometimes elicit /p/ judgments in response to stimuli in which all other features of /b/ are presumably intact. For buzz to be perceptually relevant as a /b/ cue, it should have a level relative to that of the preceding stressed vowel of not less than -35 dB, and it should fill enough of the closure interval so that, at most, about 80 msec of the closure is acoustically blank. Whether the intensity boundary value of -35 dB relative to the preceding vowel remains constant with change in vowel intensity is still to be determined, as is the extent to which the buzz-silence balance may vary for stimuli derived from spoken words whose medial stops and preceding vowels vary widely in their naturally produced durations.

REFERENCES

Halle, M., G. W. Hughes and J.-P. A. Radley. (1957) Acoustic properties of stop consonants. _Journal of the Acoustical Society of America_ _29_, 107-116.

Jakobson, R., C. G. M. Fant and M. Halle. (1952) Preliminaries to speech analysis: The distinctive features and their correlates. _Technical Report No. 13_. (Cambridge, Mass.: M.I.T Acoustics Laboratory).

Lisker, L. (1957) Closure duration and the intervocalic voiced-voiceless distinction in English. _Language_ _33_, 42-49.

Lisker, L., A. S. Abramson, F. S. Cooper and M. H. Schvey. (1969) Transillumination of the larynx in running speech. _Journal of the Acoustical Society of America_ _45_, 1544-1546.

Port, R. (1978) Effects of word-internal versus word-external tempo on the voicing boundary for medial stop closure. _Journal of the Acoustical Society of America_ _63_, S1(A).

Stetson, R. H. (1951) _Motor Phonetics_, 2nd ed. (Amsterdam: North-Holland).

Suen, C. Y. and M. P. Beddoes. (1974) The silent interval of stop consonants. _Language and Speech_ _17_, 126-134.

Effects of Word-Internal vs. Word-External Tempo on the Voicing Boundary for Medial Stop Closure*

Robert F. Port+

ABSTRACT

It is known that the perceived voicing of the medial stop in words like rabid and rapid can be controlled by changing the duration of the stop closure. We found in an earlier experiment that increasing the tempo of a preceding carrier sentence shortened the boundary between rabid and rapid along a continuum of silent closure durations, but by far less than the percent decrease in sentence duration. We hypothesized that timing in unaltered portions of the test word reduced the effect of the carrier tempo. This experiment directly compares the effect on this boundary of the tempo of a surrounding carrier when the tempo of the test word itself is changed. A speaker recorded "I'm trying to say rabid to you" at both fast and slow tempos. Rabid was excised from each, and two continua of test words with silent /b/ closures were prepared (from 50 to 200 msec) and embedded in both original sentences. Listeners identified the test words as either rabid or rapid. Results indicate: (1) tempo within the test word had a stronger effect on the boundary than tempo in the carrier, and (2) for a given tempo of carrier, the ratio of the boundary value of closure duration to the duration of the rab syllable was nearly constant for both rab durations.

INTRODUCTION

It has been known since Lisker's 1957 study that if the closure interval corresponding to medial stops, such as those in the words rabid and rapid, has no audible glottal pulsing, English-speaking listeners can hear the stop as either /b/ or /p/ depending on the duration of the closure. Longer closures

*A slightly expanded version of a paper delivered at the 95th Meeting of the Acoustical Society of America, Providence, Rhode Island, 16-19 May, 1978.

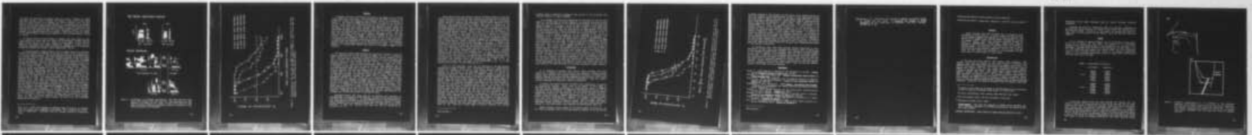
+Indiana University, Bloomington, Indiana.

Acknowledgment: Special thanks to Quentin Summerfield for help in preparation of the stimulus tapes. This research was partially funded by a PHS Biomedical Research Support Grant from the Indiana University Foundation.

[HASKINS LABORATORIES: Status Report on Speech Research SR-55/56 (1978)]

AD-A065 575 HASKINS LABS INC NEW HAVEN CONN F/G 5/7
SPEECH RESEARCH. A REPORT ON THE STATUS AND PROGRESS OF STUDIES--ETC(U)
DEC 78 A M LIBERMAN PHS-HD-01994
UNCLASSIFIED SR-55/56(1978) NL

3 OF 3
AD
A065575



are heard as voiceless and shorter ones as voiced. Studies of speech production [Lisker, (1957); Peterson and Lehiste, (1960); Port[1]] have shown that the vowel duration also changes, and varies inversely with the stop closure duration when the voicing feature of a following stop is changed. Since it is also well attested that the durations of both these intervals vary with speaking tempo [Peterson and Lehiste, (1960); Gaitenby, (1965); Port (see footnote 1)], the question arises as to what perceptual effects will follow changes in the speaking tempo of an utterance. Specifically, what will happen to the effectiveness of stop closure duration to cue the phonological voicing of a stop?

Of course, there is little reason to suppose that there might be some fixed duration in msecs that will serve as the perceptual boundary between cognate pairs such as /b/ and /p/. Rather, we would expect to find that the tempo of the context will affect perceivers' judgments. Indeed, several perceptual experiments have found an effect of speaking tempo on durational timing cues [Pickett and Decker, (1960); Summerfield, (1975)].

In one such study of tempo effects, Port (in press) was able to replicate Lisker's 1957 effect of changing the voicing of a medial stop by manipulation of the stop closure duration. He further demonstrated that the speaking tempo of a carrier sentence would shift this perceptual boundary along the closure duration continuum. In that experiment a single production of the word _rabid_ was cut and spliced so that a continuum of stop closures containing no glottal pulsing was prepared. Subjects identified the stimuli appended to both slow and fast carrier sentences. The results showed a voicing boundary between /b/ and /p/ at about 75 msec of closure duration in the slow carrier. However, when the words were appended to the fast carrier sentence, the perceptual boundary shifted significantly toward shorter values. Thus the tempo of a preceding carrier sentence _was_ able to shift the perceptual voicing boundary along this temporal continuum. On the other hand, a close comparison of the stimuli and the magnitude of the result might give some concern about the effectiveness of tempo. The stimulus sentences, that is, the carrier sentence combined with a test word, shortened by about 30 percent from the slow condition to the fast, while the perceptual boundary shortened only about 10 percent between the two conditions. Why was the effect so small? Notice that the test words in this experiment were made from a single production of _rabid_. Thus the temporal intervals immediately _adjacent_ to the closure variable were fixed throughout the experiment. If it were the case that the temporal voicing cue is primarily determined by the _ratio_ of the stop closure duration to the preceding or following syllables, then, of course, this might account for the very small effect of a carrier sentence that is external to the test word. In order to directly contrast the effectiveness of timing changes within the test word with timing outside the test word, the following experiment was conducted.

[1] Port, R. F. (1976) The influence of speaking tempo on duration of stressed vowel and medial stop in English trochee words. Ph.D. dissertation, University of Connecticut. (Available from the Indiana University Linguistics Club).

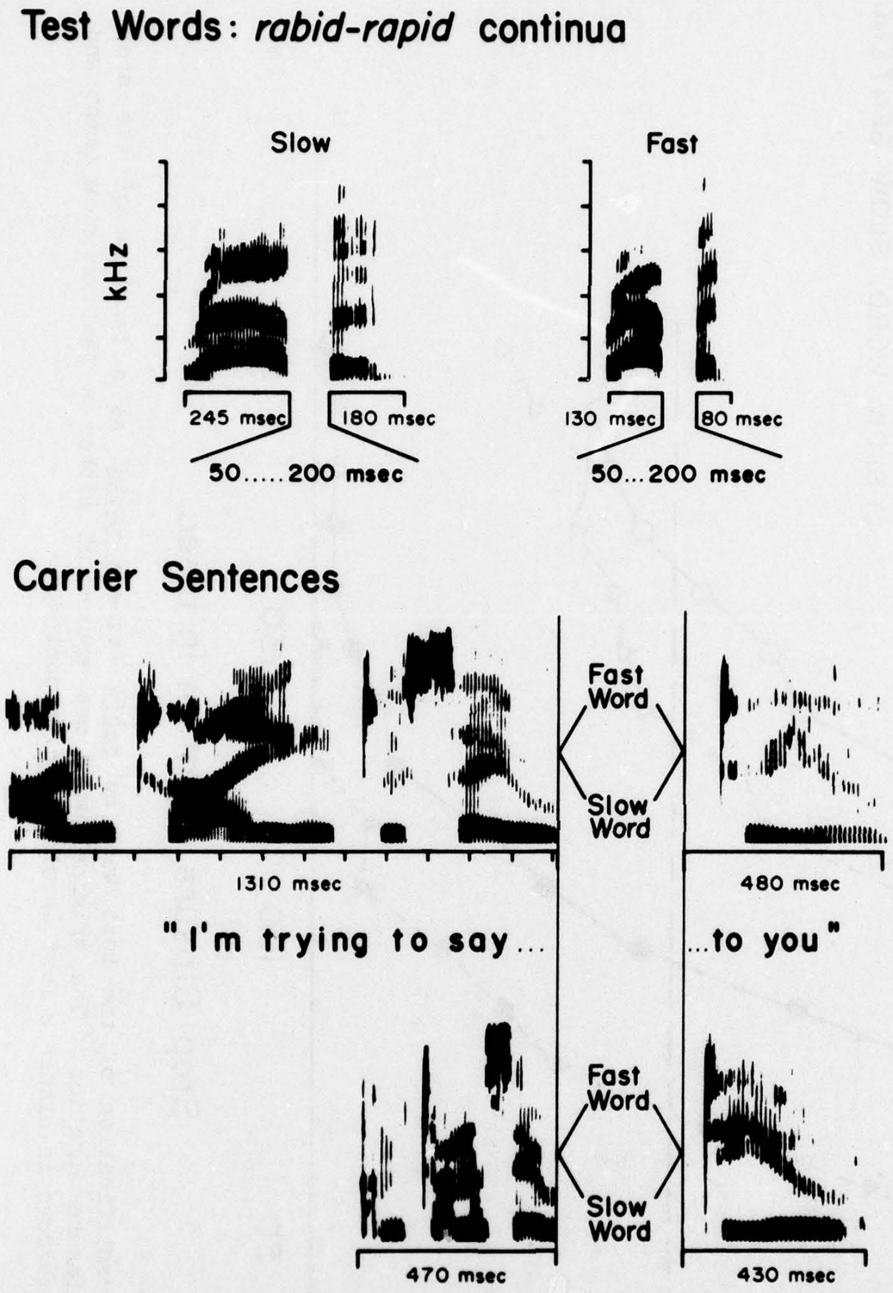


Figure 1: Illustration of stimulus construction. The Fast and Slow test words were excised from fast and slow tempo productions of the sentence "I'm trying to say 'rabid' to you." Note that the fast tempo production involved considerable phonetic reduction in the carrier portions of the sentence.

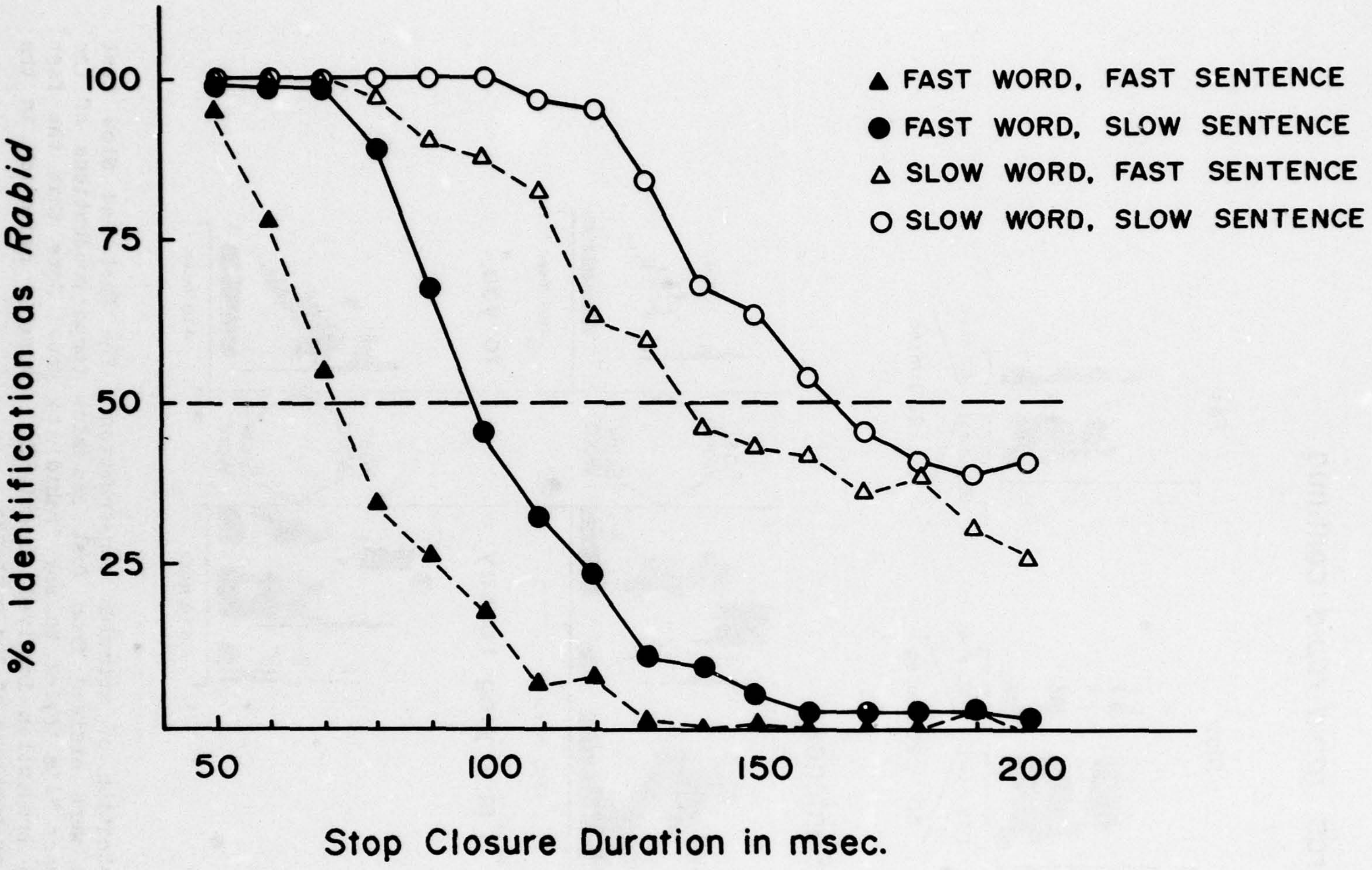


Figure 2: Identification of the test word as rabid (versus rapid) as a function of the stop closure duration. The original word was spoken at either a fast or slow tempo and embedded in either a fast or slow carrier sentence.

METHODS

A speaker recorded the sentence "I'm trying to say rabid to you" at both a fast tempo and a slow tempo. After digitization and 5 kHz filtering of the utterances, stimuli were prepared under computer control as illustrated in Figure 1. The test-word rabid was excised from each sentence by cutting at the end of the initial /r/ constriction and in the middle of the final /d/ closure. Then within each word, cuts were made immediately after closure for the medial /b/ and just before release. The glottal pulsing during the closure was discarded and a continuum was prepared by inserting silent gaps in each test word varying from 50 to 200 msec in steps of 10 msec. With the short stop closures, both test words sounded like rabid and with the long closures, both sounded like rapid. Individual members of each of these continua were then embedded in both the fast and the slow carrier sentences. A listening tape was prepared containing 10 tokens of each of these 64 stimuli in random order and presented to 12 college-educated listeners in Bloomington, Indiana, after the subjects had heard a 10-item practice tape. The listeners checked on a response sheet whether they heard the word rabid or rapid in each sentence.

RESULTS

The results of these identifications are presented in Figure 2. Here the percent identification of each stimulus as rabid, that is, as containing a medial voiced stop, is plotted as a function of the closure duration of the stop. The Fast word in the Fast sentence appears with filled triangles and the Slow word in the Slow sentence has open circles. In the Fast word-Fast sentence condition, stop closures longer than about 70 msec were heard predominantly as /p/, while in the Slow word-Slow sentence condition, the /b-p/ boundary is more than twice as long at 165 msec. What happened when the word and sentence tempos were crossed? If we start with the Slow word-Slow sentence condition, we find that changing just the carrier sentence from Slow to Fast--that is, from the open circles to the open triangles--moves the boundary about 30 msec, or 20 percent toward shorter values. This replicates our earlier experiment although the effect here is stronger than before, presumably because the test words here are sentence medial rather than sentence final. If, however, we change just the test word from Slow to Fast and leave the carrier sentence the same, (that is, from open circles to filled circles on the graph), we find that the boundary now shifts by 65 msec. This amounts to a shortening of the voicing boundary of about 40 percent--far larger than the effect of a change in the carrier sentence alone. Of course the same effects can be seen if we begin with the Fast word-Fast sentence condition: changing just the carrier moves the boundary significantly, but changing just the test word moves it considerably more.

Individual subject data showed that all 12 listeners were unanimous in providing more /b/ responses in the condition indicated in the pooled data for 5 of the 6 pairs of conditions. In the comparison of the Slow sentence-Slow word condition with the Fast sentence-Slow word condition, however, a single subject disagreed with the others by giving more /b/ responses in the latter condition. A sign test indicates that even this pair of conditions is distinct at $p < .01$.

These results imply that timing in the two syllables adjacent to the stop closure duration is considerably more important in determining the durational boundary for medial stop voicing than is the speaking tempo in a surrounding carrier sentence. Although this experiment does not permit separation of the effects of the preceding stressed syllable, rab, from the effects of the following unstressed syllable, bid, there is reason to suppose that the preceding syllable--in particular, the duration of the preceding vowel--plays the most important role in the apparent resistance of the boundary to the effects of the external carrier. First of all, we know that the duration of the preceding vowel in such words also varies as a function of following consonant voicing and, in particular, varies inversely with the stop duration itself (Port*). Second, a closer look at the classic perceptual study by Peter Denes (1955) on word final fricatives suggests an intriguing interpretation of these results. Using a synthetic pair of test words differing in the voicing of the final fricative in the words use, n. and use, v., Denes varied the duration of both the vowel and the voiceless fricative portion of his test words. He found that short final fricatives were heard as /z/ (even though technically voiceless), while longer durations were heard as /s/. He also found that the voicing boundary along his fricative duration continuum depended on the duration of the preceding vowel. In fact, when his stimulus identifications were plotted as a function of the ratio of the fricative duration to the vowel duration, it appeared that the perceptual boundary could be expressed as a constant ratio of the fricative to the preceding vowel. In particular, the criterion ratio was approximately one. That is, if the fricative was longer than the vowel, it was identified as /s/. If it was shorter than the vowel, it was identified as /z/.

Since, in this experiment, there were two durations of the rab syllable varying over a range of nearly 2 to 1, we decided to replot our identification results as a function of the ratio: stop closure duration divided by the duration of the "vowel" in the preceding syllable. For this purpose we measured the interval from the point of steepest rising F_3 slope in the transition from the /r/ into /æ/ up to the point of apparent closure for the /b/ for both Fast and Slow test words (110 msec and 200 msec respectively) and divided this into each closure duration. The replotted results appear in Figure 3 where the vertical axis remains the same as in the preceding figure. The value of 1.0 along the horizontal axis means that the stop closure is equal to the duration of the preceding vocalic interval, and values less than one mean the stop closure is shorter than the vowel. Thus, the value .5, for example, means the stop was half as long as the preceding rab syllable. In this display we notice first that the four functions no longer look as different as they did before, and that they are now grouped by carrier sentence rather than by word duration. In particular, it can be seen that in the Fast carrier sentence, the Fast and Slow test words (in filled and open triangles) cross the 50 percent boundary at almost exactly the same ratio--about .68. Similarly, in the Slow carrier sentence, the identification functions for the two test words (represented by filled and open circles) have perceptual boundaries very near each other centered around a consonant/vowel ratio of .85. Thus for a given tempo of carrier sentence, the voicing

*See footnote 1.

boundary along a continuum of ratios of stop closure to the preceding vowel interval tended to be nearly constant.

In an attempt to evaluate the significance of the effect of the test word tempo at a given tempo of carrier sentence, we examined individual subject data over just that portion of the stimulus continuum where the two test words share the range of C/V ratios. As can be seen in Figure 3, that range lies between .45 and 1.0, although the Fast word has 7 stimuli in this range while the Slow word has 12. Thus, for each subject we asked whether in the Fast carrier sentence he made a larger percentage of /b/ identifications to his 70 Fast word presentations or to the 120 Slow word identifications. Of the 12 subjects, 6 had a larger proportion of /b/ responses in the Fast word condition and 6 in the Slow condition, thereby providing no suggestion of a consistent effect. Although the pooled data in Figure 3 indicate more total /b/ responses to the Fast word than the Slow one when embedded in the Slow carrier sentence over this range, a similar sign test on individual subjects shows that only 7 of the 12 had a higher percentage of /b/ responses for the Fast word. Thus, these results suggest there is no consistent difference between the test words in the boundary value of C/V ratio at a given tempo of carrier sentence. On the other hand, even looking just at this portion of the stimulus continuum, all 12 subjects had a larger number of /b/ responses for both the Fast word and the Slow word in the Slow carrier than in the Fast carrier. In short, the differences between the carrier sentences (triangles vs. circles) in Figure 3 are highly significant, while the differences between the test words (filled vs. open) are probably not. Apparently the relative duration of a vowel and the following stop closure is a highly effective cue to the phonological voicing of the stop (a) over a range of at least 2-to-1 in vowel duration, (b) when the tempo of the surrounding speech remains constant.

DISCUSSION

It is tempting to interpret this result, which converges with Denes (1955) and is coherent with the production data showing an inverse relation between vowel duration and consonant constriction duration, as evidence that English listeners do not make voicing judgments by evaluating the stop closure duration and the preceding vowel duration independently. They are, rather, directly comparing the durations of these two intervals.

If further data should continue to support such a hypothesis, it would be of considerable interest. First, this is an example of what might be described as the temporal counterpart of coarticulation, since it suggests that the temporal information for segmental phonological features may not be localizable in any particular interval, but may inherently require several neighboring intervals for specification of its abstract temporal structure. Second, we may note that a temporal phonological cue of this type would have certain practical advantages for a language. For example, in speech production, this ratio would naturally tend to be invariant under changes of speaking tempo and changes in degree of stress. Such a cue might also permit very small timing differences in vowel duration and stop closure duration to combine in a way that yields a more prominent and perceptually useful cue.

Before closing we should comment on one further aspect of these data. If the vowel-to-consonant ratio is the most powerful temporal cue to the voicing

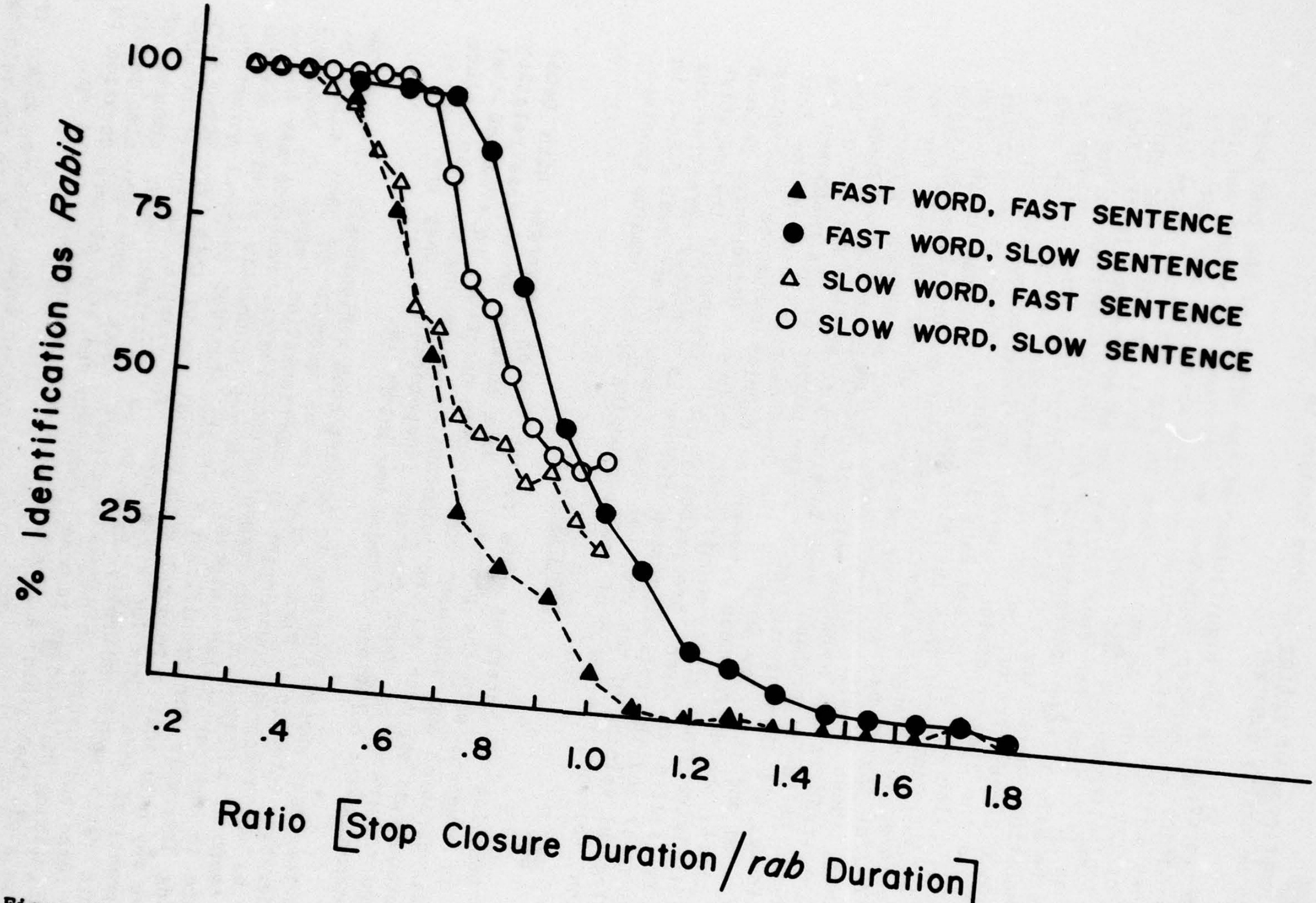


Figure 3: The same results as Figure 2 here plotted as a function of the ratio of the stop closure duration to the preceding "vowel" measured from the point of steepest slope in F_3 to /b/ closure. This interval sounds like /ræb/ when heard in isolation.

of medial consonants, what accounts for the ability of the external carrier to shift the boundary in the direction that it did? A first guess might be that the shift toward a smaller C/V ratio at the faster tempo reflects a similar shift occurring in speech production. Unfortunately, however, production data agree [Kozhevnikov and Chistovich, (1965); Klatt, (1976); Port*] that stops shorten less than vowels with a tempo increase. Thus differences in the compressibility of stops and vowels would predict a ratio shift in the opposite direction to that observed in Figure 3. On the other hand, if we assume that listeners employ the stop closure duration as a cue by making two separate comparisons, then a fairly simple model emerges. In addition to noting the relative duration of the consonant constriction to the preceding vowel, listeners may also compare the consonant duration to some intervals in the surrounding carrier sentence. When these other intervals are compressed, the stop closure boundary will also shift toward shorter values. Although we do not know what intervals might be involved in such a comparison, we do know that their effects are small relative to the more local timing effects closer to the consonant constriction.

In conclusion, then, this study revealed evidence supporting an early observation of Denes that, when listeners are forced to rely on timing cues for the voicing of a postvocalic consonant, their perception is based largely on the relative duration of the stressed vowel and the following consonant constriction. In addition, this experiment has replicated an earlier study of our own showing that the tempo of speech outside a test word will also have some effect on the perceptual boundary for voicing along a continuum of stop closures. Finally, these results clearly indicate that of these two temporal cues for voicing--one close to the stop closure itself and one more remote--the more local effect is clearly dominant.

REFERENCES

Denes, P. (1955) Effects of duration on the perception of voicing. Journal of the Acoustical Society of America 27, 761-764.

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

Klatt, D. H. (1976) The linguistic uses of segmental duration in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America 59, 1208-1221.

Kozhevnikov, V. and L. Chistovich. (1965) Speech: Articulation and Language. (Translated by Clearinghouse for Federal Technical and Scientific Information, Washington, D.C.).

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

Peterson, G. and I. Lehiste. (1960) Duration of syllabic nuclei in English. Journal of the Acoustical Society of America 32, 693-703.

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

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

*See footnote 1.

Summerfield, A. Q. (1975) How a full account of segmental perception depends on prosody and vice versa. In Structure and Process in Speech Perception, ed. by A. Cohen and S. G. Nooteboom. (New York: Springer-Verlag), pp. 51-66.

Coarticulatory Effects of Vowel Quality on Velar Function*

Fredericka Bell-Berti+, Thomas Baer, Katherine S. Harris++ and Seiji Niimi+++

ABSTRACT

Velar elevation data were collected for 12 utterance pairs contrasting in vowel quality. It is well known that velar position for any phonetic segment is determined by at least two factors: the nature of the segment itself and the phonetic environment in which the segment occurs. Thus, velar elevation increases through the series of segment-types: nasals--open vowels--close vowels--obstruents, and velar elevation for English vowels is affected by adjacent nasals. In these data, vowel quality was found to affect velar position during adjacent consonants: that is, the velum was lower for both nasals and obstruents in an environment of open vowels than in an environment of close vowels.

INTRODUCTION

It has long been known (Hilton, 1836; Bidder, 1838; Passavant, 1863; Czermak, 1869) that the position of the velum differs for different oral segments--these differences being the nature of the segment itself and of its phonetic environment. Generally, velar height differences associated with the segment vary directly with the oral cavity constriction of the segment, increasing through the series: open vowels--close vowels--obstruents (for example: Czermak, 1869; Moll, 1962; Bzoch, 1968; Lubker, 1968; Fritzell, 1969; Bell-Berti and Hirose, 1975). The influence of phonetic environment on velar position is exemplified by the effect of nasals on adjacent vowels. Velar elevation for English vowels is affected by adjacent nasals, so that the velum is lower for vowels preceding nasal consonants than for vowels preceding obstruents. Similarly, the velum is frequently lower for English vowels that

*A version of this paper was presented at the 95th Meeting of the Acoustical Society of America, Providence, Rhode Island, 16-19 May, 1978.

+On leave from Montclair State College, Upper Montclair, New Jersey.

++Also The Graduate School, The City University of New York.

+++University of Tokyo, Tokyo, Japan.

Acknowledgment: This work was supported by NINCDS grants NS-13870, NS-13617, BRSG grant RR-05596 to the Haskins Laboratories, and NINCDS fellowship grant NS-05332.

immediately follow nasal consonants than for vowels following obstruent consonants.

Although the effect of neighboring nasal and obstruent consonants in modifying the velar elevation for vowels is well known, the opposite effect has not been studied. Therefore, in this paper, we examine the effects of vowel type on velar position during adjacent nasal and obstruent consonant segments.

METHOD

A native speaker of American English (one of the authors of this paper) served as the subject for this study. An inventory of 24 two-syllable nonsense words, each beginning with /f/ and ending with /p/, was used in the experiment. The same vowel occurred in both syllables, and was either /i/ or /a/. Two consonants, one a nasal (/m/), and the other, one of six obstruents (/p,b,f,v,s,z/), occurred between the vowels in all possible combinations (Table 1). The utterances were placed in lists in random order, and the lists were read from six to ten times.

TABLE 1: Experimental utterances.

	oral/nasal	nasal/oral
	fipmip	fimpip
	fibmip	fimbip
V=/i/	fifmip	fimfip
	fivmip	fimvip
	fismip	fimsip
	fizmip	fimzip
	fapmap	fampap
	fabmap	fambap
V=/a/	fafmap	famfap
	favmap	famvap
	fasmap	famsap
	fazmap	famzap

A long thin plastic strip with grid markings was inserted into the subject's nostril and placed along the floor of the nose and over the nasal surface of the velum, to enhance the contrast between the edge of the supravelar surface and the posterior pharyngeal wall (Figure 1). A flexible fiberoptic endoscope (Olympus VF Type O) was also inserted into the subject's nostril and was positioned so that it rested on the floor of the nasal cavity with its objective lens at the posterior border of the hard palate, providing a view of the velum and lateral nasopharyngeal walls from the level of the hard palate to above the maximum elevation of the velum (observed during

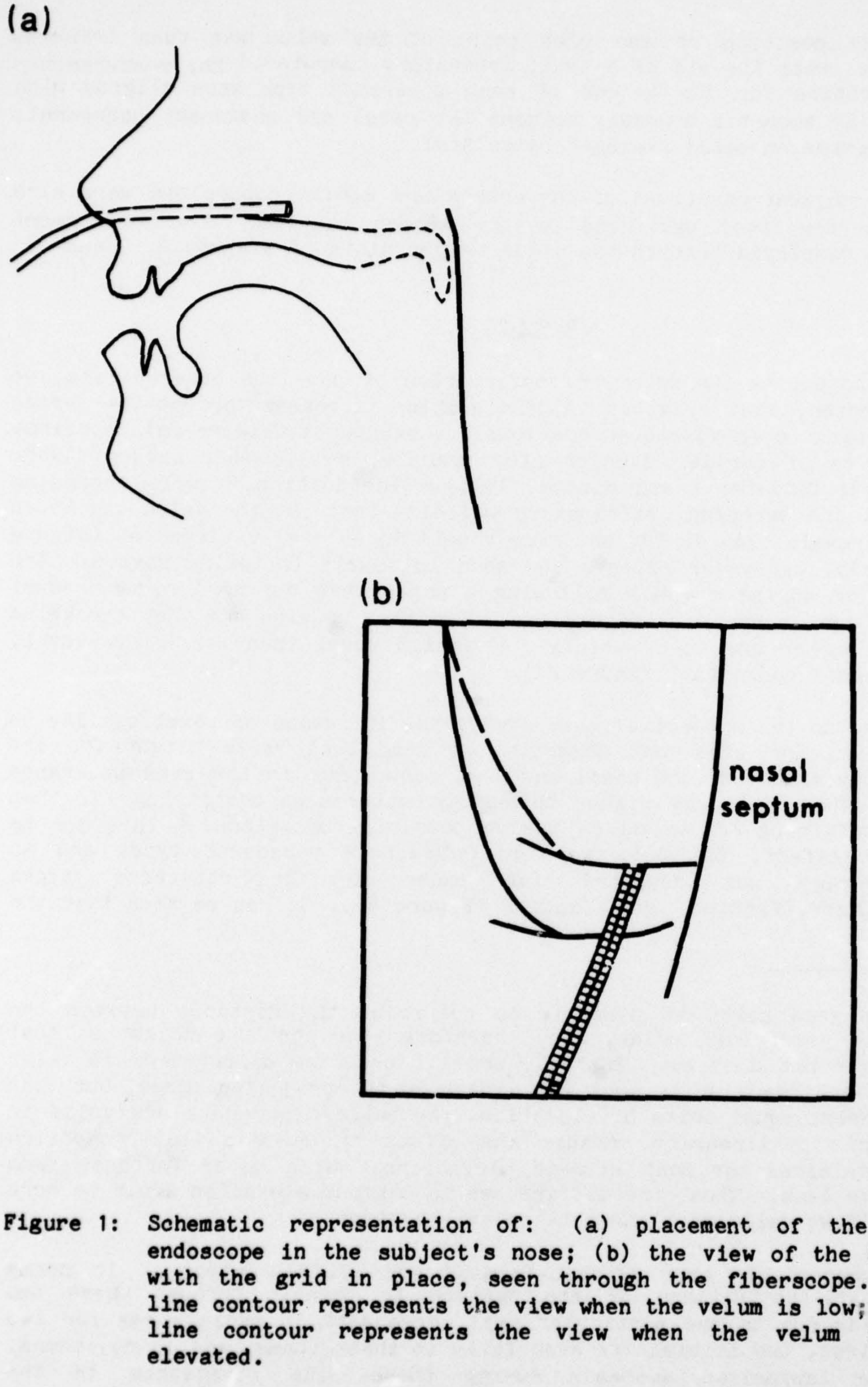


Figure 1: Schematic representation of: (a) placement of the fiberoptic endoscope in the subject's nose; (b) the view of the nasopharynx, with the grid in place, seen through the fiberscope. The solid line contour represents the view when the velum is low; the dashed-line contour represents the view when the velum is slightly elevated.

blowing). The position of the high point of the velum was then tracked, frame-by-frame, with the aid of a small laboratory computer.[1] The measurements of velar elevation for the tokens of each utterance type were aligned with reference to the acoustic boundary between the nasal and obstruent consonant, and frame-by-frame ensemble averages calculated.

Acoustic segment durations of the vowels and medial consonants were also measured from digitized waveforms of the speech samples. Average segment durations are displayed beneath the velar height plots of Figures 2, 3 and 4.

RESULTS

First, inspecting the data for confirmation of previous observations, we find, as expected, that relative velar elevation increases through the series of oral segments: open vowels--close vowels--obstruents (Figure 2a), confirming the results of earlier studies (for example, Moll, 1962; Lubker, 1968; Fritzell, 1969; Bell-Berti and Hirose, 1975). In addition, vowels preceding nasals showed the expected anticipatory effects--that is, the velum was lower during those vowels than during the same vowels in an oral environment (Figure 2b). Similarly, carryover effects are seen in vowels following nasals. The velum was lower during a vowel following a nasal than during the same vowel following an oral consonant (Figure 2c). Finally, we also see that the velum was elevated sooner and more rapidly for a high vowel than for a low vowel, following a nasal consonant (Figure 2d).

Returning to the subject of this study, the influence of vowel quality on velar elevation for consonant segments, we find that velar height for the vowel generally affected both nasal and oral consonants in the same utterance (Figure 3). The velum was higher throughout utterances containing /i/ than utterances containing /a/ in ten of twelve possible comparisons.[2] In order to clarify this effect, the data were pooled across consonant type, and an ensemble average was computed for each of the utterance types /fiCmip/,/faCmap/,/fimCip/, and /famCap/ (Figure 4). It can be seen that the

[1] Although the grid makes it possible to determine the distance between the lens and the velar high point, and, therefore, the absolute height at that point, we have not done so. Instead, we will describe differences in velar position in arbitrary units that are linear on the projected image, but that do not represent equal units of elevation. We believe our data are valid in spite of this nonlinearity because the effect of using a flat-projection reference minimizes the most extreme elevations, which occur furthest from the objective lens. Thus, the differences in maximum elevation would be more pronounced if we calculated absolute velar elevation.

[2] These two comparisons were /fizmip-fazmap/ and /fifmip-fafmap/. It seems unlikely that the failure of the pattern to persist through these two comparisons is due to the particular oral consonants in those items for two reasons. First, the initial /f/ also fails in these items, and in no others. Second, the comparison succeeds during these same consonants in the nasal/oral contrasts.

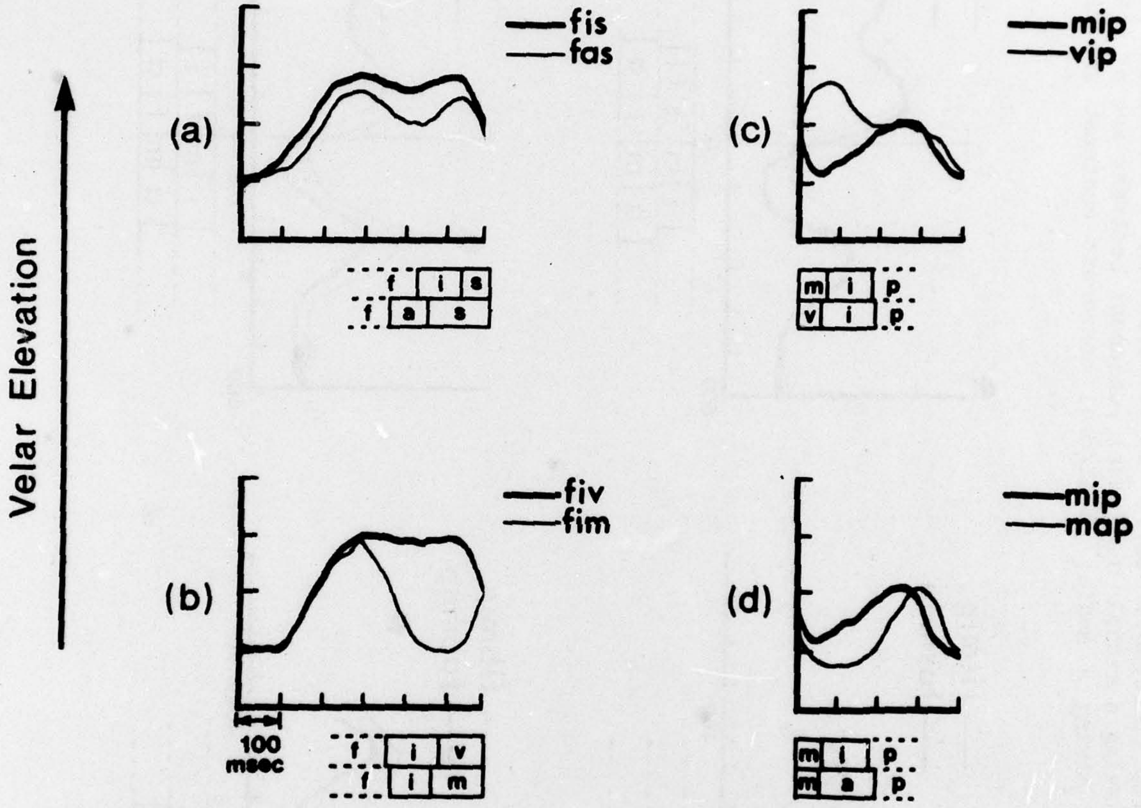
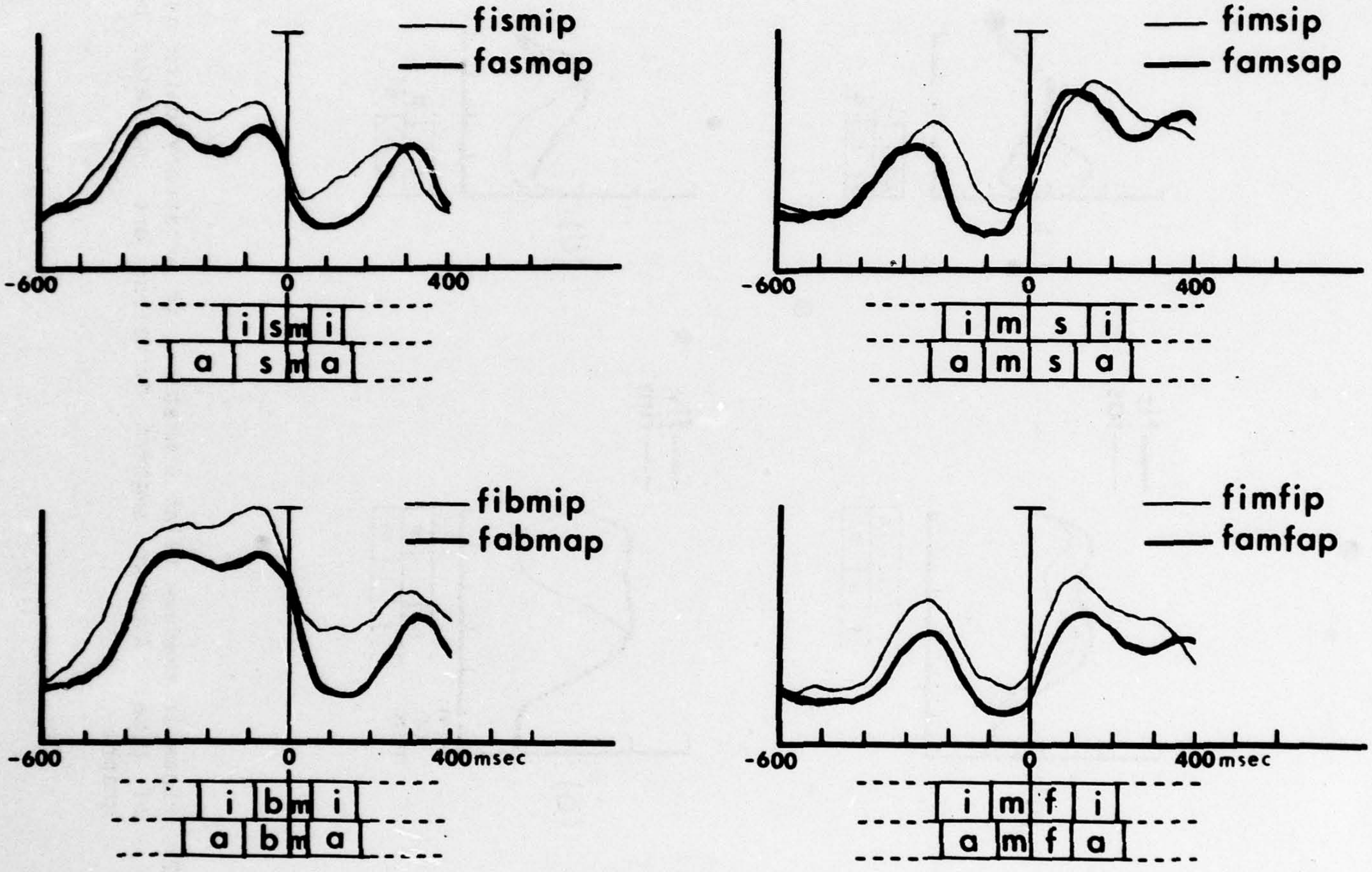


Figure 2: Plots of average velar elevation, in arbitrary units as a function of time. Average segment durations are indicated beneath each graph.

Figure 3: Plots of average velar elevation, in arbitrary units as a function of time, for utterance pairs of contrasting vowel quality. Utterances having a medial oral/nasal consonant contrast are at the left; those having a medial nasal/oral consonant contrast are at the right.



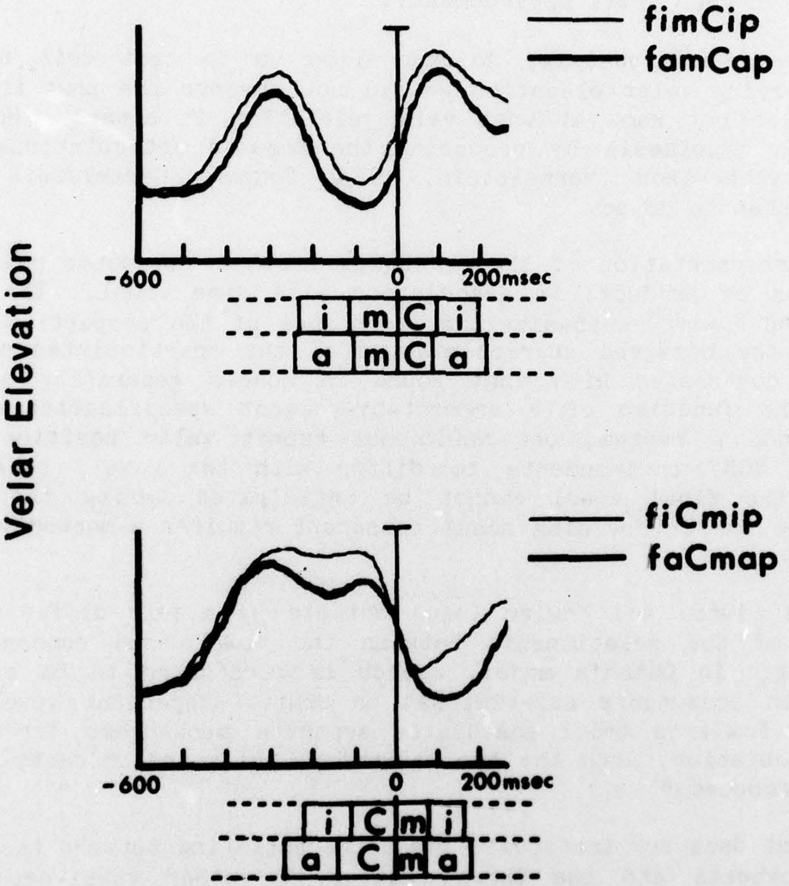


Figure 4: **Plots of average velar elevation, in arbitrary units, as a function of time, for utterance pairs of contrasting vowel quality. Data are pooled across oral consonant type. Utterances having a medial nasal/oral consonant contrast are at the top; those having a medial oral/nasal consonant contrast are at the bottom.**

overall pattern for /i/ and /a/ utterances is similar. We have no final explanation for the contrast in magnitude of differences for "nasal-first" and "oral-first" averages, though we suspect that syllable-stress may be involved.

DISCUSSION

These results may be explained in several ways. One possibility is that the acoustic requirements for consonants might differ in different environments. For example, if velopharyngeal closure is incomplete during production of an obstruent consonant, different port sizes might be required to prevent nasal coupling when the vocal tract is shaped for /i/ than when it is shaped for /a/. This would imply that velar position must be specified differently for consonants in each vowel environment.

These data, unfortunately, do not allow us to test this hypothesis, since, in observing velar elevation we did not observe the port itself, and, therefore, we do not know at what velar elevation it closed. However, we could test the hypothesis by producing the assumed articulations using an articulatory synthesizer (Mermelstein, 1973; Cooper, Mermelstein and Nye, 1977), and we plan to do so.

Another interpretation of these results is that the motor plan requires that consonants be produced in association with some vowel. The models of Ohman (1966) and Fowler[3] emphasize the importance of the properties of vowels in generating the observed characteristics of the coarticulated consonants. This view is contrasted with that found in models generating articulator position as the function of a segment-by-segment specification of feature values. In such a system, one would not expect velar position for oral consonants in VCNV environments to differ with the vowel, because velar position for the final vowel cannot be anticipated during the obstruent consonant since the intervening nasal consonant requires a markedly different velar position.

The Ohman (1966) and Fowler (see footnote 3) models differ as to the exact nature of the relationship between the vowels and consonants in a phonetic string. In Ohman's model, speech is considered to be a vowel-to-vowel act, with consonants superimposed on context-dependent vowel strings. Alternatively, Fowler's model postulates separate mechanisms for vowel and consonant articulation, with the two mechanisms activated in parallel and the segments "co-produced."

The current data are inadequate for differentiating between the acoustic-necessity hypothesis and the Ohman, Fowler, or other vowel-dependent hypotheses. Consequently, we are continuing our study of velar position in a variety of different vowel and consonant contexts.

[3] Fowler, C. A. (1977) Timing control in speech production. Unpublished Ph. D. thesis, University of Connecticut.

REFERENCES

Bell-Berti, F. and H. Hirose. (1975) Palatal activity in voicing distinctions: A simultaneous fiberoptic and electromyographic study. Journal of Phonetics 3, 69-74.

Bidder, F. H. (1838) Neue Beobachtungen ueber die Bewegungen des weichen Gaumens und ueber den Geruchssinn. (Dorpat: C. A. Kluge).

Bzoch, K. R. (1968) Variations in velopharyngeal valving: The factor of vowel changes. Cleft Palate Journal 5, 211-218.

Cooper, F. S., P. Mermelstein and P. Nye. (1977) Speech synthesis as a tool for the study of speech production. In Dynamic Aspects of Speech Production, ed. by M. Sawashima and F. S. Cooper. (Tokyo: University of Tokyo Press).

Czermak, J. N. (1869) Wesen und Bildung der Stimm- und Sprachlaute. Czermak's gesammelte Schriften, vol. 2. (Leipzig: Vilhelm Engelman).

Fritzell, B. (1969) The velopharyngeal muscles in speech: An electromyographic and cineradiographic study. Acta Otolaryngologica, Suppl. 250.

Hilton. (1836) Case of a large bony tumour in the face completely removed by spontaneous separation. Observations upon some of the functions of the soft palate and pharynx. Guy's Hospital Report 1, 493.

Lubker, J. F. (1968) An electromyographic-cineradiographic investigation of velar function during normal speech production. Cleft Palate Journal 5, 1-18.

Mermelstein, P. (1973) Articulatory model for the study of speech production. Journal of the Acoustical Society of America 53, 1070-1082.

Moll, K. L. (1962) Velopharyngeal closure on vowels. Journal of Speech and Hearing Research 5, 30-77.

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

Passavant, G. (1863) Ueber die Verschliessung des Schlundes beim Sprechen. (Frankfurt a.M.: J. D. Sauerlander).

Duration-Contingent Effects in Adaptation[*]

David Dechovitz[+] and Roland Mandler[+]

ABSTRACT

Context-dependencies in the acoustic consequences of stop consonant voicing raise critical questions about the nature of analyzers proposed for the voicing feature. In the present study, a selective adaptation method was used to address such questions. Two /da-ta/ continua were constructed, ranging in voice onset time (VOT) from 0 to 55 msec in 5 msec steps. The duration of the formant transitions was 15 msec in one series and 70 msec in the other. Effects of different magnitude were produced along the VOT dimension for within- and cross-series adaptation by an adapting stimulus from each series with the same VOT (25 msec). The present results support the view that models of voicing perception based strictly on VOT detectors are inadequate. Further, they suggest that the analysis of the voicing feature is carried out by duration-contingent channels.

The voicing distinction for initial pre-stressed stop consonants is achieved by the timing of changes in glottal aperture relative to supraglottal articulation. This relation is realized acoustically by voice onset time (VOT) (Lisker and Abramson, 1964). However, VOT values show context-dependent variability. In production, it has been observed that VOT varies inversely with the degree of constriction (and thus with the frequency of F_1) required by the vowel following the stop. For example, VOTs tend to be longer before the vowel /i/ than before /a/ [Klatt, (1975); Summerfield, (1975)]. This context-contingency is neatly paralleled in perception by the reported inverse relationship between F_1 onset frequency and the VOT boundary value; a longer value of VOT is required for the perception of a voiceless stop when F_1 has a low onset frequency [Lisker, Liberman, Erickson, Dechovitz and Mandler, (1978); Summerfield and Haggard, (1977)]. Further, measurements of spectrographic records have indicated that VOT can vary inversely with speech tempo; across speakers, VOTs in word initial /g/ and /t/ decreased as speech rate increased (Summerfield, 1972). This sensitivity is also reciprocated in perception. For example, increasing the duration of the steady-state portion of a syllable with fixed VOT increases the probability that the initial consonant will be perceived as voiced (Summerfield and Haggard, 1977). In another investigation, CV targets were perceived as more voiced following

[*] This paper was presented at the 94th meeting of the Acoustical Society of America, Miami Beach, Florida, December, 1977.

[+] Also University of Connecticut, Storrs.

[HASKINS LABORATORIES: Status Report on Speech Research SR-55/56 (1978)]

precursors with a slower syllabic rate, and more voiceless following those with a faster syllabic rate. A general explanation would be that VOT is perceived in relation to ongoing rate of speech (Summerfield, 1972).

Recent accounts of the perceptual process underlying voicing classification have included systems of feature analyzers analogous to those purported to operate in the visual and non-speech auditory modes of perception [Eimas and Corbit, (1973); Cooper, (1974)]. These devices are understood to extract sufficient acoustic information for the assignment of distinctive voicing values. However, we have reason to inquire how context-conditioned variation in the acoustic correlates of the voicing feature (as described above) is reflected in the operation of such detector systems. Some clarification comes from studies of selective adaptation.

Evidence from selective adaptation studies suggests that systems of context-sensitive detectors mediate voicing perception. For example, in investigations employing a contingent adaptation paradigm where both the voicing of the consonant and the vowel in an adapting pair of CV syllables were varied, the direction of boundary shift along voicing continua was governed by the stimulus in the adapting sequence that had the same vowel as the test series [Ades, (1974); Cooper, (1974); Miller and Eimas, (1976)]. In another investigation, adaptors varying in transition duration yielded greater boundary shifts along voicing continua with which they shared transition duration value (Diehl and Rosenberg, 1977). Though this finding was not fully consistent and, in fact, went unremarked by the investigators themselves, it suggests that voicing detectors may be selectively tuned to length of stop transition. In the following we seek further experimental support for this interpretation.

Thus, in the present experiment, an adaptation procedure was used to assess the processing of the voicing feature as a function of transition duration. Two ten-member, three formant /da-ta/ series were generated on the Haskins Laboratories parallel resonance synthesizer. Transition duration in all three formants was 15 msec in one series and 70 msec in the other. VOT was varied in 5-msec steps from 0 to +45 for the 15-msec set, and from +10 to +55 for the 70-msec set. These VOT ranges were constructed so that voiced and voiceless categories were well established in each series. For each transition duration, F_1 rose linearly from 154 Hz to a steady-state value of 769 Hz. However, since the F_1 intensity is zero until the synthesizer's periodic source is turned on, the actual F_1 onset frequency depended directly on VOT and transition duration for VOT values greater than +5 msec.

The 25 msec VOT members of the 15 msec and 70 msec series, /ta/-15 and /da/-70 respectively, were selected as adaptors for within- and cross-series adaptation. Informal listening indicated that /ta/-15 was a convincing /ta/, and similarly, that /da/-70 was a good exemplar of /da/. Spectrally, the two adaptors differed from one another in F_1 onset frequency, /da/-70 showing periodicity in F_1 at a lower value. On each of two days, the effect of adaptation with one adaptor was assessed on the two /da-ta/ continua for one group of nine phonetically naive listeners. In each session, listeners first identified 10 randomly-presented instances of each stimulus in the two continua, 2 sec separating successive items. After the initial unadapted condition, listeners were presented with two adaptation tests, each consisting of 20 consecutive adaptation trials with test items randomly selected from one test continuum. On each adaptation trial, subjects listened to 32 presentations of the adaptor with an interstimulus interval (ISI) of 300 msec, before identifying 6 test stimuli. An ISI of 2 sec intervened between the last repetition of the adaptor and the first identification item; 1.75 sec separated adjacent identification items. Subjects responded 12 times to each of the 20 test stimuli during post-adaptation identification.

The pre- and post-adaptation functions expressed as the percentage of /da/ responses averaged over subjects are given for each test continuum in Figures 1 and 2. The six identification curves were submitted to a probit analysis (Finney, 1971) (that is, fitted to cumulative normal distribution functions), and the resulting phoneme boundaries for each set were compared by t-tests for correlated observations. These analyses revealed effects contingent on transition duration between within- and cross-series conditions. Along the 15-msec series, adaptation with /ta/-15 yielded a reliable shift toward the voiceless category (t_8 = 7.679, $p < .001$), but /da/-70 did not influence perception (t_8 = .4789). Similarly, adaptation with /da/-70 resulted in a reliable shift toward the voiced category on the 70 msec set (t_8 = 10.56, $p < .001$), but /ta/-15 produced only a marginally significant shift toward the voiceless class (t_8 = 2.58, $p < .05$). In sum, along each test continuum, perception was primarily influenced by the adaptor with which it shared transition duration value. The differential effectiveness of the two adaptors cannot be accounted for by an adaptation effect that operates on the perception of the voicing feature alone. Rather, the effects observed here must be attributed to adaptation that operates in a fashion contingent on transition duration.

Examination of the pre-adaptation functions reveals that /ta/-15 was perceived as a good exemplar of /ta/, and similarly, /da/-70 as /da/ by listeners in the present study. As we would expect, then, each adaptor, when effective, shifted the phonetic boundary toward the class with which it shared voicing value. Further, an additional statistical test revealed a pre-adaptation boundary effect; the baseline crossover VOT values were significantly lower (t_8 = 37.61, $p < .001$) for the 15 msec series (18.78 msec) than for the 70 msec series (36.82 msec). These results are similar to those obtained in previous studies [Lisker et al., (1978); Summerfield and Haggard, (1977)] and provided strong initial evidence that voicing value is perceived in relation to length of transition.

There is some reason to believe that the contextual variable in the present study (that is, transition duration) reflects articulatory tempo for productions at a single place and vowel environment. For example, results of several cineradiographic [Gay and Hirose, (1973); Gay, Ushijima, Hirose and Cooper, (1974)] and spectrographic studies (Summerfield, 1972) have indicated that transition times are shortened when speaking rate becomes sufficiently rapid. Thus, recalibration of some part of the perceptual processor for variations in transition duration seems warranted.

Earlier, we noted the variability in some acoustic correlates of stop consonant voicing as a function of context. This lack of invariance argues that a phonetic description of speech is abstract, and leads us to wonder whether selective adaptation operates at a level that is abstract in the same

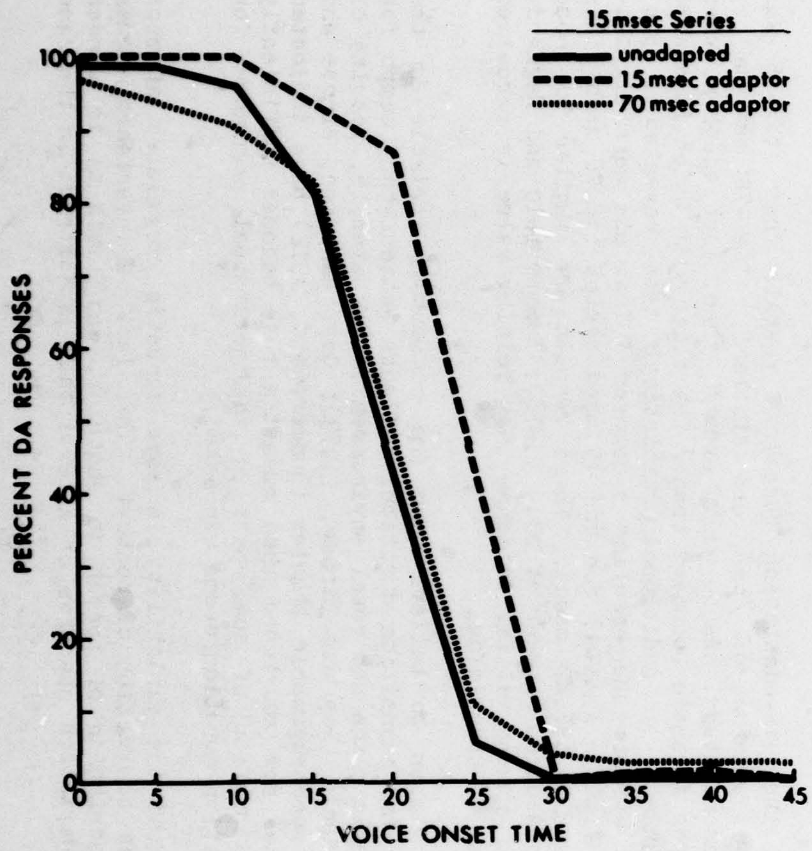


Figure 1: Pre- and post-adaptation functions for the 15-msec /da-ta/ test series.

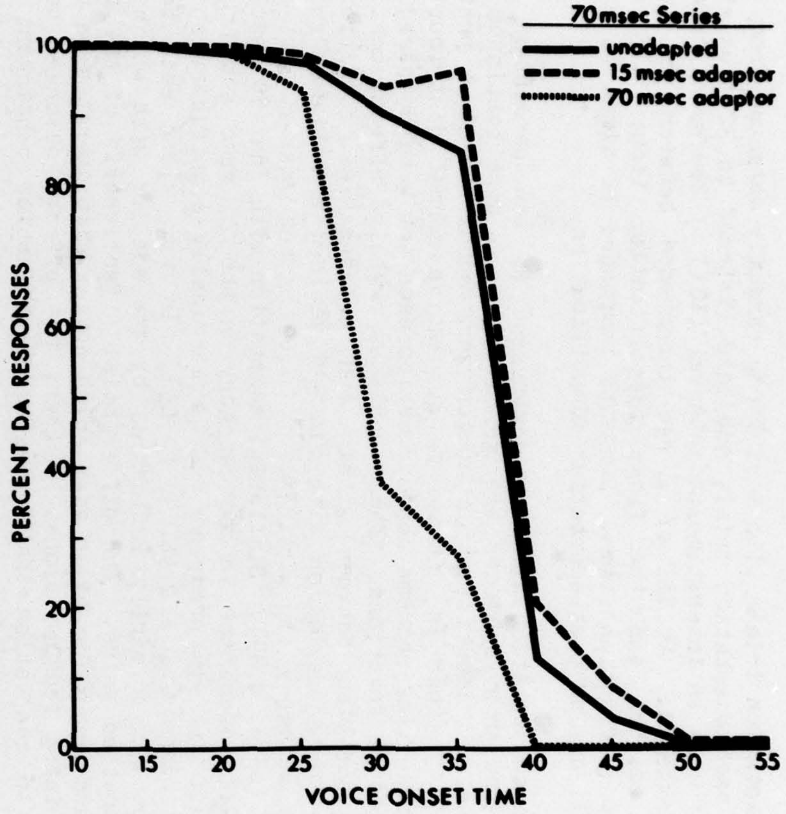


Figure 2: Pre- and post-adaptation functions for the 70-msec /da-ta/ test series.

sense. However, evidence from selective adaptation studies reveals a context-specific level of voicing perception. In the present study, the contingent nature of the adaptation effect indicates that the analysis of the voicing feature in initial stops may be carried out, at least in part, by detection channels that are dependent on transition duration.

REFERENCES

Ades, A. (1974) How phonetic is selective adaptation? Experiments on syllable position and vowel environment. Perception & Psychophysics 16, 61-66.

Cooper, W. E. (1974) Selective adaptation for acoustic cues of voicing in initial stops. Journal of Phonetics 2, 303-313.

Diehl, R. and D. Rosenberg. (1977) Acoustic feature analysis in the perception of voicing contrasts. Perception & Psychophysics 21(5), 418-422.

Eimas, P. D. and J. D. Corbit. (1973) Selective adaptation of linguistic feature detectors. Cognitive Psychology 4, 99-109.

Finney, D. J. (1971) Probit Analysis. (Cambridge: Cambridge University Press).

Gay, T. and H. Hirose. (1973) Effect of speaking rate on labial consonant production. Phonetica 27, 44-56.

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

Klatt, D. H. (1975) Voice onset time, frication, and aspiration in vowel-initial consonant clusters. Journal of Speech and Hearing Research 18, 686-706.

Lisker, L. and A. S. Abramson. (1964) A cross-language study of voicing in initial stops: Acoustical measurements. Word 20, 324-422.

Lisker, L., A. M. Liberman, D. M. Erickson, D. R. Dechovitz and R. Mandler. (1978) On pushing the voice onset time (VOT) boundary about. Language and Speech 20(3), 209-216,

Miller, J. L. and P. D. Eimas. (1976) Studies on the selective tuning of feature detectors for speech. Journal of Phonetics 4, 119-127.

Summerfield, A. Q. (1972) Towards a detailed model for the perception of voicing contrasts. Speech Perception (Psychology Department, The Queen's University of Belfast) Series 2, no. 3, 1-14.

Summerfield, A. Q. (1975) How a full account of segmental perception depends on prosody and vice-versa. In Structure and Process in Speech Perception, ed. by A. Cohen and S. G. Nooteboom. (New York: Springer-Verlag).

Summerfield, A. Q. and M. P. Haggard. (1977) On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. Journal of the Acoustical Society of America 62, 435-448.

II. PUBLICATIONS

III. APPENDIX

[redacted]

PUBLICATIONS

Abramson, A. S. (in press) The coarticulation of tones: An acoustic study of Thai. In: *Studies in Tai and Mon-Khmer Phonetics in Honour of Eugénie J. A. Henderson*, ed. by V. Panupong, P. Kullavanijaya and T. Luangthongkam. (Bangkok: Indigenous Languages of Thailand Research Project).

Abramson, A. S. (in press) The noncategorical perception of tone categories in Thai. In: *Frontiers of Speech Communication Research*, ed. by B. Lindblom and S. Öhman. (London: Academic Press).

Abramson, A. S. (in press) Static and dynamic acoustic cues in distinctive tones. *Language and Speech 21*.

Baer, T. (1978) Effects of single-motor-unit firings on fundamental frequency of phonation. *Journal of the Acoustical Society of America 64*, S90(A).

Bell-Berti, F. and K. S. Harris. (in press) Anticipatory coarticulation: Some implications from a study of lip rounding. *Journal of the Acoustical Society of America*.

Bell-Berti, F., T. Baer, K. S. Harris and S. Niimi. (in press) Coarticulatory effects of vowel quality on velar elevation. *Phonetica*.

Borden, G. J. (in press) An interpretation of research on feedback interruption in speech. *Brain and Language*.

Borden, G. J. and T. Gay. (1978) On the production of low tongue tip /s/: A case report. *Journal of Communication Disorders 11*, 425-431.

Borden, G. J. and T. Gay. (in press) Temporal aspects of articulatory movements for /s/-stop clusters. *Phonetica*.

Fowler, C. A. and M. T. Turvey. (1978) Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In: *Information Processing in Motor Control and Learning*, ed. by G. Stelmach. (New York: Academic Press).

Freeman, F. J., E. S. Sands and K. S. Harris. (1978) Temporal coordination of phonation and articulation in a case of verbal apraxia: A voice onset time study. *Brain and Language 6*, 106-111.

Fujimura, O., T. Baer and S. Niimi. (in press) A stereofiberscope with magnetic interlens bridge for laryngeal observation. *Journal of the Acoustical Society of America*.

Harris, K. S. (in press) Vowel duration change and its underlying physiological mechanisms. *Language and Speech*.

Healy, A. F. (1978) A Markov model for the short-term retention of spatial location information. *Journal of Verbal Learning and Verbal Behavior 17*, 295-308.

Healy, A. F. and M. Kubovy. (1978) The effects of payoffs and prior probabilities on indices of performance and cutoff location in recognition memory. *Memory & Cognition 6*, 544-553.

Healy, A. F. and A. G. Levitt. (1978) The relative accessibility of semantic and deep-structure syntactic concepts. *Memory & Cognition 6*, 518-526.

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

Kelso, J. A. S. (1978) Changing concepts of feedback and feedforward in voluntary movement control. *The Behavioral and Brain Sciences 1*, 153-154.

Kelso, J. A. S., D. Goodman, C. Hayes and C. L. Stamm. (in press) Movement coding and memory in the developmentally young. *American Journal of Mental Deficiency*.

Kelso, J. A. S., J. Pruitt and D. Goodman. (in press) The anticipatory control of movement. In: *Psychology of Motor Behavior and Sport*, ed. by K. Newell and G. E. Roberts. (Champaign, Ill.: Human Kinetics).

Kelso, J. A. S., D. Southard and D. Goodman. (in press) On programming and coordinating two-handed movements. *Journal of Experimental Psychology: Human Perception and Performance*.

Kelso, J. A. S., D. Southard and D. Goodman. (in press) On the nature of human interlimb coordination. *Science*.

Kiritani, S., H. Hirose and T. Baer. (1978) A preliminary report on the timing of consonant and vowel articulations in English. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*. (University of Tokyo) *12*, 29-34.

Mann, V. A. (1978) Different loci suggested to mediate tilt and spiral motion after-effects. *Investigative Opthalmology 17*, 903-909.

Raphael, L. J., F. Bell-Berti, R. Collier and T. Baer. (in press) Tongue position in rounded and unrounded front vowel pairs. *Language and Speech 22*(1).

Repp, B. H. (in press) Accessing phonetic information during perceptual integration of temporally distributed cues. *Journal of Phonetics*.

Repp, B. H. (in press) Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*.

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

Sands, E. S., F. J. Freeman and K. S. Harris. (1978) Progressive changes in articulatory patterns in verbal apraxia: A longitudinal case study. *Brain and Language 6*, 97-105.

Shankweiler, D. and I. Y. Liberman. (1978) Reading behavior in dyslexia: Is there a distinctive pattern? *Bulletin of the Orton Society 28*, 114-123.

APPENDIX

DDC (Defense Documentation Center) and ERIC (Educational Resources Information Center) numbers SR-21/22 to SR-55/56:

Status Report		DDC	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	**	**
SR-55/56	July - December 1978	**	**

**DDC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Behavior Abstracts, P.C. Box 22206, San Diego, California 92122.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Haskins Laboratories
270 Crown Street
New Haven, Connecticut 06510

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

N/A

3. REPORT TITLE

Haskins Laboratories Status Report on Speech Research, No. 55/56, July-December, 1978

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

December 1978

7a. TOTAL NO. OF PAGES

222

7b. NO. OF REFS

237

8a. CONTRACT OR GRANT NO.

HD-01994
V101(134)P-342
N01-HD-1-2420
RR-5596
BNS76-82023
MCS76-81034
NS13870
NS13617

9a. ORIGINATOR'S REPORT NUMBER(S)

SR-55/56 (1978)

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

None

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited*

11. SUPPLEMENTARY NOTES

N/A

12. SPONSORING MILITARY ACTIVITY

See No. 8

13. ABSTRACT

This report (1 July - 31 December) 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:

-Relative Accessibility of Semantic and Deep Structure Syntactic Concepts
-Some Relationships between Articulation and Perception
-Reflex Activation of Laryngeal Muscles by Sudden Induced Subglottal Pressure Changes
-Dynamic Aspects of Velopharyngeal Closure
-Effect of Speaking Rate on Relative Duration of Stop Closure and Fricative Noise
-Voicing in Intervocalic Stops and Fricatives in Dutch
-Insufficiency of the Target for Vowel Perception
-Syllable Timing and Vowel Perception
-Perception of Vowel Duration in Consonantal Context, Application to Vowel Duration
-Stimulus Dominance in Fused Dichotic Syllables
-Categorical Perception of Fused Dichotic Syllables
-Stimulus Dominance and Ear Dominance in Fused Dichotic Speech and Nospeech Stimuli
-On Buzzing the English /b/
-Effects of Word-Internal vs. Word-External Tempo on the Voicing Boundary, Medial Stops
-Coarticulatory Effects of Vowel Quality on Velar Function
-Duration-Contingent Effects in Adaptation

*This document contains no information not freely available to the general public. It is distributed primarily for library use.

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Semantics, Syntax						
Articulation - Speech perception						
Larynx - Reflex activation, air pressure						
Velum - Closure dynamics, coarticulation effects on vowels						
Vowel perception - Targets, syllable timing, Duration, consonantal context, identification						
Stops - Rate effects, fricatives						
Dutch, voicing, fricatives						
Voiced intervocalic						
Context effects						
Duration - Adaptation						
Dichotic Syllables - Stimulus dominance, fusion effects						
Categorical perception						
Speech, nonspeech, stimulus/ear dominance						

UNCLASSIFIED

Security Classification