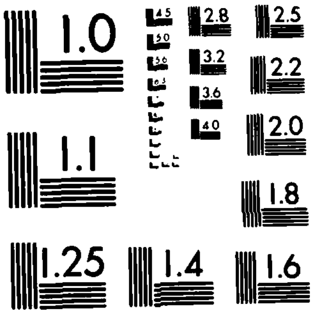
AD-A136 416 STATUS REPORT ON SPEECH RESEARCH A REPORT ON THE STATUS 1/3
AND PROGRESS OF S..(U) HASKINS LABS INC NEW HAVEN CT
A M LIBERMAN 30 SEP 83 SR-74/75(1983) N00014-83-C-0083
UNCLASSIFIED F/G 6/16 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

A136416

SR-74/75 (1983) (10)

Status Report on

SPEECH RESEARCH

April - September 1983

DTIC FILE COPY

83 12 28 067

HASKINS LABORATORIES

SR - 74/75(1983)

Status Report on

SPEECH RESEARCH

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

1 April - 30 September 1983

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06511

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

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

Michael Studdert-Kennedy, Editor-in-Chief
Nancy O'Brien, Editor
Margo Carter, Technical Illustrator
Gail Reynolds, Word Processor

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
Grant HD-16591

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

National Institutes of Health
Biomedical Research Support Grant RR-05596

National Science Foundation
Grant BNS-8111470

National Institute of Neurological and Communicative
Disorders and Stroke
Grant NS13870
Grant NS13617
Grant NS18010
Grant NS 07237
Grant NS 07196

Office of Naval Research
Contract N00014-83-C-0083

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President
Franklin S. Cooper,* Assistant to the President
Patrick W. Nye, Vice President
Michael Studdert-Kennedy,* Vice President, Research
Raymond C. Huey,* Treasurer
Bruce Martin, Controller
Alice Dadourian, Secretary

Investigators

Arthur S. Abramson*
Peter J. Alfonso*
Thomas Baer
Patrice Beddor+
Fredericka Bell-Berti*
Shlomo Bentin[1]
Catherine Best*
Gloria J. Borden*
Susan Brady*
Robert Crowder*
Laurie B. Feldman*
Carol A. Fowler*
Louis Goldstein*
Vicki L. Hanson
Katherine S. Harris*
Satoshi Horiguchi[2]
Leonard Katz*
J. A. Scott Kelso
Andrea G. Levitt*
Isabelle Y. Liberman*
Leigh Lisker*
Virginia Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald P. Shankweiler*
Sigfrid Soli*
Esther Thelan[3]
Betty Tuller*
Michael T. Turvey*
Robert Verbrugge*
Douglas Whalen

Technical and Support Staff

Michael Anstett
Margo Carter
Philip Chagnon
Elizabeth P. Clark
Vincent Gulisano
Donald Hailey
Sabina D. Koroluk
Nancy O'Brien
Gail K. Reynolds
William P. Scully
Richard S. Sharkany
Edward R. Wiley
David Zeichner

Students*

Eric Bateson
Suzanne Boyce
Jo Ann Carlisle
Andre Cooper
Tova Clayman
Patricia Ditunno
Jan Edwards
Jo Estill
Nancy Fishbein
Carole E. Gelfer
Janette Henderson
Charles Hoequist
Robert Katz
Bruce Kay
Noriko Kobayashi
Rena Krakow
Peter Kugler
Harriet Magen
Sharon Manuel
Richard McGowan
Daniel Recasens
Hyla Rubin
Judith Rubin
Arnold Shapiro
Suzanne Smith
Katyanee Svastikula
Louis Tassinary
Ben C. Watson
Deborah Wilkenfeld
David Williams

---

*Part-time
[1] Visiting from Hadassah University Hospital, Jerusalem, Israel
[2] Visiting from University of Tokyo, Japan
[3] Visiting from University of Missouri, Columbia, MO
+NIH Research Fellow

CONTENTS

I. Manuscripts and Extended Reports

I. MANUSCRIPTS AND EXTENDED REPORTS

THE ASSOCIATION BETWEEN COMPREHENSION OF SPOKEN SENTENCES AND EARLY READING ABILITY: THE ROLE OF PHONETIC REPRESENTATION*

Virginia A. Mann,+ Donald Shankweiler,++ and Suzanne T. Smith++

Abstract. When repeating spoken sentences, children who are good readers tend to be more accurate than poor readers because they are able to make more effective use of phonetic representation in the service of working memory (Mann, Liberman, & Shankweiler, 1980). This study of good and poor readers in the second grade has assessed both the repetition and comprehension of relative-clause sentences to explore more fully the association between early reading ability, spoken sentence processing, and use of phonetic representation. It was found that the poor readers did less well than the good readers on sentence comprehension as well as on sentence repetition, and that their comprehension errors reflected a greater reliance on two sentence processing strategies favored by young children: the minimum-distance principle and conjoined-clause analysis. In general, the pattern of results is consonant with a view that difficulties with phonetic representation could underlie the inferior sentence comprehension of poor beginning readers. The finding that these children place greater reliance on immature processing strategies raises the further possibility that the tempo of their syntactic development may be slower than that of good readers.

There is evidence that reading disability among children in the early elementary grades reflects some rather specific problems in the area of language. The evidence can be found in studies that have compared the performance of good and poor beginning readers on parallel language and nonlanguage tasks. Poor beginning readers are typically inferior to good beginning readers in the ability to identify spoken words that are partially masked by noise, although they are equivalent to good readers when the masked items are nonspeech environmental sounds (Brady, Shankweiler, & Mann, 1983).

---

*Also Journal of Child Language, in press.

+Also Bryn Mawr College.

++Also University of Connecticut.

Acknowledgment. We are grateful to the East Hartford school authorities for granting permission for this study. We wish to thank the principals, the reading consultants, the third grade teachers, and students at the Barnes, Center, Mayberry, Silver Lane, and Stevens Schools for their cooperation. We also thank our colleagues Vicki Hanson, Isabelle Liberman, and Ignatius Mattingly for their helpful comments on earlier versions of this manuscript. The research described herein was supported by NICHD grant HD-01994 to Haskins Laboratories, Inc.

[HASKINS LABORATORIES: Status Report on Speech Research SR-74/75 (1983)]

Likewise, they are inferior to good readers in performance on a memory task that involves recognizing printed nonsense syllables, but not when the task involves recognizing photographs of unfamiliar faces (Liberman, Mann, Shankweiler, & Werfelman, 1982). They are inferior to good readers in ordered recall of word strings, but not in ordered recall of nonverbal sequences in a block-tapping task (Mann & Liberman, in press). Finally, poor readers are inferior in ordered recall of nameable pictures, but not in ordered recall of visual patterns that do not readily lend themselves to verbal labeling (Katz, Shankweiler, & Liberman, 1981). It is thus apparent that in young children with reading disability, we do not ordinarily find a general impairment in learning and memory, or an overall retardation in language. Instead we find deficits in specific language functions.

Our attention has focused on a deficiency that we believe is basic to reading and other language skills in reading disabled children, namely, the use of phonetic representation in working memory. Poor readers' problems with verbal short-term memory are evident in their performances on a variety of tasks that require retention of ordered strings of visually-presented or spoken words and other stimuli that lend themselves to verbal labeling (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). Insight into the underlying basis of deficient memory performance is gained from the special case in which the stimulus items rhyme. Under this condition, the good readers' advantage is greatly reduced or even eliminated presumably because of interitem interference. The poor readers, in contrast, do not show much interference as a result of rhyme. This result, originally demonstrated for randomly ordered material, also obtains for spoken sentences. It is apparent that in children who are good readers, but not in those who are poor readers, memory performance depends critically on the phonologic properties of the stimulus material. The discrepancy between the two groups in response to rhyming and nonrhyming items, together with the poor readers' inferior performance on the latter, suggests that poor readers are somehow impaired in their ability to retain the full phonetic representation in working memory.

In addition to the studies of working memory, additional research conducted in our laboratory indicates that poor readers also perform less adequately than good readers on other tasks (for example, certain speech perception tasks, Brady et al., 1983, and tests of object naming, Katz, 1982) that involve accessing a phonetic representation. These further findings support the view that the basic deficit involves primarily the phonological component of language.

The research we report here is concerned with ramifications of this problem for processing sentences. It was motivated by the suggestion of some of our colleagues (Liberman, Mattingly, & Turvey, 1972) that, owing to its role as a vehicle for working memory, phonetic representation has a crucial role in sentence processing. Previous research has shown that poor readers fail to repeat spoken sentences as accurately as good readers do (Perfetti & Goldman, 1976; Weinstein & Rabinovitch, 1971; Wiig & Roach, 1975). Our research (Mann et al., 1980) confirms these findings and further reveals a difference between good and poor readers that is dependent on the makeup of the test sentences. In particular, we have found that while manipulations of syntactic structure and meaningfulness of sentences affected the performance of both good and poor readers equally, manipulations of phonetic confusability affected good readers more strongly than poor readers (Mann et al., 1980). The poor readers' performance was unaffected by the presence of a high density of phonetically-confusable words in the test sentence being repeated--a condition that so extensively penalizes good readers as to make their repetition performance equivalent to that of poor readers. We have argued that the observed tendency of poor readers toward inaccurate repetition of normal sentences is an expression of the same underlying deficit that makes them relatively tolerant of a high density of rhyme in sentences and word strings. In other words, their difficulties with repeating a sentence reflect their failure to make effective use of the phonetic structure of that sentence as a means of retaining a verbatim representation of it in working memory. Out of this failure comes a difficulty with retention not only of the words themselves, but also of their order of occurrence.

The issue we raise in the present study is whether difficulties with phonetic representation penalize the comprehension of a sentence as well as its repetition. Certainly in the case of a language such as English, in which the sequential order of words tends to convey syntactic structure, an ineffective use of phonetic representation could, in principle, lead to difficulty in sentence comprehension. The literature does, in fact, contain evidence that poor readers do not comprehend certain classes of spoken sentences as well as good readers (Byrne, 1981a; Satz, Taylor, Friel, & Fletcher, 1978). Our concern is with the extent to which the comprehension difficulties of these children can be understood as a product of an ineffective phonetic representation, and the extent to which the difficulties reflect problems with syntactic structure, as such. Certainly, poor readers may fail to comprehend certain sentences because they fail to remember the component words sufficiently and for that reason fail to recover syntactic structure. But in addition, their comprehension might also be limited by a deficient ability to apprehend the structure (Byrne, 1981a, 1981b).

In the present study we have sought to confirm that differences in comprehension of spoken sentences can indeed distinguish good and poor beginning readers. We have also attempted to discover the extent to which such differences, provided they are reliable, turn primarily on effectiveness of phonetic representation, and the extent to which they reflect differences in syntactic competence as such. Our approach has been to study the repetition and comprehension of several types of sentences among a population of good and poor third-grade readers. A preliminary study (in preparation) assessed the performance of these children on an oral sentence comprehension test, the Token Test of De Renzi and Vignolo (1962), which has proved to be a sensitive diagnostic of even minor disturbances of sentence comprehension associated with aphasia in adults (see, for example: De Renzi & Faglioni, 1978; De Renzi & Vignolo, 1962; Orgass & Poeck, 1966; Poeck, Orgass, Kerschensteiner, & Hartje, 1974). We found that the good readers surpassed the poor readers on comprehension of those later Token Test items that could be expected to tax working memory. Thus it was established that poor readers do indeed exhibit a greater degree of difficulty in comprehension of certain spoken sentences than good readers. However, we found nothing to suggest that the poor readers' errors on the Token Test items involved a syntactic deficit as such. In general those sentences that proved difficult for the poor readers also proved difficult for the good readers.

A second study (in preparation), using the same group of children, focused on the repetition and comprehension of sentences containing reflexive pronouns, such as those in 1a and 1b. These, like the Token Test items, have proven difficult for aphasic adults to comprehend (Blumstein, Goodglass, Statlender, & Biber, 1983):

1a. The clown watched the boy spill paint on himself.

1b. The clown watching the boy spilled paint on himself.

In such sentences, syntactic structure rigidly determines the antecedent of the reflexive pronoun, and by probing for subjects' comprehension of that antecedent, one can assess their ability to recover syntactic structure. Whereas our good readers surpassed the poor readers in repeating sentences like 1a and 1b, they did not surpass them on a picture-verification test of comprehension that required them to choose a drawing whose meaning best matched that of a spoken sentence. Children in both groups made few errors in identifying the antecedents of pronouns in single-clause sentences. They also made fewer errors on sentences like 1a than on sentences like 1b, in which the anaphoric referent could not be correctly assigned by adopting a minimum distance strategy. However, the number and pattern of errors were similar for good and poor readers, suggesting that they had equal mastery--or lack of mastery--of at least this aspect of syntactic structure.

Thus far, then, our findings give no reason to postulate a specific syntactic competence problem on the part of poor readers. Yet, we must be cautious about reaching a more general conclusion with regard to syntactic competence because in our earlier research we employed only a very limited set of syntactic constructions. Therefore, as a follow-up to our previous study, we studied the repetition and comprehension of a new set of spoken sentences. In choosing materials for this study, we were guided in part by research on language acquisition. Embedded constructions having a basic Subject-Verb-Object (SVO) construction and either a subject-relative or object-relative embedded clause are of special interest to students of syntactic development. Examples of such sentences appear in 2a-2d, where the first code letter refers to the role of the relativized noun in the matrix clause, and the second letter refers to the role of the head noun within the relative clause itself:

2a. (SS) The dog that chased the sheep stood on the turtle.

2b. (SO) The dog that the sheep chased stood on the turtle.

2c. (OS) The dog stood on the turtle that chased the sheep.

2d. (OO) The dog stood on the turtle that the sheep chased.

Each of these four sentences contains the same ten words; thus any differences in their meanings must be marked by word order and such phonological features as pitch contour, the juncture pause between words, and the stress on individual words. Because sensitivity to word order and phonological features might be expected to place a certain demand on the use of phonetic representation as a means of temporarily holding an utterance in working memory, we speculated that comprehension of sentences like those in 2a-2d might distinguish good and poor readers.

We were additionally interested in such sentences, moreover, because of the wealth of evidence about the errors young children tend to make, and because of current views about the emerging syntactic competence that those errors may reflect. Let us briefly consider some of that evidence. Many investigators have found that young children in the three- to eight-year-old range tend to make more comprehension errors on SO constructions than on types SS, OS, or OO (deVilliers, Tager-Flusberg, Hakuta, & Cohen, 1979; Sheldon, 1974; Tavakolian, 1981). A few investigators have also claimed that performance on OS constructions is poorer than on SS ones (Brown, 1971; Sheldon, 1974; Tavakolian, 1981). Smith (1974) attributes the relative difficulty of SO to the fact that it violates two common properties of English sentence configuration, notably the "SVO configuration" (Bever, 1970) that holds that the sequence "N-V-N" is typically "subject-verb-object," and the "minimum-distance principle" (Chomsky, 1969; Rosenbaum, 1967) that holds that the missing subject of a given verb is the noun most proximal to it. In contrast to SO, the SS construction violates only the minimum distance principle, OO violates only the SVO configuration, and OS violates neither.

One might note, however, that superior performance on SS as compared to OS cannot be explained in terms of the number of violations of expected sentence configuration, since SS violates one expectation, whereas OS violates none. A solution to this difficulty was proposed by Tavakolian (1981), who suggested that children tend to treat the two clauses of sentences such as 2a-2d as being conjoined clauses rather than as a relative clause embedded within a matrix clause (Tavakolian, 1981). Such a "conjoined clause analysis" predicts that both sentences 2a and 2c will be interpreted as meaning "The dog stood on the turtle and chased the sheep,"--a strategy that leaves the meaning of 2a intact, but alters the meaning of 2c so that it becomes equivalent to 2a. When young children act out the meaning of sentences with relative clauses like those in 2a-2d, their responses meet with this and other predictions of a conjoined-clause analysis (Tavakolian, 1981).

These accounts of children's erroneous responses to relative-clause sentences are highly germane to our interest in the sentence processing skills of good and poor beginning readers. Certainly ineffective phonetic representation might lead to impaired sentence comprehension because neither the words nor the order of occurrence are available for correct parsing. A child may assume, therefore, that the subject of a recently heard verb is the most proximal noun because of an impoverished representation of the words and their order, and thus adhere to the minimum-distance principle. However, ineffective phonetic representation, in and of itself, would not necessarily lead a child to link a verb to a noun that occurred at some remove in the sentence, as happens in a conjoined-clause analysis. We therefore anticipated that the poor readers' inefficient phonetic processing and their consequent weakness in short-term retention might lead them to make more errors than good readers that reflect adherence to the minimum-distance principle. If, further, the poor readers were to make both more minimum-distance errors and also more conjoined-clause analysis errors than the good readers, then it might be argued from the fact that such errors are typical of younger children that the poor readers are indeed on a slower schedule of syntactic development (Byrne, 1981a, 1981b; Satz et al., 1978), even though the trend of the development might be normal. If, on the other hand, poor readers make errors that are qualitatively different from those of good readers and other young children, we would have strong reason to entertain the possibility of a primary deficiency in syntactic competence as such. A finding that the pattern of poor readers' performance across the four different constructions exemplified in 2a-2d is different from that of good readers likewise would also suggest that in addition to problems involving the working memory, there is further an underlying syntactic deficiency.

METHOD

Subjects

The subjects were third-grade pupils attending public schools in East Hartford, Connecticut. All were native speakers of English with no known speech or hearing impairment and had an intelligence quotient of 90 or greater (as measured by the Peabody Picture Vocabulary Test; Dunn, 1965). Inclusion in the experiment was based jointly on teacher evaluations of reading ability and scores on the verbal comprehension subtest of the Iowa Test of Basic Skills (Hieronymus & Lindquist, 1978), which had been administered four months previously. The 18 good readers included three boys and fifteen girls (mean Iowa grade-equivalent score 4.59; range 4.1 - 5.2). The 17 poor readers included nine boys and eight girls (mean grade-equivalent score 2.32; range 1.7 - 2.6). The mean IQ for the good readers (109.3) was not significantly greater than that of the poor readers (107.7). The poor readers (mean age 9.21 years) were slightly (but not significantly) older than the good readers (mean age 8.95 years) at the time of testing.

Materials

The test materials consisted of eight tokens of each of the nonrestrictive relative clause constructions illustrated in 2a-2d. These four constructions represent the orthogonal variation of two parameters: the role of the relativized noun in the main (matrix) clause--i.e., whether the clause was subject-relative (S-) or object relative (O-)--and the role of the relative agent (the head noun) within the relative clause--i.e., whether it was the subject (-S) or the object (-O). They include:

SS--a center embedded construction of the form "N1 that V1 N2 V2 N3," in which the subject of the main clause is also the subject of the relative clause.

SO--a center embedded construction of the form "N1 that N2 V1 V2 N3," in which the subject of the main clause is the object of the relative clause.

OS--a right-branching construction of the form "N1 V1 N2 that V2 N3," in which the object of the main clause was the subject of the relative clause.

OO--a right-branching construction of the form "N1 V1 N2 that N3 V2," in which the object of the main clause is also the object of the relative clause.

Eight common animal names served as nouns: turtle, owl, alligator, horse, dog, gorilla, cat, and sheep. Their position and occurrence were randomized within each sentence type with the restriction that cat and dog never occur in the same sentence, since their stereotypical roles might bias children's response. Eight easily-depicted action verbs were used: hit, kick, run after, chase, jump on, kiss, stand on, and push. Their position and occurrence within each set of sentences was randomized with the restriction that actions that could be visually confusing to the test administrator did not occur in the same sentence (i.e., hit and kick, or hit and push). To further facilitate the scoring, none of the nouns and verbs in a sentence began with the same letter.

The test sentences were randomized and recorded on audio tape by a male native speaker of English who used natural intonation at a comfortable rate of delivery. At the time of recording, each sentence was preceded by an alerting signal (a bell). Small plastic animals were used for the toy manipulation task that provided the measure of sentence comprehension.

Procedure

Each subject was tested individually in two thirty-minute sessions during which the previously mentioned experiments were also conducted. The first session began with the experimenter placing the small plastic animals in a row on the table in front of the subject, and requesting the subject to name each one. Any incorrect or nonstandard response, such as calling the cat a "kitty," was corrected. The experimenter then read three single-clause sentences to the subject, who was asked to enact each one. These practice items included three of the eight test verbs along with the names of any animals that had been misnamed. Successful completion of the practice items was followed by presentation of the pre-recorded test materials over a loudspeaker. Before playing each test sentence, the experimenter selected the appropriate trio of animals and placed them in a predetermined random order, two inches apart, on the table in front of the subject. The subject was instructed to listen carefully to the entire tape-recorded sentence, which would be preceded by a bell, and then to act out its meaning. Emphasis was placed on listening to the entire sentence before starting to respond. Sentences were repeated only on the subject's request, and the incidence of repetitions was noted. The subject's manipulation of the animals was transcribed in terms of which animal did what action to whom.

In the second session, which was conducted at least one week after the first, the subject was instructed to listen to the sentence and to repeat it into a microphone. Each test sentence was presented only once. Responses were transcribed by the examiner, and were also recorded on audio tape for further analysis.

RESULTS

This experiment was designed to corroborate previous findings that indicated that good and poor readers tend to differ both in use of phonetic representation during sentence repetition and in spoken sentence comprehension. Further we sought to determine whether good and poor readers differ in their ability both to repeat and to comprehend a given set of spoken sentences, and to clarify the basis of any comprehension differences that were found. In order to accomplish this aim, error scores were obtained, and separate analyses performed on the data from the sentence repetition and sentence comprehension tests.

Sentence Repetition

In scoring the data from the sentence repetition task, we considered any response that departed from the test sentence as incorrect. The number of incorrect sentences (out of a maximum of eight) was then computed for each construction (SS, SO, OS, and OO); mean values for good and poor readers appear in Table 1. We found, as expected, that good readers made fewer errors than poor readers, $\underline{F}(1,33) = 4.84$, $\underline{p} < .03$. There was, however, no significant effect of either orthogonal variation in sentence structure—the role of the relativized agent in the main clause (i.e., S- vs. O-), and the role of the head noun in the relative clause (i.e., -S vs. -O). Moreover, there was no interaction of reading ability with either structural variation. As can be seen in Table 1, error scores are relatively constant across the four different types of structure, as is the extent of difference between good and poor readers. A further analysis of the pattern of children's errors within each sentence also fails to reveal any qualitative differences between good and poor readers. As can be seen in Table 2, where mean errors appear for nouns and verbs as a function of their order of occurrence in the sentence, children in both groups were more likely to repeat later parts of the sentence incorrectly, $\underline{F}(2,66) = 6.95$, $\underline{p} < .002$ for nouns, and $\underline{F}(1,33) = 16.11$, $\underline{p} < .005$ for verbs. While good readers made fewer errors than poor readers both on nouns $\underline{F}(1,33) = 4.26$, $\underline{p} < .05$, and verbs $\underline{F}(1,33) = 4.53$, $\underline{p} < .05$, there was no interaction of word position and reading ability.

Sentence Comprehension

Having confirmed that good readers made fewer errors in recall of the test sentences than poor readers, we now turn to the results of the toy manipulation task, which was our measure of sentence comprehension. These data consist of the experimenter's transcriptions of the responses each child made in manipulating the various toy animals. A response was scored as correct if each of the three nouns had been assigned its proper role(s) as subject or object of the appropriate verb, otherwise it was scored as incorrect. Each child's comprehension error score is the total number of incorrect sentences. These scores proved to be positively correlated with error scores on the sentence repetition test, $\underline{r}(35) = .40$, $\underline{p} < .02$. They are also significantly correlated with the grade-equivalent scores on the Iowa Reading Test, $\underline{r}(35) = -.43$, $\underline{p} < .01$.

Individual error scores on the four different sentence types (i.e., SS, SO, OS and OO) were computed and incorporated into an analysis of variance that included the factors reading level, role of relativized noun in the matrix clause, and role of the head noun in the relative clause. The results are displayed in Figure 1, and may be summarized as follows: The role of the relativized noun in the matrix clause had no main effect, although the effect of the role of the head noun was significant, $\underline{F}(1,33) = 21.8$, $\underline{p} < .005$, as was the interaction between these two structural factors, $\underline{F}(1,33) = 17.58$, $\underline{p} < .005$. These results agree with previous findings insofar as performance

---

Table 1

Mean Number of Incorrect Sentences on the Sentence Repetition Test
(Maximum number of possible errors equals eight)

Sentence Type	Good Readers	Poor Readers
SS	2.22	3.71
SO	2.67	3.94
OS	2.39	3.71
OO	1.78	3.65

Table 2

Mean Number of Incorrect Words During Sentence Repetition as a Function of Word Class and Word Position

Class:	Noun			Verb	
Position:	1	2	3	1	2
Good readers	1.89	2.67	2.72	1.22	3.11
Poor readers	3.29	5.06	5.59	3.18	4.24

---

on SS items was superior to that on OS and SO (Brown, 1971; Sheldon, 1974; Tavakolian, 1981). However, contrary to what others have found (deVilliers et al., 1979, Sheldon, 1974; Tavakolian, 1981), SO was not more difficult than OO. The discrepancy between our results and previous ones could reflect age differences: Other studies have employed subjects aged three to eight; ours were all aged eight and older.

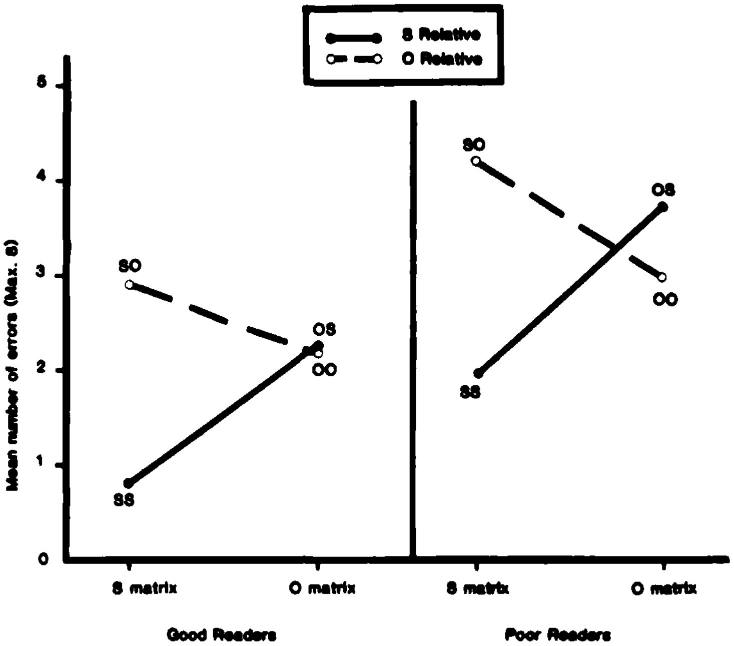


Figure 1. The performance of good and poor readers on comprehension of relative clause sentences, plotted in terms of the number of incorrect sentences as a function of the role of the relativized noun in the matrix clause (S matrix vs. O matrix) and the role of the head noun within the relative clause (S relative vs. O relative).

Of central importance is the comparison of children in the two reading groups. The poor readers, as we had anticipated, made more incorrect responses than the good readers, $F(1,33) = 9.41$, $p < .01$, yet the relative difficulty of the four different constructions was the same for good and poor readers. Thus there is no significant interaction between reading ability and the influence of matrix clause or relative clause structure. Responses to SS items were significantly more often correct than those to OS items, both for good readers, $t(34) = 5.15$, $p < .005$ and poor readers, $t(32) = 3.41$, $p < .005$; although both groups tended to miss SO items more often than OO and OS, the differences failed to reach significance.

These initial analyses were supplemented by a more detailed analysis of the responses in search of some measure that might distinguish between the good and poor readers. Using the procedure described by Tavakolian (1981), children's toy manipulation responses were coded with respect to the linear order of the three nouns in the sentence, so as to denote which nouns were chosen as subject and object of each verb. When coded this way, the response to each sentence is represented by two double-number sequences, the first indicating the nouns taken as subject and object, respectively, of the first verb, and the second indicating those taken as subject and object of the second verb. The correct response to an SS sentence is thus represented as 12,13; that for SO, is 21,13; for OS, 12,23; and for OO, 12,32.

Two classes of errors are of primary interest: those that reflect a conjoined-clause analysis, as discussed by Tavakolian (1981), and those that reflect application of a minimum-distance principle (Chomsky, 1969; Rosenbaum, 1967) in which the noun closest to a verb is chosen as its subject. As outlined in Tavakolian (1981), a conjoined-clause analysis would yield the correct response to SS sentences, but an incorrect response of 12,13 to OS, incorrect responses of either 21,23 or 12,13 to SO, and an incorrect response of 12,13 to OO sentences. An incorrect response of 12,31 to OO sentences, as discussed by Tavokolian, is also consistent with a conjoined-clause analysis. We computed for each subject the total number of errors on SO, OS, and OO that fell into these categories and thus could be taken as evidence for reliance on conjoined-clause analysis. The results, given in Table 3, reveal that for children in both groups, the number of such errors was considerable. Poor readers, however, made significantly more errors of this type than good readers, $\underline{t}(33) = 2.08$, $\underline{p} < .05$.

Table 3

Distribution of Errors on the Sentence Comprehension Test
(Mean number of errors)

Basis of Error:	Good Readers	Poor Readers
Minimum-Distance Principle (Maximum = 8)	0.33	1.59
Conjoined-Clause Analysis (Maximum = 24)	4.50	7.32
"SOV" Configuration (Maximum = 16)	0.72	1.35
Other (Maximum = 32)	2.00	3.76

Application of the minimum-distance principle, as opposed to a conjoined-clause analysis, would yield a correct response to OS constructions, but an erroneous response of 12,23 to SS constructions. When the number of erroneous responses of this type was computed and averaged across subjects, we discovered, as shown in Table 3, that the poor readers made significantly more such errors than the good readers, $\underline{t}(33) = 2.58$; $\underline{p} < .02$. For neither group, however, was the raw number of errors involving the minimum-distance principle as great as the raw number reflecting a conjoined-clause analysis, $\underline{t}(17) = 4.6$, $\underline{p} < .001$ for good readers; $\underline{t}(16) = 5.24$, $\underline{p} < .001$ for poor readers. However, when raw scores are adjusted for the difference in the number of opportunities for errors of each type, only the good readers made significantly more conjoined-clause errors than errors involving the minimum-distance principle, $t(17) = 3.8$, $\underline{p} < .005$.

Finally, we computed the number of errors made by each child that could not be accounted for either by the application of a minimum-distance principle or a conjoined-clause analysis. Children in both groups made an appreciable number of erroneous responses of 12,23 on OO and SO sentences, perhaps because they tended to interpret the configuration "NNV" that appears in such sentences as "subject-object-verb." The mean number of errors of this type appears in Table 3 under the heading "SOV" configuration, and we note that any difference between good and poor readers fails to reach significance. The remaining errors failed to follow any particular pattern. The mean number of such "other" errors is also given in Table 3. Here also, good and poor readers did not differ significantly ($\underline{p} > .05$).

DISCUSSION

Our review of the literature on language-related problems in poor readers led us to conclude that these children tend to perform at a disadvantage on many tasks that require temporary retention of verbal material, including repetition of spoken sentences. We have presented evidence that the working memory problems of poor readers, including their sentence repetition difficulties, are traceable to their failure to make effective use of phonetic representation. The present study explored the prediction that ineffective phonetic representation will also give rise to comprehension difficulties whenever language processing stresses working memory. The study employed an extensive set of relative clause constructions to assess the suggestion (Byrne, 1981a, 1981b; Satz et al., 1978) that reading-disabled children are less proficient than children who are good readers in comprehension of certain spoken sentence constructions that are mastered comparatively late. We chose this set of constructions for two reasons. First, we wished to control for sentence length and vocabulary as we ascertained whether good and poor readers could make equal use of word order and phonological structure as cues to sentence meaning. Second, we were aware of regularities in young children's errors in acting out relative-clause constructions, and of interpretations in the literature regarding the emerging syntactic competence that these errors reflect. Given that we found poor readers' comprehension of relative clause constructions to be less accurate than that of good readers, we could then attempt to clarify the precise reasons for the differences.

In an earlier study, we had tested the same groups of third-grade children on two tests of comprehension, the Token Test and a picture-verification test involving sentences with reflexive pronouns. The poor readers performed significantly worse on the more difficult items from the Token Test, which tend to stress working memory, but the test of comprehension of reflexive pronouns did not differentiate the groups, possibly because the use of pictorial cues in the latter test considerably reduces the demands on working memory. Because the Token Test results did support our expectations, it seemed worthwhile to take another approach to the assessment of sentence comprehension in these children.

The present study of relative-clause constructions assessed good and poor readers' ability to repeat test sentences, and it further compared their comprehension of the same sentence structures, noting both the quantity and nature of the errors that occurred in acting out sentence content. Our primary interest was to discover whether the comprehension difficulties of the poor readers may be regarded as a manifestation of problems with using phonetic representation to store the words of a sentence in some temporary working memory. Alternatively, the difficulties could imply an inability to analyze certain kinds of syntactic structures.

In regard to the test of sentence repetition, the results of this study are in agreement with our previous research (Mann et al., 1980), in finding good and poor readers were distinguished in the number of errors made on immediate recall but not in the types of errors. The poor readers, then, appear to have had a less effective means of retaining the words of sentences in working memory. The particulars of sentence structure turned out to have little effect on the number of errors made in repetition: Whether the relative clause modified the subject or object of the matrix clause, or whether the relativized noun phrase was the subject or object of the relative clause, did not systematically influence the accuracy of children's performance. Moreover, these variations did not affect the magnitude of the difference between the performance of good and poor readers. The poor readers were simply worse in general. This accords well with the view that phonetic memory limitation is an important factor governing difficulty of sentence repetition in poor readers.

Most importantly, the present test of comprehension successfully differentiated between good and poor readers. Poor readers made more errors than good readers, not only in repeating the words of the test sentences, but also in acting out the meaning of these same sentences. In the case of comprehension, however, the type of sentence structure significantly influenced the accuracy of performance: Sentences with subject-relative clauses in which the relativized noun phrase also serves as the subject (SS) proved the easiest structure both for good and poor readers, whereas the remaining three sentence types (SO, OS and OO) were equally difficult. Yet for present purposes, the important point is that the relative difficulty of the different types of test sentences was the same for good and poor readers. Thus, while the poor readers made consistently more mistakes than the good readers in their acting out of these sentences, they did so to an equal extent on all four of the constructions. Both in repetition and in comprehension, then, the good and poor readers differed in the number of errors made but they failed to differ in susceptibility to variations in syntactic structure. This we regard as a major outcome of the experiment.

As to the question we raised concerning the basis of the comprehension differences between the good and poor readers, such an across-the-board decrement as we have observed on the part of poor readers is as one would expect, given the assumption that their phonetic representations of the words of the sentence are less effective than those of good readers. In interpreting these findings, we should stress that the good readers' and poor readers' performance was affected by the experimental variables in the same way. We can probably assume, therefore, that they employ much the same sentence processing strategies, although the extent of their reliance on a given strategy may differ. What, then, accounts for the overall inferior performance of the poor readers? Given the moderate correlation between sentence repetition performance and sentence comprehension, and our previous demonstration of the importance of phonetic representation in poor readers' sentence repetition (Mann et al., 1980), we can assume that effectiveness of phonetic representation is certainly one factor behind the comprehension differences of good and poor readers. But, as we anticipated both in the introductory section of this paper and elsewhere (Liberman, Liberman, Mattingly, & Shankweiler, 1980; Mann & Liberman, in press), it is not necessarily the only factor. We might explain preferences for strategies based on the minimum-distance principle by reference to limitations of working memory, but limited memory capacity cannot be invoked to account for every aspect of the error pattern on the comprehension test. Indeed, the frequent adherence of children in both groups to a conjoined-clause analysis, which requires assimilation of words from well-separated portions of the sentence, does not readily lend itself to a memory interpretation.

The occurrence of both kinds of errors, those reflecting use of the minimum-distance principle, and those reflecting a conjoined-clause analysis, has been well documented among normal young children (Chomsky, 1969; Smith, 1974; Tavakolian, 1981), and their occurrence among poor readers fits well with the hypothesis that children who encounter reading difficulties may exhibit a maturational lag in language abilities (Byrne, 1981a, 1981b; Satz et al., 1978). This hypothesis receives support from a study by Byrne (1981a) that we find particularly relevant, since it involved an assessment of good and poor readers' comprehension of relative clause constructions like 3a and 3b:

3a. The bird that the rat is eating is blue.

3b. The bird that the worm is eating is yellow.

Byrne reports that when children are asked to decide which of two pictures correctly depicts the meaning of a sentence, poor readers perform as well as good readers on "semantically reversible" sentences like 3a, but do less well on "implausible" sentences like 3b. Thus it would seem that poor readers place a greater reliance on extra-linguistic cues than do good readers. In a discussion of this and another finding involving poor readers' difficulty with sentences such as "John is easy to please," Byrne (1981a) concludes that a deficient use of phonetic memory coding is not the factor responsible for poor readers' sentence comprehension difficulties. In his view:

> A better characterization is one that places poor readers further down on the linguistic development scale, relatively dependent upon

> strategies acquired in early language mastery...upon heuristic devices, including knowledge of what is usual in the world. (p. 210)

We agree with Byrne that the notion of maturational lag may be an apt way of conceptualizing the problem in many cases of early reading disability, and we have adopted this viewpoint in our studies of linguistic awareness and its relation to reading (Liberman et al., 1980; Mann & Liberman, in press). However, though it is true, as we noted, that working memory problems do not account for all of poor readers' errors in sentence processing, we cannot accept Byrne's conclusion that deficiencies in use of a phonetic memory code are not relevant to the sentence comprehension difficulties of poor readers. Our research leads us to believe that one of the factors underlying the dependency of poor readers (and, perhaps, of young children in general) on an immature grammar and world-knowledge heuristics is that their phonetic representation of the words of a lengthy sentence is often insufficient to support full recovery of syntactic structure. The successful language learner must somehow assess large portions of the phonetic structure of the utterance at hand, and rely on word order and certain phonological features to establish the correct syntactic structure and thus the correct meaning of the utterance. It is for this purpose, we suspect, that phonetic representation in working memory exists in the first place. Thus a deficient capacity to form phonetic representations may limit the development of syntactic competence. In light of these considerations, we are led to speculate further that ineffective phonetic representation may serve to retard the tempo of syntactic development among children who are poor readers. Although we do not wish to exclude prematurely the possibility that poor readers may also have a specific syntactic deficiency, we find nothing in the data that would specifically indicate such a deficiency. Rather, we would note that the language tasks that best distinguish good and poor readers are most often precisely those that place special demands on phonetic representation.

REFERENCES

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), _Cognition and the development of language_. New York: Wiley.

Blumstein, S., Goodglass, H., Statlender, S., & Biber, C. (1983). Comprehension strategies determining reference in aphasia: A study of reflexivization. _Brain and Language_, _18_, 115-127.

Brady, S., Shankweiler, D., & Mann, V. A. (1983). Speech perception and memory coding in relation to reading ability. _Journal of Experimental Child Psychology_, _35_, 345-367.

Brown, H. D. (1971). Children's comprehension of relativized English sentences. _Child Development_, _42_, 1923-1926.

Byrne, B. (1981a). Deficient syntactic control in poor readers: Is a weak phonetic memory code responsible? _Applied Psycholinguistics_, _2_, 201-212.

Byrne, B. (1981b). Reading disability, linguistic access and short-term memory: Comments prompted by Jorm's review of developmental dyslexia. _Australian Journal of Psychology_, _33_, 83-95.

Byrne, B., & Shea, P. (1979). Semantic and phonetic memory codes in beginning readers. _Memory & Cognition_, _7_, 333-338.

Chomsky, C. (1969). *The acquisition of syntax in children from five to ten.* Cambridge, MA: MIT Press.

De Renzi, E., & Faglioni, P. (1978). Normative data and screening power of a shortened version of the Token Test. *Cortex*, *14*, 41-49.

De Renzi, E., & Vignolo, L. A. (1962). The Token Test: A sensitive test to detect receptive disturbances in aphasia. *Brain*, *85*, 665-678.

deVilliers, J. G., Tager-Flusberg, H. B., Hakuta, K., & Cohen, M. (1979). Children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, *8*, 499-518.

Dunn, L. M. (1965). *Peabody picture vocabulary test.* Circle Pines, MN: American Guidance Service.

Hieronymus, A. N., & Lindquist, E. F. (1978). *Iowa Test of Basic Skills.* Boston: Houghton Mifflin Co.

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

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

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

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

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

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

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

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

Orgass, B., & Poeck, K. (1966). Clinical validation of a new test of aphasia: An experimental study on the Token Test. *Cortex*, *2*, 222-243.

Perfetti, C. A., & Goldman, S. (1976). Discourse memory and reading comprehension skill. *Journal of Verbal Learning and Verbal Behavior*, *14*, 33-42.

Poeck, K., Orgass, B., Kerschensteiner, M., & Hartje, W. (1974). A qualitative study of Token Test performance in aphasic and non-aphasic brain-damaged patients. *Neuropsychologia*, 49-54.

Rosenbaum, P. S. (1967). *The grammar of English predicate constructions.* Cambridge, MA: MIT Press.

Satz, P. Taylor, H. G., Friel, J., & Fletcher, J. (1978). Some developmental and predictive precursors of reading disabilities: A six year follow-up. In A. L. Benton & D. Pearl (Eds.), Dyslexia: An appraisal of current knowledge. New York: Oxford University Press.

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

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. Journal of Verbal Learning and Verbal Behavior, 13, 272-281.

Smith, M. (1974). Relative clause formation between 29-36 months: A preliminary report. In Papers and Reports on Child Language Development (Vol. 8, pp. 104-110). Committee on Linguistics, Stanford University.

Tavakolian, S. L. (1981). The conjoined-clause analysis of relative clauses. In S. L. Tavakolian (Ed.), Language acquisition and linguistic theory. Cambridge, MA: MIT Press.

Weinstein, R., & Rabinovitch, M. S. (1971). Sentence structure and retention in good and poor readers. Journal of Educational Psychology, 62, 25-30.

Wiig, E. H., & Roach, M. A. (1975). Immediate recall of semantically varied sentences by learning disabled readers. Perceptual and Motor Skills, 40, 119-125.

PHONETIC CODING AND ORDER MEMORY IN RELATION TO READING PROFICIENCY: A COMPARISON OF SHORT-TERM MEMORY FOR TEMPORAL AND SPATIAL ORDER INFORMATION*

Robert B. Katz+, Alice F. Healy,++ and Donald Shankweiler+

Abstract. Since children with reading disability are known to have problems using a phonetic memory strategy, it was expected that their recall of order would be inferior to that of good readers in situations where a phonetic strategy is optimal, that is, when temporal order recall, but not necessarily spatial order recall, is required. On separate tests for retention of temporal sequence and spatial location, the good readers were better than the poor readers on the temporal order task as expected, but contrary to expectation, they maintained their superiority on the spatial task as well. Nevertheless, differences in the error patterns of the good and the poor readers are supportive of earlier evidence that links poor readers' short-term memory deficiencies to reduced effectiveness of phonetic representation.

Indications in the research literature suggest that reading problems in young children tend to be associated with poor memory for the order of items in a series (Bakker, 1972; Benton, 1975; Corkin, 1974; Mason, Katz, & Wicklund, 1975; Noelker & Schumsky, 1973; Stanley, Kaplan, & Poole, 1975). Shankweiler, Liberman, Mark, Fowler, and Fischer (1979) have supposed that difficulties with order recall may reflect a deficiency in the working memory system that supports comprehension of sentences both in speech and in reading. It has been argued that the working memory system used in processing connected discourse relies on phonetic coding for its operation (Liberman, Mattingly, & Turvey, 1972), and moreover, that the retention of item order is facilitated by the use of a phonetic memory strategy (Baddeley, 1978; Crowder, 1978). One

---

*Also Applied Psycholinguistics, in press.

+Also University of Connecticut.

++Also Yale University.

Acknowledgment. This research was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories and NSF Grant BNS80-00263 to Yale University. Alice F. Healy was supported by a Senior Faculty Fellowship from Yale University during the initial preparation of this manuscript. We are grateful to Frances Klein, supervisor of reading for the East Hartford, Connecticut, public school system, and the principals and teachers of the Barnes School and Stevens School for allowing us to work with the second-grade classes. We are also grateful to the children who participated and their parents for their cooperation. Finally, we would like to thank Robert Crowder, Ovid Tzeng, Vicki Hanson, and Shlomo Bentin for their helpful comments on an earlier version of this manuscript. Alice F. Healy is now at the University of Colorado, Boulder.

[HASKINS LABORATORIES: Status Report on Speech Research SR-74/75 (1983)]

of the mechanisms responsible for this facilitating effect of phonetic coding may be the rehearsal loop proposed by Baddeley (1979). Since it has been shown that poor beginning readers tend to depend less on phonetic coding than good readers on some laboratory memory tasks (Byrne & Shea, 1979; Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Mann, Liberman, & Shankweiler, 1980; Mark, Shankweiler, Liberman, & Fowler, 1977; Shankweiler et al., 1979), we may ask whether poor readers' difficulties in remembering order may be attributed to their failure to make appropriate use of phonetic codes in working memory.

If retention of order is indeed dependent on the use of phonetic codes, we might expect matched groups of good and poor beginning readers to differ in memory for item order only when the items to be remembered can easily be named, thereby allowing them to be held in phonetically-based working memory. When the items to be held in memory cannot easily be named, there is no clear basis for expecting good and poor beginning readers to differ. A recent study by Katz, Shankweiler, and Liberman (1981) supports this possibility, finding good and poor beginning readers not significantly different in their ability to reproduce the order of an array of figures that are difficult to label (Kimura's, 1963, nonsense drawings). When these subjects were tested for retention of the order of line drawings of common objects, however, the poor readers were deficient. Thus, it is clear that the poor readers' difficulty with memory for order applied specifically to remembering the order of items that could easily be coded linguistically and held in phonetic working memory. Comparable results were obtained by Holmes and McKeever (1979) in tests of memory for the order of photographed faces and printed words with adolescent good and poor readers. Neither study, however, provided direct evidence of the memory strategy the subjects actually used. Although it has been assumed that the subjects retained the easily named items by using phonetic codes, other aspects of the stimuli could have been used, e.g., semantic aspects or visual imagery. Moreover, ordering items with readily available names by memory has been found to be easier than ordering items that are difficult to name (Katz et al., 1981), making a direct comparison of the two tasks difficult. It is therefore important to address the question raised by Katz et al. by means of an experimental paradigm that avoids these difficulties, but in which, as before, the level of success in retaining item order could be expected to depend on the use of phonetic coding. Such a paradigm has been used by Healy (1975, 1977) for testing memory for order.

Healy (1975, 1977) has shown that two aspects of memory for order can usefully be distinguished: memory for temporal sequence and memory for spatial location. In most situations outside the laboratory, the two aspects of order memory are confounded, since they vary simultaneously. Healy has devised a technique for experimentally dissociating temporal and spatial order in a way that also allows us to infer the coding strategy used in the retention of each. (see Berch, 1979, for a discussion of this and related techniques.) Moreover, to the point of our present interest, her work with adult subjects has shown that memory for temporal sequence ordinarily depends strongly on the use of phonetic coding whereas retention of spatial location does not. Instead, spatial order recall depends on the retention of the temporal-spatial pattern of the stimulus display (Healy, 1975, 1977, 1978, 1982).

If by using this method we were able to dissociate the two aspects of order memory in children, we should be well placed to infer the memory strategies actually adopted by good and poor readers and to compare directly the strategies favored by each group. Thus, we would be in a position to pinpoint more definitely than heretofore the poor readers' difficulty in retaining each type of order information by showing whether it is tied to the use of phonetic coding.

The technique used by Healy (1975, 1977) involves successive visual presentations of a set of stimulus items whose order is to be remembered. On each trial, the same set of items, which is known to the subjects beforehand, is always used. Therefore, there is essentially no requirement for remembering the items themselves, but only their order of presentation. In the temporal order recall condition, the spatial order of the items is kept constant, whereas in the spatial order recall condition, temporal order does not vary. By using conditions that are completely parallel, this methodology separately assesses the two aspects of order memory in a comparable manner. Inasmuch as the original technique had been designed for adult subjects, it was necessary to modify it to make it suitable for use with children. The memory load on each trial was reduced from four to three items and the rate of stimulus presentation was slowed. These changes were introduced in order to ensure that the least successful subjects would perform above chance, allowing us to assess their preferred memory strategy.

We expected to find evidence that the good readers would use a phonetic strategy more than the poor readers in those situations where phonetic coding is feasible. Furthermore, we expected the good readers' memory for order to be better than that of the poor readers whenever a phonetic strategy is optimal for the task. It would follow, then, that the good readers should have an advantage over the poor readers in recall of temporal order. Moreover, it ought to be possible to demonstrate greater use of phonetic coding by the good readers than by the poor readers when temporal order recall is tested. Possibly, the poor readers would prefer to use an alternative memory strategy, such as temporal-spatial pattern coding, on this task. (See Healy, 1975, for evidence that adult subjects use this strategy when phonetic coding is hampered.) For spatial order recall, on the other hand, we had no clear basis for expecting performance to vary with reading ability, because Healy has shown that phonetic coding is not the preferred strategy when this aspect of order memory is tested. On this task, we expected to find evidence that all subjects retained the temporal-spatial pattern of the stimulus display.

Method

Task

Both the Temporal Order and the Spatial Order Recall conditions required successive presentations of items. A trial consisted of a presentation of three letters followed by a list of digits, to be used as a distractor task. In the Temporal Order Recall condition, the subjects retained the temporal sequence of the three letters; the spatial locations of the letters, known to the subjects in advance, were kept constant. Likewise, in the Spatial Order Recall condition, the subjects retained the spatial locations of the letters;

the subjects were aware of the constant temporal letter sequence. During the presentation of the digits, the subjects were required to perform one of two distractor tasks. In the Digit Name task, they read the names of the digits aloud; in the Digit Position task, the subjects indicated each digit's spatial location by raising their fingers.

Subjects

The subjects were selected from four second-grade classes in the East Hartford, Connecticut, public school system. The children were of middle-class socioeconomic status and attended a neighborhood school. Candidates for the poor reader group were selected for screening if they were so designated by their teachers or if they scored below grade level on either the vocabulary or comprehension subtest of the *Gates-MacGinitie Reading Tests* (1978), which had been administered in the eighth and ninth months of the second grade. Candidates for the good reader group either received a superior evaluation or scored more than one year above grade level on one of the subtests.

The subjects selected for screening were administered the *Peabody Picture Vocabulary Test* (Dunn, 1959) and the word identification and word attack subtests of the *Woodcock Reading Mastery Tests* (Woodcock, 1973) in the ninth month of the school year. The subjects with extreme IQ scores (below 90 or above 135) were ineligible for further testing. The final good reader group consisted of the 16 subjects (8 females, 8 males) who attained the highest combined raw scores on the two Woodcock subtests, whereas the poor reader group included the 16 subjects (9 females, 7 males) with the lowest combined scores. All of the poor readers were achieving below local norms, and all of them lagged substantially behind their peers. The good readers had a mean age of 7 years, 11 months compared with the poor readers' mean age of 8 years, $t(30) = 0.3$, $p > .5$ (two-tailed). The good readers had a mean IQ of 109.1, whereas the poor readers had a mean IQ of 102.2, $t(30) = 2.1$, $p = .044$ (two-tailed). The mean combined raw score on the Woodcock was 144.4 for the good readers (range: 134 to 161) and 80.3 for the poor readers (range: 64 to 104), $t(30) = 18.3$, $p < .001$ (two-tailed).

Stimuli and Apparatus

A memory drum was used for presentation of the stimuli, which were typed onto a paper tape. The stimuli were successively presented in the display window of the memory drum. The duration of each display was 1/2 sec and the interdisplay interval was 1/2 sec.

Four different 24-trial sequences were devised. A trial consisted of a 3-letter stimulus followed by a retention interval of 3 or 12 intervening digits. The letters and digits were presented successively, each in a different one of three spatial positions that formed a horizontal array. The remaining two positions were occupied by dashes.

The letters presented were permutations of the set F, P, and V typed in capitals. These letters were chosen because F and P are visually, but not phonetically, confusable, whereas P and V have phonetically confusable names, but are not visually confusable. For the two sequences in the Temporal Order Recall condition, each of the six permutations of the three letters appeared twice at each of the two retention intervals as the temporal order of the letters, the spatial order being held constant over all 24 trials. In one of the sequences, the constant spatial order was FPV; in the other, it was VPF. For the Spatial Order Recall condition, each permutation occurred twice at each retention interval as the spatial order of the letters, while the temporal order was held constant. In one of the sequences, the constant temporal order was FPV, and in the other, it was VPF. For example, in the Temporal Order Recall condition when the constant spatial order was FPV, F would always be presented in the left position of the memory drum display, P in the middle, and V in the right position. Only the temporal order of the letters would vary. Likewise, in the Spatial Order Recall condition when the constant temporal order was FPV, F was always shown first, followed by P, then V. Only the spatial order of the letters varied across trials. Within a sequence, the presentation order of the trials was random with these three constraints: Each of the six permutations of the three letters must appear twice in every block of 12 trials, once at each of the two retention intervals; in every subset of six trials each retention interval must occur three times; a given permutation must not appear on two successive trials.

The intervening digits were selected from the set: 4, 6, 8. Selection was random with the constraints that no digit occur on two successive displays and that each digit occur equally often in every group of 15 digits. By using a mapping of the three digits to the three spatial positions, the digits that were selected for the retention intervals of the first 12 trials determined the positions, in reverse order, of the digits in the final 12 trials; the digits of the final 12 trials determined the positions in reverse order of the digits of the first 12 trials. A practice sequence of 15 digits was devised by the same method.

Response cards were prepared by typing the three letters F, P, and V in the center of white, 3 x 5-inch cards, one letter per card.

Procedure

The subjects were tested individually in two 20-min sessions. Each session was devoted to one recall condition. The order of the two conditions was counterbalanced so that half the members of each reading group participated in the Temporal Order Recall condition in the first session and in Spatial Order Recall in the second. The order of the conditions was reversed for the other subjects. Half the members of each group were tested on the sequence in which the constant temporal order was FPV and the sequence in which the constant spatial order was FPV. The remaining subjects were tested on the two sequences in which the constant order was VPF.

At the beginning of each session, the subjects were informed of the condition in which they were participating and the task was explained. For the Temporal Order Recall condition, the subjects were told the constant spatial order. Thus, the subjects had to remember only the temporal order, since they were aware of the stimulus items and their spatial locations. For the Spatial Order Recall condition, the subjects were told the constant temporal order and had to remember only the spatial order. As letters were displayed, the subjects read them aloud. As digits were presented, the subjects were required to perform one of two interpolated tasks for the first

12-trial block and the other task for the final 12-trial block. In the Digit Name task, the subjects read the digits aloud as they appeared. In the Digit Position task, the subjects raised their fingers as digits appeared, with the number of fingers raised indicating the spatial location of the presented digit. When the digit appeared in the left position, one finger was raised; two were raised for the middle position; either three or five fingers were raised for the right position, depending on which was more comfortable for the individual subject. The order of the distractor tasks was the same for each subject within both sessions, but was counterbalanced within reading groups. Before each block of 12 trials, the subjects were given practice on the appropriate distractor task using the practice sequences. During these trials, the presentation rate of the digits was manually controlled by the experimenter so that it could be increased as the subjects became more proficient at the task.

The end of a trial was signaled by the appearance of three dashes in the memory drum display window. The subjects in the Spatial Order Recall condition then attempted to reproduce the spatial order of the letters as seen in that trial by arranging the response cards into a horizontal array. The subjects in the Temporal Order Recall condition arranged the cards into a vertical array such that the top card had typed on it the letter first seen and the bottom card depicted the letter last seen.

RESULTS

The number of stimulus items incorrectly ordered by each subject for each condition was tallied. An item was considered incorrect if it was not placed in the serial position that corresponded to its position in the memory drum display. For the Temporal Order Recall condition, the serial positions refer to the temporal sequence of the items from first seen to last seen. For the Spatial Order Recall condition, the serial positions correspond to the spatial locations from left to right. Preliminary to examining the experimental predictions, we tested whether there were sex differences associated with order memory. For this test, the total number of errors was calculated for each child. These data were subjected to an analysis of variance (unweighted means analysis) with two between-groups measures (sex of child and reading ability). The results indicated that reading ability was a significant factor in order memory, $F(1,28) = 8.9$, $p = .006$, whereas sex was not, $F < 1$. The interaction of reading ability and sex was nonsignificant, $F(1,28) = 1.2$, $p > .05$. Since sex differences were not found, this factor was not included in the principal analyses of the data.

Subsequently, the data were subjected to an analysis of variance with one between-groups measure (reading ability) and four within-groups measures (recall type, distractor type, retention interval, and serial position). Significant effects involving the serial position factor were verified using a procedure by Box (1954). This procedure insured that the obtained effects were not artifacts of inhomogeneous variances and covariances. The full data set, converted to percentages, is presented in Table 1. Each percentage is based on a maximum of six errors per subject. A summary of the results of the analysis of variance is presented in Table 2 under the column labeled Absolute Errors.

Table 1

Percentages of Incorrect Placements

(Standard Deviations are Shown in Parentheses)

	3 Digits			12 Digits		
Good Readers	Pos 1	Pos 2	Pos 3	Pos 1	Pos 2	Pos 3
Temporal Order Recall						
Digit Name	20 (20)	34 (14)	32 (18)	40 (18)	41 (23)	39 (21)
Digit Position	19 (15)	30 (21)	29 (20)	29 (21)	33 (19)	34 (16)
Spatial Order Recall						
Digit Name	43 (17)	48 (18)	48 (18)	43 (29)	44 (28)	43 (26)
Digit Position	49 (23)	52 (25)	53 (20)	53 (23)	50 (20)	49 (26)
Poor Readers						
Temporal Order Recall						
Digit Name	31 (22)	43 (27)	38 (24)	36 (20)	48 (20)	51 (19)
Digit Position	30 (21)	38 (17)	40 (19)	39 (20)	54 (22)	52 (14)
Spatial Order Recall						
Digit Name	46 (24)	54 (20)	55 (26)	56 (20)	48 (23)	54 (16)
Digit Position	52 (18)	51 (16)	59 (17)	59 (24)	68 (21)	60 (23)

Table 2

Summary of Analyses of Variance

Factor	df	Absolute Errors *F*	Conditional Phonetic Errors *F*	Conditional Visual Errors *F*
Reading	1,30	8.3**	4.0	1.0
Recall	1,30	31.7***	0.5	2.3
Distractor	1,30	1.3	1.9	0.
Retention Interval	1,30	6.1*	0.7	0.
Serial Position	2,60	12.9***	0.1	1.3
Reading x Recall	1,30	0.2	1.1	2.7
Reading x Distractor	1,30	0.6	1.0	19.4***
Reading x Retention Interval	1,30	0.9	0.1	3.3
Reading x Serial Position	2,60	0.9	0.7	0.7
Recall x Distractor	1,30	4.4*	0.2	0.2
Recall x Retention Interval	1,30	3.2	7.4*	6.0*
Recall x Serial Position	2,60	7.3**	1.1	0.8
Distractor x Retention Interval	1,30	0.4	1.4	1.6
Distractor x Serial Position	2,60	0.	1.3	0.2
Retention Interval x Serial Position	2,60	1.3	1.2	0.1
Recall x Retention Interval x Serial Position[a]	2,60	0.6	1.5	4.6*

*p < .05[b]
**p < .01
***p < .001

[a] All other three-way interactions and all higher-order interactions were nonsignificant.

[b] Considering the number of factors involved in these analyses, it is conceivable that the true risk of a Type I error is greater than .05.

Good vs. Poor Readers

The expectation of an interaction between reading ability and recall type was based on past evidence of good and poor readers' differential proficiency for using phonetic codes. Temporal order recall has been found to depend usually on the retention of phonetic memory codes, with which poor readers are known to be deficient. Thus, the good readers should perform better than the poor readers on temporal order recall. No such expectation can be made for spatial order recall, however. Since retention of spatial order has not been shown to depend on phonetic coding, the performances of the good and poor readers were not expected to differ.

The percentage of incorrect placements on the two recall tasks by each reading group is shown in Table 3. It is clear that the good readers made fewer errors than the poor readers in both conditions. The analysis of variance indicated that the good readers' performance was significantly better than that of the poor readers. To control for IQ differences between the members of the two reading groups, an analysis of covariance was conducted using IQ as the covariate. (See Crowder, in press, for a discussion of the rationale for this procedure.) With IQ controlled, the two reading groups were again distinguished, $F(1,29) = 11.8$, $p = .002$. The superiority of the good readers' order memory extended both to temporal order recall and to spatial order recall; the interaction between reading ability and recall type did not approach significance.

Table 3

Error Percentages for Each Reading Group by Recall Condition

	Recall Condition	
Reading Ability	Temporal Order	Spatial Order
Good Readers	32	48
Poor Readers	42	55

Thus, the anticipated interaction between type of recall task and reading ability did not occur. It is important to ask, therefore, whether this outcome may nevertheless reflect a tendency for the good and poor readers to use different coding strategies. An examination of confusion errors was carried out in order to investigate this possibility. As in the previous studies with adults (e.g., Healy, 1982), we examined the relative percentages with which phonetic confusions and visual confusions occurred (i.e., the conditional percentages of each type of confusion error given that an error was made), rather than the absolute percentages of confusion errors. We took as evidence for phonetic coding an indication that the conditional percentages of phonetic confusion errors were greater than would be expected on the basis

Table 4

Conditional Percentage of Phonetic Errors (Standard Deviations are Shown in Parentheses)

	3 Digits			12 Digits		
Good Readers	Pos 1	Pos 2	Pos 3	Pos 1	Pos 2	Pos 3
Temporal Order Recall						
Digit Name	48 (31)	51 (42)	60 (35)	30 (26)	33 (23)	51 (31)
Digit Position	47 (36)	48 (32)	37 (37)	48 (39)	33 (33)	31 (42)
Spatial Order Recall						
Digit Name	38 (31)	45 (27)	34 (28)	40 (28)	40 (32)	57 (33)
Digit Position	24 (24)	32 (24)	30 (24)	36 (27)	36 (32)	44 (30)
Poor Readers						
Temporal Order Recall						
Digit Name	22 (36)	42 (28)	44 (31)	35 (41)	48 (21)	25 (19)
Digit Position	33 (33)	27 (28)	33 (35)	29 (30)	35 (24)	28 (26)
Spatial Order Recall						
Digit Name	35 (39)	30 (32)	24 (19)	31 (25)	36 (34)	38 (29)
Digit Position	38 (33)	41 (28)	30 (23)	29 (20)	42 (27)	37 (28)

of chance alone. The conditional percentage of phonetic errors was found by determining the ratio of the number of confusions of the letters P and V to the total number of errors for each subject for each condition. The full set of conditional percentages is shown in Table 4. The mean conditional percentage of phonetic errors for each recall type is shown in the left half of Table 5. Although the good readers made fewer errors than the poor readers overall (see Table 3), when they made an error, it can be seen that the good readers were more likely than the poor readers to confuse the phonetically similar letters. The mean conditional percentage expected by chance alone is 33%, since there were three possible types of confusions---F with P, F with V, and P with V---only one of which was a phonetic confusion. The mean conditional percentage of phonetic confusion errors tended to be greater than the chance level for the good readers on temporal order recall, $t(15) = 2.2$, $p < .05$ (two-tailed), but not on spatial order recall, $t(15) = 2.0$, $p = .07$ (two-tailed). In contrast, for the poor readers, the conditional percentages were essentially equal to the chance level, $0 < t < 1$ in both cases.

Table 5

Mean Conditional Percentage of Phonetic (P-V) Errors and Visual (P-F) Errors Given that an Error was Made for Each Reading Group

	Phonetic Errors			Visual Errors		
Reading Ability	*Temp.*	*Spat.*	*Avg.*	Temp.	Spat.	Avg.
Good Readers	43	38	40	28	36	32
Poor Readers	33	34	34	36	35	36

The phonetic error data were subjected to an analysis of variance with one between-groups measure (reading ability) and four within-groups measures (recall type, distractor type, retention interval, and serial position). The results of this analysis are summarized in Table 2 under the heading Conditional Phonetic Errors. This analysis indicated that the main effect of reading ability was marginally significant. With IQ controlled in an analysis of covariance, the reading groups were distinguished, $F(1,29) = 4.8$, $p = .038$. When an error was made, it was more likely to be a phonetic error for the good readers than for the poor readers on both temporal order recall and spatial order recall, as the interaction between reading ability and recall type was not significant. Thus, it would seem that on both tasks the good readers, more often than the poor readers, were coding in a phonetic manner.

Because of the constraints on proportions, we carried out an additional analysis of the phonetic error data after subjecting them to an arcsine transformation. This analysis fully corroborated the results of the initial one: All effects that were significant in the analysis of untransformed proportions remained significant; all other effects remained nonsignificant.

Finding that the conditional percentage of phonetic confusions was as large in spatial order recall as in temporal order recall is contrary to the expectation generated by Healy's (1975, 1977) research with adult subjects. Why phonetic coding was used in spatial order recall in this experiment might have the following explanation: In Healy's experiments, four items were presented at a rate of one per 400 ms. In contrast, we presented three stimulus items at a rate of one per sec. It is likely that in modifying Healy's paradigm for use with children, the presentation rate was kept slow enough to permit the subjects to recode phonetically in the spatial order recall condition as well as the temporal order recall condition. Apparently, good readers were better able to take advantage of this opportunity. Good readers, then, seem to adopt a phonetic memory strategy more often than poor readers. Though contrary to our original expectation, this strategy was apparently used for spatial order, as well as for temporal order, recall.

To ascertain directly whether the poor readers made greater use than the good readers of a visual coding strategy based on the shapes of the stimulus items, we computed the conditional percentage of visual errors (i.e., confusions of F and P) given that an error was made. The full set of conditional percentages is shown in Table 6. The mean percentage for each recall type is shown in the right half of Table 5. Again, the mean conditional percentage expected by chance alone is 33%, since there were three possible types of confusions, only one of which was a visual confusion. These mean percentages did not significantly differ from chance for either the good readers or the poor readers. An analysis of variance, analogous to that conducted on the conditional percentages of phonetic errors, was performed on the conditional percentages of visual errors and is summarized in Table 2 under the heading Conditional Visual Errors. The procedure of Box (1954) was used to insure that the triple interaction involving serial position was not an artifact of inhomogeneity of variances and covariances. Again, applying an arcsine transform to the data and redoing the analysis of variance did not change the results.

The mean conditional percentage of visual errors did not differ with reading ability. However, there was a highly significant interaction between reading ability and distractor type. This interaction is evidence for different coding strategies in the two reading groups. If a subject is retaining visual codes, a high percentage of visual confusion errors would be expected unless the distractor task disrupts the visual mode of processing through interference. In fact, for the poor readers the conditional percentage of visual errors was large, and significantly different from chance, $t(15) = 2.2$, $p < .05$ (two-tailed), with the Digit Name distractor task that demanded phonetic processing (41%), but was reduced considerably, and was essentially at chance, $t(15) = -1.2$, $p > .05$ (two-tailed), with the Digit Position distractor task that demanded the processing of spatial location information (30%). This difference between the two distractor types proved significant in a post hoc analysis using Fisher's protected t-test (Cohen & Cohen, 1975), $t(15) = 2.8$, $p = .013$ (two-tailed). (The protected t-test, also known as the LSD test, is an ordinary t-test performed on group means that significantly vary according to an overall F value. This test preserves the power of the t-test, while efficiently protecting against an inflated Type I error rate.) Thus, the pattern of visual errors for the poor readers suggests that they do code the to-be-remembered letters in terms of their visual features but that this coding is disrupted by the requirement to monitor the

Table 6

Conditional Percentage of Visual Errors (Standard Deviations are Shown in Parentheses)

	3 Digits			12 Digits		
Good Readers	Pos 1	Pos 2	Pos 3	Pos 1	Pos 2	Pos 3
Temporal Order Recall						
Digit Name	17 (18)	28 (34)	13 (22)	31 (30)	28 (23)	24 (18)
Digit Position	40 (32)	31 (26)	28 (35)	23 (34)	41 (39)	37 (38)
Spatial Order Recall						
Digit Name	38 (24)	32 (30)	34 (33)	28 (26)	32 (32)	17 (22)
Digit Position	54 (20)	47 (30)	49 (28)	41 (29)	32 (30)	31 (32)
Poor Readers						
Temporal Order Recall						
Digit Name	36 (34)	37 (30)	35 (26)	48 (34)	45 (33)	48 (29)
Digit Position	40 (35)	25 (24)	19 (21)	24 (35)	34 (26)	37 (18)
Spatial Order Recall						
Digit Name	40 (29)	38 (30)	40 (33)	54 (29)	31 (29)	38 (22)
Digit Position	25 (29)	34 (26)	32 (27)	36 (24)	26 (20)	31 (23)

spatial positions of the interpolated digits. In contrast, for the good readers, the conditional percentage of visual errors was actually smaller with the Digit Name task (27%) than with the Digit Position task (38%), protected $t(15) = -3.5$, $p = .004$ (two-tailed). The error percentage on the Digit Name task was significantly below chance, $t(15) = -2.6$, $p < .02$ (two-tailed), whereas the percentage on the Digit Position task was essentially at chance, $t(15) = 1.5$, $p > .05$ (two-tailed).

In summary, the good readers made a greater proportion of phonetic errors than visual errors, but the poor readers actually showed a small difference in the opposite direction. Moreover, for the poor readers, the proportion of visual errors was particularly large when they were not forced to attend to the spatial locations of the digits. These analyses of confusion errors suggest that the good readers adopt consistently a phonetic coding strategy whereas the poor readers at times code information about the visual properties of the letters.

In addition to coding the forms of the individual letters, there is another nonphonetic strategy that might be adopted as an aid to recall: retention of the temporal-spatial pattern in which items were presented and using the remembered pattern to reconstruct the order. The six patterns are illustrated in Figure 1. The experiment was designed so that each pattern occurred twice at each retention interval in each condition. On any given trial, if the subject retains the pattern and the constant order, the to-be-remembered order can be inferred. For example, in the Temporal Order Recall condition, if the subject knows that the stimulus items were presented according to pattern 2 and that the constant spatial order is FPV, then the *temporal order* FVP can be determined. Likewise, in the Spatial Order Recall condition, if the pattern and constant temporal order are known, then the spatial order can be reconstructed.

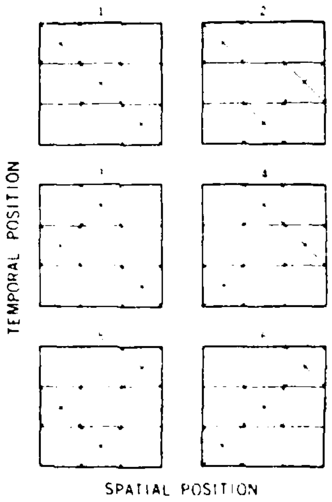


Figure 1. Temporal-spatial patterns of letter presentations. The spatial positions are shown horizontally and the temporal positions are shown vertically. For example, in pattern number 4, the subject first sees a letter in the second spatial position, then a letter in the third position, and then a letter in the first position.

Table 7

Error Percentages Committed on Each Temporal-Spatial Pattern as a Function of Reading Ability, Recall Condition, and Distractor Type

(Standard Deviations are Shown in Parentheses)

	Pattern					
Good Readers	1	2	3	4	5	6
Temporal order recall						
Digit Name	38	47	62	38	53	41
	(41)	(37)	(33)	(33)	(33)	(32)
Digit Position	19	50	47	44	34	38
	(24)	(40)	(37)	(39)	(34)	(38)
Spatial order recall						
Digit Name	28	72	66	53	72	59
	(35)	(30)	(38)	(33)	(39)	(36)
Digit Position	47	62	69	66	88	59
	(37)	(38)	(35)	(34)	(22)	(36)
Poor Readers						
Temporal order recall						
Digit Name	31	59	69	56	69	44
	(35)	(40)	(35)	(30)	(35)	(35)
Digit Position	38	50	66	53	78	47
	(33)	(31)	(34)	(41)	(30)	(37)
Spatial order recall						
Digit Name	44	75	78	72	66	72
	(39)	(31)	(35)	(30)	(38)	(35)
Digit Position	56	75	75	78	81	66
	(30)	(35)	(31)	(35)	(24)	(34)

To examine the extent to which pattern coding was used, we looked for a consistent effect of pattern over the two recall conditions, which were subdivided by distractor type. For each of these four blocks of trials, the number of incorrect trials was tallied for each of the six patterns, scoring each trial as either completely correct or as incorrect. Pattern scores were obtained by averaging the number incorrect for each pattern over the subjects in each reading group. The percentage of errors for each pattern is shown in Table 7. Inspection of the table shows that a lower percentage of errors occurred on the more regular patterns (such as patterns 1 and 6) than on the others. Also, it can be seen that the consistency of these percentages over the six patterns is, apparently, relatively large for the poor readers. To discover whether this is a statistically significant trend, the six pattern scores for each block of trials were correlated with the six scores in each of the other blocks. The use of pattern coding in any two blocks of trials should be reflected by a high correlation, since patterns that are difficult to code should result in an increase in errors in each block, whereas patterns that are easy to code will result in fewer errors. In previous research with adults (Healy, 1975, 1977), high correlations were found between pattern scores for spatial order recall conditions, implicating the use of pattern coding, but low correlations were found between scores on temporal order recall conditions. The Pearson Product-Moment correlations for each reading group are listed in Table 8. The correlations for the good readers range from .37 to .78. None is statistically significant, although all are positive. The correlations for the poor readers range from .44 to .93, and two of these are significant. Moreover, one of the significant correlations for the poor readers reflects the relationship between pattern scores on the two temporal

---

Table 8

Pearson Product-Moment Correlations for Good and Poor Readers among Error Scores on the Temporal-Spatial Patterns as a Function of Recall Type (Temporal Order or Spatial Order) and Distractor Type (Digit Name or Digit Position)

<u>Good Readers</u>	Temp. Name	Temp. Pos.	Spat. Name	Spat. Pos.
Temp.-Name	---	.39	.62	.57
Temp.-Pos.		---	.78	.37
Spat.-Name			---	.76
Spat.-Pos.				---
<u>Poor Readers</u>				
Temp.-Name	---	.89*	.73	.93**
Temp.-Pos.		---	.44	.81
Spat.-Name			---	.71
Spat.-Pos.				---

*p < .05 (two-tailed)
**p < .01 (two-tailed)

---

order recall conditions. The significant correlations for the poor readers suggest that they tended to use pattern coding for both temporal order recall and spatial order recall whereas the good readers may not have adopted this strategy.

The pattern correlations are particularly interesting because the poor readers showed a great degree of regularity on this measure, despite the fact that by several other measures their performance was less regular then that of the good readers and more nearly random: The overall performance level of the poor readers was lower than that of the good readers (see Table 3), and the conditional percentages of phonetic confusion errors were closer to the chance level for the poor readers than for the good readers (see Table 5).

Temporal Order vs. Spatial Order Recall

Whereas a comparison of the recall levels of good and poor readers was the major aim of the present experiment, an ancillary goal was to attempt to reproduce with children the effects previously found in tests of adults' memory for order (Healy, 1975, 1977). The analysis of variance examining incorrect placements indicated that the present experiment using children's data did indeed reproduce several of the effects found by Healy (1975, 1977) but failed to reproduce one. Examining the main effects, we note first a significant effect for retention interval. Not surprisingly, performance declined with the long interval of 12 digits compared with the short interval of 3 digits. Second, serial position proved significant, as performance was better on the first position than on either the second position, protected $\underline{t}(31) = 4.5$, $\underline{p} < .001$ (two-tailed), or the third position, protected $\underline{t}(31) = 4.5$, $\underline{p} < .001$ (two-tailed). Third, we found that performance on temporal order recall was generally better than on spatial order recall. Healy (1977), on the contrary, found that temporal order recall was superior only with certain interpolated distractor tasks or at certain retention intervals. Under some conditions, spatial order recall was as good as, or better than, temporal order recall.

Turning to the interactions that were reproduced with child subjects, we note a significant interaction between recall type and distractor type. As shown in Table 9, for the Temporal Order Recall condition, the Digit Name distractor, a phonetic task, resulted in a nonsignificant decrement in performance compared with the effect of the Digit Position distractor, a spatial task, $0 <$ protected $\underline{t} < 1$. This pattern of results differed in the Spatial Order Recall condition where it was found that performance was worse with the Digit Position distractor task, protected $\underline{t}(31) = 2.2$, $\underline{p} < .04$ (two-tailed). Second, it may be noted that different serial position curves for the two recall tasks are reflected in the interaction between recall type and serial position. As is evident in Table 10, for spatial order recall, the serial position curve is relatively flat; the differences between the means for any two positions are nonsignificant. In contrast, the curve for temporal order recall shows a marked superiority in performance at the first serial position compared with either the second position, protected $\underline{t}(31) = 5.6$, $\underline{p} < .001$ (two-tailed), or the third position, protected $\underline{t}(31) = 7.8$, $\underline{p} < .001$ (two-tailed).

The major departure from Healy's previous findings with adults was our finding of the use of phonetic coding for spatial order recall. (In the

Table 9

Error Percentages in Each Recall Condition by Distractor Type

Recall Type	Distractor Type	
	Digit Name	Digit Position
Temporal Order	38	36
Spatial Order	48	55

Table 10

Error Percentages in Each Recall Condition by Serial Position[a]

Recall type	Position		
	1	2	3
Temporal Order	30	40	39
Spatial Order	50	52	53

[a] For temporal order recall, the serial positions refer to the temporal sequence of the items from first seen to last seen; for spatial order recall, the serial positions correspond to the spatial locations from left to right.

present experiment, the conditional percentage of phonetic errors did not differ for temporal and spatial order recall.) As explained earlier, we attribute this difference to a slow stimulus presentation rate that allowed the subjects enough time to recode the spatial positions into phonetic form. This explanation receives additional support upon examining the results of the analysis of variance for the conditional percentage of phonetic errors. Here we found an interaction between recall type and retention interval. At the short retention interval, the results were as expected: When an error was made, it was more likely to be a phonetic error for temporal order recall (43%) than for spatial order recall (33%), protected $\underline{t}(31) = 2.1$, $\underline{p} < .05$ (two-tailed). At the long retention interval, the percentage of phonetic errors was nonsignificantly greater for spatial order recall (39%) than for temporal order recall (33%), protected $\underline{t}(31) = -1.8$, $\underline{p} < .09$ (two-tailed). The comparable percentages for spatial order recall and temporal order recall at the long retention interval suggest that the long interval allowed enough time for the subjects to recode the spatial positions linguistically.

The opposite interaction was found upon examining the conditional percentage of visual errors. In this case, the conditional percentage of errors was greater for spatial order recall (39%) than for temporal order recall (29%) at the short retention interval, protected $\underline{t}(31) = 2.2$, $\underline{p} < .04$ (two-tailed). At the long interval, percentages of visual errors for temporal order recall (35%) and for spatial order recall (33%) were not significantly different, $0 <$ protected $\underline{t} < 1$. Since visual and phonetic errors are complementary to some extent (as the conditional percentages of phonetic, visual, and other errors must sum to 100%), this pattern for visual errors may possibly be explained solely in terms of the pattern for phonetic errors.

The triple interaction of recall type, retention interval, and serial position for the conditional percentage of visual errors indicates that the increase in the percentage of visual errors on temporal order recall on the long retention interval compared with the short interval was significant on only the third serial position: in two-tailed tests, first position, $0 <$ protected $\underline{t} < 1$; second position, protected $\underline{t}(31) = -1.5$, $\underline{p} > .05$; third position, protected $\underline{t}(31) = -2.3$, $\underline{p} = .008$. On spatial order recall, in contrast, there was a decrease in the percentage of errors on the long interval at the third serial position: in two-tailed tests, first position, $-1 <$ protected $\underline{t} < 0$; second position, protected $\underline{t}(31) = 1.6$, $\underline{p} > .05$; third position, protected $\underline{t}(31) = 2.1$, $\underline{p} < .05$. This triple interaction was unexpected and is not readily interpretable.

DISCUSSION

The impetus for this study arose from a question originally addressed by Katz et al. (1981): Can we understand poor beginning readers' characteristic difficulties in remembering order as a consequence of deficient use of a phonetic memory strategy? This issue was previously approached by comparing good and poor readers' memory for the order of items in an array. In one condition, the items had readily available names that could easily be coded phonetically, whereas in a second condition, this was not the case, since the items were nonrepresentational designs. The failure to find a difference between good and poor readers in remembering the nonsense designs encouraged us to press the issue by undertaking a more analytic study of memory for order. To investigate whether, in some circumstances, good and poor beginning readers preferred to use different memory strategies, we adopted a new approach that would allow us to infer the strategy that subjects actually used.

We were able to infer the memory strategies adopted by good and poor readers in an experimental task that allowed us to assess memory for temporal order and memory for spatial order separately. Previous research using this experimental procedure (Healy, 1975, 1977) with adult subjects indicated that purely temporal order recall normally relies on phonetic coding, whereas purely spatial order recall does not. Since poor beginning readers have known deficiencies in their use of phonetic codes, we expected that their performance relative to good readers on temporal order recall might be impaired. However, no such impairment was predicted for spatial order recall, on which a nonphonetic strategy is presumably used. Moreover, we expected to find evidence for greater use of phonetic codes among good readers than poor readers whenever a phonetic strategy was possible. Therefore, basing our prediction on Healy's previous research, we expected the phonetic strategy to be evident only on temporal order recall.

The results confirmed our expectation that the good readers would use a phonetic strategy more often and more effectively than the poor readers even though the expected dissociation in memory coding for temporal and spatial order was not obtained. The data suggested that in adapting Healy's paradigm for use with children, the modifications (lengthening the stimulus presentation times and reducing the number of stimulus items per trial) had the effect of permitting phonetic coding to occur for spatial order recall as well as for temporal order recall. Thus, the procedure did not force the use of divergent strategies for the two tasks as we had intended. But in spite of this limitation, the findings supported our expectation that the good readers would use phonetic codes whenever it was possible to do so and that poor readers would attempt to use other strategies. The results indicate that the good readers preferred to use phonetic codes more than the poor readers even in spatial order recall. The poor readers, on the other hand, tended to make greater use of an alternative to the phonetic coding strategy, presumably in order to evade the difficulties they have in using phonetic codes. Thus, the poor readers, in contrast to the good readers of the present study and Healy's normal adult subjects, coded information about the visual features of the letters and elected to retain temporal-spatial patterns for the temporal order recall condition. Furthermore, they persisted in using this memory strategy for the spatial order recall condition even though a phonetic strategy was both feasible and efficient for the task, as indicated by the good readers' performance. Thus, it was found in the present study, as in the experiment of Katz et al. (1981), that in those task situations in which phonetic coding is possible, the good readers' performance was superior to that of the poor readers.

By using a paradigm that varied the task (temporal order or spatial order recall) while always using the same stimulus material, the present study provides independent support for the view that poor beginning readers' problems remembering order are linked to deficient use of phonetic coding in working memory. The present results are also consistent with the results of previous studies that found that good readers make greater use than poor readers of phonetic codes on tasks requiring recall of both item identity and item order (Liberman et al., 1977; Mann et al., 1980; Shankweiler et al., 1979). In those studies, which compared good and poor readers' ordered recall of rhyming and nonrhyming linguistic material, it was found that only the good readers' performance was detrimentally affected by the rhyming (phonetically confusable) items. Furthermore, Shankweiler et al. (1979) conducted an analysis (unpublished) of the actual substitutions committed by their subjects. This indicated that good readers made a significantly higher proportion of phonetic errors than poor readers. The present experiment permitted us to examine short-term retention of item order with no requirement for retaining item identity. At the same time, it allowed the subjects the opportunity to make either phonetic or visual errors. Again, we found that the good readers' errors were more likely to be phonetic than were those of the poor readers.

The literature points to a high degree of consensus on the failure of poor beginning readers to use phonetic strategies effectively. (The tests that distinguish good and poor readers in the early school years may not serve to differentiate older children and adults who differ in reading ability; see, for example, Johnston, 1982; Olson, Davidson, Kliegl, & Davies, in press; and Siegel & Linder, in press.) On the other hand, there is no agreement regarding the comparative levels of spatial abilities characteristic of good and poor readers. In one recent study (Symmes & Rapoport, 1972), poor readers were found to be actually better than good readers on certain spatial tasks. Thus, on one view, the poor readers of the present study would have been expected to do better on spatial order recall than the good readers and, possibly, to retain temporal-spatial patterns more often in both recall conditions. The opposite expectations, however, can be generated on the basis of the finding that poor readers are less sensitive than good readers to letter position frequencies (Mason & Katz, 1976; Mason et al., 1975). Our findings do not unequivocally support either position. Although we did find that the poor readers tended to adopt a strategy of retaining temporal-spatial patterns, they were, nevertheless, not able to perform at levels comparable to the good readers on spatial order recall. Perhaps, a better test of these conflicting hypotheses, and of our expectation of equal performances for good and poor readers on spatial order recall, would require the elimination of the opportunity for phonetic coding for spatial order recall. At all events, our expectation that poor readers would tend to use an alternative strategy, in preference to the phonetic memory strategy with which they have difficulty, draws support from the findings.

Evidence that poor beginning readers tend to prefer nonphonetic memory strategies in some situations has been previously noted. Byrne and Shea (1979), for example, reported that poor readers tended to code words semantically for retention in memory, whereas good readers tended to rely on phonetic codes. However, when the task required subjects to remember pseudowords, poor readers resorted to phonetic strategies, since those stimuli offered no option of semantic coding. Even in this case, it should be noted, the poor readers' performance was deficient. Thus, poor readers can use phonetic codes when the task requires it, but even then, they do so less efficiently than good readers. Under the particular conditions of the present experiment, neither the spatial order recall task nor the temporal order recall task logically required the use of phonetic codes. As explained earlier, it was possible to do either task by retaining temporal-spatial patterns. However, the requirement that the subjects read stimulus items aloud may have been expected to dispose them toward a phonetic memory strategy (Torgesen & Goldman, 1977). It should be remarked that in spite of this possibly biasing factor the poor readers in the present study tended to adopt the nonphonetic strategy, as did those of Byrne and Shea (1979).

In sum, the present findings, like those of Katz et al. (1981), support the view that the poor reader's problem in retaining order is linked to deficient use of phonetic codes in working memory. Thus, poor readers' inferior memory for order should not be viewed as an independent disorder. Rather, it may be considered as one manifestation of a deficiency in the domain of language, involving the use of phonetic coding in working memory.

REFERENCES

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

Baddeley, A. D. (1979). Working memory and reading. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), The processing of visible language. New York: Plenum Press.

Bakker, D. J. (1972). Temporal order in disturbed reading. Rotterdam: Rotterdam University.

Benton, A. L. (1975). Developmental dyslexia: Neurological aspects. In W. J. Friedlander (Ed.), Advances in neurology (Vol. 7). New York: Raven Press.

Berch, D. B. (1979). Coding of spatial and temporal information in episodic memory. Advances in Child Development and Behavior, 13, 1-46.

Box, G. E. P. (1954). Some theories on quadratic forms applied in the study of analysis of variance problems: II. Effects of inequality of variance and covariance between errors in the two-way classification. Annals of Mathematical Statistics, 25, 484-498.

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

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

Corkin, S. (1974). Serial-ordering deficits in inferior readers. Neuropsychologia, 12, 347-354.

Crowder, R. G. (1978). Language and memory. In J. F. Kavanagh & W. Strange (Eds.), Speech and language in the laboratory, school, and clinic. Cambridge, MA: MIT Press.

Crowder, R. G. (in press). Is it just reading? In T. Tighe & B. Schiff (Eds.), Development of perception and cognition.

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

Gates-MacGinitie reading tests (1978). (2nd Edition). Boston: Houghton Mifflin Co.

Healy, A. F. (1975). Coding of temporal-spatial patterns in short-term memory. Journal of Verbal Learning and Verbal Behavior, 14, 481-495.

Healy, A. F. (1977). Pattern coding of spatial order information in short-term memory. Journal of Verbal Learning and Verbal Behavior, 16, 419-437.

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. (1982). Short-term memory for order information. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 16). New York: Academic Press.

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

Johnston, R. S. (1982). Phonological coding in dyslexic readers. British Journal of Psychology, 73, 455-460.

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

Kimura, D. (1963). Right temporal-lobe damage. Archives of Neurology, 8, 264-271.

Liberman, A. M., Mattingly, I. G., & Turvey, M. T. (1972). Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), Coding processes in human memory. New York: Wiley.

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

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

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

Mason, M., & Katz, L. (1976). Visual processing of nonlinguistic strings: Redundancy effects and reading ability. Journal of Experimental Psychology: General, 105, 338-348.

Mason, M., Katz, L., & Wicklund, D. A. (1975). Immediate spatial order memory and item memory in sixth-grade children as a function of reader ability. Journal of Educational Psychology, 67, 610-616.

Noelker, R. W., & Schumsky, D. A. (1973). Memory for sequence, form, and position as related to the identification of reading retardates. Journal of Educational Psychology, 64, 22-25.

Olson, R. K., Davidson, B. J., Kliegl, R., & Davies, S. E. (in press). Development of phonetic memory in disabled and normal readers. Journal of Experimental Child Psychology.

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

Siegel, L. S., & Linder, B. A. (in press). Short-term memory processes in children with reading disabilities. Developmental Psychology.

Stanley, G., Kaplan, I., & Poole, C. (1975). Cognitive and nonverbal perceptual processing in dyslexics. Journal of General Psychology, 93, 67-72.

Symmes, J. S., & Rapoport, J. L. (1972). Unexpected reading failure. American Journal of Orthopsychiatry, 42, 82-91.

Torgesen, J., & Goldman, T. (1977). Verbal rehearsal and short-term memory in reading-disabled children. Child Development, 48, 56-60.

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

EXPLORING THE ORAL AND WRITTEN LANGUAGE ERRORS MADE BY LANGUAGE DISABLED CHILDREN*

Hyla Rubin+ and Isabelle Y. Liberman+

Clinical observation of children exhibiting both oral and written language disabilities has suggested that there may be parallels in their error patterns in speaking, reading, and writing that merit further investigation. The similarities are apparent in the problems these children have in many aspects of linguistic function--in word retrieval, morphology, phonology, and syntax. Thus, these children substitute "potato" for tomato in speaking, reading, and writing. They omit grammatical tense or plural markers when speaking and do the same when reading and writing. They order the sounds incorrectly when speaking certain words and also when reading and writing them. The word order they use is often faulty across these tasks. Functor words are used incorrectly whether they are spoken, read, or written. Similar observations have been made by other investigators who have noted that oral language deficits are often reflected in the written language behavior of language disabled children (Cicci, 1980). However, the nature of such a relationship has yet to be systematically investigated.

This study is the initial step in such an investigation. It proposes to analyze the errors in naming pictured objects made by language disabled children and to examine the relationship of these errors to their performance on written language tasks. Picture naming was selected as the stimulus material since research with other populations (Denckla & Rudel, 1976; Goodglass, 1980; Jansky & deHirsch, 1972; Katz, 1982; Wolf, 1981) has found it to be an informative starting point.

Because the field is relatively uncharted, it was first necessary to determine whether a naming problem indeed existed in these children. It was considered that if they were able to point to pictured objects that were named for them ("Show me the stethoscope") but were unable to name the pictures themselves at age-appropriate levels, a naming problem could be assumed. If, on the other hand, they were unable even to point to the pictured objects that they could not name, a general vocabulary deficit, rather than a specific

---

*To appear in Annals of Dyslexia, 1982.

+Also Department of Educational Psychology, University of Connecticut.

Acknowledgment. This paper was presented at the 33rd Annual Conference of the Orton Dyslexia Society, Baltimore, Maryland, November, 1982. The research is supported by NICHD Grant HD-01994 to Haskins Laboratories. We thank Ignatius Mattingly and Robert Katz for their valuable help in data analysis and in reviewing the manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-74/75 (1983)]

deficit in naming, would more accurately account for their pattern of performance.

Having determined by this procedure that there may be a naming problem in these children, it was then necessary to develop a system of analysis that would characterize the naming errors accurately and that would facilitate an explanation of their nature. Finally, the system of analysis thus derived was applied to the errors these children made in written language in the expectation that it should be equally useful in interpreting those error patterns.

METHOD

Subjects

Thirty-four children, ranging in age from 4,3 to 12,7, who were enrolled in a self-contained public school language disability program, were the subjects in this study. They demonstrated intelligence in the average range on either the Wechsler Intelligence Scale for Children-Revised or the Stanford-Binet Intelligence Scale and all had normal vision and hearing. Although they represented three ethnic groups (Black, Caucasian, and Hispanic), English was the dominant language for all and ethnic group was not a statistically significant factor in data analysis. All exhibited at least a two-year deficit on standardized expressive language and academic (or readiness) tests. Their receptive language levels were close to chronological age.

Materials

All the items included in the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1976) were used for the naming and recognition tasks. This instrument, standardized on children aged 6 through 14, consists of 85 individual line drawings of objects that are ranked in difficulty according to the frequency with which naming errors occurred in the standardization group. Some of the pictures were later selected for the spelling task. The Wide Range Achievement Test (Jastak & Jastak, 1965) was used to determine reading and spelling achievement levels.

Procedures

Subjects were tested individually for picture naming, recognition, and achievement, and in a group for spelling. In the picture naming task, they were asked to give the best name for each of the pictured objects. In the recognition task, they were asked to point to the picture named by the examiner. Here the pictures were grouped into sets of four of the same difficulty level. Every set was presented four times in randomized order; each time a different picture was named by the examiner. In the spelling task, nine subjects (with second to fifth grade achievement levels) were shown 25 individual pictures (selected by their mid-range difficulty level for naming) and were asked to spell the name of each one. Achievement in reading and spelling was tested by the appropriate subtests of the Wide Range Achievement Test (Jastak & Jastak, 1965). These subtests were given to only 25 subjects since it was not appropriate to test the nine preschool subjects for school achievement.

RESULTS AND DISCUSSION

What Is the Normal Naming Process?

In order to discuss the naming errors made by these children meaningfully, it is necessary first to consider what might take place in the normal process of picture naming. When presented with a pictured object, we access its name, which has been stored phonologically (Barton, 1971; Brown & McNeill, 1966; Fay & Cutler, 1977). Having accessed this phonological representation, we must remember it until we actually produce the word. For this purpose, we hold onto the name in a phonological buffer zone, that is, in short term or working memory, while planning the production. Substitutions such as /gog/ for /dog/ and /nunch/ for /lunch/ that occur in early language acquisition provide direct evidence of a pre-production planning stage; it is more than coincidental that phonemes that have not yet been produced are substituted for others earlier in the word (Clark & Clark, 1977). Finally, we produce the name through coordinated articulatory movements.

Is There a Naming Problem?

The pattern of results indicates a problem specifically with naming, rather than a more general vocabulary deficit. The subjects recognized an average of 71% of the pictured objects, but were able to name only 21% of the same pictures. Since it would not be meaningful to examine naming errors for pictures that were not recognized, nonrecognized items were not analyzed further. Of those that were recognized, 34% were correctly named.

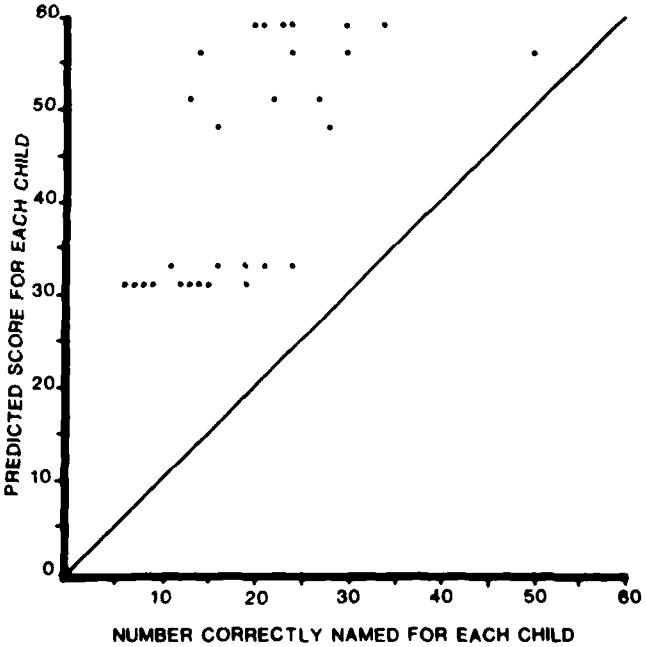


Figure 1. Scores (based on age) predicted by Boston Naming Test compared with scores obtained by language disabled children.

Since all children are able to recognize more pictured objects than they can name, it was necessary to compare the obtained scores with age-appropriate predicted scores. Figure 1 illustrates where these children stand in relation to age-matched controls, according to the norms provided by the Boston Naming Test (Kaplan et al., 1976). The number of correctly named items which were predicted and obtained for each child were significantly different, according to a one-sample t-test of the scores, $\underline{p} < .0001$. Thus, not only do these children demonstrate a gap between the number of pictured objects they recognize and the number they name, they also name significantly fewer items than age-matched controls.

What Are the Error Types and Frequencies?

The primary goal in developing an analysis system is to provide a means for examining the naming problem through an accurate and well-conceived description of error performance. Errors were characterized as phonetic, semantic, or circumlocutory. An error was considered to be phonetic if it shared 50% of the phonemes or one free morpheme with the target word. Four types of phonetic errors were delineated:

1. PH1 errors - real-word substitutions that were not semantically related to the target word, such as "sister" for scissors and "acorn" for unicorn;

2. PH2 errors - nonword substitutions for the target, such as "preztl" for pretzel and "helidakter" for helicopter;

3. PH3 errors - semantically and phonetically real-word substitutions, such as "elevator" for escalator and "tornado" for volcano;

4. PH4 errors - semantically related real-word substitutions that are also phonetically defective, such as "narrow" for dart and "kaminal" for rhinoceros.

An error was considered to be semantic if it was related only in meaning to the target word, such as "airplane" for helicopter and "stairs" for escalator. A circumlocution is a combination of words which attempts to describe the target word, such as "thing to sit at when you hurt" for wheelchair. Table 1 provides examples and frequencies of these error types.

Semantic substitutions, representing 59% of the incorrect names, are by far the most frequent error type. Semantic substitutions that are phonetically deficient (PH4, "narrow" for dart) account for another 6% of the incorrect names.

Real-word phonetic errors that are not semantically related to the target word (PH1, "acorn" for unicorn) represent only 4% of the incorrect names, the smallest proportion of the phonetic errors. Nonword phonetic errors (PH2, "preztl" for pretzel) represent 6% of the incorrect names. Real word substitutions that are phonetically and semantically related to the target word (PH3, "elevator" for escalator), or "tip of the tongue" errors (Brown & McNeill, 1966), represent 11% of the incorrect names. Circumlocutions account for another 13% of the incorrect names.

Table 1

Examples and Frequencies of Error Types

PH1 = Real word phonetic error, not semantically related 4%

- sister/scissors
- saucer/saw
- acorn/unicorn
- candle/camel
- hammer/hanger
- bathroom/mushroom
- telescope/stethoscope
- wrench/bench

PH2 = Nonword phonetic error 6%

- kalmkeno/volcano
- helican/pelican
- helidakter/helicopter
- preztl/pretzel
- maks/mask
- ocoputs/octopus

PH3 = Semantically and phonetically related 11%

- elevator/escalator
- popcorn/acorn
- clam/camel
- snake/snail
- basket/racket
- toothpick/toothbrush
- steering wheel/wheelchair
- tornado/volcano

PH4 = Semantically, then phonetically, related 6%

- narrow/dart
- kaminal/rhinoceros
- speps/escalator
- row/dart
- evevetor/escalator
- must/acorn
- bed/toboggan
- wheel/seahorse

Semantic 59%

- airplane/helicopter
- clothes/hanger
- tennis/racket
- cap/visor
- stairs/escalator
- donkey/camel
- boat/canoe
- bookbag/briefcase

Circumlocutions	Target Word	13%
put it on a clothes	hanger	
thing to sit at when you hurt	wheelchair	
it call a chair, it greens	bench	
that you turn arounds	globe	
a pirate thing for looking something	telescope	

What Do These Error Types Mean?

The present analysis system can afford possible explanations for the incorrect names that are produced. It is conceivable that the reason an incorrect name is produced is that the correct name is not stored in the lexicon. However, since the errors being analyzed here occurred in naming pictured objects that were correctly identified when named by the examiner, storage per se does not seem to be at issue. The accuracy of the stored representation may tell a more revealing story, however.

The phonological representation of a word may not be accurate enough to allow for its successful access and preservation in short term memory prior to actual production. It has been suggested (Brown & McNeill, 1966) that as we acquire new words, we first store their "generic" characteristics, such as the first phoneme, number of syllables, and stress pattern. With repeated exposure to the word, we complete this skeletal representation, supplying the final consonants, then filling in the medial segments of the word. It is this completed phonological representation that we access easily in the normal naming process.

To the extent that the generic characteristics of the target word are preserved in the actual production, we can be confident that the word was in fact accessed and held in short term memory. Table 2 presents some generic characteristics of the incorrect names produced by the children. It is clear from Table 2 that the phonetic errors retain the generic characteristics of the target words much more frequently than do the semantic errors. This trend is supported by the figures for syllable and initial phoneme agreement: 54% of the phonetic errors had the same number of syllables as the target word, as compared to only 25% of the semantic errors; 55% of the phonetic errors had the same initial phoneme as the target word, as compared to only 3% of the semantic errors.

Table 2

Generic Characteristics of Naming Errors

	Phonetic Errors (PH1-PH4)	Semantic Errors
Syllable Agreement Between Error and Target Word	54%	25%
Same Initial Phoneme in Error and in Target Word	55%	3%
Fewer Syllables in Error than in Target Word	25%	55%

In the case of phonetic errors, which tend to preserve these generic characteristics, it appears that the phonological representations of these names are either stored or held in short term memory more accurately than in the case of semantic errors, which do not tend to retain the basic phonological shape of the target word. To determine the breakdown point for both phonetic and semantic errors, we would need a more taxing recognition test to sort out whether the problem is really accuracy of storage or efficiency in short term memory coding. The present results, however, allow the conclusion that the target word has in fact been accessed when a phonetic error is made, because the generic characteristics are so frequently retained. This conclusion cannot be made about the semantic errors, since the retention of generic characteristics is so infrequent. For example, it is fair to assume that the child who says "capricorn" for unicorn has accessed the target word but no such assumption can be made about the child who says "horse" for unicorn. Further support for this position can be found in Table 2; 55% of semantic errors contain fewer syllables than the target word whereas only 25% of phonetic errors demonstrate this pattern. These syllabically less complex substitutions are usually higher frequency words, like "horse" for unicorn and "cap" for visor. Thus, again, the semantic error more often suggests that the target word has not in fact been accessed, possibly because its phonological representation is too weak. Since children who are poor readers have been shown to demonstrate phonological deficits (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Vellutino, 1977), it may be that a semantic naming error reflects a problem of that kind as well. Perhaps, then, the substitution that is similar only in meaning is not indicative of higher cognitive functioning, as might be assumed, but rather serves as a disguise for a phonological deficit affecting both oral and written language performance.

Is There a Relationship Between Naming Performance and Reading Performance?

Reading levels ranged from kindergarten to fifth grade for the 25 subjects whose achievement was tested. These children demonstrated a positive and significant relationship, $r = .54$, $p < .005$, between their reading performance and their picture-naming performance. It is interesting to note that although these children demonstrate severe deficits in both oral and written language, the relationship between naming and reading found here is similar to that found in good and poor reader groups (Jansky & deHirsch, 1972; Katz, 1982; Wolf, 1981).

What might account for this consistent pattern is the fact that the same critical components are required in the naming and reading processes (Katz, 1982). As we noted earlier, in naming, we proceed from the phonological representation of the name that best fits the picture to a phonological buffer in which we hold the representation until we actually produce the word. In reading, we decode the word, translating it into its phonological representation, and hold this representation in the phonological buffer until it is mapped onto its stored counterpart in the lexicon. Therefore, naming and reading are both linguistic processes that depend on accurate phonological representations and short term memory coding.

Is There a Relationship Between Naming Performance and Spelling Performance?

Spelling the name of a pictured object requires orthographic rule knowledge in addition to all of the previously outlined constituents of the naming process. Considering this additional requirement, it is not surprising that there was virtually no relationship, $r = .24$, between correctly named and correctly spelled items. In contrast, there is a high positive correlation, $r = .81$, $p < .008$, between the number of items that have been accessed in naming ("preztl" for pretzel) and the number that have been accessed in spelling ("cml" for camel). Similarly, there is a high positive relationship, $r = .78$, $p < .01$, between the number of semantic errors in oral naming and in spelling of a pictured item. Such correlations provide strong preliminary support for the hypothesis that similar error patterns are found across spoken and written language tasks.

CONCLUSIONS

Role of Phonological Processing

Phonological deficiencies in the accuracy of stored representations and in short term memory coding are proposed as a likely explanation of naming, or word retrieval, problems in this group of language disabled children and in other poor reader groups (Katz, 1982; Wolf, 1981). The critical facet of this explanation is the short term memory function; efficient phonetic coding seems crucial for both initial storage and eventual production of language segments. Initial acquisition of lexical items requires phonetic short term memory coding to insure storage of an accurate phonological representation, first of generic and then of additional segmental information. Successful retrieval of stored names for production depends on both the accuracy of the initial representation and the efficiency of the phonetic short term memory coding. In turn both storage and production of language segments depend on accurate and efficient perception of speech sounds. The perception of speech sounds has been found to be deficient in poor readers (Brady, Shankweiler, & Mann, 1983). Considering the evidence for the role of phonological coding in the reading process, it is anticipated that future research studies may also demonstrate a phonological basis for syntactical and morphological deficits in children with oral and written language disabilities.

Implications for Assessment and Instruction

Results of the error analysis developed here suggest that a phonetic error reflects a higher level of phonological competence than does a semantic error. Such a position is in agreement with research studies that have repeatedly demonstrated that poor readers are less sensitive to phonetic structure and less efficient in phonetic processing than are good readers (Stanovich, 1982). Diagnostically, this explanation suggests that phonetic naming errors represent more advanced phonological processing than do errors that do not bear any phonetic resemblance to the target word. It is expected that such a pattern will prove to be diagnostically significant in oral reading errors and written formulation errors as well. It would seem reasonable to suppose that substitutions that represent only a semantic association with the target word, as in reading or spelling "cat" for dog will indicate not higher cognitive functioning but rather a guessing strategy that may be masking a phonological deficiency. Furthermore, the present interpretation of error production makes questionable the commonly used instructional technique of providing semantic prompts such as category, location, or function, to facilitate attempts at naming, reading, or written formulation. Instead, it would seem more appropriate to provide phonetic prompts, such as the initial phoneme, number of syllables, or stress pattern.

Future Research

The next stage in this investigation should be the development of a more sensitive recognition task to determine the breakdown point for errors in oral and written language productions. Specifically, it is necessary to differentiate a linguistic deficit due to an inaccurate phonological representation from one due to inefficient phonetic coding in short term memory. It is anticipated that different error types result from deficiencies at different points in the process, but that such breakdown points will remain constant across oral and written language tasks. It is also anticipated that the results of this proposed next step will shed further light on appropriate diagnostic and instructional strategies.

REFERENCES

Barton, M. I. (1971). Recall of generic properties of words in aphasic patients. *Cortex*, *11*, 73-82.

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

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*, 325-337.

Cicci, R. (1980). Written language disorders. *Bulletin of the Orton Society*, *30*, 240-251.

Clark, H., & Clark, E. (1977). *Psychology and language*. New York: Harcourt Brace Jovanovich, Inc..

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

Fay, D., & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, *8*, 505-520.

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

Jansky, J., & deHirsch, K. (1972). *Preventing reading failure*. New York: Harper and Row.

Jastak, J. F., & Jastak, S. R. (1965). *Wide Range Achievement Test*. Wilmington, DE: Guidance Associates.

Kaplan, E., Goodglass, H., & Weintraub, S. (1976). *Boston naming test*. Boston: Boston University School of Medicine.

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

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

Stanovich, K. (1982). Individual differences in the cognitive processes of reading: I. Word decoding. Journal of Learning Disabilities, 15, 449-512.

Vellutino, F. R. (1977). Alternative conceptualizations of dyslexia: Evidence in support of a verbal-deficit hypothesis. Harvard Educational Review, 47, 334-354.

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

PERCEIVING PHONETIC EVENTS*

Michael Studdert-Kennedy+

In her report on the auditory processing of speech, prepared for the Ninth International Congress of Phonetic Sciences in Copenhagen, Chistovich wrote of herself and her colleagues at the Pavlov Institute in Leningrad: "We believe that the only way to describe human speech perception is to describe not the perception itself, but the artificial speech understanding system which is most compatible with the experimental data obtained in speech perception research" (Chistovich, 1980, p. 71). Chistovich went on to doubt that psychologists would agree with her, but I suspect that many may find her view quite reasonable. However, they would probably not find the view reasonable if we were to replace the words "speech perception" and "artifical speech understanding system" with the words "speech production" and "speech synthesis system." Perhaps that is because even an articulatory synthesizer does not look like a vocal tract, while our image of what goes on in the head is so vague that we can seriously entertain the notion that a network of inorganic plastic and wire might be made to operate on the same general principles as an organic network of blood and nerves.

Of course, this is impossible, not only because the physics and chemistry of organic and inorganic substances are different, but also because machines and animals have different origins. A machine is an artifact. Its maker designs the parts for particular functions and assembles them according to a plan. The machine then operates on principles that its maker knows and has made explicit in the plan. The development of an animal is just the reverse. There is no plan. The animal exists before its parts and the parts emerge by differentiation. In the human fetus, a hand (say) buds from the emerging arm, swells and gradually, by cell-death and other processes, differentiates into digits. There is no reason to suppose that the principles of behavioral development are different from those of morphological development. On the contrary, structure and function are deeply intertwined in both evolution and

---

*To appear in W. H. Warren, and R. E. Shaw, (Eds.). (in press). Persistence and change: Proceedings of the first international conference on event perception. Hillsdale, NJ: Erlbaum.

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

Acknowledgment. Preparation of this chapter was supported in part by NICHD Grant No. HD-01994 to Haskins Laboratories. I thank my teachers in these matters: Carol Fowler, Scott Kelso, Alvin Liberman, Björn Lindblom, Sverker Runeson, Quentin Summerfield, and Michael Turvey. My thanks also to Bruno Repp for much stimulating discussion, and to Nancy Frishberg for providing a simultaneous translation in American Sign Language of the first portion of the paper, as it was read at the conference.

ontogeny. Behavior emerges by differentiation, according to principles implicit in the animal's form and substance.

In short, the appropriate constraints on a model of human speech perception are biological. The model must be compatible with what we know not only of speech perception and production, but also of speech acquisition. What the infant hears determines, in part, what the infant says; and if perception is to guide production, the two processes must be, in some sense, isomorphic.

An artificial speech understanding system is therefore of limited interest to the student of human speech perception. Such a device necessarily develops in the opposite direction to the human that it is intended to mimic. For while the human infant must discover the segments of its language--words, syllables, phonemes--from their specification in the signal, the machine is granted these segments *a priori* by its makers. As a model of speech perception, the machine is tautologous and empty of explanatory content, because it necessarily contains only what its makers put in. Unfortunately, all our models of speech perception are essentially machine models.

What theories of event perception have to offer to the study of language, in general, and of speech perception, in particular, is a framework for a biological alternative to such models. Three aspects of the approach seem promising. First is the commitment to discovering the physical invariances that support perception, with an emphasis on the time-varying properties of events. Second is the view of event perception as amodal, independent of the sensory system by which information is gathered. This is important for several reasons, not least for the light it may throw on the bases of imitation and on the underlying capacities common to the perception of signed and spoken language. The third aspect is the general commitment to deriving cognitive process from physical principles and thus, for language, to understanding how its structure emerges from and is constrained by its modes of production and perception.

None of these viewpoints is entirely new to the study of speech perception. What is new is their possible combination in a unified approach. I will briefly discuss each aspect, but before I do, I must lay out certain general properties of language and central problems of speech perception.

LANGUAGE STRUCTURE

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

The two levels of patterning are phonology and syntax. The first permits us to develop a large lexicon, the second permits us to deploy the lexicon in predicating relations among objects and events (Liberman & Studdert-Kennedy, 1978; Studdert-Kennedy, 1981). My present concern is entirely with the first level. A six-year-old middle-class American child already recognizes some 13,000 words (Templin, 1957), while an adult's recognition vocabulary may be well over 100,000. Every language, however primitive the culture of its speakers by Western standards, deploys a large lexicon. This is possible because the phonology, or sound pattern, of a language draws on a small set (roughly between 20 and 100 elements) of meaningless units--consonants and vowels--to construct a very large set of meaningful units, words (or morphemes). These meaningless units may themselves be described in terms of a smaller set of recurrent, contrasting phonetic properties or distinctive features. Evidently, there emerged in our hominid ancestors a combinatorial principle (later, perhaps, extended into syntax) by which a finite set of articulatory gestures could be repeatedly permuted to produce a very large number of distinctively different patterns.

Let me note, in passing, that manual sign languages have an analogous dual structure. I do not have the space to discuss this matter in any detail. However, we have learned over the past 10 to 15 years that American Sign Language (ASL) (the first language of over 100,000 deaf persons, and the fourth most common language in the United States [Mayberry, 1978]) is a fully independent language with its own characteristic formational ("phonological") structure and syntax (Klima & Bellugi, 1979). Whether signed language is merely an analog of spoken language (related as the bat's wing to the bird's) or a true homolog, drawing on the same underlying neural structures, we do not know. But there can be no doubt that as we come to understand the structure, function, acquisition, and neuropsychological underpinnings of sign language, what we learn will profoundly condition our view of the biological status of language, in general.

Here, returning to my theme, I note simply that each ASL sign is formed by combining four intrinsically meaningless components: a hand configuration, a palm orientation, a place in the body space where it is formed, and a movement. There are some fifty values, or "primes," distributed across these four dimensions; their combination in a sign follows "phonological rules," analogous to those that constrain the structure of a syllable in spoken languages. In short, both spoken and signed languages exploit combinatorial principles of lexical formation. Their sublexical structures seem to "...provide a kind of impedance match between an open-ended set of meaningful symbols and a decidedly limited set of signaling devices" (Studdert-Kennedy & Lane, 1980, p. 35).

THE ANISOMORPHISM PARADOX

If words are indeed formed from strings of consonants and vowels, and signs from simultaneous combinations of primes, we must suppose that the listener, or viewer, somehow finds these elements in the signal. Yet from the first spectrographic descriptions of speech (Joos, 1948), two puzzling facts have been known. First, the signal cannot be divided into a neat sequence of units corresponding to the consonants and vowels of the message: at every instant, the form of the signal is determined by gestures associated with several neighboring elements. Second, as an automatic consequence of this, the acoustic patterns associated with a particular segment vary with their phonetic context. The apparent lack of invariant segments in the signal matching the invariant segments of perception constitutes the anisomorphism paradox.

The recalcitrance of the problem is reflected in the current states of the arts of speech synthesis and automatic speech recognition. Weaving a coherent, continuous pattern from a set of discrete instructions is evidently easier than recovering the discrete instructions from a continuous pattern. Speech synthesis has thus developed to a point where a variety of systems, taking a sequence of discrete phonetic symbols as input and offering a coherent, perceptually tolerable sequence of words as output, is already in use. By contrast, automatic speech recognition is still, after thirty years of research, at its beginning. Current devices recognize limited vocabularies of no more than about a thousand words. Moreover, the words must be spoken carefully, usually by a single speaker, in a small set of syntactic frames, and be confined to a limited topic of discourse. None of these devices approaches within orders of magnitude the performance of a normal human listener.

We may gain insight into why automatic speech recognition has so far failed from the corollary fact that no one has yet succeeded in devising an acceptable acoustic substitute for speech. In the burst of technological enthusiasm that followed World War II, a characteristic endeavor was to construct a sound alphabet that might substitute for spoken sounds in a reading machine for the blind. Of the dozens of codes tested, none was more successful than Morse Code, which a highly skilled operator can follow at a rate of about 35 words a minute, as against the 150-200 words a minute of normal speech. Yet with a visual alphabet, reading rates of 300-400 words a minute are commonplace. Why should this be?

Part of the answer perhaps lies in differences between seeing and hearing. Eyes comfortably scan a spatial array of static, discrete objects for information; ears are attuned to dynamic patterns of spectral change over time rather than to the abrupt "dots and dashes" of an arbitrary code. Speech has evidently evolved to distribute the acoustic information that specifies its discrete phonetic segments in patterns of change that match the ear's capacities. Yet, ironically, theories of speech perception, like the models implicit in automatic speech recognition devices, have all assumed that the signal is a collection of more or less discrete cues or properties. Not surprisingly, with this crypto-alphabetic assumption, these theories then have difficulty in recovering an integrated percept.

RESOLVING THE PARADOX

There are two possible lines of resolution of the paradox. We may reformulate our definition of the perceptual units or we may recast our description of the acoustic signal. In what follows, I will briefly sketch two current approaches that, extended and combined, may lead toward a resolution along both these lines.

Note, first, that we cannot abandon the concept of the phoneme-sized phonetic segment, and the features that describe it, without abandoning the sound structure and dual pattern on which language is premised. Moreover, there is ample evidence from historical patterns of sound change (e.g., Lehmann, 1973), errors in production (Fromkin, 1980), errors in perception (Bond & Garnes, 1980), aphasic deficit (Blumstein, 1978) and, not least, the existence of the alphabet, that the phoneme is a functional element in both speaking and listening (for fuller discussion, see Liberman & Studdert-Kennedy, 1978). What we can abandon, however, is the notion of the phoneme-sized phonetic segment as a static, timeless unit. We can attempt to recast it as a synergistic pattern of articulatory gesture, specified in the acoustic signal by spectrally and temporally distributed patterns of change.

Here, it may be useful to distinguish between the information in a spoken utterance and in its written counterpart (a similar distinction is drawn in another context by Carello, Turvey, Kugler, & Shaw, in press). Both speech and writing may serve to control a speaker's output: We may ask a subject either to repeat the words he/she hears or to read aloud their alphabetic transcription, and the two spoken outcomes will be essentially identical. But the information that subjects use to control their output is quite different in the two cases.

The form of the spoken utterance is not arbitrary: Its acoustic structure is a necessary consequence of the articulatory gestures that shaped it. In other words, its acoustic structure specifies those gestures, and the human listener has no difficulty in reading out the specifications, and thus organizing his own articulations to accord with those of the utterance. By contrast, the form of the written transcription is an arbitrary convention that specifies nothing. Rather, it is a set of instructions that indicate to the reader what he is to do, but do not specify how he is to do it (Carello, et al., in press; Turvey, personal communication). A road sign indicates "Stop," a tennis coach instructs us, "Keep your eye on the ball," but neither tells us how to do it. Their instructions are chosen to symbolize actions presumed to be in the repertoires of motorists and tennis players. If these actions were not in their repertoires, the instructions would be useless. Similarly, the elements of a transcription--whether words, syllables, or phonemes--are chosen to symbolize actions presumed to be in the repertoires of speakers. If they were not in their repertoires, the instructions would be useless. Our task is therefore to describe those actions and to understand how they are specified in the flow of speech.

Thirty years of research with synthetic speech have demonstrated that the speech signal is replete with independently manipulable "cues," which, if varied appropriately, change the phonetic percept. Two puzzling facts emerge from this work. (See Repp, 1982, for an extensive review.) First, every phonetic distinction seems to be signaled by many different cues. Therefore, to demonstrate that a particular cue is effective, we must set other cues in the synthesis program at neutral (that is, ambiguous) values. We then discover the second puzzle, namely, that equivalent, indiscriminable percepts may arise from quite different combinations of contexts and cues. Thus, Bailey and Summerfield (1980) showed that perceived place of articulation of an English stop consonant /p, t, k/, induced by a brief silence between /s/ and a following vowel (as in /spu/ or /ski/), depends on the length of the silence, on spectral properties at the offset of /s/, and on the relation between those properties and those of the following vowel. How are we to understand the perceptual equivalence of variations in the spectral structure of a vowel and in the duration of the silence that precedes it? More importantly, how are we to understand the integration of many spectrally and temporally scattered cues into a unitary percept?

The quandary was recognized and a rationale for its solution proposed some years ago by Lisker and Abramson (1964, 1971). They pointed out that the diverse array of cues that separate so-called voiced and voiceless initial stop consonants in many languages--plosive release energy, aspiration energy, first formant onset frequency--were all consequences of variations in timing of the onset of laryngeal vibration with respect to plosive release, that is, voice onset time (VOT).

> "Laryngeal vibration provides the periodic or quasi-periodic carrier that we call voicing. Voicing yields harmonic excitation of a low frequency band during closure, and of full formant pattern after release of the stop. Should the onset of voicing be delayed until some time after the release, however, there will be an interval between release and voicing onset when the relatively unimpeded air rushing through the glottis will provide the turbulent excitation of a voiceless carrier commonly called aspiration. This aspiration is accompanied by considerable attenuation of the first formant, an effect presumably to be ascribed to the presence of the tracheal tube below the open glottis. Finally, the intensity of the burst, that is, the transient shock excitation of the oral cavity upon release of the stop, may vary depending on the pressures developed behind the stop closure. Thus it seems reasonable to suppose that all these acoustic features, despite their physical dissimilarities, can be ascribed ultimately to actions of the laryngeal mechanism." (Abramson & Lisker, 1965, p. 1).

If, now, we extend this principle of articulatory coherence to other collections of cues for other phonetic features--for which, to be sure, the details have not yet been worked out--we can, at least, see how the cues may originate, and may even cohere perceptually as recurrent acoustic patterns. Moreover, we have a view of the perceptual object--consistent with Gibson's (1966, 1979) principles--as an event that modulates acoustic energy. In other words, the perceptual object is a pattern of gesture perceived directly by means of its radiated sound, or, if we are watching the movements of a signing hand, by means of a pattern of reflected light. This view, developed at Haskins Laboratories over the past thirty years, takes a step toward resolving the anisomorphism paradox by treating the perceptual object as a dynamic event rather than a static unit, but does nothing to address the problems of invariance and segmentation in the acoustic signal. For this we must turn to the work of Stevens (1972, 1975) and of Stevens and Blumstein (1978; Blumstein & Stevens, 1979, 1980).

Stevens' (1972, 1975) approach is entirely consistent with Gibson's view that "Phonemes are in the air" (Gibson, 1966, p. 94), in other words, that the acoustic signal carries invariant segments isomorphic with our phonetic percepts. For Stevens, the perceptual elements are the features of distinctive feature theory (Jakobson, Fant, & Halle, 1963). He has adopted an explicitly evolutionary approach to the link between production and perception by positing that features have come to occupy those acoustic spaces where, by calculations from a vocal tract model, relatively large articulatory variations have little acoustic effect, and to be bounded by regions where small articulatory changes have a large acoustic effect. (As a simple example, the reader might test the acoustic consequences of whispering the word east, moving slowly from the high front vowel [i] through the alveolar fricative [s] to the alveolar stop [t].)

Most of Stevens' work in recent years has been concerned with acoustic properties that specify place of articulation in stop consonants, for the good reason that the acoustic correlates of this feature have seemed particularly labile and subject to contextual variation (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). For example, in a well-known series of studies (Stevens & Blumstein, 1978; Blumstein & Stevens, 1979, 1980), Stevens and Blumstein derived by acoustic analysis a set of three "templates," characterizing the gross spectral structure at onset, integrated over the first 26 ms after stop release, for the three syllable-initial, English stop consonants, [b,d,g]. They described the templates in the terminology of distinctive feature theory as diffuse-falling for [b], diffuse-rising for [d], compact for [g]. They tested the perceptual effectiveness of these brief, static spectra by synthesis, before or as part of either steady or moving formant transitions in three vowel environments, [i,a,u]. The studies are too complex and subtly devised for summary here, but the general outcome was that most subjects were able to identify the stops with 80%-100% accuracy from the first 20-30 ms after consonant onset. Nonetheless, accuracy did vary with vowel environment and, in some syllables, subjects evidently made use of what Blumstein and Stevens term "secondary" properties, such as formant transitions, to identify the consonants.

Before we examine the implications of this last fact, we should note three important aspects of this approach to the invariance problem. First, in accord with distinctive feature theory and with the acoustic analyses of Fant (1960, 1973), Stevens and Blumstein assume that phonetic information is primarily given in the entire spectral array. "Cues" are not extracted; rather, the phonetic segment is directly specified by the signal. Second, the weight assigned to the spectrum at onset is justified by recent evidence from auditory physiology (cf. Chistovich et al., 1982; e.g., Delgutte, 1982; Kiang, 1980) that the (cat) ear is particularly sensitive to abrupt spectral discontinuities, and that the number of fibers responding to the input is increased immediately following such a discontinuity. Third, Stevens and Blumstein acknowledge the role of "secondary"--and potentially context-dependent--sources of information in patterns of spectral change (i.e., formant transitions), but attempt to exclude them by positing innate property detectors. These detectors filter out the secondary properties, it is said, and enable an infant to extract the "primary" invariances, leaving the secondary properties to be learned from their co-occurrence with the primary (Stevens & Blumstein, 1978, p. 1367).

Here, in this third aspect, we see that Stevens and Blumstein have not, in fact, completely freed their theory of perceptual atomism. By dividing the properties into "primary" and "secondary," they slip back into requiring some process of perceptual integration, accomplished, they propose, by the tautological process of "co-occurrence" or association. Moreover, the detectors themselves are purely ad hoc, tautologous entities (or processes) for which there is no independent evidence: Their existence is inferred from the fact that infants and adults respond in a particular way to stimuli that may be described as having certain properties. If we have learned nothing else from behaviorist philosophy, we should at least have learned to eschew the "Conceptual Nervous System."

Yet the detectors are supererogatory to the enterprise that Stevens and Blumstein are launched upon. The importance of their work is that they have taken the first systematic, psycholinguistically motivated, steps toward describing the invariant acoustic properties of a notoriously context-dependent class of phonetic segments. What is missing from their approach is not an imaginary physiological device, but a recognition that the signal is no more a sequence of static spectral sections than it is a collection of isolated cues. Rather the signal reflects a dynamic articulatory event of which the invariances must lie in a pattern of change.

And, indeed, moves toward this recognition have already begun. Kewley-Port (1980, 1983) has shown that an invariant pattern may be found in running spectra at stop consonant onset, and that identification accuracy for synthetic stop syllables improves, if they are synthesized from running spectra, updated at 5 ms intervals, rather than from static spectra sustained over 26 ms (Kewley-Port, Pisoni, & Studdert-Kennedy, 1983). Blumstein, Isaacs, and Mertus (1982) have found that the perceptually effective invariant may lie, not in the gross spectral shape, as originally hypothesized, but in the pattern of formant frequencies at onset. This suggests that characteristic formant shifts of the kind described in the earliest synthetic speech studies (e.g., Liberman, Cooper, Delattre, & Gerstman, 1954) may yet prove to play a role: for example, an upward shift in the low frequencies for labials, a downward shift in the high frequencies for alveolars. In fact, Lahiri and Blumstein (1982) report a cross-language (English, French, Malayalam) acoustic analysis of labial, dental, and alveolar stops that seems to be consistent with this hypothesis. The distinctions were carried by maintenance or shift in the relative weights of high and low frequencies from consonant release over the first three glottal pulses at voicing onset. All these studies move toward a dynamic rather than a static description of speech invariants.

We may see then, in (distant) prospect, a fruitful merger, consistent with theories of event perception, by which invariances in the acoustic signal are discovered as coherent patterns of spectral change, specifying a synergism of underlying articulatory gestures. From such a resolution of the invariance paradox there would follow a resolution of the segmentation paradox. For implicit in a view of the perceptual object as a coherent event is a view of "cues," "features," and, indeed, "phonemes" as descriptors rather than substantive categories of speech. The utility of features and phonemes for describing the structure of spoken languages would remain, as would--in some not yet clearly formulated sense--the functional role of the phoneme-sized phonetic segment in the organization of an utterance. But phonemes and features in perception would be seen, in origin, not as substantive categories, formed by specialized categorical mechanisms, but as emergent properties of recurrent acoustic pattern. As we will see later, this view of perception is coordinate with current research into the origins of phonological systems.

IMITATION AND THE AMODALITY OF SPEECH PERCEPTION

Let us turn now to another body of research that encourages a view of speech perception as a particular type of event perception: research on lip reading in adults and infants. The importance of this work is that it promises to throw light on imitation, a process fundamental to the acquisition of speech.

The story begins with the discovery by McGurk and MacDonald (1976; MacDonald & McGurk, 1978) that subjects' perceptions of a spoken syllable often change, if they simultaneously watch a video display of a speaker pronouncing a different syllable. For example, if subjects hear the syllable /ba/ repeated four times, while watching a synchronized video display of a speaker articulating /ba, va, ða, da/, they will typically report the latter sequence. This is not simply a matter of visual dominance in a sensory hierarchy, familiar from many intermodal studies (Marks, 1978). Nor is it a matter of combining phonetic features independently extracted from acoustic and optic displays--for example, voicing from the acoustic, place of articulation from the optic. For, although voicing is indeed specified acoustically, place of articulation may be specified both optically and acoustically, as when subjects report a consonant cluster or some merged element. Thus, presented with acoustic /ba/ and optic /ga/, subjects often report /b'ga/, /g'ba/ or a merger, /da/. (See Summerfield, 1979, for fuller discussion).

The latter effect was used in an ingenious experiment by Roberts and Summerfield (1981) to demonstrate that speech adaptation is an auditory not a phonetic process, and, more importantly, for the present discussion, to show that auditory and phonetic processes in perception can be dissociated. The standard adaptation paradigm, devised by Eimas and Corbit (1973), asks listeners to classify syllables drawn from a synthetic acoustic continuum, stretching from, say [ba] to [da], or [ba] to [pa], both before and after repeated exposure to (that is, adaptation with) one or other of the endpoint syllables. The effect of adaptation, reported in several dozen studies (see Eimas & Miller, 1978, for review), is that listeners perceive significantly fewer tokens from the continuum as instances of the syllable with which they have been adapted.

Roberts and Summerfield (1981) followed this paradigm with a series of synthetic syllables ranging from [bɛ] to [dɛ]. Their novel twist was to include a condition in which subjects were adapted audiovisually by an acoustic [bɛ], synchronized with an optic [gɛ], intended to be perceived phonetically as [dɛ]. In the event, six of their twelve subjects reported the adapting syllable as either [dɛ] or [ðɛ], four as [klɛ], one as [flɛ], one as [ma]. Not a single subject reported the phonetic event corresponding to the adapting acoustic syllable actually presented, [bɛ]. Yet, after adaptation, every subject displayed a drop in the number of tokens identified as [bɛ], roughly equal to the drop for the control condition in which acoustic [bɛ] was presented alone. Thus, while subjects' auditory systems were normally adapted by the acoustic input, their conscious phonetic percepts were specified intermodally by a blend of acoustic and optic information.

We might extend the demonstration that phonetic perception is intermodal (or, better, amodal) by citing the Tadoma method in which the deaf-blind learn to perceive speech by touch, with fingers on the lips and neck of the speaker. Tactile information may even help to guide a deaf-blind individual's own articulation (Norton, Schultz, Reed, Braida, Durlach, Rabinowitz, & Chomsky, 1977). But the lip-reading studies alone suffice to raise the question of the dimensions of the phonetic percept. The acoustic information is presumably carried by the familar pattern of formants, friction noise, plosive release, harmonic variation and so on; the optic information is carried by varying configurations of the lips and, perhaps, of the tongue and teeth (Summerfield, 1979). But how are these qualitatively distinct patterns of light and sound combined to yield an integrated percept? What we need is some underlying metric common to both the light reflected and the sound radiated from mouth and lips (Summerfield, 1979). Such a notion will hardly surprise students of action and of event-perception (e.g., Fowler, Rubin, Remez, & Turvey, 1980; Runeson & Frykholm, 1981; Summerfield, 1980). But, as I have already suggested, it is worth pursuing a little further for the light that it may throw on the bases of imitation.

Consider, first, that infants are also sensitive to structural correspondences between the acoustic and optic specifications of an event. Spelke (1976) showed that 4-month-old infants preferred to watch the film (of a woman playing "peekaboo," or of a hand rhythmically striking a wood block and a tambourine with a baton) that matched the sound track they were hearing. Dodd (1978) showed that 4-month-old infants watched the face of a woman reading nursery rhymes more attentively when her voice was synchronized with her facial movements than when it was delayed by 400 ms. If these preferences were merely for synchrony, we might expect infants to be satisfied with any acoustic-optic pattern in which moments of abrupt change are arbitrarily synchronized. Thus, in speech they might be no less attentive to an articulating face whose closed mouth was synchronized with syllable amplitude peaks and open mouth with amplitude troughs than to the (natural) reverse. However, Kuhl and Meltzoff (1982) showed that 4- to 5-month-old infants looked longer at the face of a woman articulating the vowel they were hearing (either [i] or [a]) than at the same face articulating the other vowel <u>in synchrony</u>. Moreover, the preference disappeared when the signals were pure tones, matched in amplitude and duration to the vowels, so that the infant preference was evidently for a match between a mouth shape and a particular spectral structure. Similarly, MacKain, Studdert-Kennedy, Spieker, and Stern (1983) showed that 5- to 6-month-old infants preferred to look at the face of a woman repeating the disyllable they were hearing (e.g., [zuzi]) than at the synchronized face of the same woman repeating another disyllable (e.g., [vava]). In both these studies, the infants' preferences were for natural structural correspondences between acoustic and optic information.

Interestingly, in the study by MacKain et al. (1983), the infants' preferences were only statistically significant when the infants were looking to their right sides. Kinsbourne (1973) has proposed that attention to one side of the body activates the contralateral hemisphere and facilitates processes for which that hemisphere is specialized. Given the well-known specialization of the left hemisphere for motor control of speech, we might suspect that these infants were displaying a left-hemisphere sensitivity to intermodal correspondences that could play a role in learning to speak. This hypothesis would gain support if we could establish that the underlying metric of auditory-visual correspondence was the same as that of the auditory-motor correspondence required for an individual to repeat or "imitate" the utterances of another.

To this end we may note, first, the visual-motor link evidenced in the capacity to imitate facial expression and, second, the association across many primate species between facial expression and pattern of vocalization (Hooff, 1976; Marler, 1975; Ohala, in press). Recently, Field, Woodson, Greenberg, and Cohen (1982) reported that 36-hour-old infants could imitate the "happy, sad and surprised" expressions of a model. However, these are relatively stereotyped emotional responses that might be evoked without recourse to the visual-motor link required for imitation of novel movements. More striking is the work of Meltzoff and Moore (1977) who showed that 12- to 21-day-old infants could imitate both arbitrary mouth movements, such as tongue protrusion and mouth opening, and (of particular interest for the acquisition of ASL) arbitrary hand movements, such as opening and closing the hand by serially moving the fingers. Here mouth opening was elicited without vocalization; but had vocalization occurred, its structure would, of course, have reflected the shape of the mouth. Kuhl and Meltzoff (1982) do, in fact, report as an incidental finding of their study of intermodal preferences, that 10 of their 32 4- to 5-month-old infants "...produced sounds that resembled the adult female's vowels. They seemed to be imitating the female talker, 'taking turns' by alternating their vocalizations with hers" (p. 1140). If we accept the evidence that the infants of this study were recognizing acoustic-optic correspondences, and add to it the results of the adult lip-reading studies, calling for a metric in which acoustic and optic information are combined, then we may conclude that the perceptual structure controlling the infants' imitations was specified in this common metric.

Evidently, the desired metric must be "...closely related to that of articulatory dynamics" (Summerfield, 1979, p. 329). Following Runeson and Frykholm (1981) (see also Summerfield, 1980), we may suppose that in the visual perception of an event we perceive not simply the surface kinematics (displacement, velocity, acceleration), but also the underlying biophysical properties that define the structure being moved and the forces that move it (mass, force, momentum, elasticity, and so on). Similarly, in perceiving speech, we do not simply perceive its "kinematics," that is, the changes and rates of change in spectral structure, but the underlying dynamic forces that produce these changes. Some such formulation is demanded by the facts of imitation on which the learning of speech and language rests.

ORIGINS OF THE SOUND PATTERN OF LANGUAGE

We come finally to a third aspect of current phonetic study, compatible with theories of action and event perception. The goal of the work to be discussed may be simply stated: to derive language from non-language. The topic is broad and complex. My comments here are brief, no more than a sketch of the approach.

As we have seen, every language builds its words or signs from a small set of meaningless elements, its phonemes or primes. These elements are themselves constructed from a small set of contrasting properties or distinctive features. For modern phonology, phonemes (or syllables) and their constitutive features are axiomatic primitives that require no explanation (Chomsky & Halle, 1968; Jakobson, Fant, & Halle, 1963). A central goal of linguistic study is to describe a small set of 15-20 "given' or "universal" features that will serve to describe the phonological systems of every known language. The goal has proved difficult to achieve, in large part because the various sets of features that have been proposed as potential systemic components have lacked external constraints--for example, physiological constraints on their combination (Ladefoged, 1971).

If there is indeed a universal set of linguistic features that owes nothing to the non-linguistic capacities of talkers and listeners, their biological origin must be due to some quantal evolutionary jump, a structure-producing mutation. While modern biologists may look more favorably on evolutionary discontinuities than did Darwin (e.g., Gould, 1982), we are not justified in accepting discontinuity until we have ruled continuity out. This has not been done. On the contrary, the primacy of linguistic form has been a cardinal, untested assumption of modern phonology--with the result that phonology is sustained in grand isolation from its surrounding disciplines (Lindblom, 1980).

An alternative approach is to suppose that features and phonemes reflect prior organismic constraints from articulation, perception, memory, and learning. Thus, F. S. Cooper proposed that features were shaped by the articulatory machinery. Typical speaking rates of 10 to 15 phonemes per second could "...be achieved only if separate parts of the articulatory machinery--muscles of the lips, tongue, velum, etc.--can be separately controlled, and if...a change of state for any one of these articulatory entities, taken together with the current state of others, is a change to...another phoneme...It is this kind of parallel processing that makes it possible to get high-speed performance with low-speed machinery" (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967, p. 446). A similar view was elaborated by Studdert-Kennedy and Lane (1980) for both signed and spoken language.

The most concerted attack along these lines has been developed over the past decade by Lindblom and his colleagues (e.g., Liljencrants & Lindblom, 1972; Lindblom, 1972, 1980, in press). Their goal has been not simply to specify the articulatory and acoustic correlates of certain distinctive features (as in the work of Stevens and Blumstein, discussed above), but to show how a self-organizing system of features and phonemes may arise from perceptual and motoric constraints.

The early work (Lindblom, 1972) began by specifying a possible vowel as a point in acoustic space, defined by the set of formant frequencies associated with states of the lips, tongue, jaw, and larynx. A computer was programmed to search the space for *k* maximally distinct vowels according to a least squares criterion. The vowels found were then compared with those observed in languages having *k* vowels: Despite certain obvious deficiencies, the fit of the predicted to the observed data was remarkably good. Later studies (e.g., Lindblom, 1983) have improved the fit by incorporating the results of work in auditory psychophysics (cf. Bladon & Lindblom, 1981), together with certain articulatory constraints, and by relaxing the search criterion to one of "sufficient" rather than maximum distinctness. The last move permits more than one solution for a *k*-vowel system, as indeed the observed language data require. For the present discussion, the most interesting outcome is that the derived sets of vowels form systems that invite description in terms of standard features, despite the fact that the notion "feature" was never at any point introduced into the derivation.

Recently, Lindblom has extended the procedure to derive the phoneme from sets of consonant-vowel trajectories through the acoustic space between consonant and vowel loci (Lindblom, MacNeilage, & Studdert-Kennedy, forthcoming). This work brings to bear both talker constraints (sensory discriminability, preference for less extreme articulation) and listener constraints (perceptual distance, perceptual salience) to select the syllable trajectories. Again, the interesting outcome is that when a set of trajectories is selected from a large number of possible trajectories, the syllables are not, as they might well have been, holistically distinct: Each chosen syllable does not differ from every other chosen syllable in both consonant and vowel. Rather, a few consonants and a slightly larger number of vowels occur repeatedly, while other consonants and vowel combinations do not occur at all. Thus, just as the feature emerges as a byproduct of phoneme selection, so the phoneme emerges as a byproduct of syllable selection.

This work rests on a number of assumptions that might be challenged (for example, the precise nature of talker- and listener-based constraints) and on a wealth of phonetic detail that might be questioned. Its importance does not rest on the correctness of its assumptions nor on the accuracy of its predictions--both may, and surely will, be improved in the future. Its importance lies in its style of approach: substance-based rather than formal. For if we are to do the biology of language at all, it will have to be done by tracing language to its roots in the anatomy, physiology, and social environment of its users. Only in this way can we hope to arrive at an account of language perception and production fitted to animals rather than machines.

REFERENCES

Abramson, A. S., & Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. *Proceedings of the 5th International Congress of Acoustics*. Liege: Imp. G. Thone.

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

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

Blumstein, S. E. (1978). The perception of speech in pathology and ontogeny. In A. Caramazza & E. B. Zurif (Eds.), *Language acquisition and language breakdown* (pp. 3-23). Baltimore: Johns Hopkins.

Blumstein, S. E., Isaacs, E., & Mertus, J. (1982). The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, *72*, 43-50.

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production. *Journal of the Acoustical Society of America*, *66*, 1001-1017.

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

Bond, Z. S., & Garnes, S. (1980). Misperceptions of fluent speech. In R. A. Cole (Ed.), Perception and production of fluent speech. Hillsdale, NJ: Erlbaum.

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

Chistovich, L. A. (1980). Auditory processing of speech. Language and Speech, 23, 67-75.

Chistovich, L. A., Lublinskaya, V. V., Malinnikova, T. G., Ogorodnikova, E. A., Stoljarova, E. I., & Zhukov, S. Ja. (1982). Temporal processing of peripheral auditory patterns. In R. Carlson & B. Granstrom (Eds.), The representation of speech in the peripheral auditory system (pp. 165-180). New York: Elsevier Biomedical Press.

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

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

Delgutte, B. (1982). Some correlates of phonetic distinctions at the level of auditory nerve. In R. Carlson & B. Granstrom (Eds.), The representation of speech in the peripheral auditory system (pp. 131-149). New York: Elsevier Biomedical Press.

Dodd, B. (1978). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. Cognitive Psychology, 11, 478-484.

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

Eimas, P. D., & Miller, J. L. (1978). Effects of selective adaptation on the perception of speech and visual patterns: Evidence for feature detectors. In R. D. Walk & H. L. Pick, Jr. (Eds.), Perception and experience. New York: Plenum.

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

Fant, G. (1973). Speech sounds and features. Cambridge, MA: MIT Press.

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

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

Fromkin, V. (Ed.) (1980). Errors in linguistic performance: Slips of the tongue, ear, pen and hand. New York: Academic Press.

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

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

Gould, S. J. (1982). Darwinism and the expansion of evolutionary theory. Science, 216, 380-387.

Halle, M., Hughes, G. W., & Radley, J.-P. A. (1957). Acoustic properties of stop consonants. Journal of the Acoustical Society of America, 29, 107-116.

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

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

Jakobson, R., Fant, C. G. M., & Halle, M. (1963). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Joos, M. (1948). Acoustic phonetics. *Language Monograph No. 23*, Vol. 24, Suppl.

Kewley-Port, D. (1980). Representations of spectral change as cues to place of articulation in stop consonants. *Research in Speech Perception. Technical Report No. 3*. Bloomington, IN: Indiana University, Department of Psychology.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *73*, 322-335.

Kewley-Port, D., Pisoni, D. B., & Studdert-Kennedy, M. (1983). Perception of static and dynamic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, *73*, 1779-1793.

Kiang, N. Y. S. (1980). Processing of speech by the auditory nervous system. *Journal of the Acoustical Society of America*, *68*, 830-835.

Kinsbourne, M. (1973). The control of attention by interaction between the hemispheres. In S. Kornblum (Ed.), *Attention and performance* (IV). New York: Academic Press.

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

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

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. Chicago: University of Chicago Press.

Lahiri, A., & Blumstein, S. E. (1981, November-December). *A reconsideration of acoustic invariance for place of articulation in stop consonants: Evidence from cross-language studies*. Paper presented at 102nd Meeting of the Acoustical Society of America. Miami Beach, FL.

Lehmann, W. P. (1973). *Historical linguistics*. New York: Holt, Rinehart, and Winston.

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

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

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

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*, 839-862.

Lindblom, B. (1972). Phonetics and the description of language. In A. Rigault & R. Charbonneau (Eds.), *Proceedings of the 7th International Phonetics Congress* (pp. 63-97). The Hague: Mouton.

Lindblom, B. (1980). The goal of phonetics, its unification and application. *Phonetica*, *37*, 7-26.

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

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Universals workshop*. The Hague: Mouton.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (forthcoming). *The biological bases of spoken language*. San Francisco: Academic Press.

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

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

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24*, 253-257.

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

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

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

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

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

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

Norton, S. J., Schultz, M. C., Reed, C. M., Braida, L. D., Durlach, N. I., Rabinowitz, W. M., & Chomsky, C. (1977). Analytic study of the Tadoma method: Background and preliminary results. *Journal of Speech and Hearing Research, 20*, 574-595.

Ohala, J. (in press). Cross-language uses of pitch. *Phonetica*.

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

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics, 30*, 309-314.

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

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

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

Stevens, K. N. (1975). The potential role of property detectors in the perception of consonants. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 303-330). New York: Academic Press.

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

Studdert-Kennedy, M. (1981). The beginnings of speech. In G. B. Barlow, K. Immelmann, M. Main, & P. Petrinovich (Eds.), *Behavioral development: The Bielefeld Interdisciplinary Project*. New York: Cambridge University Press.

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

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

Summerfield, Q. (1980). The structuring of language by the requirements of motor control and perception. In U. Bellugi & M. Studdert-Kennedy (Eds.), *Signed and spoken language: Biological constraints on linguistic form* (pp. 89-114). Deerfield Park, FL: Verlag Chemie.

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

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

CONVERGING EVIDENCE IN SUPPORT OF COMMON DYNAMICAL PRINCIPLES FOR SPEECH AND MOVEMENT COORDINATION*

J. A. Scott Kelso+ and Betty Tuller++

Abstract. We suggest that a principled analysis of language and action should begin with an understanding of the rate-dependent, dynamical processes that underlie their implementation. Here we present a summary of our ongoing speech production research that reveals some striking similarities with other work on limb movements. Four design themes emerge for articulatory systems: 1) They are functionally, rather than anatomically, specific in the way they work; 2) They exhibit equifinality and in doing so fall under the generic category of dynamical systems called point attractors; 3) Across transformations they preserve a relationally invariant topology; 4) This, combined with their stable cyclic nature, suggests they can function as nonlinear, limit cycle oscillators (periodic attractors). This brief inventory of regularities, though not meant to be inclusive, hints strongly that speech and other movements share a common, dynamical mode of operation.

Our work has been, and is, directed toward understanding control and coordination in so-called complex systems composed of many degrees of freedom. In brief, we want to find out how order and regularity arise in systems whose component structures are non-homogeneous. In a non-trivial sense we view the task as one of understanding the emergence of (kinetic) form, since we take our inspiration from the Soviet physiologist Bernstein (1967) who viewed movement "as a living morphological object" (p. 68). He too chose speech production as paradigmatic of the problem, for even the "simplest" of speech gestures requires cooperation among respiratory, laryngeal, and supralaryngeal structures. Nature has solved this coordination problem, but science is a long way from doing so.

At the Lake Arrowhead conference the participants spent a good deal of time discussing properties that language and movement may have in common. This issue and many others (e.g., origins, neural bases, development) are addressed in several of the papers (cf. Bellman & Goldberg, Iberall, Poizner,

---

*For Crump Institute conference (1982, March). Language and movement. Lake Arrowhead, CA. To appear in American Journal of Physiology: Regulatory, Integrative and Comparative Physiology.

+Also University of Connecticut, Storrs, CT.

++Also Cornell University Medical College, New York, NY.

Acknowledgment. This work was supported by NINCDS grant NS-13617, U.S. Office of Naval Research Contract No. N00014-83-C-00833, and BRS Grant RR-05596 to Haskins Laboratories and NINCDS grant NS-17778 to Cornell University Medical Center.

Bellugi and Iragui, Kent). In this note our aim is a bit more parochial. We wish to present briefly four sets of findings that relate the production of speech to the control and coordination of other activities such as reaching and locomoting. We believe that these observations suggest rather strongly that speech and other motor skills share a similar dynamical organization. We hasten to add that this claim is far from universally accepted; in fact, at a recent conference on speech motor control in Stockholm it constituted a major source of controversy (cf. Grillner, Lindblom, Lubker, & Persson, 1982), although in his concluding remarks, the Nobel Laureate Ragnar Granit remarked provocatively that "The motor marionette is what neurophysiology has in common with speech motoricity..." (Granit, 1982, p. 271). The problem as we see it, however, is to unpack the "motor marionette". Indeed, it is to strip away, as much as possible, the puppeteer who is pulling the strings.

In short, we resist any tendency to assume that the order and regularity we observe when people talk or move about in their environment is contained in, or prescribed by some device (such as the programs and reference levels common in machine-type theories) that embodies said order and regularity. Rather we wish to understand the generation of pattern and form without assuming a priori that there is a generator that possesses some kind of representation, neural or mental, of the pattern before it appears. This strategy applies as much to language as it does to action. Taking such a strategy seriously means, first and foremost, a commitment to understanding the rate-dependent, dynamical processes that underlie the implementation of language and action. In adopting this stance we do not mean to reject entirely the abstract, symbolic mode of operation that seems to be a hallmark of language and action. But Nature employs the symbolic mode of operation only minimally (cf. Iberall's paper) and so, at least for us, a principled analysis of language and action must begin with an account of the dynamics of speech and movement. Along with several of the participants and others, notably Pattee (1972, 1977), we wonder how it might be that discrete, rate-independent symbol strings could arise from dynamic, biological processes (cf. Kugler, Kelso, & Turvey, 1982). As far as language and action are concerned, we believe that until the dynamics have been explored more fully the question is moot. Here, we simply present some recent results that, when interpreted from a dynamical perspective, suggest there are common principles governing speech and other movements.

1. On the Functional (not Anatomical) Specificity of Motor Systems

For some time it has seemed to us (and others, e.g., Boylls, 1975; Greene, 1982; Szentagothai & Arbib, 1974; Turvey, 1977) that it is extremely unlikely that the degrees of freedom of any articulatory system are individually regulated during purposive activity (as the marionette image or earlier keyboard metaphors might suggest; for discussion see Turvey, Fitch, & Tuller, 1982). Instead, in many multi-joint movements, ensembles of muscles and joints exhibit a unitary structuring--a preservation of internal relations among muscles and kinematic components of a particular task that is stable across scalar changes in such parameters as rate and force (e.g., Kelso, Southard, & Goodman, 1979a, 1979b; see Fowler, Rubin, Remez, & Turvey, 1980; Kelso, 1981, for reviews, and Section 3 for details regarding the form that the internal "topology" takes). For us, then, the significant units of control and coordination are functional groupings of muscles and joints (which the Russians call functional synergies and we call coordinative structures) that are constrained to act as a unit to accomplish a task. One of our goals has been to try to ground this claim firmly and at the same time contrast it with notions that 'units of action' consist of anatomical arrangements such as hard-wired reflex connections or servomechanisms (see for example Gallistel's "new synthesis" of action and commentaries, 1981; also Kelso & Reed, 1981). Biological systems, as emphasized by Iberall and Yates (e.g., Iberall, 1972, 1978; Yates, 1980, 1982), are not "hard-wired, hard-geared or hard-molded," although in exhibiting the functions they do, they might appear to be so. But for us at least, biological things share no genuine likeness to machines: instead, they organize themselves to meet task demands with whatever components are available to them.

How might one establish the "soft," functional nature of muscle-joint linkages composed of many degrees of freedom? One way is to poke them around, perturb them and then examine how the potentially free variables reconfigure themselves. An instructive experiment on speech by Folkins and Abbs (1975) loaded the jaw unexpectedly during the closure movement for the first /p/ in the utterance "a /hae 'paep/ again". Lip closure was attained in all cases, apparently by exaggerated displacements and velocities of the lip closing gestures, particularly of the upper lip. Although the interpretation of this result has been uneven [initially accounted for by online feedback processing (Folkins & Abbs, 1975), later as supporting open-loop feedforward control processes (Abbs & Cole, 1982)], its impact for us as a paradigm is that anatomical structures not directly coupled to the perturbed articulator are the ones that compensate. The lips and the jaw in this case seem to constitute a functional unit, an 'equation of constraint' as it were (Saltzman, 1979); when one part is altered, other, distally linked parts automatically adapt to preserve the constraint. To us these data can hardly be accounted for by either complete preplanning (open-loop control) or fixed input-output feedback loops. But to show this, we need to demonstrate that the pattern of coupling among the articulators observed in response to the same perturbation shifts with the functional requirements of the act. For example, coordinative structure theory would predict that if the jaw is halted in its raising action during the transition into the final /b/ in /baeb/, then the lips will compensate but the tongue will not. In contrast, for a different utterance such as /baez/ the tongue will perform the primary compensation and not the lips. In short, the effects will not be fixed in reaction to the perturbation; rather, the pattern of coordination will be functionally specific to the requirements of the spoken act.

Our data (Kelso, Tuller, Bateson, & Fowler, in preparation; Kelso, Tuller, & Fowler, 1982) bear this prediction out. In one experiment, a load (5.88 Newtons, 1.5 sec duration) was applied to the subject's jaw unexpectedly (on 25% of the trials) via a DC Brushless torque motor. Movement was monitored by an optical tracking system (modified SELSPOT) that detected infrared light emitting diodes attached to the subject's lips and jaw at the midline. In addition, EMG potentials from lip and tongue muscles were obtained from paint-on and bipolar hooked wire electrodes, respectively.

The movement results were clear. The upper and lower lips preserve the timing of closure for the final /b/ in /baeb/ in the perturbed condition (like the data of Folkins and Abbs) by increasing their displacement and velocity.

But this is not a fixed, "triggered reaction" on the part of the lips to jaw perturbations. When the jaw is perturbed in exactly the same place but this time /z/ frication is required as in /baez/, there is no active lip compensation. Instead, because the jaw is lower than usual, the tongue moves further (as manifested in highly amplified tongue muscle activity) to achieve the tongue-palate relationship appropriate to frication. Like the lips in /baeb/, the tongue in /baez/ responds remarkably quickly and is time locked to the torque applied to the jaw (15-30 ms latency).

The coordinative patterns we observe in these speech experiments are highly distinctive and anything but inflexible. In this they parallel work on other movements such as cat locomotion. For example, when light touch or a weak electrical shock is applied to the cat's paw during the flexion phase of the step cycle, an abrupt withdrawal reaction occurs--as if the cat were trying to lift its leg over an obstacle. When the *same* stimulus is applied during the stance phase of the cycle, the flexion response (which would make the animal fall over) is inhibited, and the cat reacts with enhanced extension (Forssberg, Grillner, & Rossignol, 1975). Just as these reactions are non-stereotypic and functionally suited to the requirements of locomotion, so the patterns we have observed are fashioned to meet the linguistic requirements of the spoken act in unique and specific ways. The flexible patterning observed in response to perturbations in different phonetic contexts strongly speaks against either a fixed response organization (of a reflex or servo type) or a completely pre-programmed mode of control. Rather we are talking about a softly coupled system of articulators that is constrained to act, temporarily, in a unitary fashion. The cooperativity evident in the tongue-jaw-lip ensemble is specific, not to any particular articulatory target configuration, but to the production of the required sound. *The relationship is many to one; there is no isomorphism between the exact state of the articulators and the utterance that is produced.* As we will suggest next, the latter constitutes an *attractor field* (in the nomenclature of dynamical systems theory, see Abraham & Shaw, 1982; Rosen, 1970) to which articulatory trajectories converge, regardless of contextual variation (and the multiple meanings of words?).

2. *On the Equifinality Property of Motor Systems*

The spatiotemporal adjustments that occur in structures (often far removed from the structure perturbed) are constrained by the task that is performed. Seen in another light, they guarantee the task's accomplishment provided that biomechanical limits are not exceeded. This phenomenon of 'goal' achievement in spite of ever-changing postural and biomechanical rearrangements and through a wide variety of kinematic trajectories has been called motor equivalence (Hebb, 1949) or equifinality (von Bertalanffy, 1973).

We have observed equifinality in our studies of limb targeting behavior in single degree of freedom movements. Briefly, we have shown that a given target angle can be achieved despite changes in initial conditions of the limb, and despite unforeseen perturbations to the movement trajectory imposed en route to the target. This is the case in functionally deafferented humans (Kelso, 1977; Kelso & Holt, 1980; Roy & Williams, 1979) and individuals who have had the joint capsules of the index finger surgically removed, thus eliminating the seat of joint mechanoreceptors (Kelso, Holt, & Flatt, 1980).

Very similar findings have been reported in normal and deafferented monkeys for both head (Bizzi, Polit, & Morasso, 1976) and arm movements (Polit & Bizzi, 1978). Interestingly, a recent paper by Poizner, Newkirk, and Bellugi (1983) shows how 'final position control' is exploited by the linguistic system of American Sign Language, both in its lexical structure and its grammar.

Recently we have examined the production of the vowels /i/, /a/, and /u/ in isolation and in a dynamic speech context, e.g., "its a peep again." In one condition the vowels were produced normally; in another, rather extreme manipulation we artificially altered the normal configuration of the articulators by fixing the mandible using a bite block, and at the same time removed as much tactile, proprioceptive and auditory information as possible. The temporomandibular joint was anesthetized bilaterally; tactile information from oral mucosa was reduced by application of topical anesthetic (to the extent of, in some cases, eliminating the gag reflex) and audition was masked by white noise (Kelso & Tuller, 1983). Though we recognize that it was probably impossible to deprive the subject of sensory information completely, the level of performance was nevertheless quite remarkable. Measuring the vowel's acoustic spectrum at the first glottal pitch pulse, we found (in five naive subjects) no differences between normal and deprived conditions in the values of the first and second formant frequencies. Thus, in spite of the changed articulator geometry and in spite of rather drastic sensory reduction, the vocal tract accommodated to produce a normal acoustic output. Cinefluorographic work has shown that the new articulatory configuration (often involving changes in tongue and pharynx shape) preserves regions of maximum constriction between, say, the tongue and the palate for the vowel /i/ (Gay, Lindblom, & Lubker, 1981). In addition, we have recently shown in an x-ray study of bite-block speech that compensatory movements occur in a similar fashion for one adventitiously and two congenitally deaf subjects (Tye, Zimmermann, & Kelso, 1983).

What kind of system is defined when elements of the motor apparatus cooperate in an apparently complex manner to exhibit equifinality? Rosen (1970) suggests a strategy for dealing with complexity that has received only spasmodic use over the years by physiology and neuroscience, in spite of its effectiveness historically in other scientific domains. In brief, he argues that modeling complex behavior involves abstracting what the system's functional organization is rather than (or at least before) focusing on its material structure. Often complex systems have a propensity for turning themselves into rather simple, special-purpose devices to meet functional requirements.

There is now a good deal of support for the notion that 'targeting' movements are controlled by an organization dynamically similar to a (nonlinear) mass-spring system (e.g., Fel'dman, 1966; Fel'dman & Latash, 1982; Kelso, 1977; Kelso, Holt, Kugler, & Turvey, 1980). Such systems are intrinsically self-equilibrating in the sense that the "end-point" or the "target" of the system is achieved regardless of initial conditions. For us, the appeal of this model is that the "target" is not achieved by conventional closed-loop control with its processes of feedback, error detection, and comparison. Instead, it arises as an equilibrium operating point determined by the system's dynamic parameters (e.g., mass, stiffness). Kinematic variations in displacement, velocity, and trajectory are consequences of the parameters specified, not "controlled" variables (see Stein, 1982, and commentaries). Importantly, kinematics (or dynamics for that matter) are nowhere represented in the system and sensory feedback, at least in the conventional, computational sense is not required (cf. Fitch & Turvey, 1978). We are not saying that information is unimportant for the regulation and control of movement, but that it is unlikely to be provided in terms of receptor codes specific to the movement's kinematic dimensions (cf. Kelso, Holt, & Flatt, 1980). Rather, as proposed by Kugler, Kelso, & Turvey (1980) a conception of information is required that is unique and specific to the state of the system's dynamics, given perhaps geometrically in the form of gradients and equilibrium points in a potential energy manifold (see also Hogan, 1980). This is admittedly a very general description that has yet to be fully explored; it follows Thom's (1972) view of information as topologically specified in the system's dynamic qualities and offers an alternative to simplistic coding schemes in which receptor signals on a single dimension are fed back to a setpoint. In fact, we have questioned all along (as have others such as Wiener, 1965, in his last paper; Cecchini, Melbin, & Noordergraaf, 1981; Fowler et al., 1980; Iberall, 1972; Kelso, Holt, Kugler, & Turvey, 1980; Kugler et al., 1980; Yates, 1980) the appropriateness of the setpoint concept in biological processes.

We should stress again one very important point that can be misinterpreted (e.g., Bizzi et al., 1982; Soechting & Lacquaniti, 1981). The role of the mass-spring model of equifinality as we propose it is to characterize an abstract functional organization, not a unique mechanism. As we have emphasized here, it accounts for the qualitative dynamical behavior of a wide variety of materially different systems. As a style of description it has more in common with, say, Gibbs' phase rule for lawfully describing the behavior of matter as it undergoes changes in phase, e.g., from liquid to gas, regardless of chemical composition, than it has in common with, say, the details of an isolated muscle's length-tension curve. The approach here is truly dynamic: complex systems--in performing goal-directed functions--can behave as abstract, task-defined special-purpose devices such as a mass-spring. Dynamicists classify such devices as belonging generically to a taxonomic category called point attractors. We think (Saltzman & Kelso, 1983) and have preliminary evidence showing that when point attractor dynamics are expressed in task rather than articulatory coordinates, the degrees of freedom at the muscle-joint level can be wrapped up in those situations when the system displays equifinality (Saltzman & Kelso, 1983).

3. On the Topological Nature of Motor Systems

Bernstein (1967) placed great emphasis on the predominance of topological categories over metric ones in biological processes. He states "that the totality of the topological and metrical characteristics of the relations between movements and external space can be generalized under the term motor field" (italics his), and further, "that the immediate task of physiology is to analyse the properties of this field" (p. 48).

In our own experiments and in our analyses of other work, we have asked the question: what variables, or relations among variables, are preserved in the face of relevant transformations? What, if anything, remains invariant across metrical change? These questions are motivated by an approach to living systems proposed by Gelfand and Tsetlin (1962) in their theory of well-organized functions. For these authors, as for Bernstein, control and coordination are completely described by so-called non-essential ("control") variables that can effect scalar changes in the values of the function without annihilating its internal structure or topological character. The internal topology is determined by so-called "essential" variables, which elsewhere we have linked with the term "coordination" (Kugler et al., 1980).

In a wide variety of activities including locomotion, handwriting, postural balance, interlimb coordination (see Kelso, 1981; Schmidt, 1982, for reviews), we have observed a stable temporal patterning (among muscle activities or kinematic events) across scalar changes in absolute magnitude of EMG activity or kinematic components. The temporal stability often takes the form of a phase constancy among cooperating muscles as a kinematic parameter is systematically changed. Large variations in handwriting speed, for example, do not alter the intrinsic phasing among tangential velocity peaks (Viviani & Terzuolo, 1980), and, though the magnitude of acceleration pulses is much greater for a word written large than small, the timing is the same (Hollerbach, 1981). In short, the "topology" is a temporal one.

We believe that this invariant temporal structure is a fundamental "signature" of coordinated activity, including, perhaps, the production of speech. Of course, finding any kind of invariant in speech, temporal or otherwise, has been notoriously difficult. Early work at Haskins Laboratories (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; MacNeilage & DeClerk, 1969) underscored the problem in both the acoustic and physiologic domain; suprasegmental variables (such as prosodic variations and changes in speaking rate), as well as contextual (coarticulatory) effects, were shown to affect the acoustic and physiologic realization of the segment. For example, when a consonant-vowel-consonant syllable is spoken with primary stress, the muscle activity associated with production of the vowel is of longer duration and greater amplitude than it would be in an unstressed environment. The acoustic duration of the stressed vowel is also longer and the formant frequencies more extreme, than when the same vowel is produced without primary stress. Thus, although the metrics of speech shift constantly, segmental identity is somehow preserved. How can this be?

In our work (Tuller, Harris, & Kelso, 1982; Tuller, Kelso, & Harris, 1982a), we hoped that by applying two transformations that are believed to be particularly important for speech--changing syllable stress and speaking rate--we might uncover motoric variables, or relations among variables, that remain unaltered. We approached the problem initially by examining electromyographic (EMG) and acoustic recordings of speakers' productions of utterances in which syllable stress and speaking rate were orthogonally varied. Native speakers of English produced two-syllable utterances of the form pV1pV2p, where Vn was either /i/ (as in "peep") or /a/ (as in "pop"). Each utterance was spoken with primary stress placed on either the first or second syllable. The subjects read lists of these utterances at two self-selected speaking rates, "slow" (conversational) and "fast."

EMG recordings were obtained from five muscles known to be active during production of the speech sounds that we used: 1) Orbicularis oris participates in bringing the lips together for /p/. 2) Genioglossus bunches the main body of the tongue and brings it forward for the production of the vowel /i/. 3) The anterior belly of digastric and 4) the inferior head of lateral pterygoid are associated wtih jaw lowering during speech, while 5) medial pterygoid acts to raise the jaw (Tuller, Harris, & Gross, 1981).

When subjects increased speaking rate or decreased syllable stress, the acoustic duration of their utterances decreased as expected, and the magnitude and duration of activity in individual muscles changed markedly. In general, EMG activity was of longer duration and greater magnitude for production of stressed than unstressed syllables. EMG activity was of shorter duration or increased amplitude in syllables spoken quickly compared with those spoken slowly.

In order to evaluate possible phasing relations among muscle events that might be stable across such large individual variations, we looked at period durations (e.g., between the onsets of muscle activity for vowel 1 and vowel 2) and latencies of corresponding consonantal events relative to such periods. We examined all possible muscle combinations across each of the four speaking conditions, i.e., conversational or fast rate with first or second syllable stressed. One very consistent result emerged, namely, an invariant linear relationship between duration of the vocalic cycle (onset of muscle activity for V1 to onset of muscle activity for V2) and the latency between V1 onset and the intervening consonant. Thus, timing of consonant production relative to vowel production was invariant over substantial changes in the period of the vocalic cycle. New kinematic results in which articulatory movements corresponding to vowel and consonantal gestures were examined have confirmed this result (Tuller, Kelso, & Harris, 1982b, in press), implicating a functionally significant vowel-to-vowel cyclicity in English (see also Fowler, in press).

In short, these data not only provide evidence for relational invariance in timing among articulatory events in speech, but also share a close correspondence to results obtained in many other motor activities. To use Winfree's (1980) term, the preservation of "temporal morphology" across scalar variation may be a design feature of all motor systems and may be Nature's way of solving the problem of coordinating complex systems, like speech, whose degrees of freedom are many. It will not be lost on the reader that this design may arise from the (thermodynamic) requirement that biological systems, to persist, must be cyclical in nature. In our final comment, we turn to a discussion of the fundamental rhythmicity that characterizes many articulatory activities, and perhaps even language itself.

4. On the Fundamental Cyclicity of Motor Systems

The ubiquitous cyclicity in biological processes at many scales of analysis needs little comment here (see Winfree, 1980, for a good review). As for the neural basis of rhythmic motor behavior, Delcomyn (1980) remarks that the big questions no longer concern central versus peripheral control, but rather what kind of oscillatory processes are involved and how they interact to effect coordination in an animal. He goes on to say that "Recognition that systems of oscillators are universal will lead to a better understanding of motor control...and...bring neuroscientists much closer to the goal of understanding how nervous systems function" (p. 498). Similarly, Grillner (1977,

1982) and others (Kelso, Tuller, & Harris, 1983) have argued that rhythmic generation in locomotion, respiration, and mastication share a common neural design logic.

Though speech certainly uses many of the same body parts as chewing, its rhythmic basis is much less secure, in spite of the fact that linguists have long claimed that languages are rhythmical and people perceive them to be so. Moreover, the timing data discussed in Section 3 were also suggestive of some basic rhythmical structure underlying the maintenance of temporal order across transformation. Lenneberg (1967) reviewed some indirect evidence on psychological and physiological "clocks" that led him to posit a basic speech periodicity of 6±1 cycles per second. To test the idea, Lenneberg suggested using a computer to monitor some easily isolable speech event associated with syllable onset, and plotting its frequency distribution over an extended period of running speech. The suggestion was taken up seriously by Ohala (1975) who measured some 10,000 successive jaw opening gestures during a 1.5 hour reading period, but to little avail: An extremely wide variance band accompanied a dominant, but ill-defined periodicity of 250 ms. According to Ohala (1975) his findings gave "no support to the claim that there is any isochronic principle underlying speech, at least the speech of this particular speaker" (p. 434), who, parenthetically, was himself. In addition, there have been many acoustic studies of speech rhythm, most of which have reported large departures from measured isochrony (see Fowler, in press, for review and also a fresh look on the issue).

Part of the problem in establishing the existence of an articulatory rhythm rests on the measurement process (as it apparently does in the acoustic domain as well; Fowler & Tassinary, 1981). Speech production is inherently multidimensional; during running speech different articulators are involved to different degrees and the temporal overlap, "coarticulation," among articulators is considerable. Confronted with so many co-occuring events, there is little chance of identifying a basic rhythm, even though our perceptual impressions lead us to suppose that there is one.

We have adopted an experimental paradigm that may provide some insight (Kelso & Bateson, 1983). Briefly, we asked subjects to speak "reiterantly," that is, to substitute the syllable /ba/ or /ma/ for the real syllable in an utterance yet still maintain the utterance's normal prosodic structure. Thus a sentence "When the sunlight strikes raindrops in the air" would be produced "ba ba <u>ba</u> ba <u>ba</u> <u>ba</u> ba ba ba <u>ba</u>" where the underlining indicates an idealized (and simplified) stress pattern. A previous acoustic study by M. Liberman and Streeter (1978) found that the segmental makeup of target utterances had little or no effect on the duration of the substituted nonsense syllables, which were principally determined by stress and constituent structure. For example, the acoustic duration of reiterant syllables in "cunning scholars deciphered the tablets" was identical to "thirteen teachers were furloughed in August."

The benefit of the reiterant technique is that the removal of segmental factors (besides having minimal effects on the metrical pattern) allows one to measure the movements of the <u>primary</u> supralaryngeal articulators, in our case the lips and jaw involved in /ba/ and /ma/. Figure 1 (left) shows displacement-time profiles of the jaw and lower lip plus for one such sentence.

Although there are clear effects of stress on the space-time behavior of articulatory gestures (e.g., a tendency for large amplitudes and longer durations for stressed syllables), the overall periodicity is very stable indeed. Coefficients of variation in cycle duration (lip closure-to-lip closure or jaw opening-to-jaw opening) were in the region of 15 to 20%. This relatively narrow band variance, concentrated around a cyclicity of approximately 5 Hz, contrasts sharply with Ohala's (1975) earlier work, which for reasons discussed previously was likely subject to contaminating factors. When segmental variation is removed and measurements confined to the action of primary articulators, it is possible to identify (as we have here we think, for the first time) an articulatory cyclicity in its "purest" form. Clearly the periodicity we observe is not perfectly isochronous: unless one were dealing with an ideal totally conservative harmonic oscillator (which exists only in textbooks) one would not expect it to be. Nevertheless, as shown in phase-portrait form in Figure 1 (right), the trajectories do exhibit stable

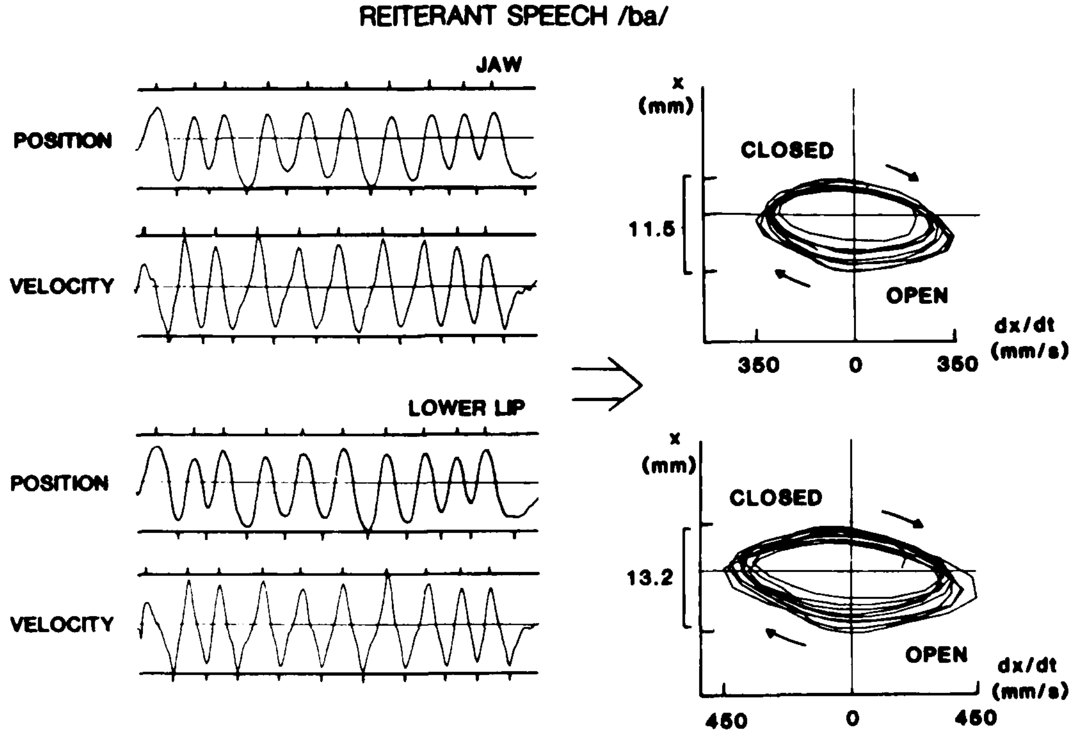

REITERANT SPEECH /ba/

Figure 1. Left. Position-time and corresponding velocity-time profiles of the jaw and lower lip (plus jaw) of the sentence "When the sunlight strikes raindrops in the air" spoken reiterantly with the syllable /ba/ interjected for the real syllables (see text for details). Right. Phase-portraits corresponding to the articulatory profiles shown on the left. Closed refers to the portion of the trajectory in which the articulator is moving into and out of closure for the bilabial consonant. Open refers to the vocalic portion of the syllable. Ordinate is position, x in mm, abscissa is velocity, dx/dt in mm/sec.

orbits and single-peaked velocity profiles regardless of stress and rate variations and small changes in initial conditions. (There are, in fact, some interesting differences in the microstructure of the stressed and unstressed syllables when viewed on the phase plane that space does not permit us to discuss here.) These trajectories describe the behavior of the articulatory system for this task: their topology is characteristic of nonlinear, limit cycle oscillations, which are the only predicted temporal stability for biological processes (Iberall, 1978; Yates, 1982). In a limit cycle, any dissipative losses that occur during a cycle are compensated for by a forcing function (an "escapement" pulse), which, in the case of speech, is precisely tuned to the required stress level. As suggested for locomotion (Shik & Orlovskii, 1965) we might expect each cycle to be instituted *de novo* in speech, in order to satisfy local phonetic and more global, suprasegmental constraints.

Elsewhere, we have identified muscle linkages in general with nonlinear oscillatory processes (see previous section) and demonstrated their entrainment properties both within (Kelso, Holt, Rubin, & Kugler, 1981) and across anatomically separate subsystems (Kelso, Tuller, & Harris, 1983). By this reasoning, which is consistent with homeokinetic theory, any persistent motion must exhibit limit-cycling behavior (Kugler et al., 1980). Speech cannot be granted exempt status. An ensemble of functioning muscles is first and foremost a "thermodynamic engine" (Bloch & Iberall, 1982; Kugler et al., 1980) whose dissipative cyclic motions are sustained through the capability to draw on a source of potential energy. Thus, such functional units share not only common sources of afferent and efferent information (Boylls, 1975) but also a vascular, metabolic network as well (Bloch & Iberall, 1982).

Though we have yet to test this idea, we might expect a complex system like speech to consist of different nested periodicities; the cycling we have observed here, for example, may well be coupled into the respiratory cycle in a harmonically-related fashion, just as the locomotory motions of many animals are (Bramble & Carrier, 1983). Indeed, in a preliminary study of continuous limb movements in which the subject chooses a preferred frequency and amplitude and we record movements over an extended period of time (≃90 sec), spectral analysis reveals two dominant peaks--one at the preferred frequency (≃2 Hz) and the other at ≃.25 Hz, corresponding to the respiration rate. In this case, as in speech, shorter term cyclicities may cohere under a longer term power-cycle such as the inspiration-expiration-inspiration cycle.

The present data on speech, then, combined with evidence from many other motor activities are strongly suggestive of a temporal organization of the limit cycle type. We have begun to identify the cyclicities and to show that they can be functionally significant, following the methods of biospectroscopy (Bloch et al., 1971). A good beginning has been made with physiological tremor (Goodman & Kelso, 1983).

CONCLUSIONS

We recognize that this inventory of parallels between speech and other motor behaviors is incomplete. We have omitted, for example, any detailed discussion of coarticulation, which recent evidence suggests is a faculty not restricted to human speech. Thus the grooming behavior of mice can be modified by its relation to actions that occur before or after it in an overall sequence (Fentress, in press). "Motor marionette" theories that posit a discrete organization of elements of behavior do not handle such findings very well. We recognize also that our results may indicate only analogies, and that the stronger claim--that they arise from common dynamical principles--is very risky. But it is precisely these functional similarities existing in structurally very different systems that allow us to identify them as belonging to the same set. The regularities we see in speech and movement, and the laws that underlie them, may have more in common than the particular structures that embody the laws. Indeed, the strategy adopted here--of identifying functional organizations common to materially very different systems--was central to Rashevsky's (1954) early attempts at formulating the field of relational biology and remains at the core of dynamical systems theory (e.g., Abraham & Shaw, 1982; Rosen, 1970). The same sentiment has recently been expressed by Eigen and Winkler (1981, p. 252). Our tentative, but non-trivial claim, then, is that speech and other articulator movements are dynamically alike with respect to the way they are controlled and coordinated.

REFERENCES

Abbs, J. H. & Cole, K.J. (1982). Consideration of bulbar and suprabulbar afferent influences upon speech motor coordination and programming. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson, (Eds.), Speech motor control. Oxford: Pergamon Press.

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

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

Bertalanffy, L. von. (1973). General system theory. Harmondsworth, Eng.: Penguin University Press.

Bizzi, E., Chapple, W., & Hogan, N. (1982). Mechanical properties of muscle: Implications for motor control. Trends in Neuroscience, 5, 395-398.

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

Bloch, E. H., & Iberall, A. S. (1982). Toward a concept of the functional unit of mammalian skeletal muscle. American Journal of Physiology, 1982, 242, R411-R420.

Bloch, E., Cardon, S., Iberall, A., Jacobowitz, D., Kornacker, K., Lipitz, L., McCulloch, W., Urquhart, J., Weinberg, M., & Yates, F. E. (1971, February). Introduction to a biological systems science. NASA Contractors Report No. 1720, Washington, D. C.

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

Bramble, D. M. & Carrier, D. R. (1983). Running and breathing in mammals. Science, 219, 251-256.

Cecchini, A. B. P., Melbin, J., & Noordergraaf, A. (1981). Setpoint: Is it a distinct structural entity in biological control? Journal of Theoretical Biology, 93, 387-394.

Delcomyn, F. (1980). Neural basis of rhythmic behavior in animals. Science, 1980, 210, 492-498.

Eigen, M., & Winkler, R. (1981). *Laws of the game*. New York: Alfred A. Knopf.

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

Fel'dman, A. G., & Latash, M. L. (1982). Interaction of afferent and efferent signals underlying joint position sense: Empirical and theoretical approaches. *Journal of Motor Behavior*, *14*, 174-193.

Fentress, J. C. (in press). Hierarchical motor control. *Neurosciences Research Program Bulletin*.

Fitch, H., & Turvey, M. T. (1978). On the control of activity: Some remarks from an ecological point of view. In R. Christina (Ed.), *Psychology of motor behavior and sports*. Urbana, IL: Human Kinetics.

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

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

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

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

Fowler, C. A., & Tassinary, L. (1981). Natural measurement criteria for speech: The anisochrony illusion. In J. Long & A. Baddeley (Eds.), *Attention and performance IX*. Hillsdale, NJ: Erlbaum.

Gallistel, C. R. (1981). The organization of action: A new synthesis. *The Behavioral and Brain Sciences*, *4*, 609-619.

Gay, T., Lindblom, B., & Lubker, J. (1981). Production of bite-block vowels: Acoustic equivalence by selective compensation. *Journal of the Acoustical Society of America*, *69*, 802-810.

Gelfand, I. M., & Tsetlin, M. L. (1962). Some methods of control for complex systems. *Russian Mathematical Surveys*, *17*, 95-116.

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

Granit, R. (1982). Some remarks from the viewpoint of neurophysiology. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), *Speech motor control*. Oxford: Pergamon Press.

Greene, P. H. (1982). Why is it easy to control your arms? *Journal of Motor Behavior*, *4*, 260-286.

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

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

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

Hebb, D. O. (1949). The organization of behavior. New York: Wiley.

Hogan, N. (1980, August). Mechanical impedance control in assistive devices and manipulators. Proceedings of the Joint Automatic Control Conferences, San Francisco, Vol. 1, TA-108.

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

Iberall, A. S. (1972). Toward a general science of viable systems. New York: McGraw-Hill.

Iberall, A. S. (1978). Cybernetics offers a (hydrodynamic) thermodynamic view of brain activities. An alternative to reflexology. In F. Brambilla, P. K. Bridges, E. Endroczi, & C. Heusep (Eds.), Perspectives in endocrine psychobiology. New York: Wiley.

Kelso, J. A. S. (1977). Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 3, 529-543.

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

Kelso, J. A. S. & Bateson, E. V. (1983). On the cyclical basis of speech production. Journal of the Acoustical Society of America, 73 (Suppl. 1), S67 (Abstract).

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

Kelso, J. A. S., Holt, K. G., & Flatt, A. E. (1980). The role of proprioception in the perception and control of human movement: Toward a theoretical reassessment. Perception & Psychophysics, 28, 45-52.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. (1980). On the concept of coordinative structures in dissipative structures: II. Empirical lines of convergence. In G. E. Stelmach (Ed.), Tutorials in motor behavior. New York: North-Holland.

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

Kelso, J. A. S., & Reed, E. (1981). Motivating muscles: The problem of action. Contemporary Psychology, 26, 181-182.

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

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

Kelso, J. A. S., & Tuller, B. (1983). "Compensatory articulation" under reduced conditions of afferent information: A dynamic formulation. Journal of Speech and Hearing Research, 26, 217-214.

Kelso, J. A. S., Tuller, B., & Fowler, C. A. (1982). The functional specificity of articulatory control and coordination. Journal of the Acoustical Soiety of America, 72 (Suppl. 1), S103 (Abstract).

Kelso, J. A. S., Tuller, B., Bateson, E., & Fowler, C. A. (ms. in preparation).

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

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 1-47). New York: North-Holland.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1982). On control and coordination in naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), The development of movement control and coordination, New York: Academic Press.

Lenneberg, E. H. (1967). Biological foundations of language. New York: Wiley.

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

Liberman, M., & Streeter, L. A. (1978). Use of nonsense-syllable mimicry in the study of prosodic phenomena. Journal of the Acoustical Society of America, 63, 231-233.

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

Ohala, J. (1975). The temporal regulation of speeh. In Fant, G., & Tatham, M. A. A. (Eds.) Auditory analysis and perception of speech. London: Academic Press.

Pattee, H. H. (1972). Laws and constraints, symbols and language. In C. H. Waddington (Ed.), Towards a theoretical biology. Chicago: Aldine.

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

Poizner, H., Newkirk, D., & Bellugi, U. (1983). Processes controlling human movement: Neuromotor constraints on American sign language. Journal of Motor Behavior, 15, 2-18.

Polit, A., & Bizzi, E. (1978). Processes controlling arm movements in monkeys. Science, 1978, 201, 1235-1237.

Rashevksy, N. (1954). Topology and life: In search of general mathematical principles in biology and sociology. Bulletin of Mathematics and Biophysics, 16, 317-348.

Rosen, R. (1970). Dynamical system theory in biology: Vol. 1. Stability theory and its applications. New York: Wiley.

Roy, E. A., & Williams, I. D. (1979). Memory for location and extent: The influence of reduction of joint feedback information. In G. C. Roberts & K. M. Newell (Eds.), Psychology of motor behavior and sport--1978. Champaign, IL: Human Kinetics.

Saltzman, E. (1979). Levels of sensorimotor representation. Journal of Mathematical Psychology, 20, 91-163.

Saltzman, E. L., & Kelso, J. A. S. (1983). Toward a dynamical assessment of motor memory and control. In R. A. Magill (Ed.), Memory and control of action. Amsterdam: North Holland.

Saltzman, E. L., & Kelso, J. A. S. (1983). Skilled action: A task dynamic approach. Unpublished manuscript.

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

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

Soechting, J. E., & Lacquaniti, F. (1981). Invariant characteristics of a pointing movement in man. Journal of Neuroscience, 1, 710-720.

Stein, R. B. (1982). What muscle variables does the central nervous system control? The Behavioral and Brain Sciences, 4, 535-578.

Szentagothai, J., & Arbib, M. A. (Eds.). (1974). Conceptual models of neural organization. Neurosciences Research Program Bulletin, 12(3).

Thom, R. (1975). Structural stability and morphogenesis. (D. J. Fowler, Trans.) Reading, MA: Benjamin, Inc. (Original work published 1972).

Tuller, B., Harris, K. S., & Gross, B. (1981). Electromyographic study of the jaw muscles during speech. Journal of Phonetics, 9, 175-188.

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

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

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1982b). On the kinematics of articulatory control as a funtion of stress and rate. Haskins Laboratories Status Reports on Speech Research, SR-71/72, 81-88.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (in press). Converging evidence for the role of relative timing in speech. Journal of Experimental Psychology: Human Perception and Performance.

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

Turvey, M. T., Fitch, H. L., & Tuller, B. (1982). The problems of degrees of freedom and context-conditioned variability. In J. A. S. Kelso (Ed.), Understanding human motor control. Champaign, IL: Human Kinetics.

Tye, N., Zimmermann, G., & Kelso, J. A. S. (1983). Compensatory articulation in hearing-impaired speakers: A cinefluorographic study. Journal of Phonetics, 11, 101-115.

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

Wiener, N. (1965). Perspectives in cybernetics. In N. Wiener & J. P. Schade (Eds.), Cybernetics of the nervous system, (Progress in brain research, Vol. 17). Amsterdam: Elsevier.

Winfree, A. T. (1980). The geometry of biological time. New York: Springer-Verlag.

Yates, F. E. (1980). Physical causality and brain theories. American Journal of Physiology, 238, R277-R290.

Yates, F. E. (1982). Outline of a physical theory of physiological systems. Canadian Journal of Physiology and Pharmacology, 60, 217-248.

PHASE TRANSITIONS AND CRITICAL BEHAVIOR IN HUMAN BIMANUAL COORDINATION*

J. A. Scott Kelso+

Abstract. The conditions that give rise to phase shifts among the limbs when an animal changes gait are poorly understood. Often a 'switch mechanism' is invoked whose neural basis remains speculative. Abrupt phase transitions also occur between the two hands in humans when movement cycling frequency is continuously increased. The asymmetrical, out-of-phase-mode shifts suddenly to a symmetrical, in-phase mode involving simultaneous activation of homologous muscle groups. The boundary between the two coordinative states is indexed by a dimensionless critical number, which remains constant regardless of whether the hands move freely or are subject to resistive loading. Coordinative shifts appear to arise because of continuous scaling influences that render the existing mode unstable. Then, at a critical point, bifurcation occurs and a new stable (and perhaps energetically more efficient) mode emerges.

It is well known that when quadrupeds change their mode of gait from a trot to a gallop, the phase relations of the limbs are altered abruptly from a roughly out-of-phase, asymmetric mode to an in-phase, symmetric mode. Although such discontinuous changes in coordination are not well understood, it is frequently assumed that central pattern generators exist (often equated with motor programs) whose role is to select the desired spatiotemporal pattern of muscle activities (Brooks & Thach, 1981; Gallistel, 1980; Keele, 1981; MacKay, 1980; Schmidt, 1982). In the case of so-called stereotypic activities like locomotion, the basic programs are hypothesized to be innately given (Grillner, 1977; Thelen, Bradshaw, & Ward, 1981). We report here, however, that under certain conditions phase transitions also exist in voluntary cyclical movements of the two hands. Under instructions to increase frequency of cycling progressively, a sudden and spontaneous shift occurs from an asymmetrical, 180 degree out-of-phase mode in which one wrist flexes as the other extends, into a symmetrical, in-phase mode that involves simultaneous

---

*Also American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, in press.

+Also Departments of Psychology and Biobehavioral Sciences, University of Connecticut

Acknowledgment. This work was supported by NIH grant NS-13617, Biomedical Support Research Grant RR-05596 and Contract No. N00014-83-C-0083 from the U.S. Office of Naval Research. The data were presented in preliminary form at a Kroc Foundation Conference on Nonlinear Mechanisms in Brain Function (Santa Barbara, California, March 1-5, 1982) organized by A. Mandell and F. E. Yates. Drs. E. Saltzman, M. Studdert-Kennedy, B. Tuller and M. Turvey provided very useful comments on earlier drafts. A basic software package developed by Dr. P. Rubin was adapted, with the help of G. Lame, for the analysis conducted here.

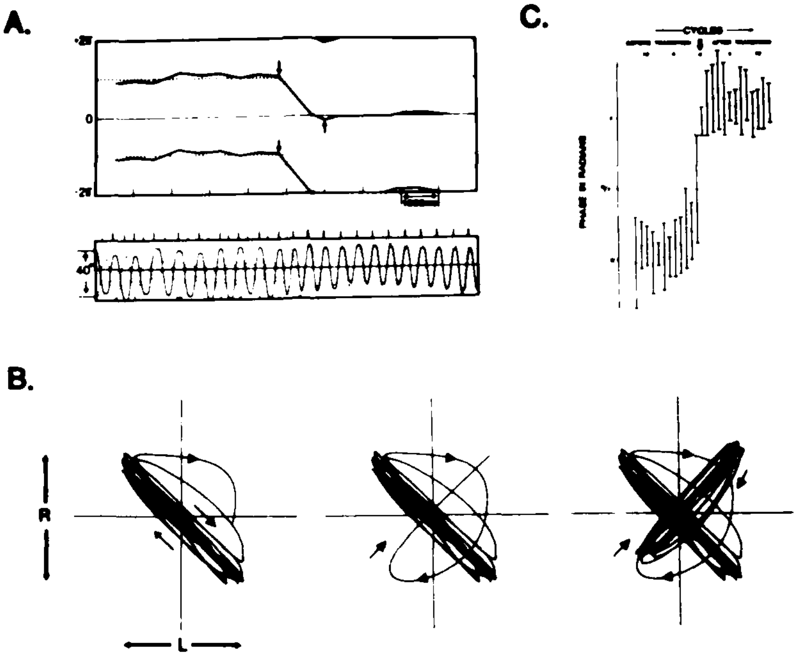
[HASKINS LABORATORIES: Status Report on Speech Research SR-74/75 (1983)]

activation of homologous muscle groups. When the transition is allowed to occur naturally, the critical frequency is predictable from the preferred frequency regardless of whether the hands move freely or are subjected to resistive loading. We take these data to support the notion (Kugler, Kelso, & Turvey, 1980) that phase transitions in movement may follow the same laws as the phase transitions and critical behavior described for many other natural phenomena (e.g., Fleury, 1981; Haken, 1975, 1978; Iberall & Soodak, 1978; Riste, 1975).

The basic experiments reported here required subjects to cycle the hands at the wrist in the horizontal plane in an asymmetical mode, that is, one in which flexion (extension) of one wrist was accompanied by extension (flexion) of the other. Similar experiments have been carried out using movements of the index fingers. A preliminary presentation of the finger movement data, whose results were basically identical to the present studies has been presented (Kelso, 1981; see also Kelso & Tuller, in press). The subjects, seated with forearms firmly supported in a position parallel to the ground, grasped a freely rotating handle with each hand, the positions of which were converted to DC voltages by potentiometers mounted over the respective axes of motion. A full description of the apparatus appears in Kelso and Holt (1980). These signals were recorded on FM tape and later subjected to analog-to-digital conversion at a sampling frequency of 200 Hz. Time-domain displacement tracings were obtained that could be displayed and analyzed on a computer graphics terminal. Instructions to subjects were to commence cycling the hands slowly and then to increase rate of cycling either in response to a verbal cue provided by the experimenter at 15 sec intervals or by a metronome whose interpulse interval could be adjusted in 100 ms increments every 15 sec. Driving frequencies in the metronome case ranged from 1-5 Hz. In another experiment subjects performed a series of trials under identical instructions but with a resistive load applied to both limbs. In this case the vertical rods leading to the potentiometers were clamped between fixed wooden blocks, thus providing a frictional damping force throughout the range of motion for each limb of approximately 5.9 Newtons.

Before the experimental manipulation, baseline measures of subjects' preferred frequency and amplitude in both asymmetrical and symmetrical modes were obtained under free and resistive loading conditions. Subjects were instructed to choose their preferred frequencies and amplitudes in such a way that they "could perform the task all day," if required to do so. Movements of each limb were then continuously sampled at 200 Hz for 30 sec. Measures of frequency (in Hz), amplitude (in deg.) and interlimb phase (in rad.) were obtained for each limb on every cycle. In addition, by assuming an approximately sinusoidal motion, we estimated the total mechanical energy expended per unit moment of inertia per cycle (proportional to the square of a given cycle's peak velocity).

The results were unequivocal for all the six subjects' data analyzed. In Figure 1a we show the movement trajectories of the two limbs for one subject as the rate increased. The rapid shift in phase is obvious. Figure 1b shows the same data on the Lissajous plane with one limb's displacement plotted against the other. It can be seen that the phase relations between the limbs are initially very stable. Were the two motions perfectly sinusoidal with phase equal to radians, a straight line would be observed. As frequency



A. A computer generated display of displacement-time profiles of left (solid line) and right (dashed line) hands plotted against each other and accompanying phase relationship between the two. The peaks of one hand movement act as a "target" file and their phase position is calculated continuously relative to the peak-to-peak period of the other "reference" file. The display repeats the phase curve so that phase lags and leads can be noted. The subject in this case is simply increasing the frequency of cycling in an asymmetric mode in response to a verbal cue from the experimenter. B. Identical data to those shown in Figure 1a, except displayed on the Lissajous plane. Positions of the left and right hands are displayed on the ordinate and abscissa, respectively. Viewed from left to right, the hands first preserve a quite stable out-of-phase relation that becomes more variable (less stable) over time as evident in the widening of the Lissajous phase portrait. Eventually the hands jump into the next mode, which remains quite stable thereafter. C. The average value of phase plotted over cycles before and after the transition. Bars correspond to standard deviations. Each point is the average of 19 different phase transition experiments (11 free and 8 resisted). The abrupt phase shift is apparent.

Figure 1

increases, the phase difference between the limbs becomes more variable, evident in the widening of the Lissajous phase portrait. Following the transition, phase becomes stable once again. The overall picture of phasing between the limbs as a function of cycles is shown in Figure 1c. Each point on the phase diagram corresponds to the mean of 19 different phase transition experiments (11 free and 8 resisted). In all cases, an abrupt change in phase was observed. Usually, the jump from one mode to the other occurred within a cycle; seldom did the transition require more than 2 or 3 cycles. On two occasions, both in the same subject, temporary transitions occurred in which the limbs moved from an asymmetrical to symmetrical pattern, and then returned to an asymmetrical pattern. Eventually, however, a permanent transition to the symmetrical, in-phase mode was observed.

Others as well as ourselves have shown that in bimanual finger movement tasks only two modes--symmetrical and asymmetrical--are stable regardless of whether the subjects are naive or whether they are skilled musicians (Kelso, Holt, Rubin, & Kugler, 1981; Yamanishi, Kawato, & Suzuki, 1980). This is not to say that other phase relations are not possible, only that they tend to be much more variable. Skilled pianists, as well as those who study their motor performance (Shaffer, 1980), have long recognized the difficulty in performing complex bimanual rhythms. In fact, characteristic "errors" often occur--manifested as tendencies to produce in-phase and out-of-phase patterns--and are only avoided through much practice.

The present data indicate that when cycling frequency is increased, one mode becomes unstable only to disappear and be replaced by another stable mode. In this they share a likeness to studies of locomotion in decerebrate cats (Shik, Severin, & Orlovskii, 1966) that demonstrated that a steady increase in electrical stimulation applied to the midbrain region was associated with increases in rate of locomotion. Moreover, transitions in gait occurred when sufficiently strong current was employed. Like some of our data, unstable regions were also noted in which the animal sometimes trotted and sometimes galloped. Above a certain value of current (80 μA), however, only galloping occurred. Our results, similar to these findings on gait, suggest that changes in coordination may be ordered by changes in the magnitude of a single parameter.

We have some reason to suppose that the 'new' stable mode is energetically more favorable at a given frequency than its predecessor. In the free unloaded experiments, cycle frequency increased significantly across the transition (from an average of 2.26 Hz over the 5 cycles before the transient phase to 2.50 Hz averaged over 5 cycles after the transient, $t(10) = 3.45$, $p = .006$), but cycle amplitude and energy dropped across the transition, $t(10) = 2.59$ and 2.11; $p = .03$ and $p = .06$, respectively. The pattern was similar in the eight resistive loading experiments: frequency increased significantly across the transition, while amplitude and energy dropped slightly but not significantly. It should be emphasized that under both resistive and nonresisted conditions, cycle energy was always substantially greater before the transition than in either of the corresponding preferred mode conditions ($p < .01$).

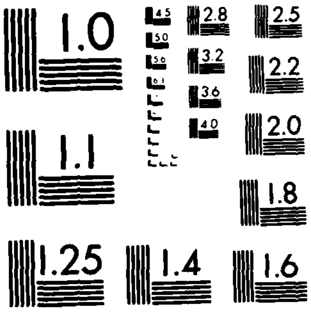
Systematic relationships between energy utilization and modal behavior have also been reported in studies of gait in horses (Hoyt & Taylor, 1981), and gnus (Pennycuick, 1975). Horses locomoting in a free environment, for example, elect only those ranges of speed within a gait that correspond to regions of minimum oxygen expenditure (Hoyt & Taylor, 1981). Moreover, when horses are forced to maintain a given gait at a speed other than preferred, metabolic costs increase dramatically, until, at some threshold value a shift into the next most economical mode occurs. Shifts in locomotory modes are not hard-wired or deterministic (except perhaps at the very limits of stability): Horses can trot at speeds at which they normally gallop or walk, but it is metabolically expensive to do so.

It is also possible to delay the phase transition observed in these experiments consciously. The critical value at which the transition occurs naturally, however, (that is, without a purposeful effort to resist it), is highly predictable. Though the absolute values of frequency, amplitude, and energy (measured over the last five consecutive cycles before the transient phase) vary considerably between and within subjects, one *relative* measure does not. When the frequency at transition is scaled to the individual's preferred frequency in the out-of-phase mode a highly linear relationship is observed.

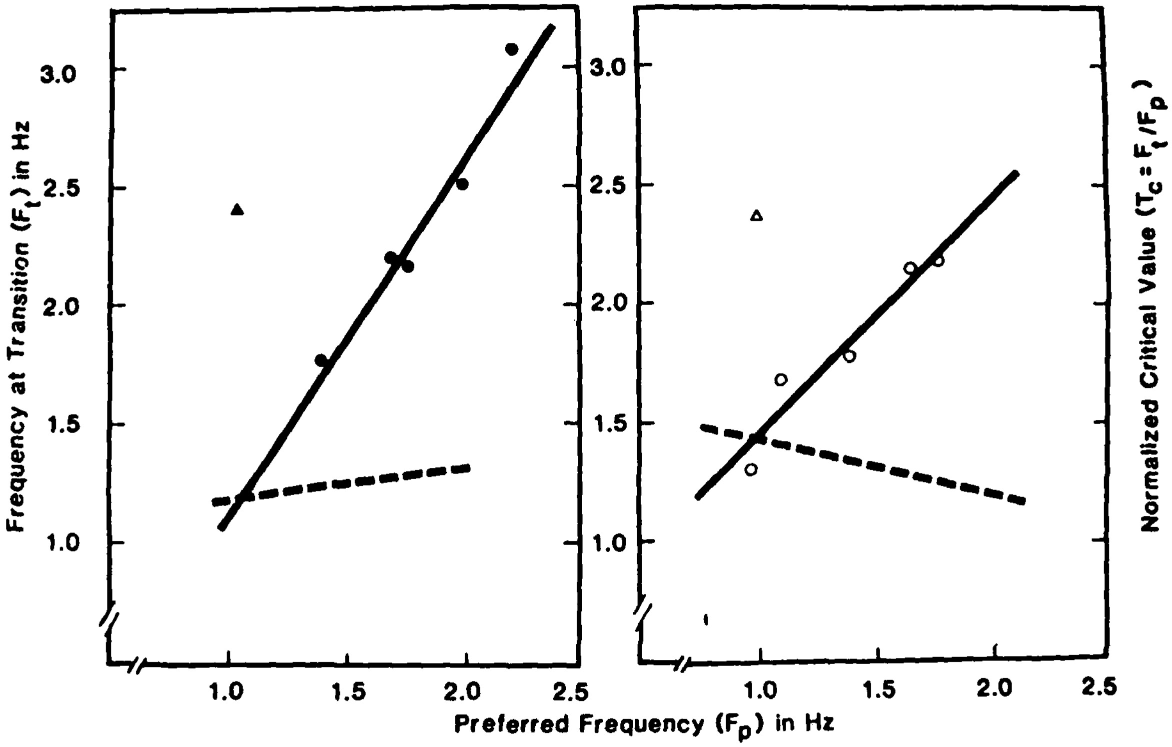
This relationship, along with least squares regression lines, is plotted in Figure 2 for free and resistive loading experiments for five subjects (solid lines). The effect of resistive loading was to reduce both preferred frequency and transition frequency in a reliable fashion ($p < .01$). The mean preferred frequencies for free and resisted experiments were 1.81 Hz (= 552 ms, SD = 30 ms) and 1.37 Hz (= 730 ms, SD = 33 ms), respectively. The mean transition frequency for the free case was 2.34 Hz (= 427 ms, SD = 48 ms) and 1.83 Hz (= 546 ms, SD = 36 ms) for the resisted case. These findings appear to eliminate any simple interpretation in which the redundant symmetric mode (which involves homologous muscles) is chosen when the capacity limit for processing information in the asymmetric mode (which involves nonhomologous muscle) is reached (Cohen, 1971).

Although resistive loading systematically reduced transition and preferred frequency, it did not alter the relationship between the two. The slopes of the functions relating transition and preferred frequency were different from zero, $F(1,3) = 84.95$, $p < .01$ for the unloaded experiments, and $F(1,3) = 25.80$, $p < .02$, for the loaded experiments. However, the slopes were not different from each other, $F(2,6) = 2.04$, $p > .10$. Moreover, the correlations between preferred and transition frequency (equivalent to normalized regression slopes) were very similar, $r = .95$ for resisted and $r = .98$ for unresisted conditions. Thus whatever the changes in mean and variance that are introduced by parametric changes in resistance, the critical behavior--manifested in the functional relation between transition and preferred frequency--remains unchanged. In fact, when the transition frequency is expressed in units of preferred frequency, the resulting dimensionless ratio is constant across all preferred frequencies whether loaded or not. Neither of the functions shown as dotted lines in Figure 2 is significantly different from zero, $F_s(1,3) = 1.71$ and 2.83, $p_s > .10$ for free and resisted cases, respectively, or from each other, $F(2,6) = 1.67$, $p > .1$. The mean "critical value" across both conditions, with and without resistive loading, is 1.313, with a coefficient of variation of .077.

AD-A136 416 STATUS REPORT ON SPEECH RESEARCH A REPORT ON THE STATUS 2/3
AND PROGRESS OF S..(U) HASKINS LABS INC NEW HAVEN CT
A M LIBERMAN 30 SEP 83 SR-74/75(1983) N00014-83-C-0083
UNCLASSIFIED F/G 6/16 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A



Solid lines. The relationship between subjects' mean preferred frequency (F_p) in the asymmetric, out-of-phase mode and the mean transition frequency (F_t) calculated over the last five consecutive cycles before the phase shift. Solid dots refer to the free, unresisted conditions, open dots to the resistive loading experiments. For the free case the least squares linear regression line was: $F_t = 1.55F_p - .48$. For the loaded case it was: $F_t = 1.02F_p + .43$. The solid and open triangles represent data from one subject who made a deliberate effort to prevent the phase transition from occurring. In this case the subject's transition frequency is about 2.5 times greater than her preferred frequency. In the unresisted case, for which estimates of mechanical energy per unit moment of inertia per cycle are most valid, she shows by far the largest energy drop across the transition of all the subjects (mean of 30.49 "energy units"/cycle before transition to 20.51 "energy units"/cycle after transition, compared to a group average of 19.38 "energy units" before and 17.33 "energy units" after). Dashed lines. The same data replotted for the subjects' preferred frequency (F_p) against the ratio of transition frequency over preferred frequency, that is, the proposed critical transition point (T_c). The least squares regression line for the free case is: $T_c = .13F_p + 1.04$, mean value of T_c is 1.284 (S.D. = .057). For the loaded case, the regression function is: $T_c = -.23F_p + 1.66$ with a mean T_c of 1.342 (S.D. = .132). Combining both experiments, the overall regression equation is: $T_c = -.09F_p + 1.46$, mean T_c of 1.313 (S.D. = .10).

Figure 2

It may not simply be chance that if a similar normalization procedure is applied to Hoyt and Taylor's (1981) locomotion data, and a ratio calculated between the horse's preferred speed in a given gait and the speed at which the transition occurs from one gait to another, a critical value of approximately 1.33 results for both walk-trot and trot-gallop transitions. As in our data, regardless of what the preferred speed is, the transition appears to occur at some constant proportion of the preferred value. Stride frequency at the trot-gallop transition in animals ranging from mice to horses has been shown to scale to total body mass (M) raised to the power of -.14 (Heglund, McMahon, & Taylor, 1974). This exponent is in close agreement to that of $M^{-1/8}$ predicted by McMahon's (1975) model of elastic similarity in which muscle stress (tension per unit cross sectional area) is hypothesized to be the same at gait transitions in homologous muscles in animals of different size. In the present experiments, when the proposed critical value (T_c) is scaled to preferred frequency (F_p) for all observations, an exponent of -.12 results ($T_c = .14F_p^{-.12}$). If further work shows preferred frequency to be tightly coupled to mass, then it may be that the elastic similarity model can be applied not only to gait transitions but also to the modal shifts observed here.

We would be remiss if we did not mention the possibility that the pattern of results observed for hand movements here (and perhaps for gait changes as well) shares common features with other critical phenomena in nature (Fleury, 1981; Haken, 1975, 1978; Iberall & Soodak, 1978; Nicolis & Prigogine, 1977; Prigogine, 1980; Riste, 1975; Soodak & Iberall, 1978). Systems at many scales of magnitude and varying widely in material properties appear to be qualitatively similar with respect to their behavior at critical points (Fleury, 1981; Haken, 1978). For example, our findings seem consistent with certain aspects of phase transition theory in physics (Kadanoff, 1971; Stanley, 1971) one of which is that parameters adjusted in an experiment may shift the critical point (as resistance does to frequency here) without altering the critical behavior itself (for examples, see Fleury's 1981 review). Moreover, a major characteristic of many physical and biological systems is that new "modes" or spatiotemporal orderings arise when the system is scaled on certain parameters to which it is sensitive (e.g., Haken, 1978; Iberall & Soodak, 1978). In the present experiments, continuous scaling on frequency resulted in the initial modal pattern becoming unstable, until, at a critical value, bifurcation occurred and a different modal pattern appeared.

The present approach, if pursued rigorously, may rationalize currently available neurophysiological accounts of transitions in coordination that assume a "switch mechanism" mediated by "coordinating fibres" (Grillner, 1982) (neither of whose neural basis is well-defined, see Selverston, 1980, and commentaries). Instead, a careful elaboration of the conditions which give rise to *switching*, may constrain possible neural explanations of the emergence of new spatiotemporal pattern.

REFERENCES

Brooks, V. B., & Thach, W. T. (1981). Cerebellar control of posture and movement. In V. B. Brooks (Ed.), *Handbook of physiology: The nervous system* (pp. 877-946). Baltimore, MD: American Physiological Society.

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

Fleury, P. A. (1981). Phase transitions, critical phenomena, and instabilities. Science, 211, 125-131.

Gallistel, C. R. (1980). The organization of action: A new synthesis. Hillsdale, NJ: Erlbaum.

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

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

Haken, H. (1978). Synergetics, an introduction: Nonequilibrium phase transitions and self-organizations in physics, chemistry and biology. Berlin: Springer-Verlag.

Hoyt, D. F., & Taylor, C. R. (1981). Gait and the energetics of locomotion in horses. Nature, 292, 239-240.

Heglund, N., McMahon, T. A., & Taylor, C. R. (1974). Scaling stride frequency and gait to animal size: Mice to horses. Science, 186, 1112-1113.

Iberall, A. S., & Soodak, H. (1978). Physical basis for complex systems: Some propositions relating levels of organization. Collective Phenomena, 3, 9-24.

Kadanoff, L. P. (1971). Critical behavior: Universality and scaling. In M. S. Green (Ed.), Proceedings of the International School of Physics (pp. 100-117). New York: Academic Press.

Keele, S. W. (1981). Behavioral analysis of movement. In V. B. Brooks (Ed.), Handbook of physiology: The nervous system. Baltimore, MD: American Physiological Society.

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

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

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

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

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

Landauer, R. (1981). Nonlinearity, multistability and fluctuations: Reviewing the reviewers. American Journal of Physiology, 241, R107-R113.

MacKay, W. A. (1980). The motor program: Back to the computer. Trends in Neurosciences, 3, 97-100.

Margaria, R. (1976). Biomechanics and energetics of muscular exercise. Oxford: Clarendon Press.

McMahon, T. A. (1975). Using body size to understand the structural design of animals: Quadrupedal locomotion. Journal of Applied Physiology, 39, 619-627.

Nicolis, G., & Prigogine, (1977). I. Self organization in non-equilibrium systems: From dissipative structures to order through fluctuations. New York: Wiley.

Pennycuick, G. J. (1975). On the running of the gnu (Connochaetes taurinus) and other animals. Journal of Experimental Biology, 63, 775-799.

Prigogine, I. (1980). From being to becoming. San Francisco, Freeman.

Riste, T. (Ed.). (1975). Fluctuations, instabilities and phase transitions. New York: Plenum.

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

Selverston, A. I. (1980). Are central pattern generators understandable? The Behavioral and Brain Sciences, 3, 535-571.

Shaffer, L. H. (1980). Analyzing piano performance: A study of concert pianists. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 443-456). Amsterdam: North-Holland.

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

Soodak, H., & Iberall, A. S. (1978). Homeokinetics: A physical science for complex systems. Science, 201, 579-582.

Stanley, H. E. (1971). Introduction to phase transitions and critical phenomena. New York: Oxford University Press.

Thelen, E., Bradshaw, G., & Ward, J. A. (1981). Spontaneous kicking in month old infants: Manifestation of human central locomotor program. Behavioral and Neural Biology, 32, 45-53.

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

TIMING AND COARTICULATION FOR ALVEOLO-PALATALS AND SEQUENCES OF ALVEOLAR +[j] IN CATALAN

Daniel Recasens+

Abstract. General articulatory characteristics and V-to-C coarticulatory effects for alveolo-palatals [ɲ], [ʎ] vs. sequences [nj], [lj] in Catalan VCV utterances have been measured at the point of maximum alveolar contact and over time by means of dynamic palatography. Data show that the amount of V-to-C coarticulation in tongue-dorsum contact varies inversely with the duration of the temporal lag between the periods of alveolar closure and palatal closure. Results support the view that coarticulation is affected by contrasting timing constraints on articulatory activity.

INTRODUCTION

Phoneticians have characterized alveolo-palatals [ɲ] and [ʎ] as "mouillé" or palatalized sounds based on a transitory perceptual effect of a [j] nature caused by the formation of a narrow dorsopalatal channel at the release (von Essen, 1957; Grammont, 1933; Jones, 1956; Sweet, 1877). Moreover, they have argued that [ɲ] and [ʎ] contrast clearly with sequences composed of alveolars [n] and [l] followed by [j], thus, [nj] and [lj]. Such differentiation has been made on the following grounds:

1) The [j] element is more auditorily salient in sequences than in alveolo-palatals for speakers of languages that contrast the two phonetic categories (Rousselot, 1912).

2) Alveolo-palatals involve more linguopalatal contact than sequences (Chlumský, 1931).

The research reported here investigates the articulatory basis for these two differentiation criteria in Catalan in the light of data on articulatory dynamics collected by means of dynamic palatography. The use of dynamic palatography represents an improvement with respect to the use of static palatography from which those criteria were derived. While dynamic palatogra-

---

+Also University of Connecticut, Storrs.

Acknowledgment. This paper is part of a doctoral dissertation submitted to the University of Connecticut. I acknowledge the assistance of Ignatius G. Mattingly, Arthur S. Abramson, and Thomas Baer for their helpful comments and suggestions. This research was supported by NICHD Grant HD01994 and NINCDS Grant NS-13617 to Haskins Laboratories, and by a fellowship from the government of Catalonia (CIRIT Commission) to the author. A shorter version of this paper was presented at the Tenth International Congress of Phonetic Sciences in Utrecht, August, 1983.

[HASKINS LABORATORIES: Status Report on Speech Research SR-74/75 (1983)]

phy allows tracking changes in linguopalatal contact over time, static palatography allows taking only one recording of linguopalatal contact for successive articulatory events. Therefore, it cannot show at what point in time during the release of alveolo-palatals vs. sequences the [j] configuration occurs, nor whether alveolo-palatals involve more palatal contact than sequences along all the dynamic stages involved in their articulation or just at a particular moment in time.

First, this study argues that the articulatory differentiation between alveolo-palatals and sequences is brought about primarily by two contrasting timing strategies: while the periods of alveolar closure and palatal closure are produced quasi-simultaneously for alveolo-palatals, a considerable temporal lag occurs between the two periods for sequences. On the other hand, contrasting degrees of linguopalatal contact during alveolar closure and during palatal closure do not help to differentiate invariably between alveolo-palatals and sequences. If this hypothesis is tenable it implies that different timing strategies are responsible for differences in linguopalatal contact in the diachronic process of palatalization that changed Latin clusters composed of alveolar plus [j] to alveolo-palatals in Romance languages: the loss of temporal lag between alveolar and palatal closures involved, presumably, an anticipatory raising of tongue body with respect to tongue tip (Haden, 1938) with consecutive widening of tongue contact from the alveolar region towards the palatal area (Bhat, 1974; Nandris, 1952).

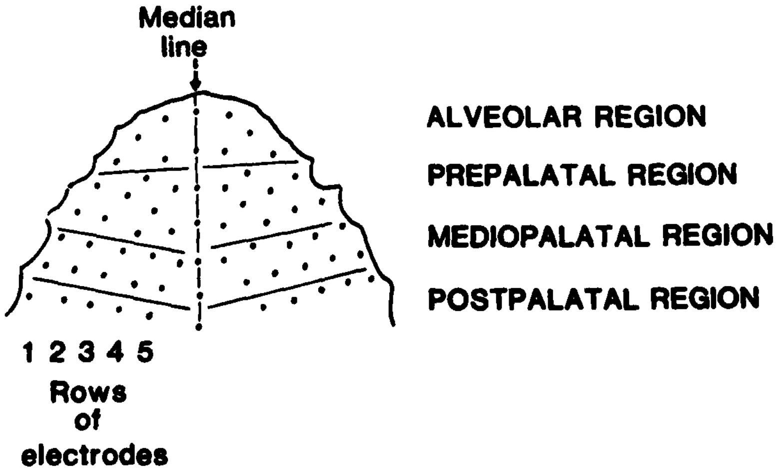


Figure 1. Electropalate.

A second purpose of this study is to show that the contrasting timing strategies for alveolo-palatals and sequences cause contrasting coarticulatory effects to occur at the period of alveolar closure in VCV utterances. In particular, the following hypothesis was tested: the amount of V-to-C coarticulation in tongue-dorsum activity during the period of alveolar closure varies inversely with the duration of the temporal lag between alveolar closure and dorsal closure. The following rationale underlies this hypothesis. Dorsopalatal [j] is, to a large extent, resistant to coarticulatory effects from the surrounding vowels (Chafcouloff, 1980; Kent & Moll, 1972; Lehiste, 1964) according to the severity of the constraints imposed upon the tongue dorsum in the constriction gesture required for the production of the consonant. On these grounds, sequences of alveolar + [j] should show smaller coarticulatory effects than alveolo-palatals since the [j] configuration has a more independent status for sequences than for alveolo-palatals.

METHOD

The artificial palate used in this study contains 63 electrodes evenly distributed over its surface and allows tracking linguopalatal dynamics over time (1 frame= 15.6 ms). Detailed information about this palatographic system (Rion Electropalatograph Model DP-01) is available in Shibata (1968) and Shibata et al. (1978).

The electrodes are arranged in five semicircular rows; for purposes of data interpretation, they have been grouped in articulatory regions and sides taking advantage of their equidistant arrangement in parallel curved rows on the artificial palate. As shown in Figure 1, the surface of the palate has been divided into four articulatory regions (alveolar, prepalatal, mediopalatal, and postpalatal) and two symmetrical sides (right and left) by a median line traced along the central range of electrodes. This division criterion established in terms of articulatory areas on the palatal surface is based on anatomical grounds (Catford, 1977).

General articulatory characteristics and coarticulatory trends were studied for utterances composed of the vowels [i], [a], [u] arranged in all possible VCV combinations for alveolo-palatals [ɲ], [ʎ] and sequences [nj], [lj]. Sequences *[Vnji], *[Vlji], which would collapse with [Vni], [Vli] since they do not occur in Catalan, were excluded. It was decided to include sequences composed of V + [n, l] + [i] (for V=[i], [a], [u]) since, as for alveolo-palatals and sequences of alveolar + [j], they show a tongue-dorsum raising gesture towards the palatal vault after the release of the alveolar closure. A speaker (the author) of Catalan (a Romance language spoken in Catalonia, Spain) with the artificial palate in place recorded all utterances with [ɲ], [ʎ], [ni] and [li] 10 times, and those with [nj] and [lj] 5 times; repetitions were averaged for data interpretation. They were embedded in a Catalan frame sentence "Sap__poc" 'He knows__just a little.'

Differences in the size of linguopalatal contact and V-to-C coarticulatory effects were analyzed at the point of maximum alveolar contact (PMCA). For alveolo-palatals, PMCA happened to be always the frame in time with the largest amount of on-electrodes all along the VCV utterance (PMC). For sequences of alveolar + [j] and sequences of alveolar + [i], two possibilities had to be accounted for:

1) PMCA coincides with PMC, as for alveolo-palatals.

2) PMCA precedes PMC. PMC occurs after the release of the alveolar closure. PMCA shows less linguopalatal contact than PMC but still the largest number of on-electrodes during the period of alveolar closure.

Temporal differences in linguopalatal dynamics between the periods of closure at the alveolar region and the palatal region were also analyzed. For this purpose, alveolo-palatals, sequences of alveolar + [j] and sequences of alveolar + [i] were lined up according to the frame that shows maximum linguopalatal contact over the surface of the palate, namely, PMC.

RESULTS

1. Point of Maximum Alveolar Contact (PMCA)

a. *Degree of linguopalatal contact*. Figure 2 shows the linguopalatal configuration at PMCA for alveolo-palatals and sequences of alveolar + [j] in symmetrical environments, except for *[inji] and *[ilji]. Sequences composed of V+[n, l]+[i] are also included for comparison. Tongue contact is represented by the area between the contour lines and the sides of the palate; the area where there is no contact is medial to the contour lines.

It can be observed that sequences of alveolar + [j] and alveolo-palatals show alveolar contact, that for sequences being more fronted (with the tongue tip) than that for alveolo-palatals (with the tongue blade). Behind the alveolar region, a larger central cavity all along the median line is found for sequences vs. alveolo-palatals, except for the postpalatal area where linguopalatal contact can be the same for the two categories (for [anja] vs. [aɲa] and for [ulju] vs. [uʎu]) or even larger for sequences than for alveolo-palatals (for [alja] vs. [aʎa]). Contact for sequences of alveolar + [j] and sequences of alveolar + [i] is highly similar at all articulatory regions: the two show a large alveolo-prepalatal cavity behind the alveolar closure and some narrowing of the constriction towards the rear of the palate due to coarticulation of tongue-dorsum activity with the following palatal articulations [j], [i]. Moreover, lateral airflow occurs through postpalatal slits at both sides of the palate for sequences [alja], [ulju] and [ali], [uli], but through a prepalatal slit on the left side for sequences with [ʎ] (only for V=[u]). (For utterances with a lateral consonant in the context [iCi], airflow takes place along a lateral channel between the teeth and the cheeks). Figure 2 also shows that the equivalent [ini], [ili] of the non-occurring sequences *[inji], *[ilji] present less contact than [iɲi], [iʎi] all over the palatal surface.

In summary, at PMCA, sequences of alveolar + [j] are produced similarly to sequences of alveolar + [i] and present less linguopalatal contact than alveolo-palatals when the whole surface of the palate is taken into consideration. However, this relation does not necessarily hold when each articulatory region is accounted for separately.

b. *Coarticulatory activity*. Figure 3 shows the linguopalatal configuration at PMCA separately for alveolo-palatals and sequences of alveolar + [j] in symmetrical VCV environments. It allows the analysis of coarticulatory

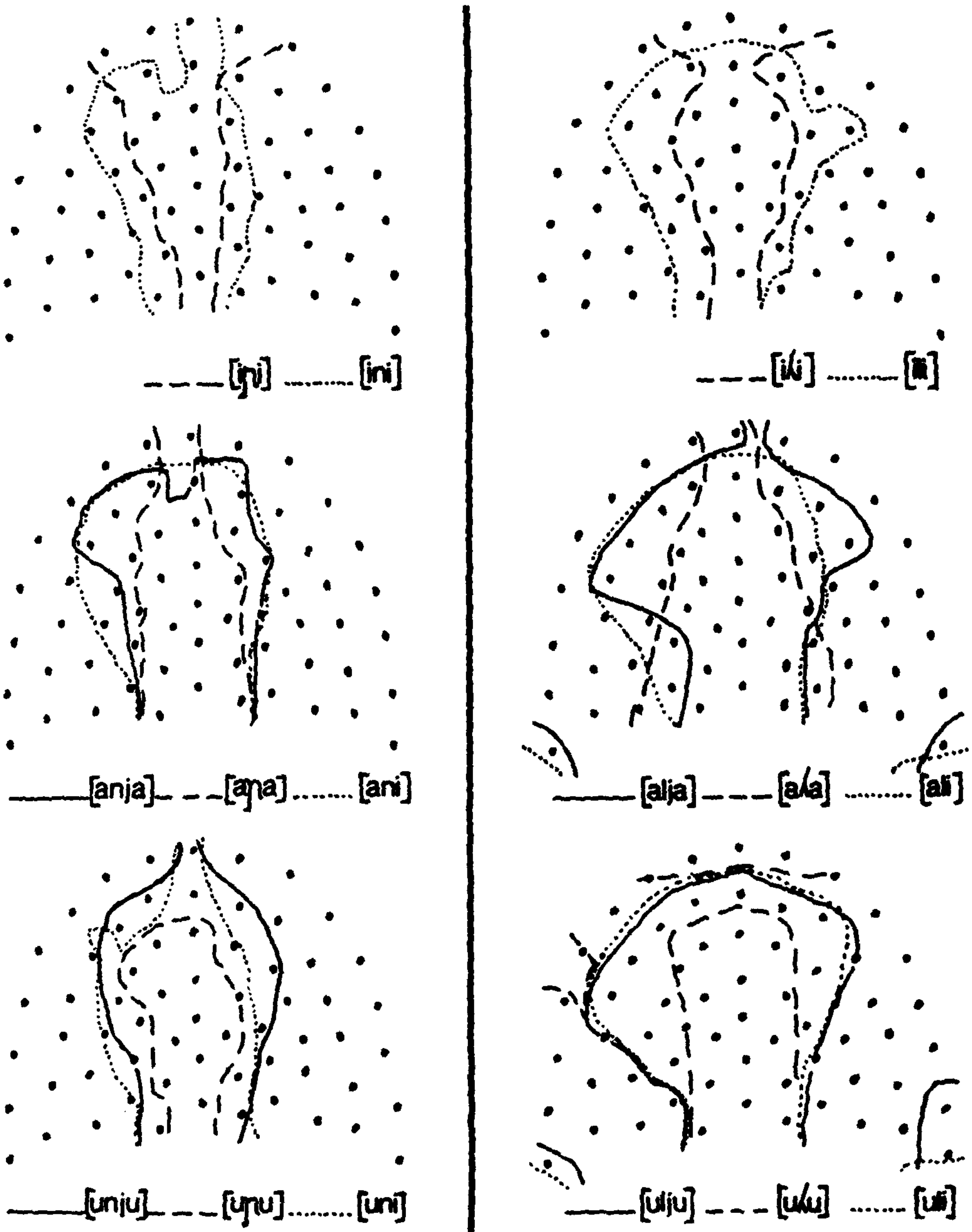


Figure 2. Linguopalatal patterns at PMCA for alveolo-palatals [ɲ], [ʎ] and sequences [nj], [lj] in symmetrical environments, and for sequences [Vni], [Vli]. They have been plotted simultaneously for comparison.

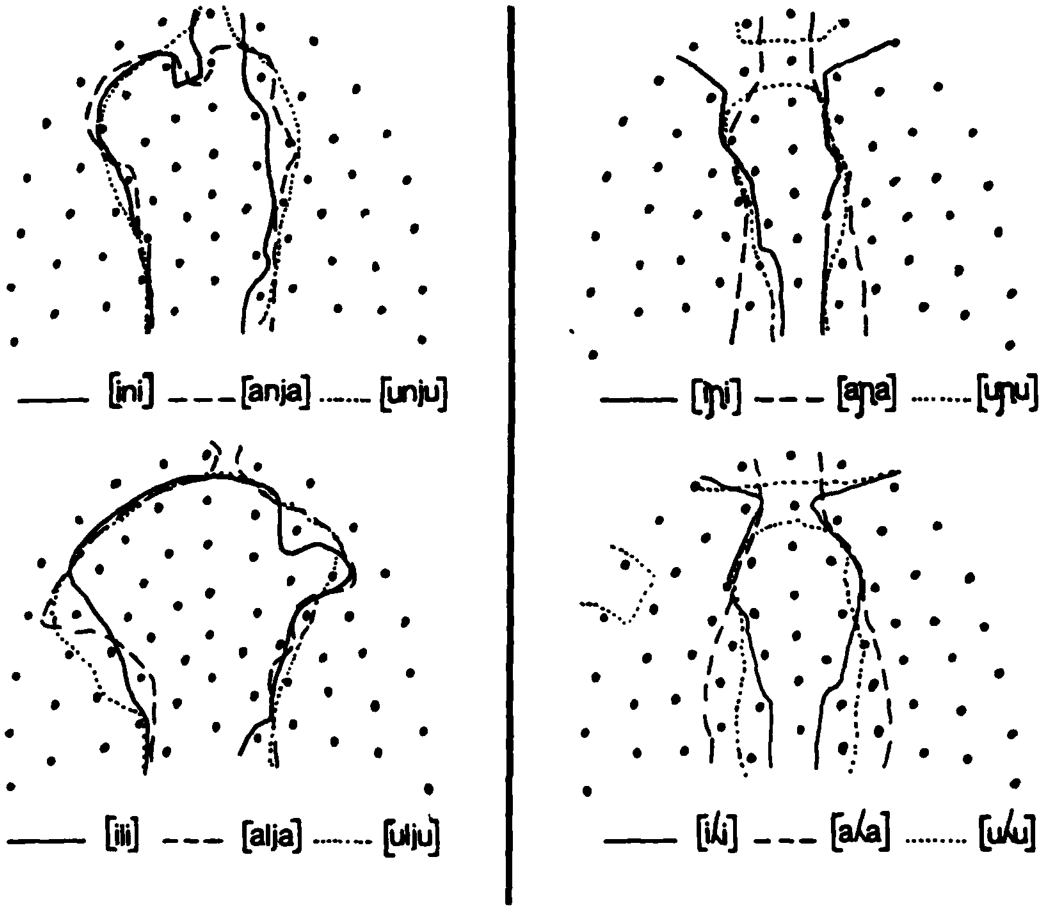


Figure 3. Coarticulatory effects at PMCA for sequences of alveolar [*n*, l] + [j, i] (left) and alveolo-palatals [ɲ], [ʎ] (right) in symmetrical environments.

trends upon tongue-dorsum activity from the surrounding vowels at PMCA when V1=V2. The non-existent symmetrical sequences *[inji], *[ilji] have been replaced by [ini], [ili].

For [ɲ], the mediopalatal and postpalatal passage shows maximal narrowing for high vowels [i] and [u], and larger opening for low vowel [a]; thus, the tongue dorsum appears to be sensitive to a large jaw opening gesture (as for [a]) but not to tongue backing dynamics (as for [u]). For [ʎ], differences in size of the mediopalatal and postpalatal passage are found for high front [i] (narrowest) and low back [a] (largest), high back [u] falling in between; thus, tongue-dorsum placement during the production of [ʎ] appears to be sensitive to degrees of tongue backing as well as jaw opening in the adjacent vowels. For sequences with [j] and [i], very small (from [i] vs. [a], [u]) effects are found from the surrounding vowels in the degree of contact at the mediopalate and postpalate.

Therefore, at PMCA, tongue-dorsum activity is far more sensitive for alveolo-palatals than for sequences to changes in articulatory configuration shown by the same surrounding vowels.

Figures 4 through 7 show coarticulatory effects at PMCA for alveolo-palatals and sequences in asymmetrical vocalic environments. Anticipatory effects (from V2) and carryover effects (from V1) in contact size at the rear of the palate are reported below.

Anticipatory effects in degree of mediopalatal and postpalatal contact are very small or non-existent for the two phonetic categories. In any case, effects for alveolo-palatals are larger than effects for sequences. Thus, as shown in Figure 4, contrasting degrees of opening are found for alveolo-palatals for V2=[a] (larger) vs. [i], [u] (smaller) mainly when V1=[a]. Effects for sequences with [j] (sequences with fixed V2=[i] show no contrasting VCV combinations for analysis of anticipatory effects) are very small and non-systematic (see Figure 5).

Carryover effects in degree of mediopalatal and postpalatal contact are found for alveolo-palatals (see Figure 6), more so for [ʎ] than for [ɲ]: for [ɲ] (left), a preceding low vowel causes less mediopalatal and postpalatal contact than a preceding high vowel; for [ʎ] (right), variability in contact size is found for V1= [i]>[u]>[a]. For the two sequence types, namely, alveolar + [j] and alveolar + [i] (see Figure 7), carryover effects in degree of contact at the rear of the palate are very small and occur for V=[i]>[a], [u].

Therefore, at PMCA, alveolo-palatals show larger anticipatory and carryover effects in degree of contact at the mediopalate and postpalate than sequences of alveolar + [j], which, in their turn, behave similarly to sequences of alveolar + [i].

2. Dynamics

Dynamic palatography allows analyzing the relative timing of alveolar closure and palatal closure and, thus, testing the hypothesis that the interval between them should be shorter for alveolo-palatals than for sequences with [j].

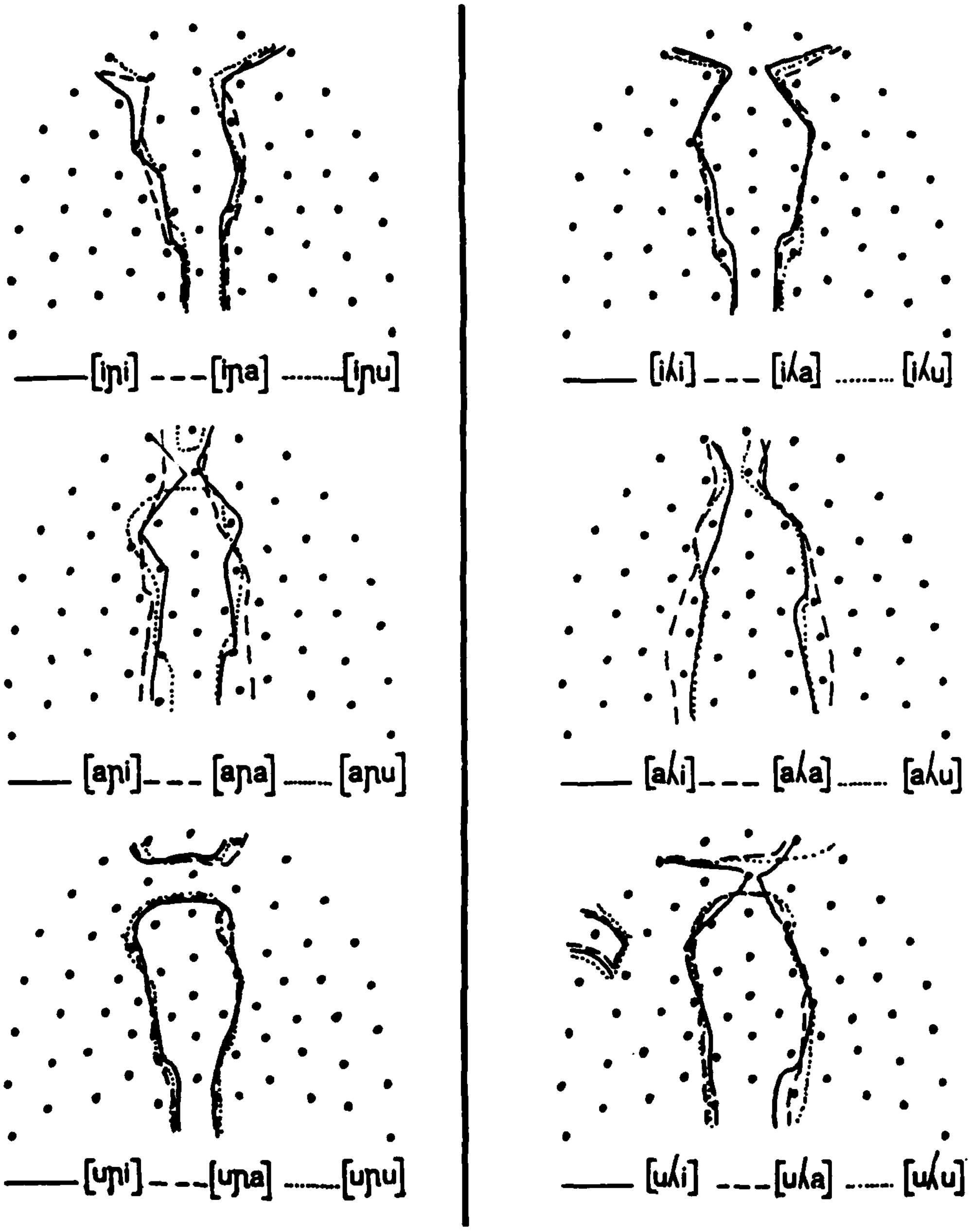


Figure 4. Anticipatory effects at PMCA for alveolo-palatals [ɲ] (left) and [ʎ] (right).

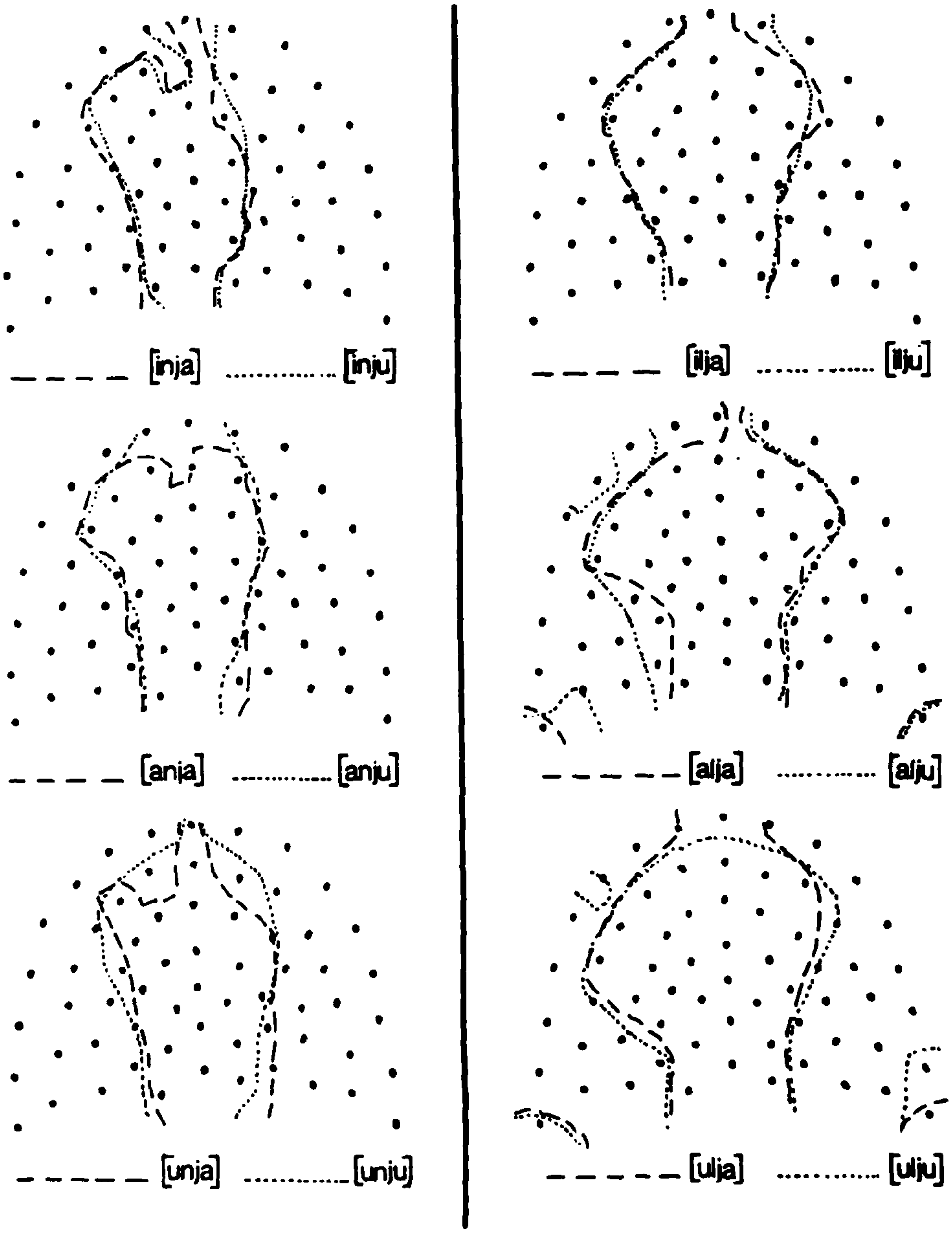


Figure 5. Anticipatory effects at PMCA for sequences [nj] (left) and [lj] (right).

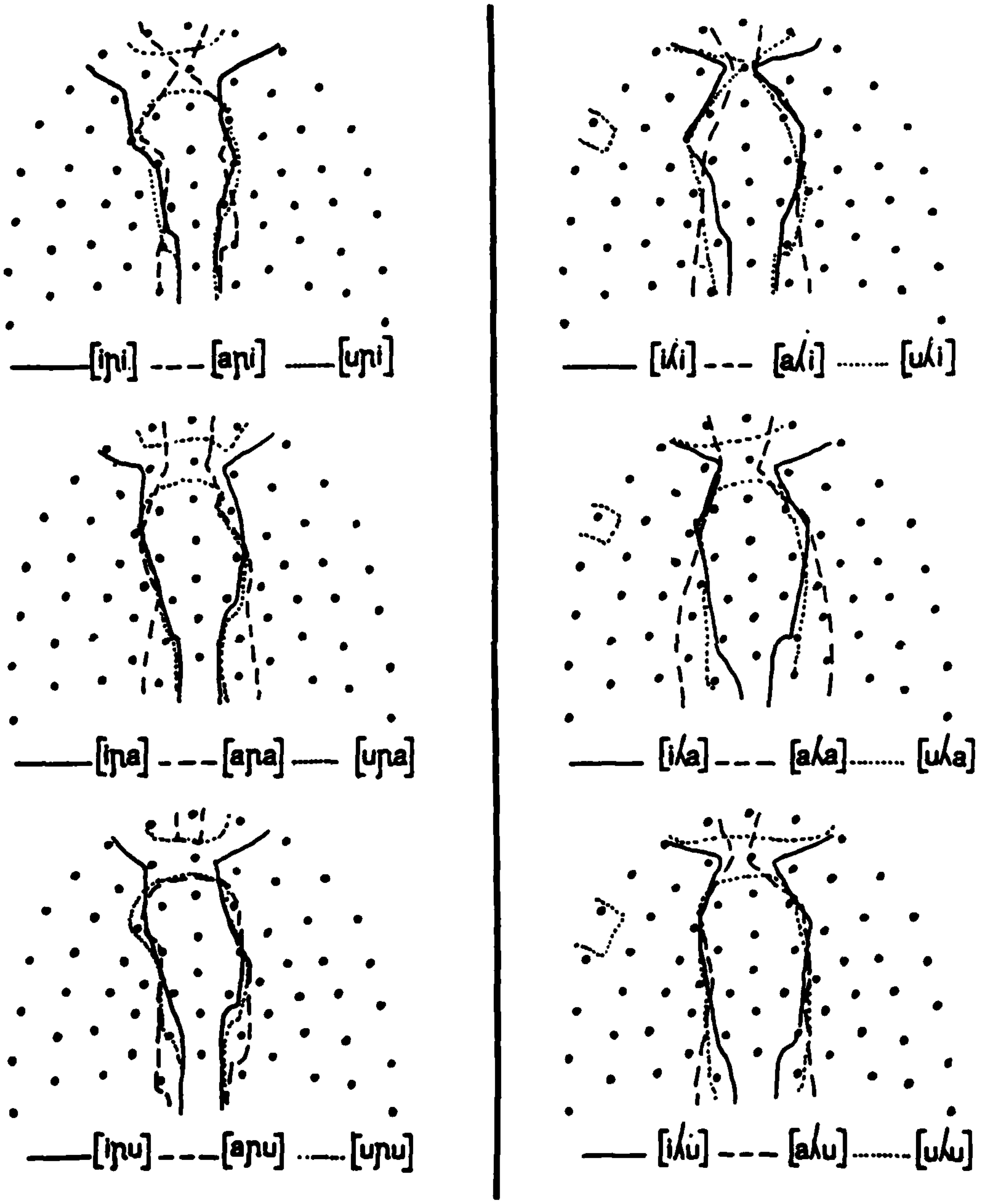


Figure 6. Carryover effects at PMCA for alveolo-palatals [ɲ] (left) and [ʎ] (right).

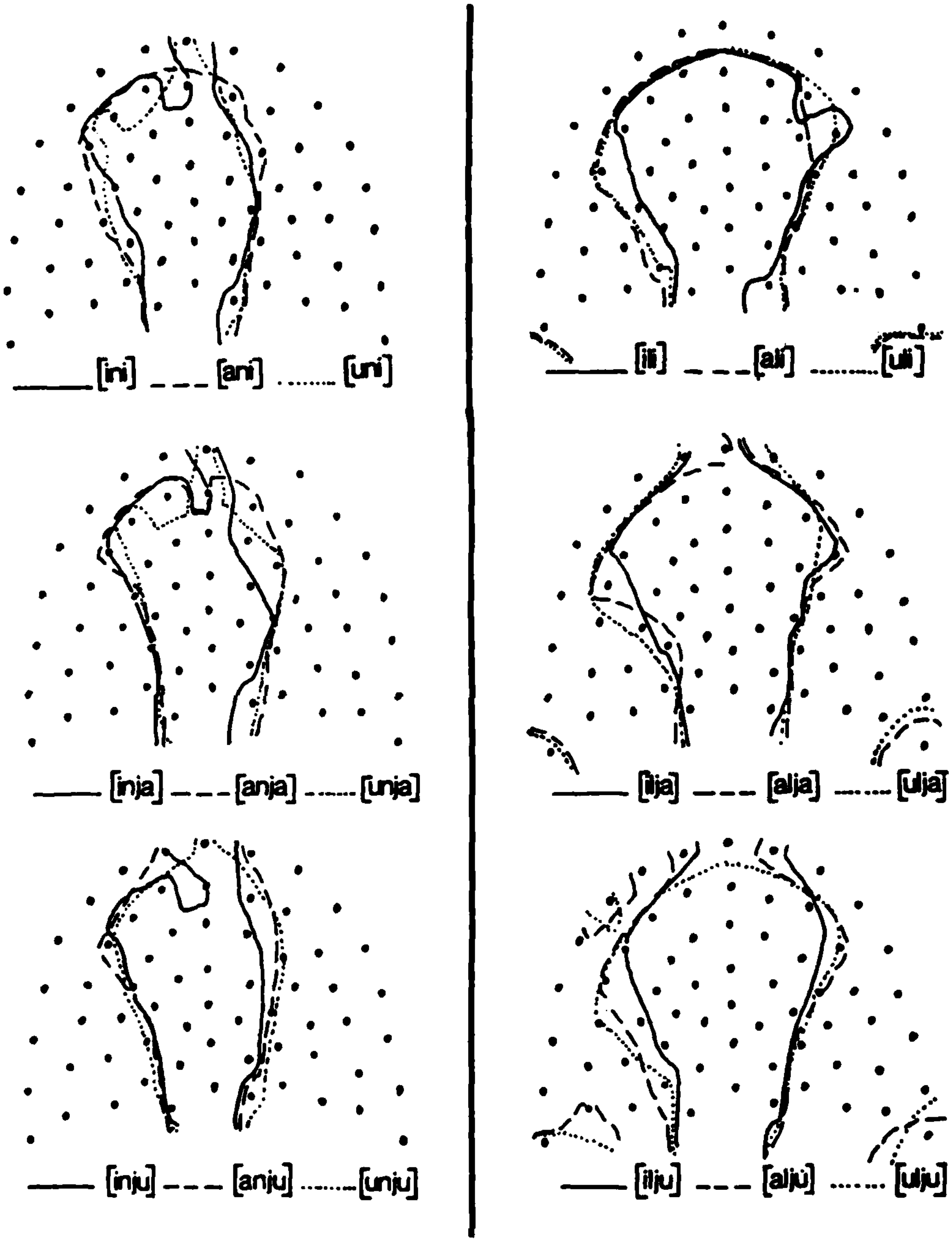


Figure 7. Carryover effects at PMCA for sequences [nj], [ni] (left) and [lj], [li] (right).

Figures 8 and 9 show linguopalatal dynamics for alveolo-palatals (above) and sequences of alveolar + [j] (middle) in symmetrical environments with V=[a], [u]. Linguopalatal dynamics for sequences [Vni] and [Vli] (below) has also been included for comparison. The line-up point is at PMC. Each panel provides data on contact for the right side of the palate at the periphery of the alveolar region (row 1) vs. the central area of the mediopalatal and postpalatal regions (row 5) over time in ms (horizontal axis). Contact data (vertical axis) are given on an electrode-by-electrode basis starting from the backmost electrode (numbered 1 in the figure) up to the frontmost one (numbered 3.5). The frontmost electrode has been counted as .5 since it is placed on the median line (see Figure 1).

For alveolo-palatals, the peak of alveolar contact (row 1) and the peak of palatal contact (row 5) are achieved simultaneously at PMC, or else, the peak of palatal closure can be achieved around 15 ms before the peak of alveolar closure (always at PMC). For sequences of alveolar + [j], the onset of maximum alveolar closure can occur between -15 and -45 ms while that of maximum palatal closure is found between 0 and +15 ms. While alveolo-palatals show no temporal lag between the two peaks, a temporal lag 15 to 45 ms long occurs for sequences ([anja] 15 ms, [unju] 30 ms, [alja] and [ulju] 45 ms). For alveolars followed by [i], the onset of maximum alveolar closure occurs between -60 and -95 ms and that of maximum palatal closure between 0 and -15 ms. A temporal lag of 60 to 95 ms occurs for sequences with [i] ([ani] and [uni] 60 ms, [ali] 80 ms and [uli] 95 ms).

Figures 8 and 9 provide information about the degree of contact at the center of the mediopalate and postpalate associated with the [j] component. Data show that the peak of tongue-dorsum activity is larger for sequences with [j] than for alveolo-palatals when laterals and nasals with [a] are taken into account; however, the opposite trend is observed for nasals with [u]. Sequences with [i], on the other hand, show a high peak of tongue-dorsum activity in all environments, analogous to or higher than that for alveolo-palatals and sequences with [j].

In summary, while alveolo-palatals show nearly simultaneous peaks of alveolar and palatal contact, sequences show a lag between the two, longer for sequences of alveolar + [i] than for sequences of alveolar + [j]. Moreover, tongue-dorsum raising activity at the release as indicated by the peak of palatal contact is greater for sequences with [i] than for sequences with [j], and generally but not always larger for sequences with [j] than for alveolo-palatals.

DISCUSSION AND CONCLUSIONS

During alveolar closure in alveolo-palatals, two commands are being actualized: tongue-blade occlusion and tongue-dorsum raising. As a result of this synergistic activity, a large degree of contact is obtained over the entire surface of the palate. During alveolar closure in sequences with [j], only one command is actualized: tongue-tip occlusion. The tongue dorsum can be said to coarticulate with [j], as shown by a progressive increase in contact towards the rear of the palatal region, analogously to sequences with [i]. The tongue blade shows contact only at the sides of the palate, thus leaving a large central cavity at the front of the palatal region. The degree

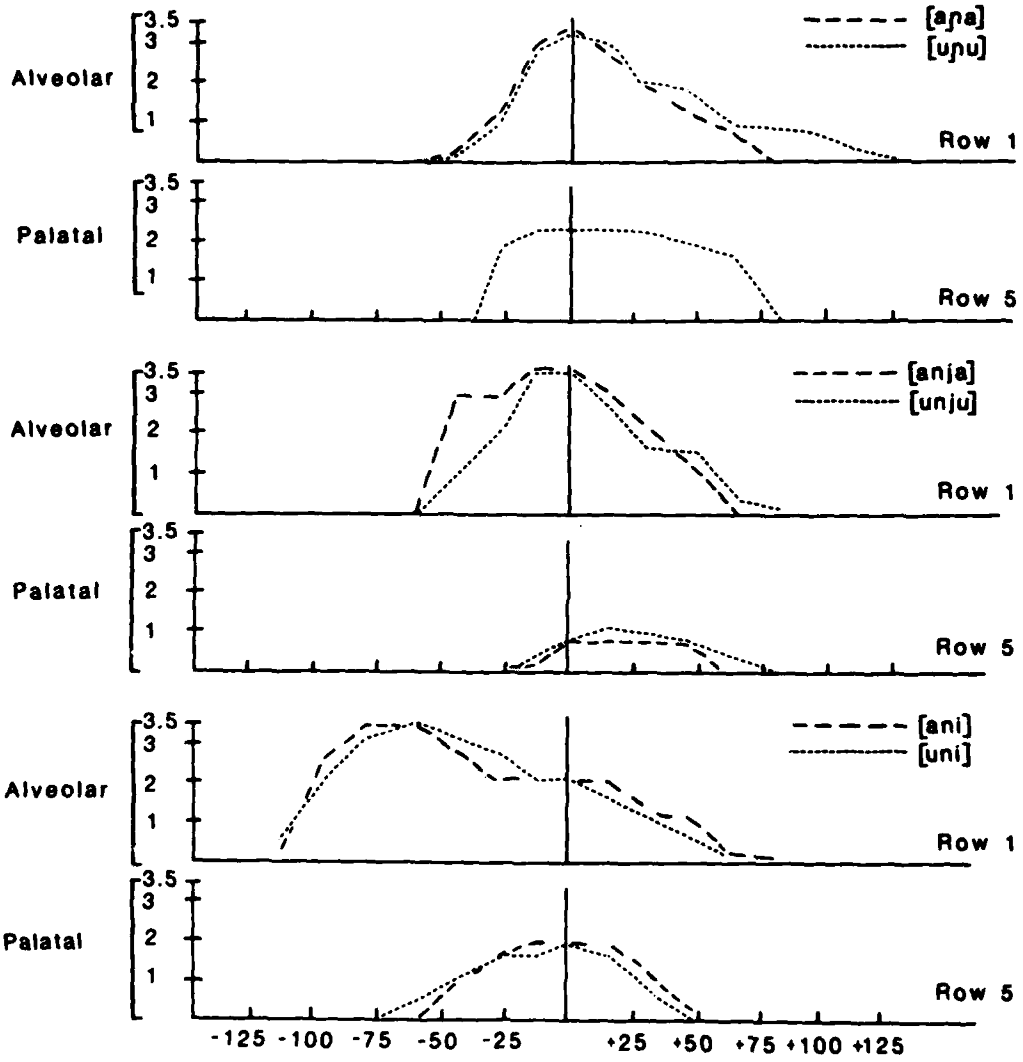


Figure 8. Linguopalatal patterns over time (ordinate: contact placement; abscissa: time in ms) for [ɲ] (top) and [nj] (middle) in symmetrical environments, and for [Vni] (bottom). The line-up point is at PMC.

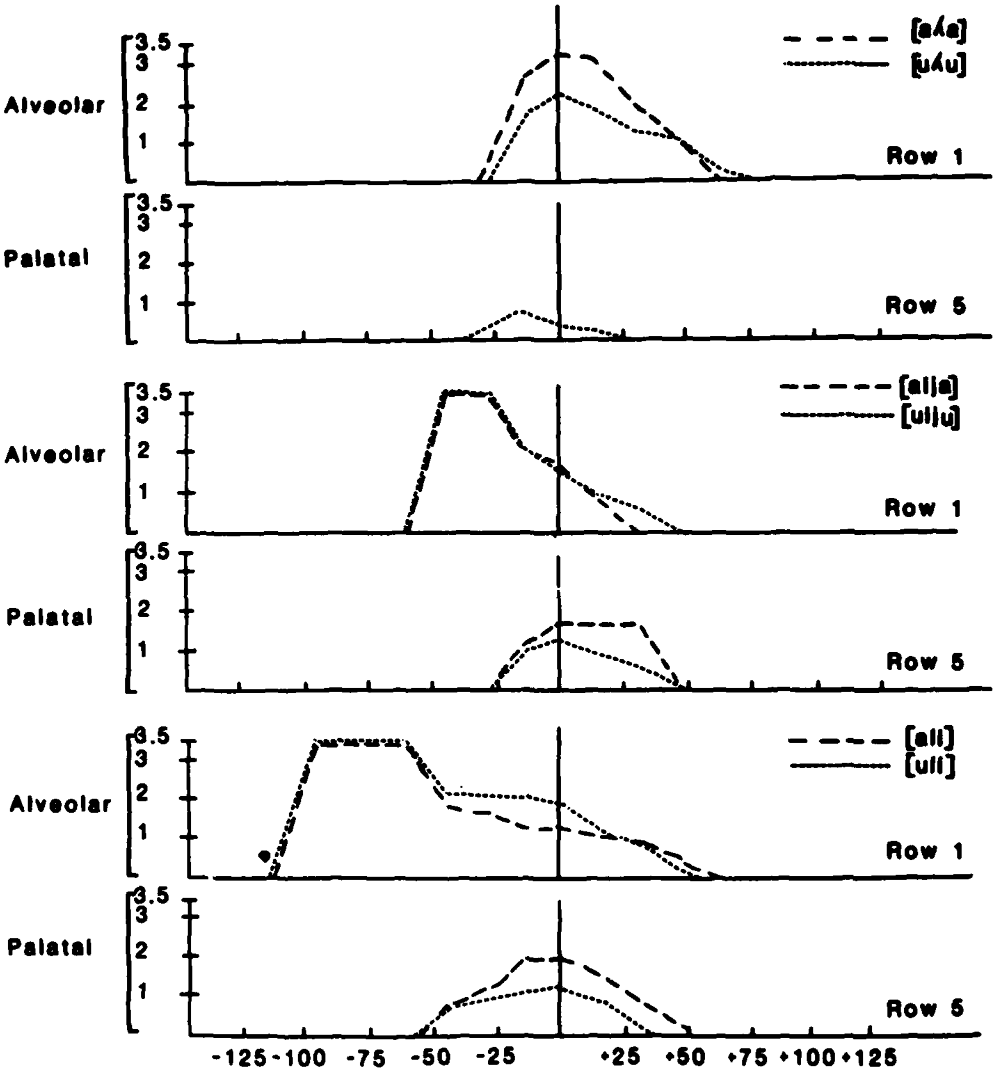


Figure 9. Linguopalatal patterns over time (ordinate: contact placement; abscissa: time in ms) for [ʎ] (top) and [lj] (middle) in symmetrical environments, and for [Vli] (bottom). The line-up point is at PMC.

of contact turns out to be invariably larger for alveolo-palatals than for sequences with [j] when the overall surface of the palate is taken into consideration, but not necessarily with respect to each articulatory region taken separately.

During palatal closure, the two consonantal types share an articulatory command for tongue-dorsum raising. For alveolo-palatals, this command is actualized together with the command for tongue-blade occlusion; for sequences with [j], it is actualized by itself at some temporal lag after alveolar closure. The degree of dorsal contact at the period of palatal closure is generally but not always larger for sequences than for alveolo-palatals. It seems that the glide component of the sequence involves less tongue-dorsum activity and less articulatory precision than expected when directed towards an articulation that involves tongue-dorsum activity as well, e.g., [u] vs. [a].

An interpretation for this set of data supports the hypothesis that presence vs. absence of a temporal lag between alveolar closure and palatal closure is an invariant constraint used by the speaker when actualizing alveolo-palatals vs. sequences with [j]. Spatial constraints in terms of degree of linguopalatal contact can be said to act as secondary articulatory traits in the task of differentiation between the two phonetic categories. On these grounds, the formation of alveolo-palatals in Romance languages from Latin sequences with [j] can be explained as a result of the loss of the temporal lag between alveolar and palatal closures and, therefore, the acquisition of a new rule of temporal constraint that generates the two simultaneously.

Coarticulation data reported in this study can be summarized as follows. Alveolo-palatals show coarticulatory effects at the point of maximum alveolar contact in symmetrical and asymmetrical vocalic environments; carryover effects are larger than anticipatory effects. Coarticulatory effects for sequences with [j] are very small and non-systematic, analogously to sequences with [i]. These contrasting coarticulatory effects can be explained with reference to the temporal constraints involved in the tongue-dorsum raising gesture during the production of alveolo-palatals vs. sequences. Thus, the palatal articulation needs to be less precise when simultaneous with alveolar contact (for alveolo-palatals vs. sequences). As a result of this contrasting articulatory mechanism, while the temporally independent [j] component in sequences blocks effects from V1 and V2, the tongue dorsum during the production of alveolo-palatals is freer to coarticulate with the surrounding vowels.

REFERENCES

Bhat, D. N. S. (1974). A general study of palatalization. *Working Papers on Language Universals*, Stanford University, *14*, 17-58.

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

Chafcouloff, M. (1980). Les caractéristiques acoustiques de [j, ɥ , w, l, r] en français. *Travaux de l'Institut de Phonétique d'Aix* (Université de Provence), *7*, 7-56.

Chlumský, J. (1931). Les sons mouillés et la théorie de l'abbé Rousselot. Actes du 1er Congrès de philologie slave, Prague, 541-547.

Essen, O. von (1957). Allgemeine und angewandte Phonetik. Berlin: Akademie-Verlag.

Grammont, M. (1933). Traité de phonétique. Paris: Delagrave.

Haden, E. F. (1938). The physiology of French consonant changes. Supplement to Language, 14.

Jones, D. (1956). An outline of English phonetics (8th ed.). Cambridge: W. Heffer and Sons.

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

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

Nandris, O. (1952). Les palatalisations romanes. Orbis, I, 136-145.

Rousselot, A. P. (1912). Dictionnaire de la prononciation française. Revue de Phonétique, II, 260-285.

Shibata, S. (1968). A study of dynamic palatography. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo), 2, 28-36.

Shibata, S., Ino, A., Yamashita, S., Hiki, S., Kiritani, S., & Sawashima, M. (1978). A new portable type unit for electropalatography. Annual Bulletin of the Research Institute of Logopedics & Phoniatrics (University of Tokyo), 12, 5-10.

Sweet, H. (1877). A handbook of phonetics. Oxford: Clarendon press series.

V-TO-C COARTICULATION IN CATALAN VCV SEQUENCES: AN ARTICULATORY AND ACOUSTICAL STUDY

Daniel Recasens+

Abstract. Electropalatographic and acoustical data on VCV sequences for Catalan consonants involving contrasting degrees of tongue-dorsum contact ([j], [ɲ], [ʎ], [n]) show that the degree of V-to-C coarticulation varies monotonically and inversely with the degree of tongue-dorsum contact and the size of the back cavity behind the place of constriction. This finding suggests that, to a large extent, coarticulation is regulated by mechanical constraints on articulatory activity. Evidence for larger carryover than anticipatory V-to-C effects is also presented.

INTRODUCTION

Little progress has been achieved in characterizing the programming of articulatory gestures used by the speaker to actualize phonetic segments in running speech (Harris, 1977). Thus, a large body of experimental evidence supports the view that no one-to-one mapping relationship is to be found between underlying phonemic units and articulatory gestures. Instead, the articulatory manifestations of phonetic segments can be said to coarticulate in running speech in the sense that articulatory gestures are inherently context-sensitive and overlap over time. Therefore, articulatory invariance is to be sought in the process of articulatory dynamics itself. Accordingly, the underlying units that control such a process can best be characterized in terms of dynamic gestures (see Fowler, Rubin, Remez, & Turvey, 1980) rather than in terms of static articulatory targets correlated with linguistic units such as phonemes or phonemic features.

A plausible view about how the production process is organized around patterns of articulatory dynamics is that taken by some researchers at Haskins Laboratories. According to Fowler (1980) and Fowler et al. (1980), this

---

+Also University of Connecticut.

Acknowledgment. I wish to thank the members of my advisory committee, Arthur S. Abramson, Ignatius G. Mattingly, Alvin M. Liberman, and Katherine S. Harris, for their valuable help during all stages of this research. Thomas Baer made very insightful comments on previous versions of the manuscript. The research was supported by NICHD Grant HD01994 and NINCDS Grant NS-13617 to Haskins Laboratories, and by a fellowship from the government of Catalonia (CIRIT Commission) to the author. An earlier version of this paper was presented at the 104th Meeting of the Acoustical Society of America in Orlando, November, 1982 (Journal of the Acoustical Society of America, 1982, 72, S101).

[HASKINS LABORATORIES: Status Report on Speech Research SR-74/75 (1983)]

process is executed by means of coordinative structures, namely, muscle groupings organized functionally to actualize linguistic units in fluent speech. The constraints on articulatory movement allowed by the coordinative structure specify those articulatory dimensions along which context-adjustment may take place. Thus, in the light of this approach, coarticulatory activity ought to be predictable from constraints on articulatory displacement.

To investigate the regularities underlying the process of coarticulation, the effect of surrounding vowels upon tongue-dorsum contact during the production of palatal and alveolar consonants was analyzed in this study. The prediction that the degree of vowel-to-consonant coarticulation varies monotonically and inversely with the degree of tongue-dorsum contact was tested. Consistent with Fowler et al. (1980), for consonants produced with contrasting degrees of constraint on the tongue dorsum to make dorsopalatal contact, more tongue-dorsum contact ought to produce less coarticulatory activity and less tongue-dorsum contact larger coarticulatory effects.

There is evidence from the literature that coarticulation on tongue-dorsum activity and degree of tongue-dorsum constriction are inversely related. In the articulatory domain, data on alveolar stops for Swedish and English (Öhman, 1966), on English alveolar fricatives (Carney & Moll, 1971), and on German alveolar stops (Butcher & Weiher, 1976) show that, during the production of these consonants, the tongue dorsum coarticulates with the surrounding vocalic environment and produces transconsonantal effects. In the acoustical domain, large effects from surrounding vowels on alveolar [l] are documented in different languages: English (Bladon & Al-Bamerni, 1976; Lehiste, 1964) Italian (Bladon & Carbonaro, 1978), French (Chafcouloff, 1980).

Data for palatal consonants show less coarticulation. Kent and Moll (1972) found no tongue-dorsum effects from the surrounding vowels during closure of English [j] in VCV sequences. Lehiste (1964) for English and Chafcouloff (1980) for French report small F2 effects from V to C in the case of [j]. According to Stevens and House (1964), the spread of F2 values at the boundaries of the vocalic portions of English VCV sequences is smaller for palatals than for consonants articulated further front in the mouth. Analogously, Bladon and Carbonaro (1978) found little or no acoustic evidence of V-to-C coarticulation for the Italian palatal [ʎ] in VCV sequences.

A comparison of coarticulatory trends for both consonantal sets according to data from the literature summarized above shows that V-to-C effects are larger for alveolars than for palatals. Such a difference is associated with contrasting strategies of tongue-dorsum activity as follows: in the case of alveolars, the tongue dorsum is left free to coarticulate with surrounding vowels; for palatals, it appears to be directly involved in the constriction gesture, thus blocking possible coarticulatory effects to a large extent.

To my knowledge the prediction that degree of tongue-dorsum contact and degree of coarticulation can be related monotonically has not been systematically investigated before. In order to test the prediction, V-to-C coarticulatory trends for palatal and alveolar consonants that involve different degrees of tongue-dorsum contact were studied here. Consonants [j], [ɲ], [ʎ] and [n] in Catalan (a Romance language spoken in Catalonia, Spain) have been chosen for this purpose. Contrasting degrees of tongue-dorsum contact are associated with these consonants for [j]>[ɲ]>[ʎ]>[n], both as traditionally described and according to a survey of palatographic recordings from the literature across different Romance languages and contextual conditions (e.g., Haden, 1938; Rousselot, 1924-1925) I performed for the present study. Thus, [j] can be characterized as a dorsopalatal approximant, leaving a narrow passage along the palatal median line; [ɲ] and [ʎ] appear to be alveolo-palatal stops produced with large linguopalatal contact over the surface of the palate with the tongue blade and the tongue dorsum (less so than for [j], and more so for [ɲ] than for [ʎ]); [n] is an alveolar consonant produced with tongue-tip occlusion and no contact with the tongue dorsum at the center of the palate.

In summary, it appears that [j], [ɲ], [ʎ], and [n] involve decreasing degrees of tongue-dorsum contact. In a language with alveolars and palatals contrasting in tongue-dorsum contact, [j], [ɲ], [ʎ] and [n] ought to show increasing degrees of V-to-C coarticulation.

METHOD

I. Articulatory Analysis

Electropalatographic (EPG) data were collected for Catalan consonants [j], [ɲ], [ʎ], and [n] in all possible VCV combinations with V= [i], [a], [u]. The utterances were embedded in a Catalan frame sentence "Sap __ poc," 'He knows __ just a little.' A single speaker of Catalan (speaker Re, the author), also fluent in Spanish, English, and French, repeated all utterances 10 times with the artificial palate in place while the electropalatographic signal and the corresponding acoustic signal were recorded on tape for later analysis.

The artificial palate used in this sudy contains 63 electrodes evenly distributed over its surface and permits tracking linguopalatal contact patterns over time (1 frame= 15.6 ms). Detailed information about this palatographic system (Rion Electropalatograph Model DP-01) is available in Shibata (1968) and Shibata et al. (1978). The electrodes are arranged in five semicircular rows; for purposes of data interpretation, they have been grouped in articulatory regions and sides taking advantage of their equidistant arrangement in parallel curved rows on the artificial palate. As shown in Figure 1, the surface of the palate has been divided into four articulatory regions (alveolar, prepalatal, mediopalatal, and postpalatal) and into two symmetrical sides (right and left) by a median line traced along the central range of electrodes. This division in terms of areas on the palatal surface is based on anatomical grounds (Catford, 1977).

For each VCV utterance, contact data were tabulated at the frame that presented the highest number of on-electrodes (that is, point of maximum contact or PMC) and averaged across repetitions for interpretation.

II. Acoustical Analysis

Four repetitions of all VCV combinations from this and two other Catalan speakers (Bo and Ca), also fluent in Spanish, were recorded for acoustical

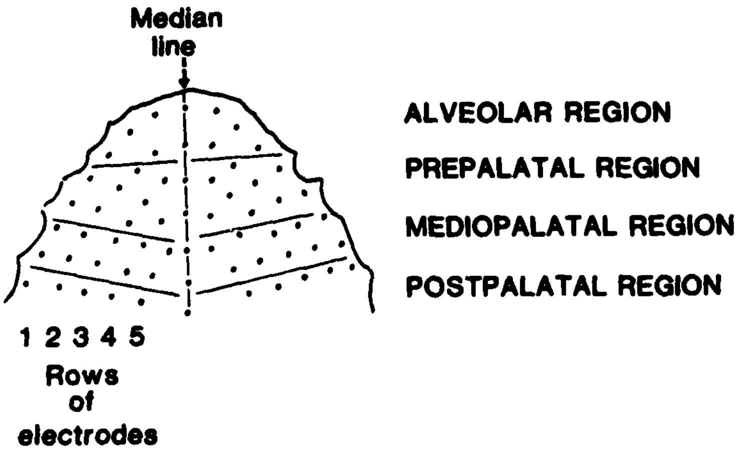


Figure 1. Electropalate.

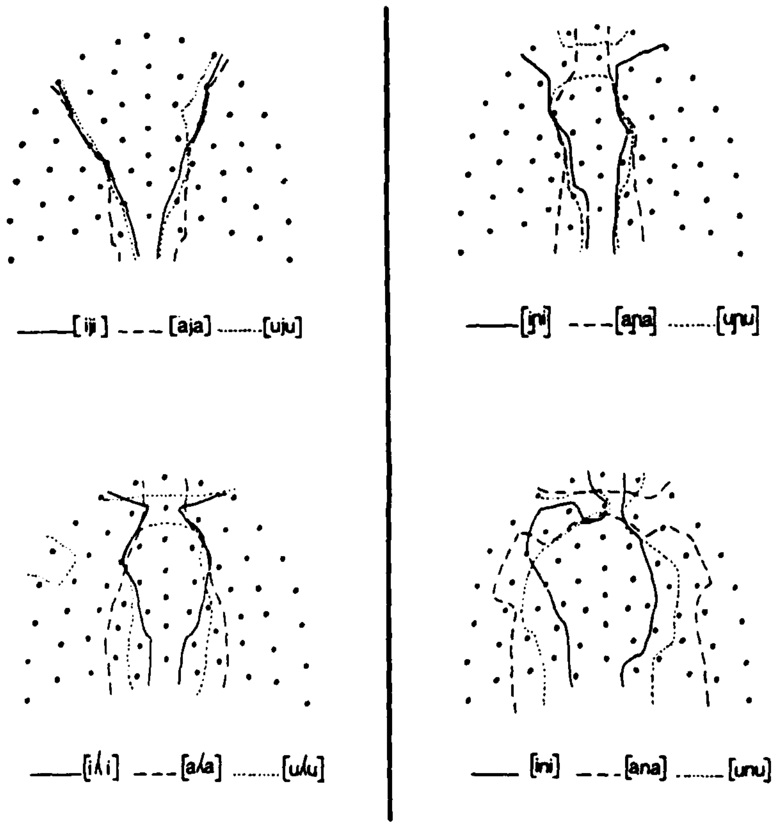


Figure 2. Linguopalatal configuration for [j], [ɲ], [ʎ] and [n] at PMC in symmetrical environments (speaker Re).

analysis. They were digitized at a sampling rate of 10 kHz, after preemphasis and low-pass filtering. An LPC (linear prediction coding) program included in the ILS (Interactive Laboratory System) package was used to measure the frequencies of the three lowest spectral peaks at PMC. To identify PMC on the acoustic wave for speaker Re, EPG data were also digitized at a sampling rate of 20 kHz, with no previous preemphasis or filtering. Labeling procedures were executed using WENDY (Haskins Laboratories Wave Editing and Display system). For speakers Bo and Ca, for whom no EPG data were available, PMC was estimated by visually identifying the F1 frequency minimum in the transition from the first vowel to the consonant. Such a point was found to match PMC satisfactorily for speaker Re. Acoustical data were averaged across repetitions for interpretation.

The prediction that degree of coarticulation varies along with changes in degree of tongue-dorsum contact will be studied according to the following procedure. For each consonant, I will present articulatory and acoustical data at PMC on general production characteristics in symmetrical VCV environments and V-to-C coarticulatory effects in symmetrical and asymmetrical VCV environments. In all cases I will concentrate exclusively on patterns of contact at the rear of the palate (mediopalate and postpalate) that reflect tongue-dorsum activity. In the acoustic domain, only data on F2 frequencies will be presented, given the affiliation between this formant with differences in back cavity size and in degree of palatal constriction for palatal and alveolar consonants (Fant, 1960).

RESULTS

I. Consonant [j]

In Figure 2, tongue contact is represented by the area between the contour lines and the sides of the palate; the area where there is no contact is medial to the contour lines. According to the figure, the dorsopalatal approximant [j] is produced with a dorsal constriction along the entire mediopalatal and postpalatal regions except for a narrow passage along the median line, and lowered tongue tip and tongue blade. High F2 values for [j] (1925-2425 Hz, according to Table 1) are dependent upon half-wavelength of the combined mouth-pharynx system behind the constriction; the small range of F2 variation (500 Hz) denotes a highly fixed and well-defined back cavity configuration independent of speaker and vocalic environment.

Figure 2 shows coarticulatory effects in symmetrical vocalic environments. They affect the width of the central passage in the mediopalatal and postpalatal areas, with analogous maximal narrowing for high vowels [i] and [u], and more opening for low vowel [a]. As shown in Table 1, observed F2 values for [j] vary in direct relationship to the degree of palatal constriction. Thus, they are found to be high for high vowels [i] and [u] (2050 to 2425 Hz) and low for low vowel [a] (1925 to 2150 Hz).

Table 1

F2 values in Hz for [j], [ɲ], [ʎ], and [n] at PMC in symmetrical environments (speakers Re, Bo, and Ca).

	[iji]	[aja]	[uju]	[iɲi]	[aɲa]	[uɲu]
Re	2350	1925	2200	2350	1775	2150
Bo	2425	2150	2425	2425	2000	2425
Ca	2150	1925	2050	2150	1575	2250

	[iʎi]	[aʎa]	[uʎu]	[ini]	[ana]	[unu]
Re	2275	1600	1900	2210	1570	1075
Bo	2400	2000	1850	2350	1675	1150
Ca	2000	1600	1750	2075	1350	1100

Figure 3 shows coarticulatory effects in asymmetrical vocalic environments. Anticipatory effects from V2 (shown on the left) and carryover effects from V1 (shown on the right) have been measured when the transconsonantal vowel is kept constant. It can be seen that patterns resulting from carryover and anticipatory effects are almost the same as for the symmetrical high-vowel environment: the effect of a high vowel is found to override that of a low vowel systematically, thus causing maximal degree of constriction at the mediopalate and postpalate, independent of coarticulatory direction.

Acoustical data on anticipatory and carryover effects are presented in Table 2. F2 values have been averaged across VCV contexts for each V2 (anticipatory effects) and V1 (carryover effects) for each speaker. Cross-vocalic ranges have also been included. In contrast to the EPG data, the acoustical data in Table 2 show larger carryover effects (from V1=[i], [u]>[a])[1] than anticipatory effects (from V2=[i]>[u]⩾[a]) for all speakers. Thus, the range of F2 values across contrasting V2 is lower (40, 105, and 110 Hz for different speakers) than that across contrasting V1 (100, 210, and 315 Hz for different speakers).

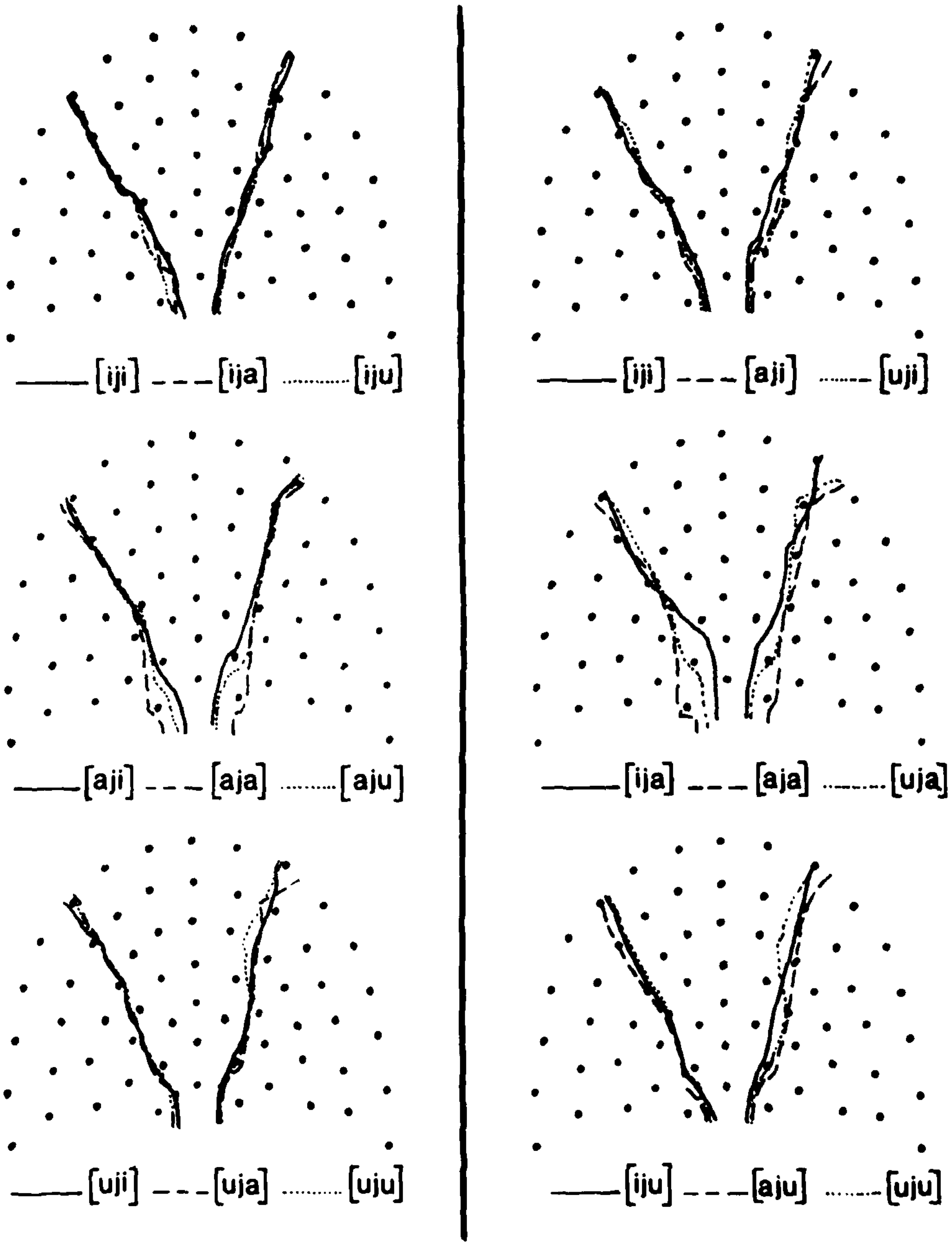


Figure 3. Anticipatory (left) and carryover (right) effects for [j] at PMC (EPG data; speaker Re).

Table 2

Anticipatory and carryover effects for [j], [ɲ], [ʎ], and [n] (F2 values in Hz; speakers Re, Bo, and Ca) over all VCV contexts for each V2 (anticipatory effects) and V1 (carryover effects).

		[j]		[ɲ]	
		Anticipatory	Carryover	Anticipatory	Carryover
	[i]	2240	2300	2185	2300
Re	[u]	2135	2225	2115	2210
	[a]	2135	1985	2085	1875
Range		105	315	100	425
	[i]	2375	2425	2335	2425
Bo	[u]	2355	2425	2300	2425
	[a]	2335	2215	2285	2065
Range		40	210	50	360
	[i]	2075	2025	2050	2085
Ca	[u]	1990	2085	2015	2065
	[a]	1965	1925	1955	1690
Range		110	100	95	395

		[ʎ]		[n]	
		Anticipatory	Carryover	Anticipatory	Carryover
	[i]	2005	2190	1655	2230
Re	[u]	1885	1935	1595	1075
	[a]	1885	1650	1640	1585
Range		120	540	60	1155
	[i]	2250	2275	1875	2260
Bo	[u]	2000	2090	1635	1285
	[a]	2140	2050	1765	1735
Range		250	225	240	975
	[i]	1885	2000	1665	2000
Ca	[u]	1835	1800	1485	1215
	[a]	1785	1700	1540	1475
Range		100	300	180	785

II. Consonant [ɲ]

The alveolo-palatal nasal [ɲ] is produced with contact all over the surface of the palate with tongue blade and tongue dorsum, except for a narrow passage along the median line (see Figure 2). At the postpalate, this passage shows equal or less (never more) contact than for [j]. F2 for [ɲ] is pharynx-cavity dependent. As shown in Table 1, the range of F2 values for [ɲ] (850 Hz) is larger and the values can be lower (1575-2425 Hz) than for [j]. This is essentially due to the fact that the postpalatal passage can show more variability and can be larger in degree of opening for [ɲ] than for [j].

Coarticulatory trends in symmetrical environments (see Figure 2) show, just as for [j], maximal narrowing of the passage at the rear of the palate for high vowels [i], [u], and larger opening for low vowel [a]. Differences in degree of postpalatal contact are larger (for [a] vs. [i], [u]) than for [j]. As shown in Table 1, F2 values for [ɲ] vary in direct relationship to the degree of palatal contact, as for [j]. Thus, they are found to be high for high vowels [i] and [u] (2150 to 2425 Hz) and low for low vowel [a] (1575 to 2000 Hz). Lower values for [a] with [ɲ] than with [j] accord well with the fact that [aɲa] shows less dorsopalatal contact at the postpalate than [aja].

Anticipatory (left) and carryover (right) effects with respect to degree of the constriction at the rear of the palate are shown in Figure 4. Carryover trends occur systematically, i.e., a preceding low vowel causes a wider passage than a preceding high vowel, independent of V2. Anticipatory trends from V2 are overriden by V1; thus, the passage width is always more open for V1=[a] than for V1=[i], [u], independent of V2. Similarly, acoustical data (see Table 2) show larger carryover effects (more so than for [j], from V1=[i]>[u]>[a]) than anticipatory effects (as for [j], from V2=[i]>[u]>[a]) for all speakers. The fact that [ɲ] shows similar anticipatory effects and larger carryover effects than [j] at the articulatory and acoustical levels results from the smaller degree of tongue-dorsum contact.

III. Consonant [ʎ]

The alveolo-palatal [ʎ] is produced with contact all over the palatal surface with tongue blade and tongue dorsum, except for a narrow passage along the median line that is larger than that for [ɲ] (see Figure 2). Therefore, [ʎ] involves a smaller degree of tongue-dorsum contact than [j] and [ɲ]. As a lateral consonant, [ʎ] is articulated so that the airstream passes out at the sides of the vocal tract. The absence of lateral slits for some utterances and the presence of only a prepalatal slit on the left side of the palate for others, suggests that the airstream passes mainly through a channel formed by the external surface of the teeth and the inner walls of the cheek.

F2 for [ʎ] shows essentially the same cavity affiliation as for [j]. According to Table 1, there is a larger range of F2 variation (800 Hz) for [ʎ] than for [j]. The result is consistent with a more variable back cavity configuration. On the other hand, F2 values can be lower (1600-2400 Hz), in accordance with a larger back cavity behind the constriction.

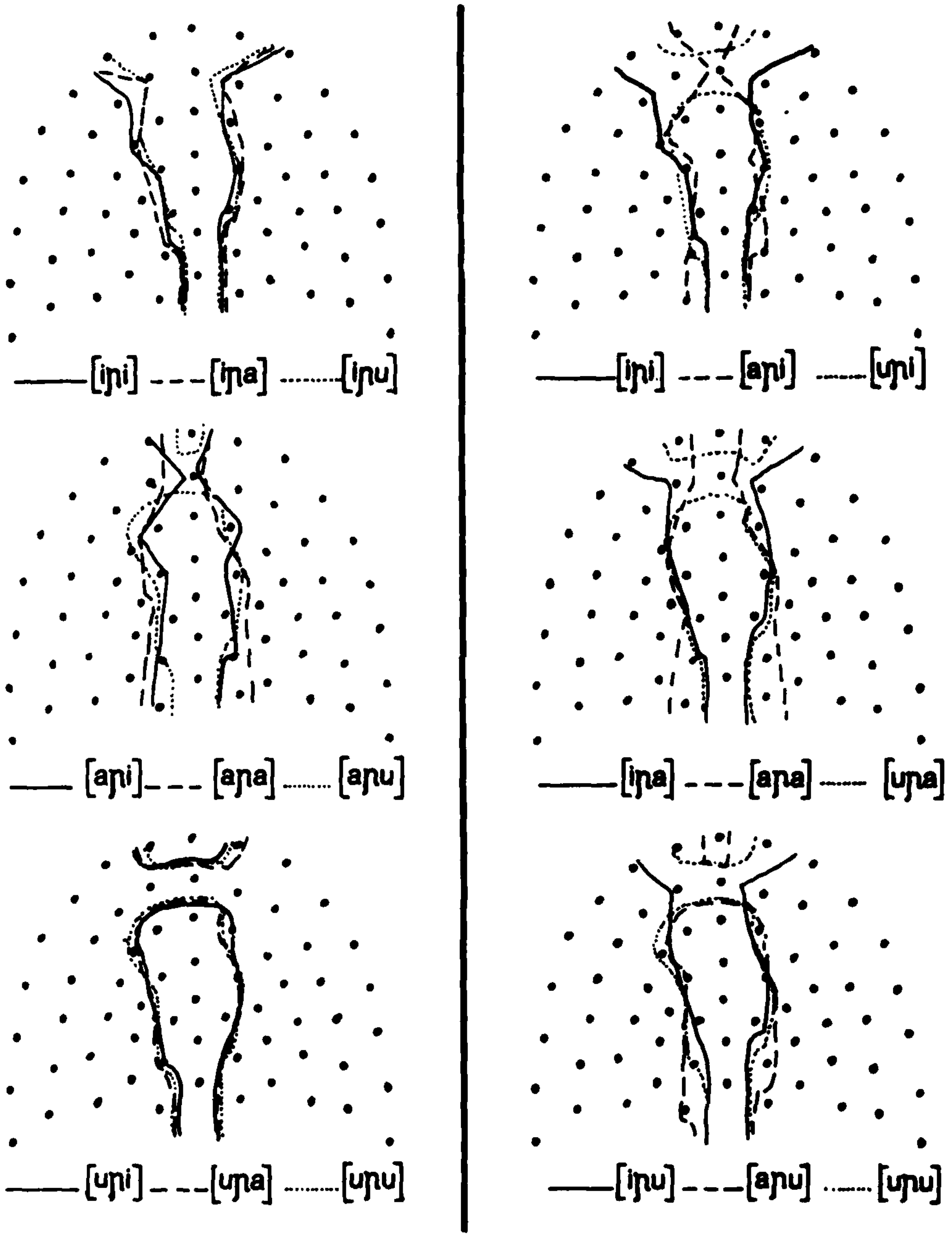


Figure 4. Anticipatory (left) and carryover (right) effects for [ɲ] at PMC (EPG data; speaker Re).

Coarticulatory trends in symmetrical environments (see Figure 2) show differences in the size of the palatal passage for high front [i] (narrowest) and low back [a] (widest), high back [u] falling in between. This pattern differs from that for [j] and [ɲ], which show no contrast between [i] and [u]. Thus, the tongue-dorsum placement during the production of [ʎ] vs. [j], [ɲ] appears to be sensitive to degrees of tongue backing as well as jaw opening in the adjacent vowels. Consistently, contrasting cross-speaker effects on F2 are found according to differences in degree of dorsal contact for [i] (2000-2400 Hz)> [u] (1750-1900 Hz)> [a] (1600-2000 Hz) (see Table 1).

Carryover effects are larger than anticipatory effects (see Figure 5). They are also larger than for [j] and [ɲ] in showing contrasting degrees of contact for V1=[i]>[u]>[a]; anticipatory effects are small or non-existent and conform always to the degree of mediopalatal and postpalatal opening appropriate for V1. Larger carryover than anticipatory effects are also observed for the articulatory traits that characterize laterality. Thus, a lateral prepalatal slit on the left side of the palate is always found when V1=[u] and is absent when V1=[i], [a], while no anticipatory effects are found in this respect.

Acoustical data (see Table 2) for F2 frequencies also show larger carryover than anticipatory effects for all speakers. Carryover trends are observed mainly from V1=[i]>[u]>[a] and anticipatory effects mainly from V2=[i]>[u], [a]. Ranges of F2 values show that anticipatory effects for [ʎ] are larger than for [j] and [ɲ] (for speakers Re and Bo but not for speaker Ca), and that carryover effects are larger than for [j] and can be larger or smaller than for [ɲ].

IV. Consonant [n]

The consonant [n] is produced with apico-alveolar constriction and complete contact all along the sides of the palate, thus leaving a large central cavity along the median line (see Figure 2). The cavity is much larger than that for palatal consonants, thus indicating a smaller degree of tongue-dorsum contact. F2 for [n] is dependent upon the pharynx cavity, as for [ɲ]. According to Table 1, it is lower (1075-2350 Hz) and shows more variability (1275 Hz) than for [ɲ], thus indicating larger pharynx-cavity size and higher degree of tongue-body adaptability to the vocalic environment.

Coarticulatory effects in symmetrical environments (see Figure 2) in degree of contact at the rear of the palate are found for [i]>[u]>[a]. The passage becomes narrower towards the postpalate for high [i] and [u] than for low [a]. Cross-vocalic differences in size of the passage are larger than for any alveolo-palatal consonant, thus reflecting higher sensitivity of tongue-dorsum activity to the surrounding vowels. As shown in Table 1, large cross-speaker F2 differences are found for [i] (2075-2350 Hz)> [a] (1350-1675 Hz)> [u] (1075-1150 Hz), as a result of important changes in pharynx-cavity size reflected by differences in the size of the passage at the mediopalatal and postpalatal areas. Lower F2 for [u] than for [a] (and not for [a] than for [u], as would be expected from differences in degree of contact at the rear of the palate) may be due to lip rounding effects.

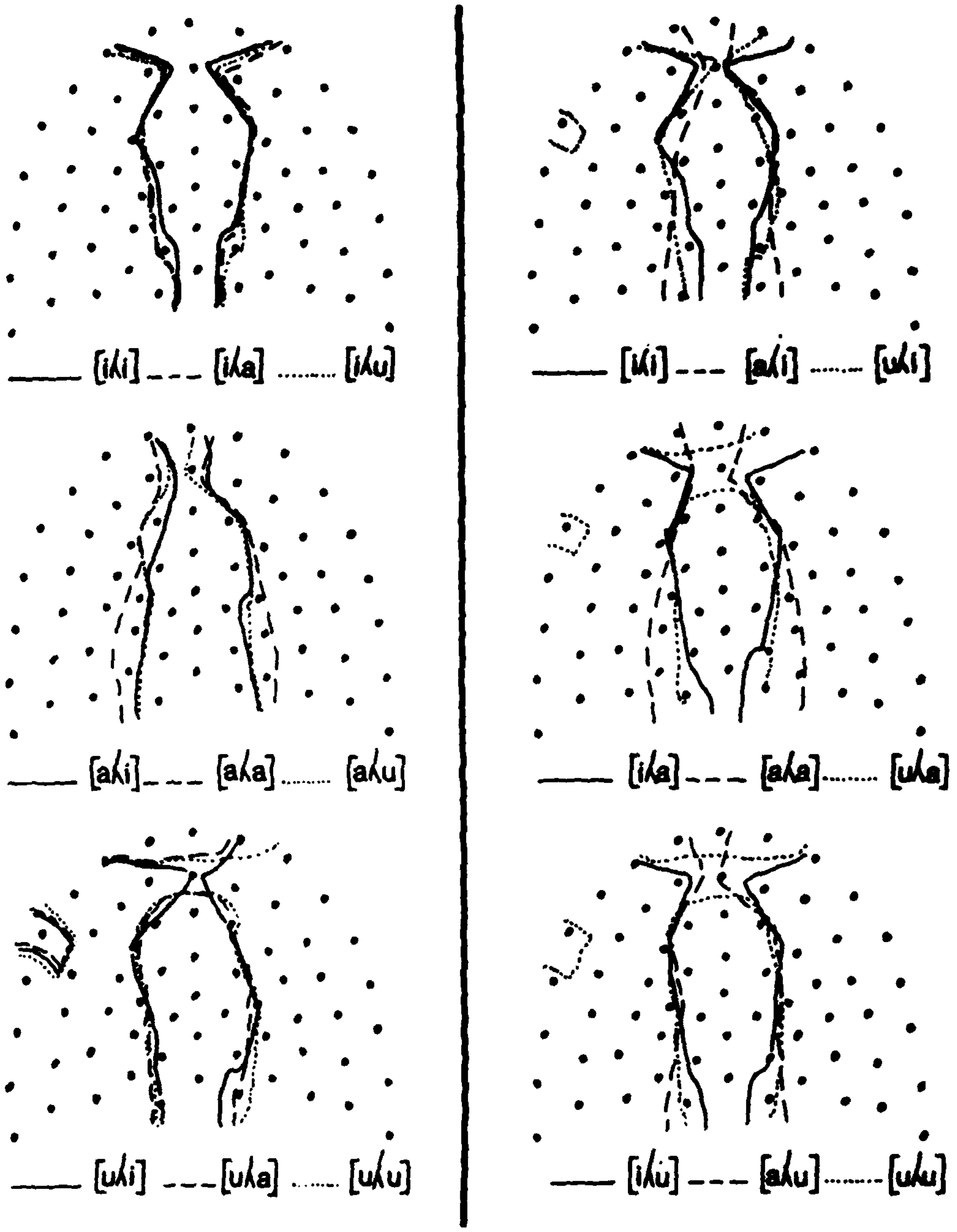


Figure 5. Anticipatory (left) and carryover (right) effects for [ʎ] at PMC (EPG data; speaker Re).

According to Figure 6, large carryover effects in the opening size of the mediopalatal and postpalatal passage are found when V2=[a], [u] (from V1=[a]>[u]>[i]) and very small effects when V2=[i] (from V1=[a], [u]>[i]). Anticipatory effects are found when V1=[a], [u] (from V2=[a]>[u]>[i]) but not when V1=[i]. Anticipatory and carryover effects in degree of tongue-dorsum contact are larger for [n] than for any palatal consonant.

Table 2 shows strong carryover effects upon F2 for all speakers from V1=[i]>[a]>[u], and much smaller anticipatory effects from V2=[i]>[a]>[u]. Ranges of F2 values show that carryover effects are always larger for [n] than for palatal consonants, and that anticipatory effects are generally but not always larger.

SUMMARY AND CONCLUSIONS

Palatographic data show that the degree of tongue-dorsum contact, on average, decreases along the series [j], [ɲ], [ʎ], [n]. Coarticulatory effects on tongue-dorsum contact for [j], [ɲ], [ʎ] and [n], measured at PMC, can be summarized as follows:

1) Dorsopalatal approximant [j]: In symmetrical environments, articulatory and acoustical effects are found from high vs. low vowels. In asymmetrical environments, anticipatory and carryover patterns of contact show that the effect of a high vowel always overrides that of a low vowel; in the light of the acoustical data, larger carryover than anticipatory effects are found mainly from high vs. low vowels.

2) Alveolo-palatal nasal [ɲ]: In symmetrical environments, articulatory and acoustical effects are found from high vs. low vowels, more so than for [j]. In asymmetrical environments, articulatory and acoustical data show carryover effects mainly from high vs. low vowels and small or non-existent anticipatory effects; overall, [ɲ] shows larger carryover effects than [j] and similar anticipatory effects.

3) Alveolo-palatal lateral [ʎ]: In symmetrical environments, larger articulatory and acoustical effects than for [j] and [ɲ] are found for high front vs. high back vs. low back vowels. In the light of articulatory data, contrasting carryover effects occur for those vowels while anticipatory effects are small or non-existent; acoustical data show larger carryover effects for the three vowels than anticipatory effects. Overall, coarticulatory effects in asymmetrical environments are larger than for [j] and [ɲ] in the articulatory and acoustical domains.

4) Alveolar nasal [n]: In symmetrical environments, articulatory and acoustical effects are found for high front vs. high back vs. low back vowels, more so than for palatal consonants. In the light of articulatory data, carryover and anticipatory effects can be large or small depending on the quality of the transconsonantal vowel; acoustical data show stronger carryover than anticipatory effects for the three vowels. Overall, coarticulatory effects in asymmetrical environments are larger than for palatal consonants.

It can be concluded that the amount of V-to-C coarticulation is dependent upon the degree of tongue-dorsum contact observed during the production of the

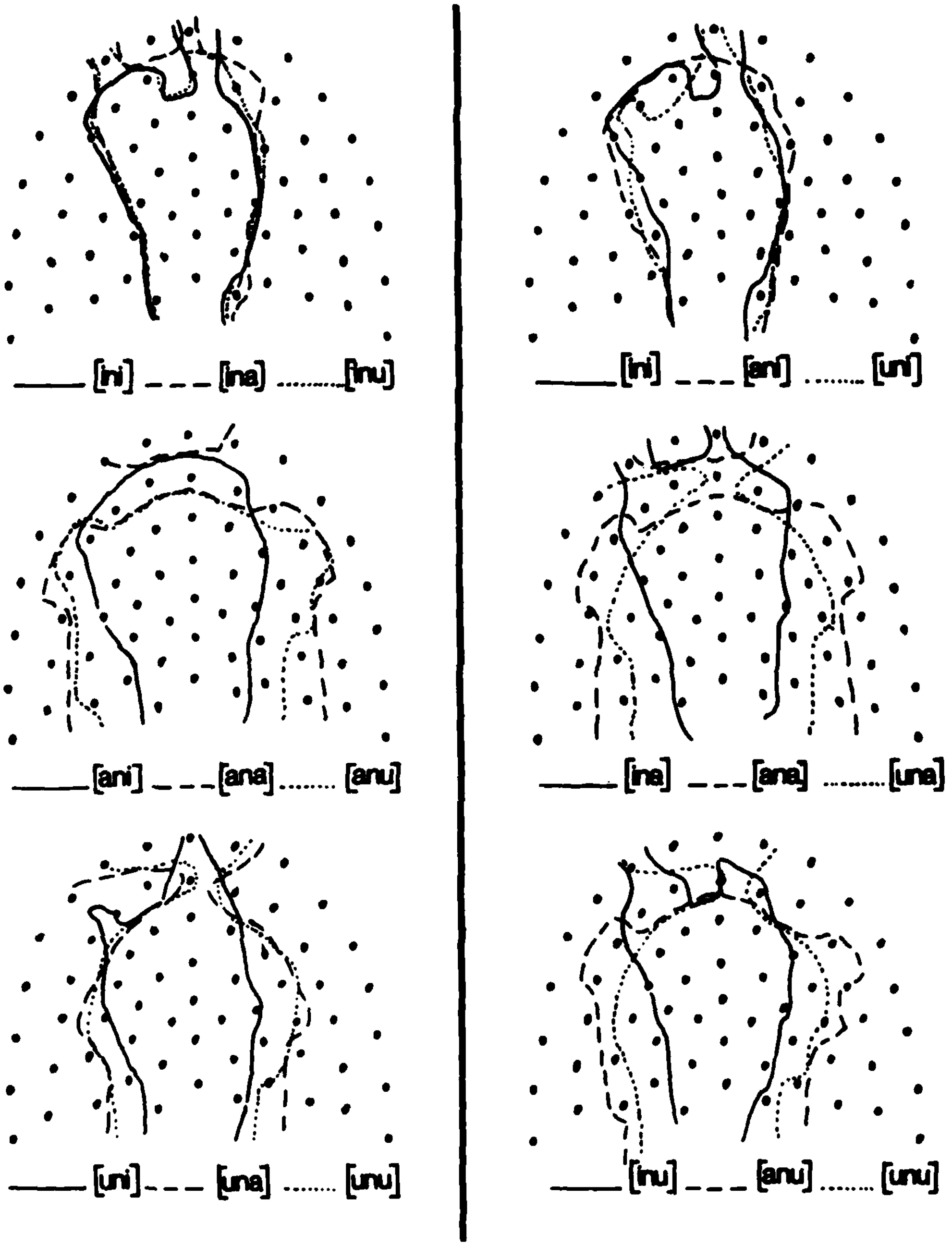


Figure 6. Anticipatory (left) and carryover (right) effects for [n] at PMC (EPG data; speaker Re).

consonant. Thus, on the one hand, a defined tongue-dorsum raising gesture towards the palatal area, as for a dorsopalatal consonant such as [j], results in little coarticulatory sensitivity to the surrounding vowels. On the other hand, alveolo-palatals such as [ɲ] and [ʎ], which show a greater degree of opening of the mediopalatal and postpalatal passage and in range of F2 values than [j], coarticulate more freely with the surrounding vocalic environment; moreover, a larger passage for [ʎ] than for [ɲ] results in larger coarticulatory effects. Finally, alveolar [n], produced with less tongue-dorsum contact than alveolo-palatals, shows the largest V-to-C coarticulatory effects of all the consonants studied here.

It is true, then, that the degree of V-to-C coarticulation varies inversely with the degree of tongue-dorsum contact required for the production of the consonant. Moreover, this variation is monotonical: a progressive decrease in degree of tongue-dorsum contact causes coarticulatory activity to vary progressively in similar amounts. Thus, for the different degrees of tongue-dorsum activity for [j]>[ɲ]>[ʎ]>[n], different degrees of coarticulatory activity are obtained for [n]>[ʎ]>[ɲ]>[j].

This systematic dependence of coarticulatory effects on the degree of linguopalatal contact suggests that, to a large extent, coarticulation is regulated by mechanical constraints on articulatory activity. Thus, a large degree of constraint on tongue dorsum results in a large amount of dorsal contact and a small degree of coarticulation; as the degree of constraint decreases, dorsal contact becomes smaller and coarticulatory effects increase. In line with Fowler et al. (1980), those may be the invariant relationships underlying the speech production mechanism.

With respect to the issue of directionality of coarticulatory effects, carryover effects have been found to be larger than anticipatory effects independent of speaker and vocalic environment. From the present study, it can be concluded that this finding reflects a language-specific property of how articulatory programming is organized in Catalan. However, evidence for the same trend has been found for English (Bell-Berti & Harris, 1976; Gay, 1974).

REFERENCES

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

Bladon, R., & Al-Bamerni, A. (1976). Coarticulatory resistance in English /l/. *Journal of Phonetics*, *4*, 135-150.

Bladon, R., & Carbonaro, E. (1978). Lateral consonants in Italian. *Journal of Italian Linguistics*, *3*, 43-55.

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

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

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

Chafcouloff, M. (1980). Les caractéristiques acoustiques de [j, ɥ, w, l, r] en français. *Travaux de l'Institut de Phonétique d'Aix, Université de Provence*, *7*, 7-56.

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

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

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

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

Haden, E. F. (1938). The physiology of French consonant changes. Supplement to Language, 14, Dissertation no. 26.

Harris, K. S. (1977). The study of articulatory organization: some negative progress. In M. Sawashima & F. S. Cooper, Dynamic aspects of speech production (pp. 71-82). Tokyo: University of Tokyo Press.

Kent, R., & Moll, K. (1972). Cinefluorographic analyses of selected lingual consonants. Journal of Speech & Hearing Research, 15, 453-73.

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

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

Rousselot, A. P. (1924-1925). Principes de phonétique expérimentale (vols. I and II). Paris: Didier.

Shibata, S. (1968). A study of dynamic palatography. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo, 2, 28-36.

Shibata, S., Ino, A., Yamashita, S., Hiki, S., Kiritani, S., & Sawashima, M. (1978). A new portable type unit for electropalatography. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo), 12, 5-10.

Stevens, K. N., & House, A. S. (1964). Analysis of formant transitions for consonants. Quarterly Progress Report (Massachusetts Institute of Technology), 73, 165-166.

FOOTNOTE

[1] This shorthand notation indicates the ordering of values as a function of vowel environment.

THE RELATIVE ROLES OF SYNTAX AND PROSODY IN THE PERCEPTION OF THE /š/-/č/ DISTINCTION*

Patti Jo Price+ and Andrea G. Levitt++

Abstract. A silent interval that cues the /š/-/č/ distinction in many contexts is less likely to do so when it coincides with certain boundaries. In natural speech these boundaries are generally marked by both prosody and syntax. We independently varied syntax and prosody to assess their contributions to the phonetic interpretation of silences occurring at these boundaries. We used a set of four sentences, four durations of silence, and two prosodic patterns (Experiment 1). We constructed sentences using three techniques that differed in the amount of prosodic control and in naturalness: synthesis by rule, concatenation of naturally produced syllables, and cross-splicing of naturally produced utterances. Silence duration had a strong effect on the perception of the /š/-/č/ contrast in all conditions. For the Synthetic Condition, we also found a strong effect of the prosodic pattern. We found no evidence of any purely syntactic effect. In Experiment 2, the two syllables surrounding the silence were excised from the sentences of Experiment 1 and presented to listeners for labeling. Prosody had a significant effect in the Synthetic Condition and in the Natural Condition. The results indicate that the local prosodic pattern (one syllable with a pitch fall and a longer duration) can be sufficient to influence listeners' perception of the /š/-/č/ contrast. There is also evidence that the prosodic information may be subject to context effects.

INTRODUCTION

The introduction of a short silent interval before an appropriate intervocalic fricative noise can change listeners' labelings from 'sh' to 'ch'

---

*Also to appear in Language and Speech, Vol. 26, 1983.

+Massachusetts Institute of Technology.
++Also Wellesley College.

Acknowledgment. We gratefully acknowledge the support of NICHD Grant No. HD-01994 to Haskins Laboratories and NIH Grant No. 2 T32 NS07040-7 to M.I.T. We thank Harriet Magen for her help in running the subjects and tallying the data. We thank Leah Larkey and Janet Pierrehumbert for useful comments and suggestions concerning our data. We also thank René Collier, Alice Healy, Alvin Liberman, Brad Rakerd, and Bruno Repp for useful comments on earlier versions of this paper and our anonymous reviewer for requesting terminological clarification.

(Dorman, Raphael, & Liberman, 1979). For example, in the utterance "say shop," the introduction of silence after the word "say" can change the percept of "shop" to "chop." Others have shown, however, that this change is much less likely to occur when the silence coincides with a sentence boundary (Rakerd, Dechovitz, & Verbrugge, 1982). Presumably, the listeners interpret the silence as a consequence of the sentence boundary, that is, as a pause, rather than as the silence associated with oral closure for /č/. Dechovitz (1980, 1981) has argued that sentence-internal clause boundaries have a similar effect on listeners' perception and that such boundaries will have an effect even when they are not marked by appropriate prosody.

Syntactic boundaries in natural speech are, however, generally associated with significant prosodic changes that may be largely, or entirely, responsible for the subject's interpretation of the silence. It is therefore important to carefully control for the role of prosody insofar as possible before attributing the effect purely to syntax. Aspects of prosody that may mark clause boundaries include a drop in F_0, a lengthening of the clause-final syllable, and a period of silence before the beginning of the next clause. By independently varying the syntax and these prosodic markers in several sentences, we can test the relative roles of syntax and prosody in influencing a listener's decision that the silence is to be attributed to oral closure for /č/ or to a pause followed by /š/.

The separation of syntactic and prosodic effects leads to an important methodological consideration: Since prosody and syntax are often correlated in natural speech, the more effectively the two are separated, the less natural the sentences begin to sound. In our attempt to deal with this problem, we have used three techniques to create the sentences so that some are more natural sounding, but less carefully controlled, and others the reverse.

EXPERIMENT 1

Method

Stimuli. Table 1 shows the two pairs of sentences used. Each pair contains an equal number of syllables and shares a large number of words: Sentences 1a and 1b differ, or "are disambiguated," before "pay," whereas sentences 2a and 2b are disambiguated after "pay." The two members of each pair differ in syntactic structure: Sentences 1a and 2a have a syntactic break after "pay"; sentences 1b and 2b do not. We used four durations of silence (0, 30 60, 90 ms) following "pay." The sentences were generated in two versions: one with a prosodic pattern appropriate for a break following "pay," and one with a pattern appropriate for no break following "pay." Patterns appropriate for a break principally involve the syllables immediately before that break. These syllables may show longer duration, a fall or fall-rise pitch pattern, a tapering off in amplitude, and a following pause. The same syllables occurring in a sentence without such a break are shorter and have flatter pitch and amplitude patterns. Here we have investigated the combined roles of pitch pattern and duration in marking the boundary; the two were not separated in this study.

We found in pilot studies that an intervocalic /š/ preceded by silence generally was perceived as /š/ unless the onset was edited to be more abrupt (cf. Rakerd et al., 1982). In order to allow silence to operate as an effective /š/-/č/ cue, we therefore had to edit the friction noise to make it more ambiguous between /š/ and /č/. We shortened the initial friction noise and gave it a sharper rise time. These changes were based on measurements of natural speech productions of /č/.[1]

Table 1

Source sentences. Subjects hear either "Shipley" or "Chipley" following "pay."

1a.
Since we have all our back pay, Shipley and I want to leave town.

1b.
He wants enough to repay Shipley, and I want to leave town.

2a.
That he could pay, Shipley reiterated.

2b.
That he could pay Shipley was a shock to me.

The dependent variable in our design was the perceptual change of "Shipley" to "Chipley." We chose proper names to minimize effects of lexical frequency and semantic expectation. The stressed open syllable "pay" can show clearly the pitch, amplitude, and duration patterns that may mark clause finality versus non-finality, and its final high front glide transitions are similar in productions of either "pay ship" or "pay chip."

The three methods used to create the sentences were:

(1) Synthesis by rule: These sentences were not very natural sounding but prosodic patterns were strictly controlled.
(2) Concatenation of syllables excised from naturally produced strings: These sentences were more natural than in the Synthetic Condition but prosodic patterns were disrupted.
(3) Cross-splicing of large pieces of naturally produced utterances: These sentences sounded natural but prosodic patterns were not strictly controlled.

Synthetic Condition. A version of each of the four sentences in Table 1 was generated using Ingemann's (1978) rules on the OVE-IIIc synthesizer at Haskins Laboratories (Liljencrants, 1968). To facilitate the perceptual change to /č/, the /š/ frication from sentence 1a was edited so that the initial fricative noise was shorter and had a sharper rise time. This frication was used in all the synthetic sentences. Though an intonation 'fall' generally occurs in sentence-final position and a 'fall-rise' pattern in phrase-final position, the rise part of the fall-rise may occur either before the break or on the first syllable after the break (see, e.g., Cooper &

Sorensen, 1977). Delattre (1965) observed that the rise part of the fall-rise pattern is generally not as important in American speech as the fall part. To sort out the relative perceptual values of these two patterns we used two 'final' versions of "pay" (one with a fall-rise F_0 pattern and the other with a falling F_0 pattern, both of equal length and amplitude), and one 'non-final' version (with a shorter duration, and a flat amplitude and F_0 pattern). Figure 1 shows the F_0 and temporal patterns for the source sentences used in this condition.

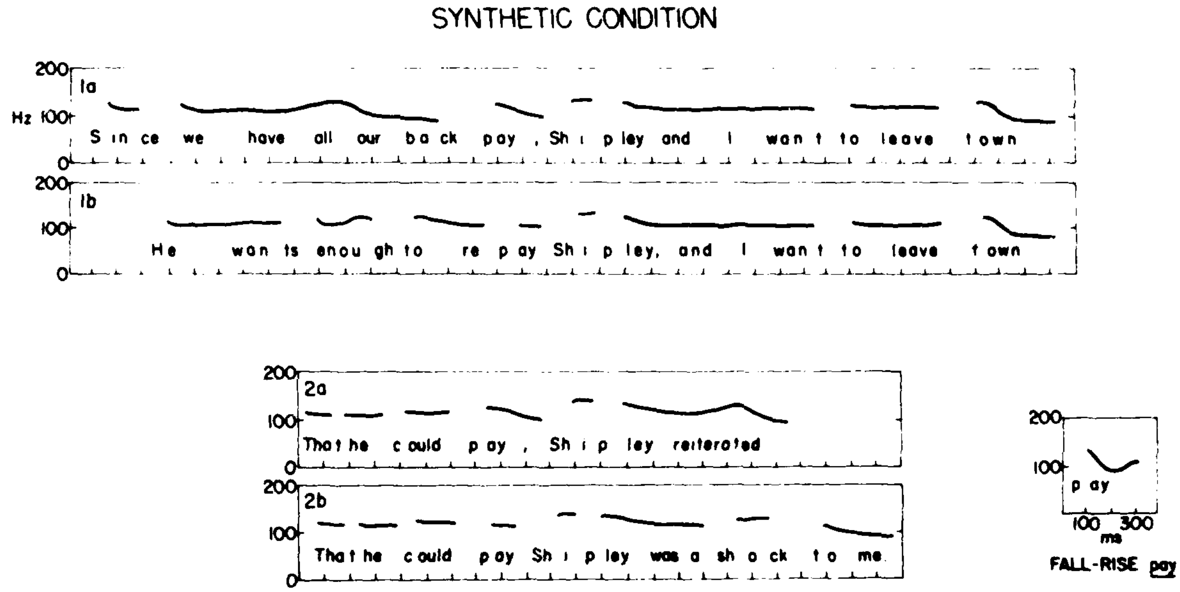

SYNTHETIC CONDITION

Figure 1. Synthetic Condition sentences with F_0 patterns. The axes at the left show frequency in Hertz. Sentences 1a and 2a contain the "pay" with a fall (final) contour. The flat (non-final) "pay" shown in sentences 2a and 2b was switched with the fall (final) "pay" shown in sentences 1a and 2a in order to control syntax and prosody independently. In the fall-rise part of the Synthetic Condition, the F_0 pattern on "pay" shown at the right was substituted everywhere for the fall pattern shown in sentences 1a and 2a at the left. Silence was inserted after "pay."

Note that sentence 1b has a syntactic and prosodic break after "Shipley," while sentence 1a does not. Since this break occurs in the part of the sentence that the two members of the pair are supposed to share, sentence 1a was edited to create a compromise version in which the duration of the final vowel in "Shipley" was *increased* by 75 ms and an amplitude contour (a symmetric fall and rise) was added. The matching parts of sentences 2a and 2b before "pay" were identical as generated and no editing was necessary.

The Synthetic Condition was divided into two blocks, each consisting of 5 separate randomizations of 32 stimuli: 4 sentences (see Table 1) X 2 "pay"s (final and non-final) X 4 silence durations (0, 30, 60, 90 ms). The two blocks differed in that the final "pay" used had either the fall-rise contour or the fall contour. Digitized versions of each sentence were created before randomization.

Concatenated Condition. In this condition the starting point was natural speech. In order to preserve as much segmental naturalness as possible, while at the same time eliminating most prosodic cues, strings of two or three syllables were recorded, digitized, edited, and then spliced together to form the sentences in Table 1. A randomized list of these strings was read with list intonation by one of the authors (PJP). The list contained the syllables of the test sentences as well as similar strings from some additional sentences. By list intonation we mean that, in general, all syllables had a pitch fall; the last syllable in a string (the prepausal syllable) fell to a lower level and was longer. An example of the strings used for sentence 2b appears in Table 2.

Table 2

Example of the strings generated for sentence 2b. The middle column in Table 2 contains the syllables to be concatenated with others to form the sentences. Sets of strings similar to these 11 strings were generated for the 4 test sentences and for an additional 28 filler sentences. The strings were randomized and read with a list intonation. The pieces in the middle column were then isolated and spliced together to form the sentences in the Concatenated Condition. The symbol # indicates a pause.

#	that	hat
heet	he	key
key	could	pould
peed	pay	shay
shay	ship	lip
leap	lee	we
we	wuh	zuh
zuh	zuh	shuh
shuh	shock	tock
tuke	to	moo
moo	me	#

Note that the syllable strings were constructed so that the syllables in the middle column were uttered in phonetic contexts similar to that of the part of the sentence into which the syllable was to be spliced. Adjustments of the transcriptions were made to condition phonological rules such as flapping. The syllable strings were low-pass filtered at 5 kHz and sampled at a rate of 10 kHz before editing. One "ship" was used in all the sentences. The friction noise at the beginning was made more ambiguous between /š/ and /č/: it was shortened and its onset made sharper. A single "pay," from the pre-pausal context, was used. LPC analysis and resynthesis were used to flatten the pitch of this syllable, and the waveform editor was used to shorten it from about 300 ms to about 200 ms, thereby creating the 'nonfinal' version of "pay." Analysis of the LPC-flattened "pay" revealed that the pitch was not flattened during the first 40 to 50 ms of the vowel. This left a sharp pitch fall at the vocalic onset, which we felt was not unreasonable for a vowel following a voiceless consonant (Hombert, Ohala, & Ewan, 1979). The sentences composed of concatenated syllables were edited further to eliminate any audible discontinuities. Figure 2 shows the F_0 and temporal patterns of the source sentences used in this conditon.

The four source sentences in Figure 2 were generated in two versions: one with prepausal "pay" (shown in sentences 1a and 2a) and one with the 'flattened' and shortened "pay" (shown in sentences 1b and 2b). We used four durations of silence (0, 30, 60, and 90 ms) between the "pay" and the "ship" of the resulting 8 sentences to create 32 stimuli. Five separate randomizations of the 32 stimuli were recorded.

Natural Condition. The four sentences of Table 1 were included in a randomized list containing 28 filler sentences. This list was read by one of the authors (PJP). Sentences with the same words (and, presumably, syntactic structure) but with (presumably) inapproprite prosodic structure were created by cross-splicing pieces of the sentences as indicated in Figure 3. The naturally occurring "ship" in each of the sentences was replaced by the single edited "ship" used in the Concatenated Condition. The resulting eight sentences were used to generate the 32 stimuli of the experiment (with 0, 30, 60, or 90 ms of silence between "pay" and "Shipley" in each). Again, five separate randomizations of the 32 stimuli were recorded. The F_0 and temporal patterns of the source sentences used in this condition are shown in Figure 3.

Subjects and procedure. Ten Yale undergraduates with no reported history of speech or hearing problems were paid to listen to the four resulting tapes (Synthetic Fall-rise, Synthetic Fall, Concatenated, and Natural) in counterbalanced order over Grason-Stadler model TDH 39-300Z headphones connected to an Ampex tape recorder. Subjects were asked to write 's' if they heard "Shipley" in the sentence and 'c' if they heard "Chipley." They were told that it was important to listen to the entire sentence before deciding.

Results

We analyzed the results of the three conditions separately to see whether prosodic pattern, syntactic structure, or silence duration affected the number of 's' responses. An analysis of variance was performed on the 's' responses in each of three conditions (Synthetic, Concatenated, and Natural). Each analysis included the factors disambiguation (before/after), syntactic context (break/no break), prosody ('final'/'non-final'), and silence duration (0, 30, 60, or 90 ms). The analysis of the Synthetic Condition had as an additional factor the pitch change that marked the break after the "pay" (fall/fall-rise).

Figure 4 (thick lines) presents the results of this experiment. The data are averaged across disambiguation (before/after), syntactic context (break/no break), and for the Synthetic Condition, the pitch marker of the break ('fall'/'fall-rise').

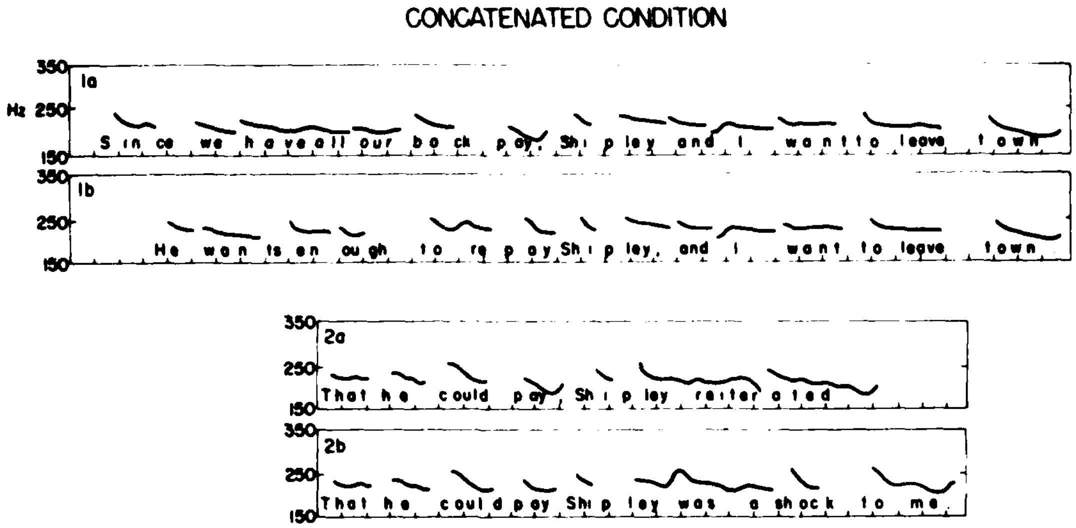


Figure 2. Concatenated Condition sentences with F_0 patterns. The axes at the left show frequency in Hertz. The 'flattened' (non-final) "pay" shown in sentences 1b and 2b was switched with the original fall-rise (final) "pay" shown in sentences 1a and 2a. Silence was inserted after "pay."

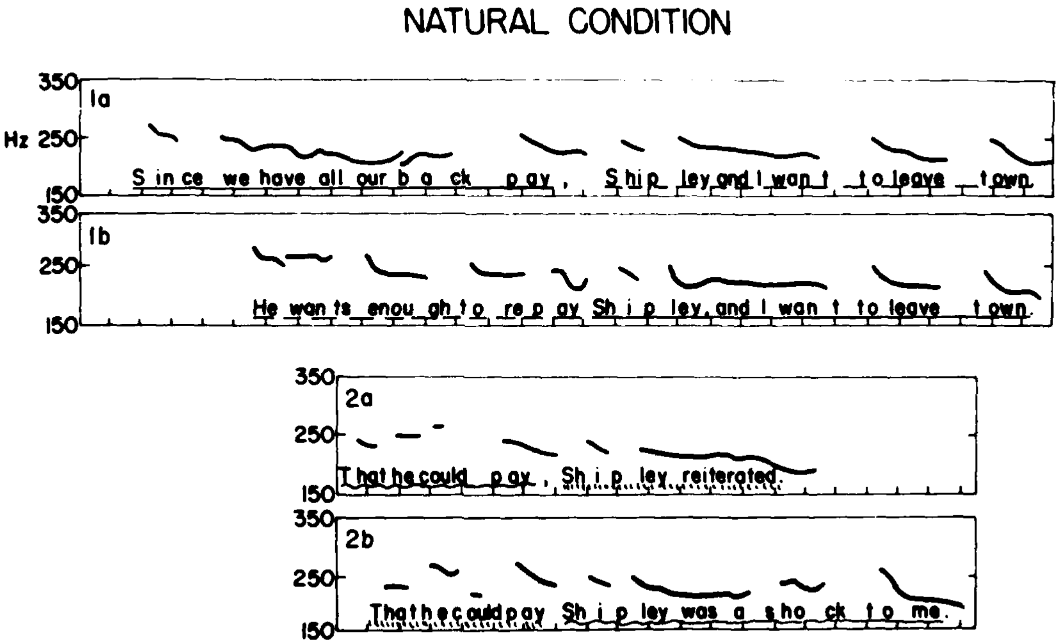


Figure 3. Natural Condition sentences with F_0 patterns. The axes at the left show frequency in Hertz. These sentences were cross-spliced as indicated: Portions of the sentences with the same underlining were joined to form the new sentences. Silence was inserted after "pay."

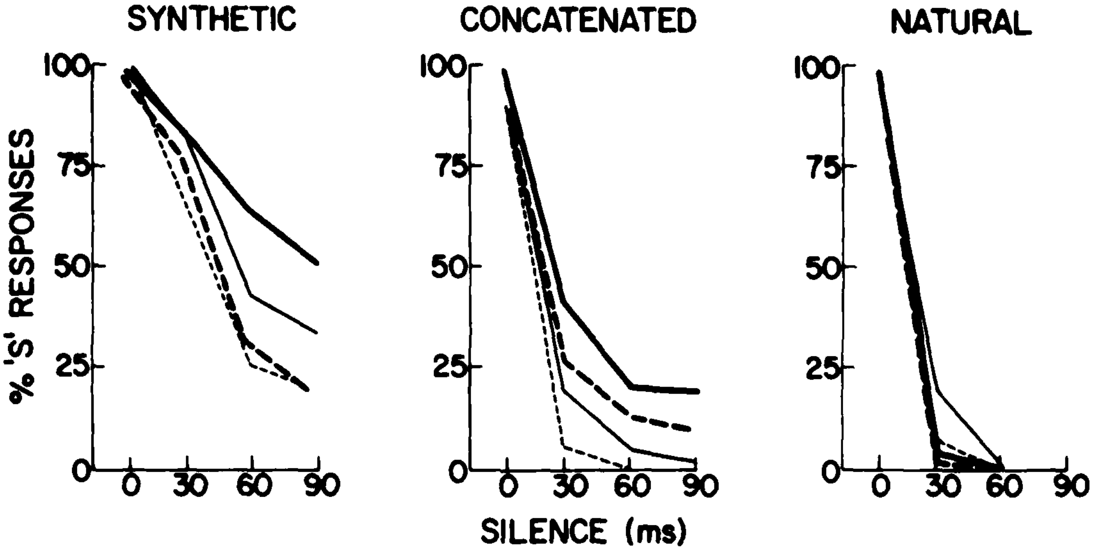


Figure 4. Percent 's' responses are plotted for the Synthetic (left), Concatenated (middle), and Natural (right) Conditions. Thick lines represent responses to sentences; thin lines to controls. Solid lines show responses to items with 'final' "pays"; dashed lines to 'non-final' "pays."

In all conditions there was a highly significant main effect of silence duration. In the Synthetic Condition there was also a significant main effect of prosody (final/non-final) (more 's' responses for 'final' "pay"s, as expected): $F(1,9) = 12.34$, $p = .0066$, and a significant interaction of prosody and silence duration, $F(3,27) = 12.84$, $p < .0001$. There was no significant difference between the two final pitch patterns on "pay" ('fall' versus 'fall-rise'). Finally, there was a significant interaction of syntax (break/no break) and prosody (final/non-final), $F(1,9) = 6.45$, $p = .0318$. When there was a 'final' contour on "pay," sentences with a syntactic break received slightly more 's' responses than sentences without such a break; whereas when there was a 'non-final' contour on "pay," sentences with a syntactic break had slightly fewer 's' responses than sentences without a break. The expectation, of course, would be that if syntax were an independent cue to the listener, sentences with a 'final' contour on "pay" and a syntactic break would show more 's' responses than would sentences without a syntactic break. This is the opposite of what was obtained. A significant interaction, $F(3,27) = 3.51$, $p = .0287$, between pitch change (fall, fall-rise) and silence duration was due to the fact that the number of 's' responses was sometimes slightly higher for fall, other times higher for fall-rise, depending on the given silence duration. There was no systematic pattern of differences. A significant three-way interaction of syntax, prosody, and disambiguation, $F(1,9) = 5.95$, p = .0374, was also present: for a 'final' contour on "pay" there were slightly more 's' responses for sentence 1a (with a break) than for sentence 1b (without a break), whereas there were slightly more 's' responses for sentence 2b (without a break) than for sentence 2a (with a break).

In the Concatenated Condition the only significant effect, besides that of silence duration, F(3,27) = 39.66, p < .0001, was a three-way interaction among disambiguation (before/after), prosody (final/non-final), and silence duration, F(3,27) = 3.08, p = .0445, due to the fact that sentences disambiguated before the break showed a slight rise in number of 's' responses for the 'final' "pay"s at the longest silence duration whereas sentences disambiguated after the break did not. Although there was no significant prosodic effect, seven of the ten subjects did show more 's' responses when the preceding "pay" had the 'final' prosodic pattern as opposed to the 'non-final' one.

In the Natural Condition the effect of silence duration was most pronounced: subjects' responses changed almost completely from 's' to 'c' with the introduction of 30 ms of silence, regardless of sentence type. There was also a significant interaction of disambiguation (before/after) and prosody (final/non-final), F(1,9) = 9.00, p = .0150: for the sentences disambiguated in their initial part, subjects showed a greater number of 's' responses for a 'final' "pay" than for a 'non-final' "pay." However, in the sentence pair that was disambiguated in its final part, subjects showed a greater number of 's' responses for a non-final "pay." There was another significant two-way interaction of prosody (final/non-final) and silence duration, F(3,27) = 3.86, p = .0203, and a significant three-way interaction, F(3,27) = 5.77, p = .0035, of those factors and disambiguation (before/after), which both seem due to the fact that each of the four individual "pay"s used in this condition produced slightly different cross-over points. Experiment 2 deals with this issue more directly.

Discussion

There was a clear phonetic effect of silence duration in all three conditions. For all subjects and all conditions the introduction of silence after "pay" caused subjects to report "Chipley" rather than "Shipley."

The effect of the pitch and duration patterns of "pay" (final/non-final) on the number of 's' ("Shipley") responses is clear in the case of the synthetically produced sentences. Subjects' responses to the two final pitch patterns on "pay" (fall and fall-rise), however, did not differ, which suggests that both were equally good at signaling a break to American English listeners. In the Natural Condition, a prosodic effect was obtained only in the sentences that were disambiguated before the syntactic break but not in the sentences disambiguated after the syntactic break. However, since the sentences in the Synthetic Condition that were disambiguated after the break show a prosodic effect, we believe that the failure to find one in the sentences disambiguated after the break in the Natural Condition reflects the fact that the overall prosodic contours of the natural sentences are not as well controlled as in the other conditions. We do not believe that site of disambiguation was the crucial factor. Finally, the Concatenated Condition showed a trend in the direction of an effect of prosodic pattern.

We found no evidence of a purely syntactic effect: Grammatical structure of the sentences independent of the prosodic patterns was not a significant factor, nor was there a trend in that direction for any of our three conditions. A negative result, of course, does not prove that such syntactic effects cannot occur. However, it is essential to disentangle possible prosodic effects from syntactic effects in order to demonstrate the latter clearly. Our results show that prosody can play an important role in the perception of the /š/-/č/ distinction.

What, then, is the domain of the prosodic effect? Is the falling pitch pattern and longer duration of the "pay" sufficient to cue a change in the number of 's' responses regardless of context? Experiment 2 addresses these questions.

EXPERIMENT 2

Method

The various "pay ship"s from the preceding experiment were isolated by waveform editing. For the Synthetic Condition, this resulted in three "pay"s (two 'final'--fall and fall-rise--and one 'non-final,' which was flat in pitch and shorter) times four durations of silence, or 12 source stimuli. For the Concatenated Condition, the two "pay"s (the original pre-pausal and its flattened and shortened version) and four silence durations resulted in eight stimuli. For the Natural Condition, the four "pay"s (one for each of the sentences in Figure 1) and four silence durations resulted in 16 stimuli. Ten randomizations of each set of stimuli were prepared, blocked by condition, and presented for labeling to twelve new subjects in counterbalanced order. Subjects were asked to write 's' if they heard "pay ship" and 'c' if they heard "pay chip."

Results

A two-way analysis of variance (prosody and silence duration) was performed on each of the conditions. In all three conditions, silence duration was highly significant. Prosody was a significant main effect in the Synthetic Condition (as in Experiment 1), $F(2,22) = 5.23$, $p = .0138$, and in the Natural Condition, $F(1,11) = 6.34$, $p = .0286$ (unlike Experiment 1 where it was part of a significant interaction), but not in the Concatenated Condition. There was an interaction of prosody and silence duration in the Synthetic Condition, $F(6,66) = 2.25$, $p = .0489$, and in the Natural Condition, $F(3,33) = 4.71$, $p = .0076$.

Figure 4 (thin lines) shows the results of Experiment 2. As before, results for the Synthetic Condition are averaged over the two final versions of "pay" used (fall/fall-rise), and results for the Natural Condition are averaged over the two tokens of the final "pay"s and over the two tokens of the non-final "pay"s.

In order to compare Experiments 1 and 2, we did an unequal N analysis of variance on the results of the two experiments for each condition.

For the Synthetic Condition, the combined analysis (Experiments 1 and 2) showed highly significant effects of prosody, $F(2,40) = 15.82$, $p < .0001$, and silence duration, $F(3,60) = 67.02$, $p < .0001$, and a highly significant interaction of prosody and silence duration, $F(6,120) = 7.51$, $p < .0001$, as had been found in each of the separate analyses. There was also a significant three-way interaction of task, prosody, and silence duration, $F(6,120) = 4.07$, $p = .0009$, reflecting a greater number of 's' responses for silence durations of 60 ms or greater in the sentences than in the "pay ships"s.

For the Concatenated Condition, we found a highly significant effect of silence duration, $F(3,60) = 187.05$, $p < .0001$, the only significant effect in each of the separate analyses, and a significant interaction of task and silence duration, $F(3,60) = 3.49$, $p = .0211$, again showing a greater number of 's' responses for the longer silence durations (here, 30 ms or longer) in the sentences than in the "pay ship"s.

For the Natural Condition, we found in the combined analysis a significant effect of silence duration, $F(3,60) = 1035.57$, $p < .0001$, as we had in the separate analyses, and a significant prosodic effect, $F(1,20) = 6.16$, $p = .0221$, as well as a significant interaction of prosody and silence duration, $F(3,60) = 6.14$, $p = .001$, as we had in Experiment 2.

Discussion

In the separate analysis of the results of Experiment 2 alone, a strong effect of silence duration was again demonstrated in each of the three conditions. In the Synthetic Condition, as in the previous experiment, prosody was a significant main effect and a significant interactive effect with silence duration. In the Concatenated Condition, prosody was not a significant effect in the sentences or in the controls. The original "pay" in this condition was from a prepausal context. Although the pitch was flattened by LPC analysis and resynthesis and the syllable was shortened, other cues to 'finality' may have remained. It is also possible that the syllable was insufficiently flattened and/or shortened. In any case, though the flattening and shortening resulted in something more like a non-final "pay," as seen by the trends in the data, the effect did not reach significance. In the Natural Condition, prosody as a main effect and its interaction with silence duration were both significant in the controls, though they were not in the sentences (Experiment 1).

When we compare the results of Experiments 1 and 2, task emerges as a significant interactive effect in both the Synthetic and the Concatenated Conditions. In both cases the interactions appear due to the fact that in the experiment with sentences, there tend to be a greater number of 's' responses at longer silence durations than in the experiment with the two syllables. In the Synthetic and the Concatenated Conditions, prosody is more controlled, but the sentences sound less natural and less coarticulated. It seems reasonable that subjects might interpret silence as a random pause (and not as closure for the affricate) in these less natural sounding sentences and therefore respond with more 's' responses. The lack of naturalness would be less salient in the experiment with the two syllables. Furthermore, since the utterances are shorter and are less likely to be heard as sentences, the silences may be less likely to be interpreted as pauses.

In sum, we see a very similar pattern of results for the two experiments. A sharp pitch fall and a longer duration seem to be sufficient to sway listener judgments towards pause plus /š/ rather than /č/, when other factors are neutralized.[2] Further, the more effectively these factors are neutralized, that is, in the Synthetic Condition, the more important these aspects of prosody can be. Of course, in actual speech communication such factors are not generally separated. That people can make reliable judgments when prosodic factors are varied and others are neutralized is evidence, we feel, that prosody is a significant factor among many that people are attuned to in speech understanding.

GENERAL DISCUSSION

The results of these experiments show a clear pattern of the effect of prosody on the perception of the /š/-/č/ distinction in a variety of contexts. Although no purely syntactic effects were found here, it is possible that a change in the subject's task would elicit such an effect. Miller (1982), for example, has suggested that variations in prosody (speaking rate, in her case) are "automatically" taken into account by the listener, whereas semantic effects only emerge when the task focuses on meaning. Semantic or syntactic structures are more likely to play a role when the task more directly demands them. We also believe that, in general, listeners use any strategies and any information available (see also Cutler, 1982). We would argue, however, that prosody is more available to the listener as an aid in initial parsing of a sentence than syntax can be at this stage.

Our data also provide evidence for the importance of the syllables immediately preceding the boundary in cueing that boundary. The same "ship" was used in the Concatenated and Natural Conditions, yet the patterns of 's' responses differ. Some context effects of domains larger than this are suggested in the comparisons of the two experiments.

A further result of our study bears on methodology. We feel that the cross-splicing of large pieces of naturally produced sentences is the least appropriate of the techniques we used. On the one hand, the fact that in these sentences the key parameters are sometimes conflicting and in general are not independently controlled make the data difficult to interpret. On the other hand, naturalness is a highly desirable feature in test stimuli.

There is much evidence that the pitch contour and temporal properties of the local environment of a break can carry a great deal of weight in marking that break (see, e.g., Cooper & Sorensen, 1977; Grosjean, 1982; Larkey, 1980; Pierrehumbert, 1980). We found that these factors can outweigh those of the syntax and semantics of a sentence. This, together with other reports of segmental and suprasegmental interactions (see, e.g., Klatt & Cooper, 1975; Lehiste, 1975; Nooteboom & Doodeman, 1980; Summerfield, 1975), suggest the possibility that listeners may use suprasegmental information to assign an initial syntactic structure before decoding the rest of the information. We see research along these lines as promising for investigations of acoustic correlates of prosodic information and of their role in marking perceptual units for the listener.

REFERENCES

Cooper, W., & Sorensen, J. (1977). Fundamental frequency contours at syntactic boundaries. Journal of the Acoustical Society of America, 62, 683-692.

Cutler, A. (1982). Prosody and sentence perception in English. In J. Mehler, E. Walker, & M. Garrett (Eds.), Perspectives on mental representation (pp. 201-260). Hillsdale, NJ: Erlbaum.

Dechovitz, D. R. (1980). Constraints imposed on phonetic processing by syntax. Journal of the Acoustical Society of America, 67, Supplement 1, S99.

Dechovitz, D. R. (1981). Phonetic integration as a potential test for controversial surface structures. Journal of the Acoustical Society of America, 69, Supplement 1, S115.

Delattre, P. (1965). Comparing the phonetic features of English, French, German, and Spanish: An interim report. Philadelphia: Chilton Books.

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

Grosjean, F. (1982). How long is the sentence? Prediction and prosody in the on-line processing of a language. Unpublished manuscript.

Hombert, J. M., Ohala, J., & Ewan, W. (1979). Phonetic explanations for the development of tones. Language, 55, 37-58.

Ingemann, F. (1978). Speech synthesis by rule using the FOVE program. Haskins Laboratories Status Report on Speech Research, SR-54, 165-174.

Klatt, D. H., & Cooper, W. E. (1975). Perception of segment duration in sentence contexts. In A. Cohen & S. E. G. Nooteboom (Eds.), Structure and process in speech perception (pp. 69-89). New York: Springer-Verlag.

Larkey, L. S. (1980). The role of prosodic information in speech perception. [microfilm]. Ann Arbor: University of Michigan.

Lehiste, I. (1975). The phonetic structure of paragraphs. In A. Cohen & S. E. G. Nooteboom (Eds.), Structure and process in speech perception (pp. 195-206). New York: Springer-Verlag.

Liljencrants, J. (1968). The OVEIII speech synthesizer. IEEE Transactions: Audio, 16, 137-140.

Miller, J. (1982). On the distinction between prosodic and semantic factors in word identification. Journal of the Acoustical Society of America, 71, Supplement 1, S95.

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

Pierrehumbert, J. (1980). The phonology and phonetics of English intonation. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

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

Summerfield, Q. (1975). How a full account of segmental perception depends on prosody and vice versa. In A. Cohen & S. E. G. Nooteboom (Eds.), Structure and process in speech perception (pp. 51-68). New York: Springer-Verlag.

FOOTNOTES

[1] It is common experimental practice to neutralize cues other than those under investigation. It is somewhat difficult to determine an appropriate neutral value for the /š/-/č/ friction noise. Our pilot studies indicated that what is neutral with respect to /š/-/č/ in utterance initial position is not neutral in vocalic contexts.

[2] That listeners continued to hear /š/ even when the edited friction noise was preceded by short intervals of silence indicates that we did not in editing eliminate all the cues that identify /š/.

INTERSECTIONS OF TONE AND INTONATION IN THAI*

Arthur S. Abramson+ and Katyanee Svastikula+

Abstract. The distinctive tones of a tone language may be said to have "ideal" pitch contours that are perhaps best seen in citation forms. Strings of tones in running speech show perturbations of the ideal contours through tonal coarticulation and the effects of segmental features. These tones intersect with sentence intonation, which also makes much use of pitch. For our research we chose Thai, a language with five phonemic tones, because much analytic and perceptual work had been done on its tones. We recorded all possible sequences of three tones on key words in sets of simple and complex declarative sentences for acoustic analysis into waveforms, overall amplitude, and fundamental frequency. We looked for "declination," i.e., a drop in fundamental frequency from beginning to end, and interaction between declination and tone. Such declination as we found is somewhat obscured, especially in short sentences, by the local effects of the lexical tones. The tones themselves remain physically distinct in all contexts examined.

INTRODUCTION

Older approaches to the study of sentence intonation, for example the important work of Trager and Smith (1951), generally tried to analyze intonation into phonological units of one kind or another. More recent work, perhaps best exemplified by Cooper and Sorensen (1981), has sought rather to correlate intonational variables with syntactic features. Since pitch is the most salient auditory aspect of intonation, it is not surprising that investigators have given most of their attention to its major physical correlate, fundamental frequency (F_0). In addition, one major observation on which there is emerging consensus is that declarative sentences show "declination," an overall fall of F_0 from the beginning to the end of the sentence. Indeed, it may be possible to predict the course of this declination by rule (Cooper & Sorensen, 1981).

---

*Also in H. Fujisaki & E. Gårding (Eds.), Proceedings of the Working Group on Intonation, XIIIth International Congress of Linguists. Dordrecht, The Netherlands: Foris Publications, in press.

+Also University of Connecticut.

Acknowledgment. This research was supported by Grant HD-01994 from the National Institute of Child Health and Human Development to Haskins Laboratories, New Haven, Connecticut, U.S.A. Preliminary work for this study was done during the first author's stay, as part of a sabbatical leave, at Chulalongkorn University, Bangkok, Thailand from August 29, 1980 to January 1, 1981.

[HASKINS LABORATORIES: Status Report on Speech Research SR-74/75 (1983)]

In phonemic tones, as in intonation, although such other features as amplitude shifts may play a role, it is generally agreed that F_0 levels and contours furnish the major phonetic underpinnings. One might think then that in a true tone language, one in which in principle every syllable in the morpheme stock bears a tone, it would be hard to use the same laryngeal and aerodynamic mechanisms to control global intonation contours while at the same time using them moment by moment to control the local F_0 patterns of the tones. Anyone with experience in speaking such a language, however, knows very well that the communicative use of sentence intonation seems to be as free as in non-tone languages.

Certain questions have motivated our present research. Is declination normal in the declarative sentences of a tone language? One study on Mandarin Chinese suggests otherwise (Lieberman & Tseng, 1980). If it is normal, what are the interactions between it and the F_0 contours of the lexical tones? That is, each tone could be, even while preserving essential aspects of its "ideal" contour, a local perturbation of the overall intonation line; or it could simply be that some or all of the tones in the system lose their distinctiveness--they become neutralized--for certain stretches of the intonational contour. We are also interested in the manifestations of tone and intonation at major syntactic boundaries within the sentence, but that is beyond the scope of this paper.

We have chosen Thai (Siamese) as the language for our study, because much work has been done on its five phonemic tones as well as other phonetic features (Abramson, 1962, 1978; Erickson, 1976). Also, more than enough work for our needs has been published on its syntax (Kuno & Wongkomthong, 1981; Panupong, 1970; Warotamasikkhadit, 1972). Finally, one of us (K.S.) is a native speaker.

PROCEDURE

For the present stage of our research we have used two speakers, a man and a woman, who are native speakers of Central Thai, the regional dialect upon which the standard language of Thailand is based. Both of them are currently graduate students at the University of Connecticut. (Of course, K.S. was not one of them.)

In experimental phonetic research there is a constant tension between the desire for perfectly relaxed vernacular speech and the need for utterances that can be easily analyzed and manipulated in the laboratory to yield a statistically satisfying data base. To what extent the understanding of phonetic phenomena has been distorted, paradoxically, by methodological constraints is not fully known. In our approach we have tried to have it both ways. Thus we used our informants to record two kinds of material, conversation as well as sentences composed by us.

After our informants became completely relaxed in the presence of the microphone, we succeeded in recording about 25 minutes of spontaneous conversation about the stresses and strains of graduate school and life in a foreign country. Because of the expected gross imbalances in the occurrences of sentence types and the five tones, we have so far made very little use of this material, although we hope to exploit it further.

For each syntactic slot in a three-word simple declarative sentence, we chose five tonally differentiated key words. All possible sequences of three words times five tones yielded 125 sentences. Because of grammatical and syntactic constraints, as well as the need to expand these sentences into longer complex sentences, we could not completely control for the immediate phonetic context of each tone, although we tried to foresee some difficulties in segmenting the key words out of the sentences. We expanded these basic sentences by inserting material between the key words. This yielded 125 complex declarative sentences of the same overall syntactic structure with each one containing an embedded relative clause. They were all of about the same length.

Each sentence was written in Thai script on an index card. Each speaker was instructed to peel one card off the top of the deck and read it in as natural, relaxed, and unemotional a fashion as possible, put the card down and then take the next card and repeat the procedure. Ultimately, each sentence was read three times by each speaker. To our ears the effect was certainly not one of spontaneous colloquial speech; nevertheless, the reading sounded like a perfectly normal Thai rendition for this special kind of speech behavior.

ANALYSIS

Using a cepstral method of F_0 extraction provided in the Interactive Laboratory System (ILS) package of computer programs, we analyzed the utterances for F_0 and overall amplitude, contours of which we displayed in synchrony with a wave form. Editing facilities on our VAX computer enabled us to enlarge selected portions of any utterance graphically and listen to them separately as outputs of our pulse-code modulation (PCM) system. We were also able to reject spurious records, especially at the onset or offset of an utterance, or to correct dubious values by making direct measurements of repetition rate on the wave form. The wave forms and amplitude displays were indispensable for setting the boundaries of the key words, especially in the complex sentences.[1]

Given the probable local shifts upward and downward of any overall F_0 intonation contour, not to speak of the tonally determined movements around the intonation contour in our data, it is necessary to choose a consistent criterion for the putative declination effect. Under the influence of the systematic, carefully reasoned and tested procedures of Cooper and Sorensen (1981), we have chosen the "top line" measurement, i.e., a line connecting the F_0 peaks at diagnostic points in the sentence.[2] In both the simple and complex sentences, we have found the highest F_0 value for each key word in first, second, and third position. The first peak is arbitrarily assigned the time of one second in order to imply that there may be speech before it and that, if so, its duration is irrelevant. The onsets of all the top lines are aligned on Peak 1 = 1 sec. For each succeeding peak, the length of time from the first peak is noted. The resulting tables of data show the extent to which any top line effect is present in our Thai material in the sense that, whatever happens in the individual utterances, each tone abstracted from the sentences ought to show declination as it is viewed through time across the three key positions. Also, it will then be possible to see whether the F_0

contours of the tones are affected by placement along any declination that may be found.

RESULTS

In this first report, we regret to say that we have not yet been able to analyze all the utterances of our two speakers. Indeed, we can only present data from a large percentage of the sentences read by our male informant.[3] A very brief look at the productions of the female informant and at the dialogue will be mentioned.

Simple Sentences

The F_0 data for all 125 tone-sequences uttered three times by our male informant are presented in Table 1. The average temporal placements of the peaks are also given, with Peak 1 arbitrarily set at 1 sec. Inspection of the table reveals no clear overall declination effect for the short declarative sentences. Since it was immediately apparent from close examination of the underlying tokens that the overall F_0 contours were being largely determined by the particular sequences of tones, we used a rather loose criterion to establish whether or not declination was present. We simply required that Peak 2 be at least 5 Hz lower than Peak 1 and Peak 3 at least 5 Hz lower than Peak 2. We found that only 12.8% of the utterances showed declination by this criterion. A very small sampling of data obtained so far from the second speaker, our female informant, does not contradict this finding.

Table 1 is arranged to show the average values for each tone as it occurs in each of three positions in the simple declarative sentences. Of the five tones only the *falling tone shows declination* by our 5-Hz criterion, although the mid tone almost makes the grade. That is, viewed across all the tonal sequences, these two tones show decreasing peaks as they move toward the end of the sentence. Even this observation is complicated by the fact, as shown in the column of grand means, that the falling tone has the highest average F_0 peak. If we look again at the tonal sequences, we find that when this tone occurs in final position, 71 out of 75 utterances show no declination. At the bottom of Table 1, the grand means for the peaks do indeed show a small decline from Peak 1 but no significant change between Peaks 2 and 3.[4]

We have decided not to do a close examination of the F_0 contours for the lexical tones in the simple sentences, because there was little or no declination to interact with them. As far as we can see, the main effects are those of coarticulation as observed in previous work (Abramson, 1979a; Gandour, 1974). That is, the "ideal" contours observed in citation forms are somewhat perturbed, particularly at their onsets and offsets, by coarticulation with neighboring tones and by the particular consonantal contexts; nevertheless, the full Thai system of five tones is preserved and each tone is readily identifiable both auditorily and graphically.

Table 1

Means and standard deviations of peak F_0 (Hz) and times of occurrence of the peaks (sec) for the key words (labeled by their tones) in the simple sentences. N=75 in each cell.

Peaks:		P1	P2	P3	F_0 Grand Means
Tones					
Mid	F_0	143.4	133.2	130.4	135.7
	SD	8.9	8.1	10.6	
	t	1.0	1.2	1.5	
	SD	0.0	0.1	0.1	
Low	F_0	131.1	122.1	130.4	127.9
	SD	12.7	10.5	7.5	
	t	1.0	1.2	1.6	
	SD	0.0	0.1	0.1	
High	F_0	150.7	147.4	150.5	149.5
	SD	10.2	6.6	9.8	
	t	1.0	1.4	1.6	
	SD	0.0	0.1	0.2	
Falling	F_0	171.8	167.2	149.4	162.8
	SD	9.9	8.2	7.5	
	t	1.0	1.3	1.6	
	SD	0.0	0.1	0.1	
Rising	F_0	140.9	132.9	139.1	137.7
	SD	13.5	7.7	9.4	
	t	1.0	1.4	1.8	
	SD	0.0	0.1	0.1	
Grand Means					
	F_0	147.6	140.6	140.0	
	t	1.0	1.3	1.6	

Complex Sentences

What with greater processing time, more segmentation problems, the necessity for separate graphic displays of the key words in addition to those of the whole sentences, and the need occasionally to redo the F_0 extraction of low-amplitude stretches, we have so far been able to examine somewhat fewer utterances of complex sentences. The F_0 data for 244 sentence tokens out of 375 (i.e., 111 tone sequences out of the expected 125) are presented in Table 2 for our male speaker. This table is organized in much the same way as Table 1, except that the grand means at the right and the bottom are weighted to reflect the uneven numbers (N) of items analyzed. Here it is to be recalled that the sequences of three tones have filler material between the key words.

By the criterion given under Simple Sentences, 38.9% of the utterances of the complex sentences showed declination. That is, for the long complex sentences we see a somewhat more overt declination for the single speaker examined so far than for the short sentences (12.8%).

Looking at Table 2 for an overall effect of declination on the peak values of the individual tones, we find in fact that all tones but the low tone (but cf. Figure 2 and Footnote 6) show lower frequency values for their peaks as they move from Peak 1 to Peaks 2 and 3. The grand means at the bottom of the table reflect this very clear trend. The column of grand means at the right shows once again that the falling tone has the highest peak value. Indeed, of the 47 utterances analyzed with the falling tone in second position, 36 show a higher peak for the second position than the first; for 10 of the remaining 11 sequences the peak of the falling tone in second position is lower than Peak 1 apparently only because the first position is also occupied by the falling tone.

Another major question of concern to us, as indicated in the Introduction, is the effect of the intonation line on the F_0 contours of the five tones of Thai. To this end, we needed an average F_0 contour for each tone in each of the three sentence positions. Of course, all tokens to be averaged first had to be normalized in time. By means of a computer program written for the purpose,[5] we obtained such displays as those shown in Figure 1. Again, for this stage of the research we had to restrict ourselves to an examination of a limited sample. Thus, only 20 of the available 50 F_0 curves for the rising tone in third position, normalized in time, are shown in the upper part of the figure. This sampling was taken at random from our computer tapes. The very small scatter in this bundle of curves suggests great stability in production. Indeed, the average of the 20 curves, which is shown in the lower part of the figure, could easily have been derived by eye.

The procedure illustrated in Figure 1 was followed for the five tones in all three positions of the complex sentences. The resulting average F_0 contours are shown in Figure 2. (The average curve in Figure 1 is, accordingly, presented in the rising-tone box with the label "3" for third position.)

Looking at the tonal shapes in Figure 2, we can make two broad observations: (1) The height of the overall tonal contour in the voice range drops progressively across the three positions.[6] (2) The contours that come closest to the ideal shapes known from earlier work (Abramson, 1962; Erickson, 1974) are best seen in the third position, which is pre-pausal.

Table 2

Means and standard deviations of peak F_0 (Hz) and times of occurrence of the peaks (sec) for the key words (labeled by their tones) in the complex sentences.

Peaks:		P1	P2	P3	F_0 Grand Means
Tones					
Mid	F_0	146.9	128.0	111.7	130.1
	SD	14.1	5.3	4.4	
	$\underline{t}$	1.0	2.2	3.1	
	$\overline{SD}$	0.0	0.2	0.3	
	N	46	45	37	
Low	F_0	140.2	111.0	118.3	122.6
	SD	11.3	4.1	9.7	
	$\underline{t}$	1.0	2.2	3.2	
	$\overline{SD}$	0.0	0.2	0.3	
	N	45	51	49	
High	F_0	160.6	142.6	132.3	144.7
	SD	5.6	5.4	10.0	
	$\underline{t}$	1.0	2.3	3.1	
	$\overline{SD}$	0.0	0.1	0.3	
	N	49	49	55	
Falling	F_0	179.1	166.9	132.5	159.4
	SD	12.6	5.0	6.7	
	$\underline{t}$	1.0	2.2	3.2	
	$\overline{SD}$	0.0	0.1	0.3	
	N	54	49	53	
Rising	F_0	149.0	134.1	111.3	131.5
	SD	7.7	6.3	6.0	
	$\underline{t}$	1.0	2.3	3.2	
	$\overline{SD}$	0.0	0.1	0.3	
	N	50	50	50	
Grand Means					
	F_0	156.0	136.6	122.1	
	$\underline{t_0}$	1.0	2.2	3.2	

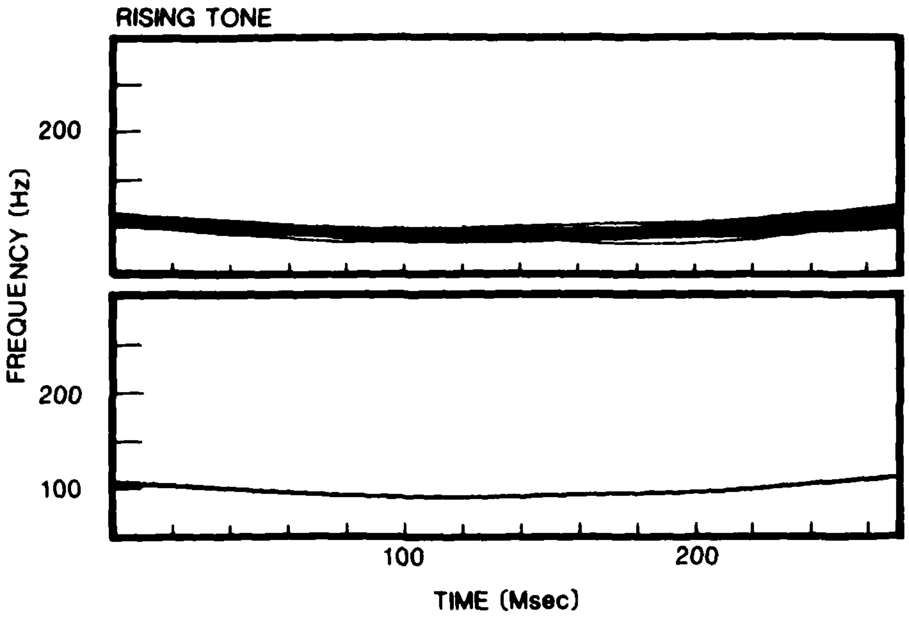


Figure 1. Rising tone in final position of the complex sentences. Above: A sample of 20 time-normalized F_0 contours. Below: The average contour of the 20 tokens.

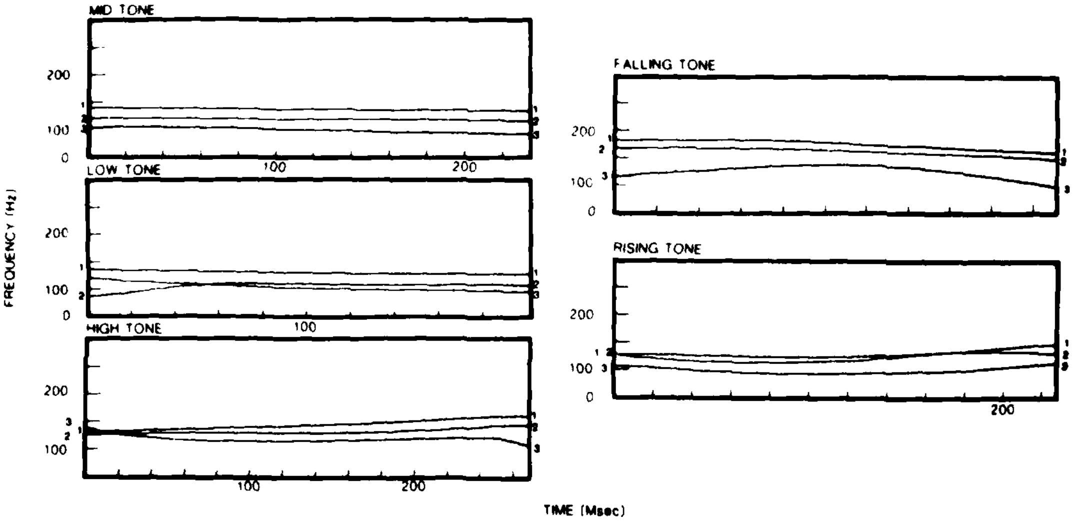


Figure 2. The average F_0 contour of 20 tokens of each of the five tones in the complex sentences. The numbers at each end of a curve show its position in the sentences.

Finally, there is the question of preservation of the five-way set of tonal distinctions. Again, as in the simple sentences, we have apparent effects of segmental and tonal coarticulation, but we also have the effects of a much more obvious declination. Even so, the full tonal system seems to be well preserved in all the key words, as shown by inspection of the time-normalized families of curves with their averages, exemplified in Figure 1, for all the tones in each position. As a matter of fact, all five tones are clearly distinct in shape, as shown in Figure 3, even when 60 randomly chosen tokens of each one are averaged across the three positions in the complex sentences. It is true that in our study there may be syntactic and prosodic factors that contribute to the maintenance of contrast. The key words in initial and final position probably have enough prominence in the sentence to discourage suspension of distinctions. The key word in second position occurs immediately after the end of an embedded relative clause where there may be a resetting of the tone-control mechanism (Erickson, 1976) even while the intonation continues to fall.

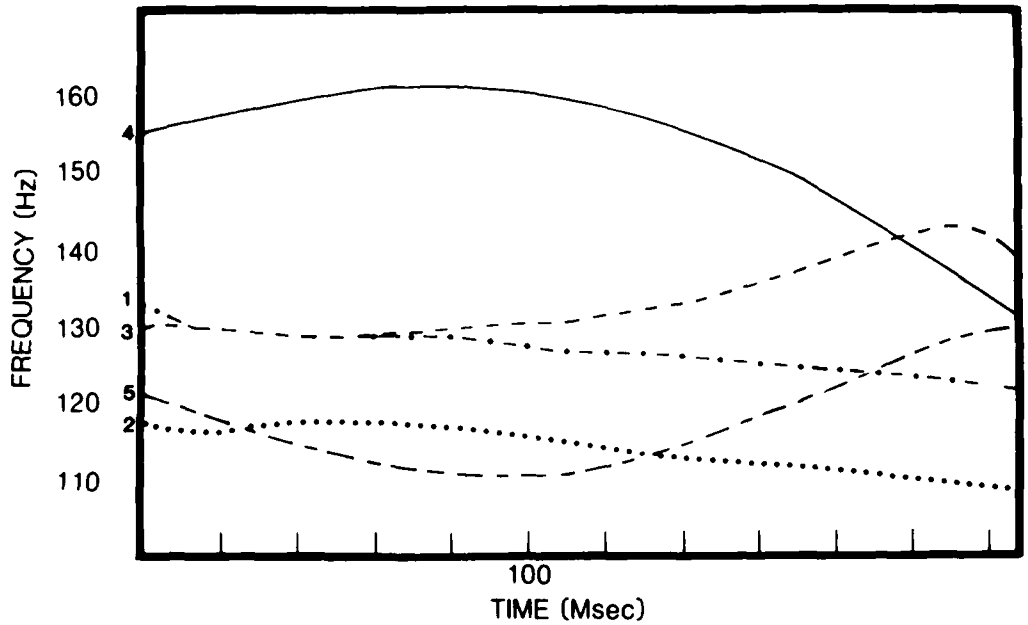


Figure 3. Average F_0 contours of the five tones in all three positions of the complex sentences. 1 = mid tone, 2 = low tone, 3 = high tone, 4 = falling tone, 5 = rising tone.

So far we have hardly taken more than a cursory look at a small portion of the conversation. Our hypothesis is that a more detailed examination will reveal that the sentence is not reliably the domain of the declination effect. Rather, declination may be most likely to occur at the end of each person's portion of the discourse before someone else takes his turn to speak. That is, its communicative value may be as a signal for turn-taking.

SUMMARY AND DISCUSSION

Our research has been built upon earlier work on Thai intonation (Abramson, 1979b; Henderson, 1949; Noss, 1972; Rudaravanija, 1965; Thongkum, 1976). Declination as a feature of sentence intonation is to be found in Thai and perhaps other tone languages, although it is much less clear-cut than in English[7] and perhaps other non-tone languages. In our short declarative sentences the perturbing effects of the local F_0 manifestations of the lexical tones are much more injurious to a global declination effect than in the long complex sentences in which the key words are separated by other speech material. Even in the long sentences, however, the effects of the tones make it very difficult, at least for now, to devise a formula, as has been done for English (Cooper & Sorensen, 1981), that would predict intermediate F_0 values of the top line. This is not surprising given the similar difficulty mentioned by Cooper and Sorensen in devising a top-line rule for Japanese, a "pitch accent" language that in the matter of moment-by-moment control of F_0 for linguistic purposes might be viewed as standing somewhere between a tone language like Thai and a non-tone language like English.

In our discussion of Table 1, our reasoning was that even though the individual tokens of simple sentences do not reliably show declination, if there is some kind of pre-programming toward this end on the part of the speaker, we might expect that a separate examination of each tone across the three positions would reveal a decline in peak values, thus manifesting declination in a more abstract way. This is not convincingly demonstrated. In the long complex sentences, on the other hand, such pre-programming is more readily apparent, although we may be dealing only with the physiological effect of coming to the end of a breath group (Lieberman, 1967), which does not reliably happen at the end of a very short sentence in Thai. Another sign of pre-programming would be somewhat higher F_0 values for Peak 1 of Table 2 for the long sentences than Peak 1 of Table 1 for the short sentences but the same values for Peak 3, an effect found for English (Cooper & Sorensen, 1981). That is, the speaker may be looking ahead to a final F_0 value and setting his onset F_0 so that his declination will be "right." Such an effect is apparent here only for Peak 1; the third peaks are in fact lower in Table 2. All in all, we may tentatively conclude that while long-range planning of the sentence exists, short-range planning plays a larger role in a tone language.

To the extent that the speaker's pre-programming of an utterance does include a certain amount of F_0 declination, the question still remains as to the domain of this feature (Umeda, 1982). Most work, including our own, has focused on the sentence as the traditional domain of intonation, yet some intonational features may go beyond the sentence to some larger unit of discourse (Lehiste, 1975). In particular, our cursory look at the very natural piece of dialogue that we succeeded in recording, suggests that a full analysis may reveal that declination is such a feature. In the reading of independent sentences in a laboratory setting, of course, the careful avoidance of list reading may yield a style in which the sentence itself is necessarily the domain of all global intonational features.

Finally, our findings show that sentence intonation, at least the kind of declarative intonation examined in this study, does not reduce the number of tonal oppositions in each key position. That is, although the absolute F_0

values of the tones move up and down with the intonation line, each of the five tones keeps its characteristic F_0 contour everywhere and, to the ear of the listener, its appropriate pitch contour.

REFERENCES

Abramson, A. S. (1962). *The vowels and tones of standard Thai: Acoustical measurements and experiments*. Bloomington, IN: Indiana University Research Center in Anthropology, Folklore and Linguistics.

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, *21*, 319-325.

Abramson, A. S. (1979a). The coarticulation of tones: An acoustic study of Thai. In T. L. Thongkum, P. Kullavanijaya, V. Panupong, & K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugénie J. A. Henderson* (pp. 1-19). Bangkok: Chulalongkorn U. Press.

Abramson, A. S. (1979b). Lexical tone and sentence prosody in Thai. *Proceedings of the 9th International Congress of Phonetic Sciences*, *2*, 380-387. Copenhagen: University of Copenhagen.

Cooper, W. E., & Sorensen, J. M. (1981). *Fundamental frequency in sentence production*. New York: Springer.

Erickson, D. (1974). Fundamental frequency contours of the tones of Standard Thai. *Pasaa*, *4*, 1-25.

Erickson, D. M. (1976). *A physiological analysis of the tones of Thai*. Unpublished doctoral dissertation, University of Connecticut.

Gandour, J. (1974). Consonant types and tone in Siamese. *Journal of Phonetics*, *2*, 337-350.

Henderson, E. J. A. (1949). Prosodies in Siamese: A study in synthesis. *Asia Major*, N.S.*1*, 189-215.

Kuno, S., & Wongkomthong, P. (1981). Relative clauses in Thai. *Studies in Language*, *5*, 195-226.

Lehiste, I. (1975). The phonetic structure of paragraphs. In A. Cohen & S. G. Nootebom (Eds.), *Structure and process in speech perception* (pp. 195-206). New York: Springer.

Lieberman, P. (1967). *Intonation, perception and language*. Cambridge, MA: MIT Press.

Lieberman, P., Landahl, K., & Ryalls, J. H. (1982). Sentence intonation in British and American English. *Journal of the Acoustical Society of America*, *71*, S112. (Abstract)

Lieberman, P., & Tseng, C. Y. (1980). On the fall of the declination theory: Breath group versus "declination" as the base form for intonation. *Journal of the Acoustical Society of America*, *67*, S63. (Abstract)

Noss, R. B. (1972). Rhythm in Thai. In J. G. Harris & R. B. Noss (Eds.), *Tai phonetics and phonology* (pp. 33-42). Bangkok: Central Institute of English Language.

Panupong, V. (1970). *Inter-sentence relations in modern conversational Thai*. Bangkok: The Siam Society.

Rudaravanija, P. (1965). *An analysis of the elements in Thai that correspond to the basic intonation patterns of English*. Unpublished doctoral dissertation, Teachers College, Columbia University.

Thongkum, T. L. (1976). Rhythm in Thai from another view point. *Pasaa*, *6*, 144-158.

Trager, G. L., & Smith, G. L. (1951). *An outline of English structure*. Norman, OK: Battenburg Press.

Umeda, N. (1982). "F_0 declination" is situation dependent. Journal of Phonetics, 10, 279-290.
Warotamasikkhadit, U. (1972). Thai syntax. The Hague: Mouton.

FOOTNOTES

[1] We are grateful to Charles Marshall for his help and advice in adjusting of the parameters of the cepstral algorithm to the voices of our speakers. We also wish to thank Stephen Eady and Louis Goldstein for the valuable special routines they designed to make the use of the computer programs so much easier.

[2] Although a baseline through F_0 minima has been used by some, Cooper and Sorensen argue (1981, p. 30) that the top line is better "because its associated F_0 peak values exhibited a variety of advantages over the bottom line, including relative ease of measurement, greater capability of being influenced by the speaker's coding of linguistic structures, and perceptual salience for the listener..."

[3] We hope to fill in the missing data in a continuation of this work.

[4] Of course, the F_0 peaks do not by themselves specify the tones. Thus it is not an anomaly to find in Table 1 that the mid and low tones have the same value for P3. Indeed, the low tone could even have a higher peak (cf. the same cell in Table 2).

[5] The program (OVERLAY) for time-normalizing curves and averaging them was written by Gerald Lame and then modified for some of our special needs by Michael Anstett.

[6] The apparent contradiction between this general tendency and the order 1, 3, 2 for the peaks of the low tone in Table 2 is to be ascribed to differences in the onsets of this tone. The peak frequencies are all at the beginning of this tone, which is best described, perhaps, as a low fall.

[7] The reliability of the declination effect in sentences, even for English, has been called into question (Lieberman, Landahl, & Ryalls, 1982).

SIMULTANEOUS MEASUREMENTS OF VOWELS PRODUCED BY A HEARING-IMPAIRED SPEAKER*

Nancy S. McGarr+ and Carole E. Gelfer+

Abstract. Perceptual judgments, acoustic measurements, and electromyographic (EMG) records were obtained for one deaf speaker producing the vowels [i, ɪ, æ , ɑ, ʊ ,u] in an [hVd] frame. Overall listener judgments were consistent with spectral measurements. In general, front vowels were perceived as more similar to targets than back vowels, and high vowels were perceived correctly more often than low vowels. Experienced and inexperienced listeners were found to differ significantly in their categorization of the point vowels [i, æ , ɑ , and u] but not for [ɪ and ʊ]. The vowel space, as determined by the formant frequency measures, was reduced with respect to normal values particularly in the region appropriate to high back vowels. However, EMG records of genioglossus and orbicularis oris do not entirely account for the perceptual and acoustic data. In particular, genioglossus activity is relatively undifferentiated across all vowels when compared to data from normals. The results of this study generally support the widespread notion of reduced vowel space secondary to a reduced range of tongue movement in this deaf speaker. The physiological records were also characterized by a significant degree of variability from token to token. In this regard, these data are different from acoustic and physiological patterns that have been previously reported for vowels produced by deaf speakers.

INTRODUCTION

Many previous studies have described the typical vowel errors produced by hearing-impaired speakers. These studies usually relied on perceptual assessments wherein experienced or inexperienced listeners transcribed the productions and the resulting error patterns were analyzed (e.g., Hudgins & Numbers, 1942; Smith, 1975). In these studies, hearing-impaired speakers were found to produce back vowels correctly more often than front vowels (Boone, 1966; Geffner, 1980; Mangan, 1961; Nober, 1967; Smith, 1975) and low vowels

---

*To appear in Language and Speech.

+Also Graduate School and University Center, The City University of New York.

Acknowledgment. We are grateful to Mrs. Margaret Hlibok for enthusiastically serving as a subject for this study, to Rosemary Szczesiul for assisting in the data analysis, and to Judith Rubin, Katherine S. Harris, and Peter Alfonso for providing helpful comments. This work was supported by NINCDS Grants NS-13617 and NS-13870, and NIH Biomedical Research Support Grant RR-05596 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-74/75 (1983)]

correctly more often than those with mid or high tongue positions (Geffner, 1980; Nober, 1967; Smith, 1975). On the other hand, Stein's (1980) cineradiographic study of five deaf speakers showed "fronting" of back vowels. Similarly, Crouter (1963) reported greater variation in tongue shape for [i] than for [u] and [ɑ] as measured by cinefluorography.

Hearing-impaired speakers also fail to distinguish between what has traditionally been referred to as the "tense-lax" distinction between vowel pairs such as [i-ɪ]. Often the substitution is to the tense member of the pair (Mangan, 1961; Monsen, 1974; Smith 1975), although other less closely related vowel substitutions have also been reported (Hudgins & Numbers, 1942; Markides, 1970).

The acoustic characteristics of vowels produced by deaf speakers have also been examined using techniques such as spectrographic analysis (Angelocci, Kopp, & Holbrook, 1964; Bush, 1981; Monsen, 1976) and linear predictive coding (LPC) (Osberger, Levitt, & Slosberg, 1979). Formant frequency measures show a reduced phonological space with formant values tending toward the neutral vowel [ʌ]. Monsen (1976) noted that the second formant of vowels produced by hearing-impaired children remained around 1800 Hz rather than varying as different vowels were articulated. Perceptual judgments and acoustic analyses have, thus, led some researchers (e.g., Angelocci et al., 1964; Horwich, 1977) to propose that hearing-impaired speakers use a limited amount of tongue movement and consequently do not achieve vowel differentiation. Some studies (Bush, 1981; Martony, 1968) suggest that deaf speakers who produce vowel distinctions do so by exaggerated variations in F_0, particularly for high vowels such as [i] and [u]. Existing physiological studies of deaf speech production--electromyography (Huntington, Harris, & Sholes, 1968; McGarr & Harris, 1983; Rothman, 1977) and cineflurography (Crouter, 1963; Stein, 1980; Zimmermann & Rettaliata, 1981)--are few and provide minimal information regarding vowel production.

Each type of investigation--descriptive, acoustic, and physiological--contributes partial insight into a deaf speaker's vowel production. However, only a few studies (cf. Huntington et al., 1968; Rothman, 1977) incorporated simultaneous acoustic and articulatory measures of production with listener judgments or phonetic transcriptions. The paucity of such studies is undoubtedly related to the considerable effort and specialized technology required to obtain such measures from deaf speakers. However, the information potentially gained from such simultaneous measures could greatly enhance our knowledge of speech organization in the deaf population.

This study was undertaken as a preliminary investigation of the hypothesis that deaf speakers fail to vary tongue position in their attempt to achieve vowel differentiation. EMG activity was recorded from the posterior genioglossus muscle and superior and inferior orbicularis oris of one deaf speaker. Listener judgments were obtained and acoustic analyses were performed in order to reconcile these measures with physiological records.

METHOD AND PROCEDURE

The pre-lingually deaf speaker (pure tone average for .5, 1, and 2 kHz = 105dB ISO) was a woman who attended an oral school for the deaf and also received remedial speech classes as an adult. Speech samples obtained from the subject were analyzed in several ways. First, a listener highly experienced with the speech of the deaf rated her spontaneous speech samples for overall intelligibility. Following the format described by Subtelny (1975), this subject was classified as difficult to understand, producing only occasional intelligible words or phrases. Second, judgments of vowel identity were obtained from five listeners experienced with the deaf and eighteen listeners who had no previous experience with deaf speech. Listeners were asked to identify the vowel they heard from a closed set of vowels and diphthongs. From these data, confusion matrices were derived. Third, narrow phonetic transcriptions were made by a phonetician. The listener judgments and phonetic transcriptions will be described further below.

Simultaneous acoustic and electromyographic recordings were made of the speaker's production of ten randomized repetitions each of the vowels: [i, ɪ, æ, ɑ, ʊ, u] in an [hVd] frame. Because of technical problems, only five repetitions of [ʊ] could be analyzed perceptually and acoustically; the EMG signals for this vowel could not be analyzed. Conventional hooked-wire electrodes were inserted into the posterior fibers of the genioglossus muscle, which elevates and bunches the main body of the tongue (Raphael & Bell-Berti, 1975; Raphael, Bell-Berti, Collier, & Baer, 1979). The electrode preparation and insertion techniques for this muscle have been reported in detail elsewhere (Hirose, 1971). Patterns of peak genioglossus activity for vowels produced by a hearing speaker are shown in Figure 1 for purposes of comparison with our data (Alfonso & Baer, 1982). This figure shows that greater muscle activity occurs for the front vowels [i] and [ɪ], and to a lesser extent, [u]; the genioglossus shows relatively little activity for [ɑ]. Thus, genioglossus appears to be active for high vowels in general and for front vowels in particular.

Measures were also made of lip-rounding activity using surface electrodes to record from the superior and inferior orbicularis oris muscles (Allen, Lubker, & Harrison, 1972). It was assumed that only [ʊ] and [u] would show significant oribicularis oris activity.

The acoustic and electromyographic (EMG) data obtained from the deaf speaker were analyzed in the following three ways. First, the experienced listeners' judgments were used to sort the production tokens into three categories: 1) perceptually correct productions (at least 4 of the 5 listeners agreed with the intent of the talker), 2) perceptually incorrect productions (4 or more listeners disagreed with the intent of the talker), and 3) perceptually equivocal (2 or 3 listeners heard the vowel as intended; the remaining heard it as incorrect). Second, spectral analyses and vowel duration measurements were performed on an interactive computer system at Haskins Laboratories. Third, the EMG signals were rectified, integrated, and further analyzed as previously described (Kewley-Port, 1973).

RESULTS

A. Listener Judgments

Table 1 shows the confusion matrices obtained from the listeners' scores. Fifty judgments were obtained from the experienced listeners (5 listeners x 10 repetitions) for each vowel; 180 judgments were obtained from the

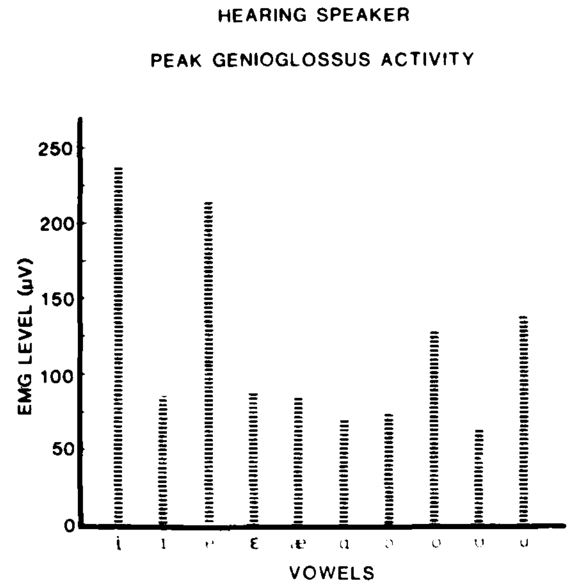


Figure 1. Peak genioglossus activity in microvolts (μ V) for vowels produced by a male speaker with normal hearing (after Alfonso & Baer, 1982).

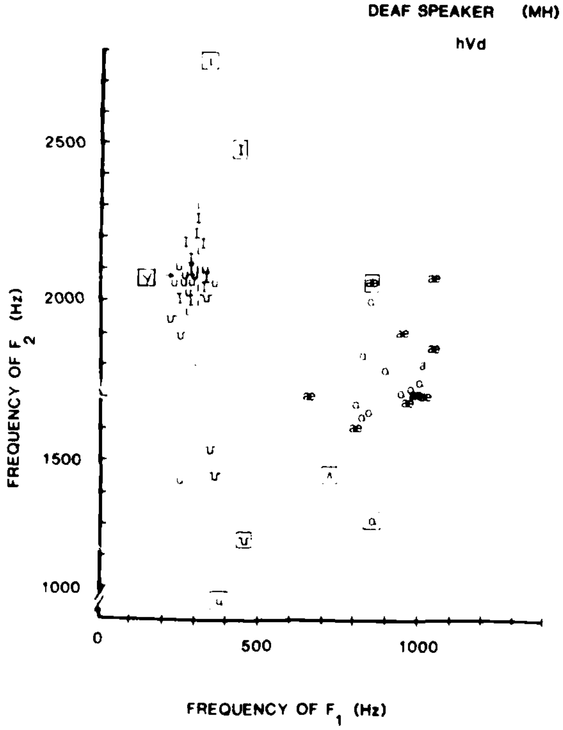


Figure 2. Formant values of F_1 and F_2 for five vowels produced by the deaf speaker. Values in squares are the average formant values for (non-deaf) women reported by Peterson and Barney (1952). Values for [y] are from Fischer-Jørgenson (1960).

Table 1

Confusion matrices of listeners' judgments for vowels produced by the deaf speaker. Scores are reported as percentages.

Experienced Listeners

Target vowel	i	ɪ	ɛ	æ	ɑ	ɔ	ʊ	u	eɪ	aɪ	ɝ	Other vowels
i	74	16										10
ɪ	48	40						2		10		
æ	2	4	32	42	10							10
ɑ	6	4	26	42	8							14
ʊ	32	12					16	32				8
u	46	24	2				8	14				6

Inexperienced Listeners

Target vowel	i	ɪ	ɛ	æ	ɑ	ɔ	ʊ	u	eɪ	aɪ	ɝ	Other vowels
i	42	21	1	1	.5	1	4	13	6	3	-	7.5
ɪ	27	34	2	3	1	1	8	4	7	6	-	7
æ	5	11	11	15	5	.5	7	5	12	12	7	9.5
ɑ	8	13	7	14	6	2	9	4	11	9	8	8
ʊ	22	17	2		1		16	25	5	1		11
u	30	24	4	2	.5	1	8	9	7	2	2	10.5

---

inexperienced listeners (18 listeners x 10 repetitions). Percentages are reported for each listening group for each vowel. In general, the pattern of correct responses is similar for the two groups of listeners. Overall, listeners perceived the front vowels [i], [ɪ], and [æ] as correct more often than the back vowels [ɑ, ʊ or u]. Confusions for the high front vowels [i] and [ɪ] were most often restricted to this tense-lax pair although this was not the case for other vowel pairs. Substitution errors occurred across the vowel space for other target vowels. Of significance is the considerable number of [i] or [ɪ] substitutions for [ʊ] or [u] targets. Percentages of correct judgments for experienced and inexperienced listeners across all vowel types (taken from Table 1), and their averages, are summarized in Table 2. Table 3 shows the ranking of the most common combined listener responses (again taken from Table 1) for each vowel. It is interesting that vowels tended to be judged as more fronted than their targets. A two-by-two Chi-square analysis was performed on the most common listener response versus all other choices in order to ascertain if the two groups of listeners differed in their categorizations. There was a significant difference between experienced and inexperienced listeners for the vowels [i] (χ^2 16.4, $p < .01$), [æ] (χ^2 17.3, $p < .01$), [ɑ] (χ^2 18.3, $p < .01$, [u] (χ^2 4.5, $p < .05$) but not for the vowels [ɪ and ʊ]. That is, both groups of listeners tended to cluster their responses for the lax vowels, while the experienced listeners' responses also clustered for the point vowels. Inexperienced listeners, on the other hand, were more scattered in their responses for the point vowels.

B. Acoustic Measures

Figure 2 shows the values for F_1 and F_2 for all tokens of all vowels. These measurements were taken at the center, and relatively steady-state, portion of the vowel. Formant values for F_1 grossly differentiate between high and low vowels, while the range of F_2 variation is restricted. These latter values imply limited backward movement of the tongue. In an attempt to produce the back vowel [ɑ], this speaker succeeds only in approaching mid range. Thus, the values for the low vowels [æ] and [ɑ] cluster, and the tendency for listeners to perceive [ɑ] as [æ] is not surprising. The F_2 values for [u] are grouped with [i] and [ɪ], and thus, an acoustic basis for the listeners' perceptual judgments becomes somewhat more apparent. Some formant values for [ʊ] are similarly found to have a high F_2, although two tokens show a more appropriate formant range.

Because these acoustic data are not totally adequate in explaining listener identification accuracy, particularly in discriminating [i] from [ɪ], it seemed reasonable to assume that some other acoustic cue must be available to the listeners. Figure 3 shows F_2 plotted against duration for all vowels. It can be seen that the vowels [i] and [ɪ] are differentiated on the basis of duration, with values for [i] considerably longer than those for [ɪ]. Differentiation of vowels such as [i] and [ɪ] on the basis of durational cues has been noted previously for deaf speakers (Angelocci et al., 1964; Levitt, Osberger, & Stromberg, 1979; Monsen, 1974). There is no clear differentiation of other vowels based on durational cues. Overall durations of vowels produced by this deaf speaker were considerably longer than those reported for normals, which is frequently observed for hearing-impaired speakers (Calvert, 1961; Osberger & Levitt, 1979).

Table 2

Percentage of correct judgments for each vowel.

	i	ɪ	æ	ɑ	ʊ	u
Inexperienced	42	34	15	6	16	9
Experienced	74	40	42	8	16	14
MEAN	58%	37%	28.5%	7%	16%	11.5%

Table 3

Listener responses for each vowel in rank order.

i	ɪ	æ	ɑ	ʊ	u
i	i	æ	æ	u	i
ɪ	ɪ	ɛ	ɛ	i	ɪ
	ʊ	ɑ	ɪ	ʊ	u
	eɪ	ɪ	ɑ	ɪ	ʊ
	aɪ	aɪ	i		ɛ

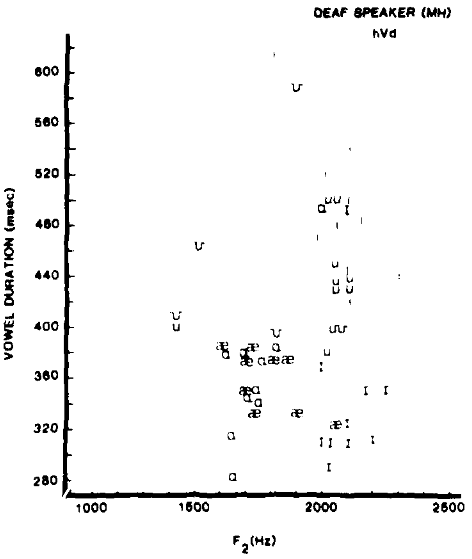


Figure 3. Plot of F_2 values and vowel duration measures for the deaf speaker's productions.

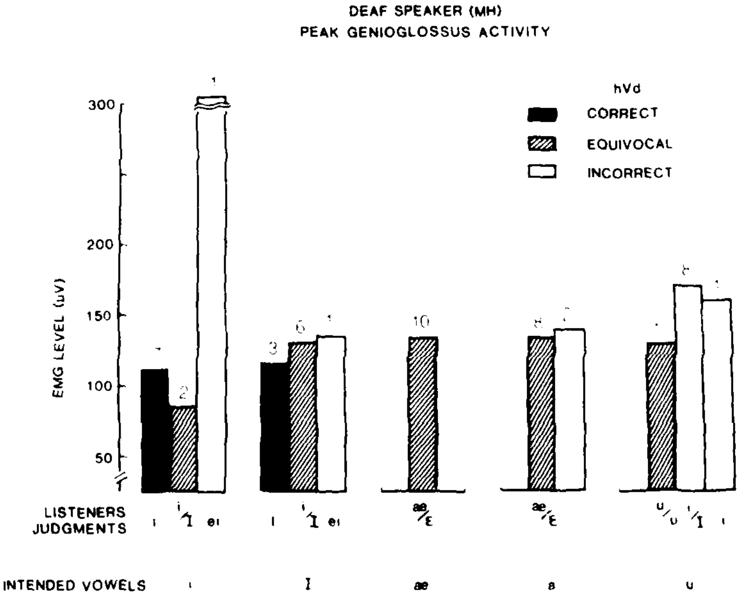


Figure 4. Peak genioglossus activity in microvolts (μV) for vowels produced by the deaf speaker. At the top of each column, the number of experienced listeners whose judgments fell into each category (perceptually correct, equivocal, or incorrect) are noted. At the bottom of each column are noted the vowel judgments assigned by the listener to the corresponding token. See text for more detailed discussion.

C. EMG Analysis

Figure 4 shows the patterns of peak posterior genioglossus activity for the five vowels analyzed for this deaf speaker. EMG activity for [ʊ] could not be analyzed due to technical problems. Perceptually correct productions, perceptually incorrect productions, and perceptually equivocal productions are plotted. The data show an obvious lack of differentiated peak genioglossus activity across nearly all vowels regardless of perceptual category. However, one would expect more genioglossus activity for [i] and [ɪ], somewhat less for [u], and still less for [æ] and [ɑ] (see Figure 1). This pattern was not observed even for this speaker's correct productions. Furthermore, peak genioglossus activity was not greater for the vowel [i] than for the vowel [ɪ], as might be observed in the productions of hearing speakers (Alfonso & Baer, 1982; Raphael & Bell-Berti, 1975). Furthermore, values of peak genioglossus activity for all incorrect categories of [u] were greater than values obtained for any perceptually correct high front vowel.

Because of this unexpected pattern of genioglossus activity for [u] as well as the number of listeners who judged the production as [i] (cf. Table 1), narrow phonetic transcriptions were obtained. Eight of the ten tokens intended as [u] were transcribed by a trained phonetician as [y], a high front rounded vowel not typical in American English. Figure 5 shows a comparison of genioglossus activity for selected tokens intended and transcribed as [i] with those intended as the vowel [u] but transcribed as [y] (and perceived as [i] by our listeners, cf. Figure 4). Both sets of tokens are distinguished by variability in the onset and offset of genioglossus activity. In some instances (e.g., token 1 for correct [i] productions), onset of genioglosssus occurs quite early, while for other tokens (e.g., token 3), the onset is considerably later. It is noteworthy that, despite token-to-token variability for both correct and incorrect productions, the overall pattern of activity for the two categories is nearly identical. That is, no single distinguishable peak of muscle activity is identifiable with production of a high front vowel.

Figure 6 shows genioglossus and orbicularis oris activity for three utterances: [i] correct, [u] equivocal, and [y/u] substitutions (i.e., [u] incorrect). There is no token that four of the five experienced listeners judged correctly as [u]. For both the equivocal productions of [u] and those transcribed as [y], there is the expected orbicularis oris activity associated with lip rounding. However, while it is difficult to state with certainty what differentiates the last two categories, in the equivocal case, orbicularis oris activity is maintained as long as that for genioglossus, while for [y] orbicularis oris activity ceases earlier and genioglossus activity begins sooner. Thus, it is possible that the temporal relationship between orbicularis oris and genioglossus represents at least one of the underlying bases for the acoustic cues that lead to different listener impressions.

DISCUSSION

The acoustic results of the present study are in general agreement with those of previous studies in demonstrating a reduced vowel space. However, the reduction appears to occur mostly in the front-back dimension, with the

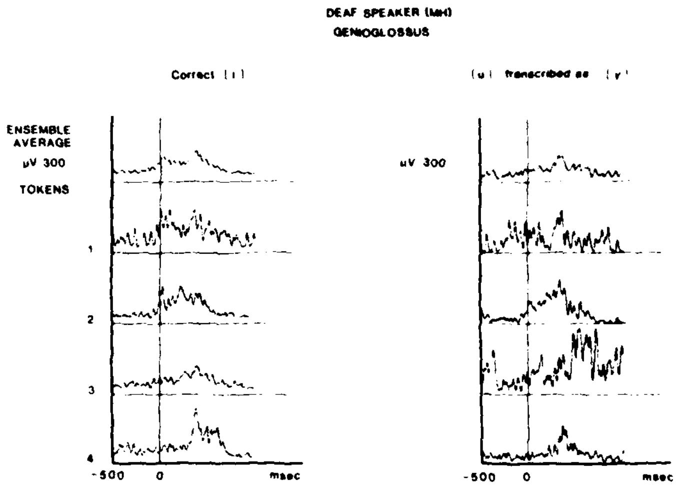


Figure 5. Genioglossus activity in microvolts (μV) for selected tokens of vowels produced by the speaker. At the left, tokens transcribed by the phonetician as a correct production of [i], at the right, tokens intended as [u] but transcribed as [y]. Data plots show the ensemble average for 7 tokens of [i], and 8 tokens of [u] for the genioglossus muscle. Four individual tokens are shown below. The vertical line, the line-up point at 0 ms for these measures, is the onset of voicing for the vowel.

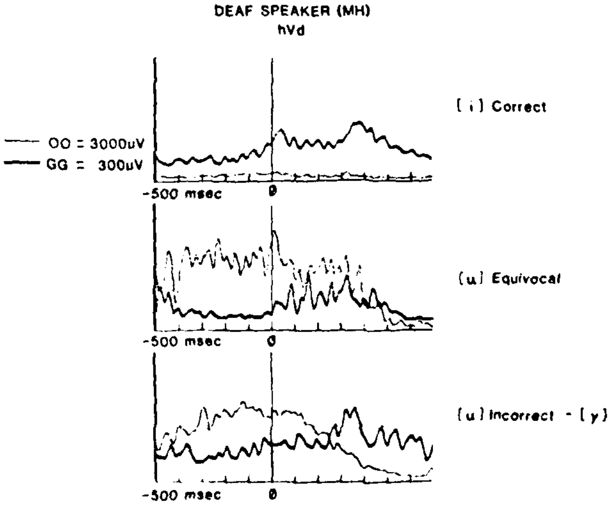


Figure 6. Selected individual tokens of the EMG potentials from the genioglossus and orbicularis oris muscles as produced by the deaf speaker. The line-up point is as in Figure 5. Offset of voicing occurs 500 ms after the line-up point. Tokens shown are [i] judged as correct, (top), [u] as equivocal (mid), [u] as [y] (bottom). See text for further discussion.

high vowels [i, ɪ , u] and some tokens of [ʊ] clustered around a high F_2 (range = 1975-2300 Hz) and the low vowels [æ , ɑ] clustered around the mid-range F_2 (range = 1600-2075 Hz). In general, the judgments of both experienced and inexperienced listeners are consistent with the acoustic measures, although the experienced listeners, on average, made more correct judgments than the inexperienced listeners (cf. Table 2). The higher scores achieved by the experienced listeners may be attributed to this group's ability to disambiguate [i] from [ɪ], and [æ] from other front vowels (cf. Table 1). The data also show that this speaker tends to produce front vowels more often than back vowels, whether correct or incorrect, and to produce high vowels correctly more often than low vowels. These data thus differ from previous descriptive studies of vowels produced by deaf speakers that report better production of back or low vowels (Boone, 1966; Geffner, 1980; Mangan, 1961; Nober, 1967; Smith, 1975), although the data concur with results obtained in cineradiographic studies (Crouter, 1963; Stein, 1980).

It is apparent that there is an acoustic basis for the listeners' judgments. Formant values for [i] and [ɪ] fall roughly in the appropriate range so that the relatively high number of correct judgments for these vowels can be explained. Similarly, formant values for this speaker's intended productions of [u] account for the high percentage of [i] and [ɪ] listener judgments and the [y] judgments of the phonetician. This speaker had considerable success in differentiating high and low vowels, although the formant values for the low vowels are inappropriate with respect to normal productions. Thus, the acoustic basis for the very low percentage of [ɑ] judgments is readily explained. In fact, overall there is a fairly straightforward relationship between the acoustic measures and the listener judgments.

We are limited in our inferences regarding the physiological basis of the acoustic data in that only one tongue muscle (posterior fibers of the genioglossus) was studied. Therefore, the implied failure to produce back tongue movements from the acoustics cannot be confirmed physiologically. However, we can address ourselves to the relative appropriateness of the degree of genioglossus activity for all vowels. As noted in Figure 3, genioglossus activity for this deaf speaker is, on average, relatively undifferentiated across all vowels studied. This is in striking contrast to the results for a normal speaker (cf. Figure 1). Furthermore, even for tokens that were perceived as correct, onset and offset of genioglossus activity was highly variable from token to token. It is not surprising, then, that there are so many equivocal and incorrect productions. Furthermore, the pattern of EMG activity also does not readily distinguish between [i] and [ɪ], so that the corresponding listener judgments seem to be based primarily on duration cues. However, the relatively uniform level of genioglossus activity for [i,ɪ,u] does explain the general tendency for F_2 values to occur in regions expected for high front vowels.

Therefore, based on acoustic and physiological measures, we conclude that this deaf speaker fails to vary tongue position, particularly in the front-back dimension, in order to achieve vowel differentiation. Although the vowel space is reduced overall, there is considerable differentiation in the high-low plane, as is evident from the ranked listener responses in Table 3. Productions of [u] differed from [i] primarily on the basis of lip-rounding, as noted in the electromyographic records of the orbicularis oris. Such a production strategy is not suprising when we consider that, owing to difficulty in perceiving acoustic cues, deaf speakers rely heavily on visual information for deriving cues to place of articulation. Examples of these would include lip-rounding, as noted above, and jaw lowering for production of low vowels. When acoustic cues can be perceived with limited residual hearing, e.g., vowel duration, the speaker employs these, as noted for the tense-lax pair [i-ɪ].

While this study is only preliminary, it provides some insight into the physiological differences between deaf and hearing speakers in vowel production. We are intrigued by the token-to-token variability noted in the onset and offset of genioglossus records. We intend to examine this issue further by examining other tongue muscles that are known to be important in vowel production, particularly the extrinsic muscles, hyoglossus and styloglossus. In addition, we will investigate the hypothesis that deaf speakers, such as our subject, who do not vary tongue position, achieve vowel differentiation by exaggerated variation in larynx height and fundamental frequency.

REFERENCES

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

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

Angelocci, A., Kopp, G., & Holbrook, A. (1964). The vowel formants of deaf and normal hearing eleven to fourteen year old boys. Journal of Speech and Hearing Disorders, 29, 156-170.

Boone, D. (1966). Modification of the voices of deaf children. Volta Review, 68, 686-692.

Bush, M. (1981). Vowel articulation and laryngeal control in the speech of the deaf. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Calvert, D. R. (1961). Some acoustic characteristics of the speech of profoundly deaf individuals. Unpublished doctoral dissertation, Stanford University.

Crouter, L. (1963). A cinefluorographic comparison of selected vowels spoken by deaf and hearing subjects. Unpublished master's thesis, University of Kansas.

Fischer-Jørgensen, E. (1960). Almen fonetik (Illustrationer). Copenhagen: Rosenkilde og Bagger.

Geffner, D. (1980). Feature characteristics of spontaneous speech production in young deaf children. Journal of Communication Disorders, 13, 443-454.

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

Horwich, E. J. (1977). An examination of the vowels of deaf girls. S.B. thesis, Massachusetts Institute of Technology.

Hudgins, C. V., & Numbers, F. C. (1942). An investigation of the intelligibility of the speech of the deaf. Genetic Psychological Monographs, 25, 289-392.

Huntington, D., Harris, K. S., & Sholes, G. (1968). An electromyographic study of consonant articulation in hearing-impaired and normal speakers. Journal of Speech and Hearing Research, 11, 147-158.

Kewley-Port, D. (1973). Computer processing of EMG signals at Haskins Laboratories. Haskins Laboratories Status Report on Speech Research, SR-33, 173-184.

Levitt, H., Osberger, M. J., & Stromberg, H. (1979). Acoustic characteristics of [i-ɪ] confusions in deaf children's speech. Journal of the Acoustical Society of America, 66, S13 (Abstract).

Mangan, K. (1961). Speech improvement through articulation testing. American Annals of the Deaf, 106, 391-396.

Markides, A. (1970). The speech of deaf and partially hearing children with special reference to factors affecting intelligibility. British Journal of Disorders in Communication, 5, 126-140.

Martony, J. (1968). On the correction of the voice pitch level for severely hard of hearing subjects. American Annals of the Deaf, 113, 195-202.

McGarr, N. S., & Harris, K. S. (1983). Articulatory control in a deaf speaker. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.), Speech of the hearing impaired: Research training and personnel preparation. Baltimore: University Park Press, 1983.

Monsen, R. B. (1974). Durational aspects of vowel production in the speech of deaf children. Journal of Speech and Hearing Research, 17, 386-398.

Monsen, R. B. (1976). Normal and reduced-phonological space: The production of English vowels by deaf adolescents. Journal of Phonetics, 4, 189-198.

Nober, H. (1967). Articulation of the deaf. Exceptional Child, 33, 611-621.

Osberger, M. J., & Levitt, H. (1979). The effect of timing errors on the intelligibility of deaf children's speech. Journal of the Acoustical Society of America, 66, 1316-1324.

Osberger, M. J., Levitt, H., & Slosberg, R. (1979). Acoustic characteristics of correctly produced vowels. Journal of the Acoustical Society of America, 66, S13. (Abstract)

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

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

Raphael, L. J., Bell-Berti, F., Collier, R., & Baer, T. (1979). Tongue position in rounded and unrounded front vowel pairs. Language and Speech, 22, 37-48.

Rothman, H. (1977). An electromyographic investigation of articulation and phonation patterns in the speech of deaf adults. Journal of Phonetics, 5, 369-376.

Smith, C. R. (1975). Residual hearing and speech production of deaf children. Journal of Speech and Hearing Research, 18, 795-811.

Stein, D. (1980). A study of articulatory characteristics of deaf talkers. Unpublished doctoral dissertation, University of Iowa.

Subtelny, J. D. (1975). Speech assessment of the deaf adult. Journal of the Academy of Rehabilitative Audiology, 8, 110-116.

Zimmermann, G., & Rettaliata, P. (1981). Articulatory patterns of an adventitiously deaf speaker: Implications for the role of auditory information in speech production. Journal of Speech and Hearing Research, 24, 169-178.

EXTENDING FORMANT TRANSITIONS MAY NOT IMPROVE APHASICS' PERCEPTION OF STOP CONSONANT PLACE OF ARTICULATION*

Karen Riedel+ and Michael Studdert-Kennedy++

Abstract. Synthetic speech stimuli were used to investigate whether aphasics' ability to perceive stop consonant place of articulation was enhanced by the extension of initial formant transitions in CV syllables. Phoneme identification and discrimination tests were administered to twelve aphasic patients, five fluent and seven non-fluent. There were no significant differences in performance due to the extended transitions, and no systematic pattern of performance due to aphasia type. In both groups, discrimination was generally high and significantly better than identification, demonstrating that auditory capacity was retained, while phonetic perception was impaired; this result is consistent with repeated demonstrations that auditory and phonetic processes may be dissociated in normal listeners. Moreover, significant rank order correlations between performances on the Token Test and on both perceptual tasks suggest that impairment on these tests may reflect a general cognitive rather than a language-specific deficit.

Some researchers have attributed speech comprehension deficits in aphasia to a defect in the processing of acoustic information in the speech signal. Tallal and Newcombe (1978) proposed a connection between nonverbal auditory processes, phonetic perception, and spoken language comprehension. They hypothesized that aphasics have a primary defect in temporal analysis affecting their ability to process rapidly changing acoustic cues. They suggested that this defect is responsible not only for failure to perceive specific phonemes, but also for a variety of other temporal processing problems

---

*Submitted to Brain and Language.

+New York University Medical Center

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

Acknowledgment. The experiments reported here are drawn from a dissertation submitted by Karen Riedel to the City University of New York. The research was conducted at the Institute of Rehabilitation Medicine, New York University Medical Center and was supported, in part, by Grant NICHD 01994 to Haskins Laboratories. The work was facilitated by the Rehabilitation Services Administration, Department of Health, Education and Welfare, under the designation of New York University as a Rehabilitation Research and Training Center 16-P-56801/2-18. The authors wish to thank Martha T. Sarno, Director, and the staff of Speech Pathology Services, Institute of Rehabilitation Medicine, for their cooperation and support. We also thank Lawrence Raphael and Donald Shankweiler for comments on the manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-74/75 (1983)]

compromising aphasics' ability to understand speech. The present study tests this hypothesis on a group of post-CVA aphasics.

Tallal and Newcombe trained a group of 10 missile-wounded, left-brain-damaged subjects to identify, with a button press, contrasting pairs of 3-formant synthetic syllables, differing in the direction of their second formant transitions. The syllables were to be identified as either /ba/ or /da/. One pair of syllables had short (40 ms) transitions on all formants, the other had extended (80 ms) transitions. Training continued to a criterion of 20 correct out of 24 consecutive responses or until 48 trials had been given. Only 4 of their 10 subjects reached criterion on the syllables with short formant transitions, but 7 out of 10 reached criterion on the syllables with extended formant transitions. The six subjects who had difficulty on the short transition syllables also made the greatest number of errors on a nonverbal sequencing task, in which they had to specify the order of two tones, presented with very brief (from 8 to 305 ms) intervals between them. Impairment on the latter task correlated highly with impairment on the Token Test (DeRenzi & Vignolo, 1962). Given these findings, Tallal and Newcombe inferred a causal chain from impairment in judgments of rapidly presented nonverbal sequences to impairment in the perception of phonetic contrasts, signaled by rapid formant transitions, to impairment in language comprehension.

We should note an ambiguity in the interpretation of the improvement in aphasics' place of articulation judgments, attributed by Tallal and Newcombe to transition extension. Research with normal listeners has demonstrated that identifications of syllable-initial stop consonants shift in manner from stop to glide when formant transitions are extended (Liberman, Delattre, Gerstman, & Cooper, 1956; Miller & Liberman, 1979). For example, an increase in the duration of bilabial transitions from 30 to 60 ms shifts judgments from predominantly /b/ to predominantly /w/: The boundary between the two manner classes averages 40 ms. Was it then the extension of formant transitions *per se* that improved aphasics' performance or was it the shift to a different phonetic contrast? This ambiguity would not have arisen if Tallal and Newcombe had blocked the manner shift by confining formant transition extension to those formants (F2 and F3) that carry place of articulation information, while leaving the formant that carries manner information (F1) unchanged.

Other experimenters have used synthetic speech to examine the speech perception abilities of aphasics (e.g., Basso, Casati, & Vignolo, 1977; Blumstein, Cooper, Zurif, & Caramazza, 1977; Kellar, 1979). This research, limited to studies of voice-onset-time (VOT) perception, has indicated that aphasics of both major diagnostic categories, nonfluent (Broca's) and fluent (Wernicke's) have unusual difficulty in reliably assigning stimuli from a VOT continuum to one of two classes. However, some aphasics who perform poorly on this phoneme identification task perform almost normally when asked to judge whether paired stimuli from the VOT continuum are the same or different. This finding shows that in aphasia, the discrimination of acoustic parameters may be functionally separable from phoneme identification. Moreover, these studies and others (e.g., Auerbach, Naeser, & Mazurski, 1981) have found little evidence of a direct connection between disorders of phonetic perception and reduced general comprehension of speech.

The goals of the present study were therefore: (1) to look for an improvement, similar to that reported by Tallal and Newcombe, in aphasics' identification and discrimination of stop consonant place of articulation, both when all three syllable-initial formant transitions were extended and when only F2 and F3 transitions were extended, and (2) to assess the relation between aphasics' performances on these tasks and their language comprehension, as measured by the Token Test.

METHOD

Test Materials

Three pairs of syllables were synthesized on the Haskins Laboratories parallel resonance synthesizer. The pairs differed from each other only in the formant patterns used to render /ba/ vs. /da/. The stimulus patterns for pairs 1 and 2 were modeled after those used by Tallal and Newcombe and described by Tallal and Piercy (1974, 1975).[1] All stimulus patterns began with 13 ms of prevoicing and were followed by a three-formant pattern. Values are listed in Table 1. The durations of all three formant transitions were 30 ms in the first pair and 82 ms in the second pair. The third pair was identical to pair 2 except that formant transition extension was confined to those formants (F2 and F3) that carry most of the place of articulation information, while the formant that carries manner information (F1) was left unchanged. Formant transitions for all pairs were followed by a steady state portion sufficient to produce an overall stimulus duration of 250 ms.

Table 1

Onset and ending values of the three pairs of formant transition patterns used for identification and discrimination

	/b/		/d/	
	Onset	Ending	Onset	Ending
F1	202	688	202	688
F2	848	1077	1535	1077
F3	2193	2527	3029	2527

Subjects

Twelve adult aphasic out-patients of the Institute of Rehabilitation Medicine, New York University Medical Center, New York City, were tested. Subjects were limited to individuals who had sustained a left hemisphere CVA, were native English speakers, and had no history of neurological impairment before the onset of aphasia. All were screened for normal peripheral hearing through the speech frequencies. The mean length of time post-onset was 3.2 years (range 1 to 6 years). Their mean age was 55 years (range 36 to 66 years). A wide range of aphasia severity was reflected in the group, from mild to severe speech/language disturbance. Subjects were categorized into two types, fluent and nonfluent, on the basis of clinical examination and an analysis of speech characteristics (Goodglass & Kaplan, 1972). Auditory comprehension impairment was assessed with the Token Test (Spreen & Benton, 1977).

General Procedure

Subjects were tested individually in an IAC soundproofed chamber. The tape recorded stimuli were played on a Wollensak 1520 tape recorder and presented free field at a comfortable loudness level.

IDENTIFICATION TESTS

These tests were designed to answer the following questions:

1. Does extension of initial stop consonant formant transitions contribute to improved phoneme identification in aphasic subjects (a) when all three formant transitions are extended and/or (b) when formant transition extension is confined to F2 and F3?

2. Is any improvement produced by extending the formant transitions of stop-vowel syllables confined to a specific subtype of aphasia?

3. Is phoneme identification performance associated with performance on the Token Test?

Identification Procedure

Subjects were told that they would hear computer-generated syllables that sounded like "ba" or "da." Sample syllables (four /ba/ and four /da/) were presented. The identification task, which consisted of marking the correct syllable on a prepared answer sheet, was demonstrated by the experimenter. To familiarize subjects with the task, twelve practice items were then presented. These were followed by a 24 item (12 tokens of each syllable) randomized phoneme identification test, with 4 seconds between items.

Each identification test was followed by two discrimination tests (described below). The entire set of identification and discrimination tests was then repeated in reverse order. Testing was accomplished in 2 to 3 one half-hour sessions.

Results

The first four data columns of Table 2 present the individual and mean percent correct for the two aphasic groups. No differences in accuracy of phoneme identification were found among the synthetic pairs. Wilcoxon Matched Pairs tests (for 1 vs. 2, 2 vs. 3, and 1 vs. the average of 2 and 3), carried out on subjects whose score on pair 1 was less than 100% (N=9) and on subjects whose score on pair 1 was less than 90% (N=6) yielded no significant differences.

Type of aphasia also had no significant effect on performance of the identification tests. Certain individuals in both groups were prone to errors in identification, but others, specifically the milder aphasics, encountered no difficulty.

Table 2 (rightmost column) lists individual Token Test scores. A significant rank order correlation between identification scores and Token Test performance was found, $r = .83$, $p < .01$.

DISCRIMINATION TESTS

These tests were designed to answer the following questions:

1. Is aphasics' discrimination of stop-vowel syllables improved (a) when all three formant transitions are extended and/or (b) when formant transition extension is confined to F2 and F3?

2. Does reducing the inter-stimulus interval (ISI) between syllables affect discrimination performance?

3. Is there a difference between aphasics' ability to identify syllables and their ability to make same-different judgments about them?

4. Is there a correlation between phoneme discrimination and Token Test performance?

Subjects

Eleven of the subjects who were tested on identification were also tested on discrimination. One aphasic failed to understand task demands even after repeated trials and therefore was eliminated from discrimination testing.

Discrimination Procedure

The stimuli were identical to those of Experiment 1. Two same-different discrimination tests for each of the three pairs were constructed. The two tests differed only in the interstimulus interval (ISI), which was 500 ms for discrimination test 1 and 50 ms for discrimination test 2. There were 4 sec between items.

Subjects were informed that they would hear the two syllables, presented previously in the identification test, in pairs, and were instructed to decide whether the two stimuli were the same or different. Four demonstration pairs

Table 2

Percent correct on identification and discrimination of synthetic syllable pairs and on Token Test

Subject	Identification Test*				Discrimination Test*								Token Test
					ISI 500 ms				ISI 50 ms				
	1	2	3	Mean	1	2	3	Mean	1	2	3	Mean	
Group: Fluent													
1	96	100	100	99	100	100	100	100	100	100	100	100	100
2	98	96	98	97	95	95	100	98	100	83	100	94	100
3	74	71	77	74	83	83	90	85	70	80	90	80	58
4	60	64	54	59	55	53	95	68	60	52	80	64	45
5	50	50	58	53	-	-	-	-	-	-	-	-	35
Mean	76	76	77	76	83	83	96	87	83	79	92	84	68
Group: Non-Fluent													
6	100	100	100	100	100	100	100	100	100	100	100	100	100
7	100	100	100	100	98	90	98	95	98	98	100	99	99
8	100	100	100	100	100	100	95	98	100	98	100	100	99
9	96	88	100	95	90	90	100	93	100	90	100	97	79
10	50	75	58	61	100	100	90	97	90	100	100	97	77
11	54	56	73	61	95	95	90	93	93	90	88	90	52
12	54	67	44	55	75	78	80	78	75	83	73	77	23
Mean	79	84	82	82	94	93	93	94	94	94	94	94	76

*1 = syllables with 30 ms transitions on all formants

2 = syllables with 82 ms transitions on all formants

3 = syllables with 30 ms transitions on F1, 82 ms on F2 or F3

were presented and the experimenter indicated the appropriate response on a prepared answer sheet. Two answer sheets were available for use depending on individual need. The primary answer sheet contained the letters S for "same" and D for "different." If, after one practice set was administered, this response form was deemed too difficult for the subject, a second sheet was provided on which simple symbols were drawn to convey the concept of "same" (two circles) and "different" (a circle and a square). A practice set of eight items was presented, followed by the 20 item discrimination test.

Results

Table 2 (columns 5-12) lists individual and mean percent correct for the two groups of aphasics. None of the appropriate Wilcoxon Matched Pairs tests showed significant improvement in discrimination of /ba/ and /da/ as a function of formant transition extension. Differences due to ISI were also not significant. Finally, although aphasic groups are too small for us to generalize from their data, there was no consistent or reliable pattern associated with aphasia type, other than a non-significant tendency in the fluent group for series 3 (F2 and F3 transitions only lengthened) to result in higher discrimination scores.

Regardless of the length of the ISI or the duration of the initial formant transition, aphasics performed significantly better on discrimination tests than on identification tests. Only one out of seven aphasics with Token Test scores below 80% reached 80% correct on the three identification tests, whereas all aphasics reached that criterion on at least two discrimination tests. Wilcoxon Matched Pairs tests between subjects' mean identification scores and mean discrimination scores across all stimulus pairs (see Table 2) for each ISI (eliminating subject 5 who did no discrimination tests, and subject 6 whose scores were 100% on every test) give $W = 10$ ($N = 10$, $p < .05$) for ISI = 500 ms, and $W = 4$ ($N = 9$, one tie, $p < .02$) for ISI = 50 ms. Again, as with the identification tests, there was a significant rank order correlation between perceptual performance and Token Test score ($r = .86$, $p < .01$).

DISCUSSION

To support the hypothesis that the basic impairment underlying speech comprehension deficits in aphasia is a failure to analyze rapidly changing acoustic events, studies should demonstrate, at least, that identification improves when spectral changes occur more slowly and/or that performance deteriorates when test syllables to be discriminated are presented at a sufficiently rapid rate. Furthermore, if rate of spectral change is the crucial factor in aphasics' phonological performance, their ability to identify should be no worse than their ability to discriminate.

The present study yields no evidence to support the hypothesis. Aphasics' identification performance did not benefit from the extension of the initial formant transitions conveying place of articulation information. The results from pairs 1 and 2, the two pairs in which stimulus patterns were closely modeled after Tallal and Piercy (1974, 1975), in no way replicate the findings reported in Tallal and Newcombe (1978).

It is noteworthy that pair 2 stimulus patterns (all three formant transitions extended) elicited a variety of identifications from aphasics. The lack of uniformity in labels given these stimulus patterns was corroborated by informal judgments from normal listeners (cf. Liberman et al., 1956; Miller & Liberman, 1979). Reported identifications included, in addition to /ba/-/da/, the labels /wa/-/la/, /bwa/-/dla/, /wa/-/da/ and /ra/-/ya/. Tallal and Newcombe do not report how subjects identified their stimuli, but if, as seems likely, similar shifts in judged manner class occurred, the improved performance of three of their ten subjects with lengthened transitions could, as we remarked in the introduction, have reflected either facilitation of auditory processing for stop consonants, as they assert, or shifts in the manner class of the phonetic segments specified by the extended formant transitions.

In any event, since stimulus patterns for pairs 1 and 2 were, as far as possible, identical to those used by Tallal and Newcombe, the difference in study outcome must be due to other variables, such as the precise experimental procedure, or the nature of the study population. Whatever the source of the difference, the present results are consistent with those of Blumstein, Tartter, Nigro, and Statlender (in press), who also found that formant transition extension had no effect on aphasics' ability to identify or discriminate place of articulation. Auerbach et al. (1981) found that benefit from extending formant transitions was confined to subjects who manifested a "word deafness" component in their speech comprehension impairment. None of the subjects tested here presented this rare unimodal deficit.

Stimulus patterns for pair 3 (extension confined to F2 and F3) were identified as /ba/ and /da/ by all subjects. Nevertheless, except for three fluent aphasics for whom correct syllable discrimination increased, improved stop consonant synthesis had no effect on performance; and these three demonstrated no consistent superiority in identification of the improved patterns, as would be required to justify the claims of Tallal and Newcombe.

The results also offer no support for the notion that aphasics with comprehension deficits discriminate poorly when the interval between stimuli to be discriminated is sharply reduced. Differences between discrimination scores when test syllables were separated by 50 ms vs. 500 ms were small and no trends could be discerned either for the group as a whole or for individual subjects. It was not unusual for a subject to show an increment on the 500 ms over the 50 ms task on one test series, no differences on the second, and a decrement on the third.

The difference in the effect of reduced ISI between the present study and that of Tallal and Newcombe is probably due to task differences. Tallal and Newcombe asked that subjects indicate the order in which two tones occurred, a task calling for both identification and ordering of the tones. The present study simply required that subjects discriminate between two syllables, clearly a less demanding task. Nonetheless, if aphasic deficit does indeed reflect a failure in the processing of rapidly presented acoustic events, the simpler task of the present study should also have reflected this failure at reduced values of ISI.

Performance deficits were not confined to, nor more severe in, one diagnostic group rather than another. Neither group was more sensitive than the other to a reduction in ISI in the discrimination tests, and both fluent and nonfluent aphasics with comprehension deficits demonstrated better discrimination than identification.

This last finding is perhaps the most striking result of the whole study since it runs directly counter to the notion, implicit in Tallal and Newcombe's hypothesis, that phonetic perception is merely an auditory process. A dissociation between discrimination and identification has been reported by others for a different phonetic contrast, voiceless unaspirated vs. voiceless aspirated English stops, signaled by variations in VOT (Blumstein et al., 1977; Kellar, 1979). Moreover, such a dissociation is precisely what we would expect from repeated demonstrations that auditory and phonetic processes may be dissociated in normal listeners (e.g., Mann & Liberman, in press; Studdert-Kennedy, 1983).

Finally, the high correlation between perceptual task performance and Token Test scores is consistent with the results of Tallal and Newcombe, but inconsistent with other investigations in which synthetic stimuli have been used to explore the connection between phonetic deficits and speech comprehension impairment in aphasia (Basso et al., 1977; Blumstein et al., 1977). Identification and discrimination deficits were confined to individuals with substantially reduced Token Test scores, i.e., scores under 80%. Individuals with high or normal Token Test scores obtained near perfect scores on all nine perceptual tests, and no aphasic with a substantially reduced Token Test score ever outperformed aphasics with little or no comprehension impairment.

Although these correlations match those reported by Tallal and Newcombe, the interpretation of the correlations must be different, since the present study found no evidence to support the temporal deficit hypothesis. As far as the identification task goes, we may note that both identification and the Token Test require subjects to perform without the advantage of the semantic context provided in naturalistic situations to support identification. Identifications of contrasting stimuli (two CV syllables, two shape or color names) tend to be labile and over time often become increasingly confused. However, this account will not explain the correlation between discrimination and Token Test performances, so that we must look for other similarities in the cognitive requirements of the tasks. We may note that both the perceptual tests and the Token Test are extremely artificial and require consistent levels of attention over relatively long periods of time. Of course, it is also possible that the tests share no common factor: The several tests may all be sensitive indices of aphasia, but for different unrelated reasons.

REFERENCES

Auerbach, S., Naeser, M., & Mazurski, P. (1981). Primary auditory cortical function and auditory comprehension in aphasia: A function approach. Paper presented at the 19th Annual Meeting of the Academy of Aphasia, London, Ontario.

Basso, A., Casati, G., & Vignolo, L. A. (1977). Phoneme identification defects in aphasia. Cortex, 13, 84-95.

Blumstein, S. E., Cooper, W. E., Zurif, E., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. Neuropsychologia, 15, 371-383.

Blumstein, S., Tartter, V., Nigro, G., & Statlender, S. (in press). Acoustic cues for the perception of place of articulation in aphasia. Brain and Language.

DeRenzi, E., & Vignolo, L. A. (1962). The Token Test: A sensitive test to detect receptive disturbances in aphasia. Brain, 85, 665-678.

Goodglass, H., & Kaplan, E. (1972). The assessment of aphasia and related disorders. Philadelphia: Lea & Febiger.

Kellar, L. A. (1979). Categorical perception of words, CV syllables, and musical stimuli in aphasia. Paper presented at the 17th Annual Meeting of the Academy of Aphasia, San Diego, CA.

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

Mann, V. A., & Liberman, A. M. (in press). Some differences between phonetic and auditory modes of perception. Cognition.

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

Spreen, O., & Benton, A. L. (1977). Neurosensory centers comprehensive examination for aphasia. (1977 revision).

Studdert-Kennedy, M. (1983). On the dissociation of auditory and phonetic perception. In R. Carlson & B. Granström (Eds.), The representation of speech in the peripheral auditory system (pp. 3-10). Amsterdam: Elsevier Biomedical Press.

Tallal, P., & Newcombe, F. (1978). Impairment of auditory perception and language comprehension in dysphasia. Brain and Language, 5, 13-24.

Tallal, P., & Piercy, M. (1974). Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. Neuropsychologia, 12, 83-93.

Tallal, P., & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended stop consonant. Neuropsychologia, 3, 69-74.

FOOTNOTE

[1] Tallal and Piercy (1974, p. 86) provide a table of F2 and F3 transition patterns for their two stimuli representing /ba/ and /da/. However, they report in a footnote to a later paper (Tallal & Piercy, 1975) that the description in their first paper was incorrect. They provide spectrograms of the corrected syllables without listing the actual formant values. Table 1 values are estimated from these spectrograms.

AGAINST A ROLE OF "CHIRP" IDENTIFICATION IN DUPLEX PERCEPTION*

Bruno H. Repp

Duplex perception occurs when a single formant transition (or a pair of such transitions) of a synthetic syllable is isolated and presented to one ear while the remainder of the syllable (the "base") is presented to the opposite ear (Rand, 1974). Listeners report hearing a nonspeech "chirp" in the ear receiving the transition and, at the same time, a syllable in the other ear; the perceived identity of the syllable-initial consonant is determined by the contralateral formant transition. Previous accounts of this phenomenon have attributed the speech percept to dichotic integration or fusion of the transition with the base (e.g., Cutting, 1976; Liberman, Isenberg, & Rakerd, 1981). The nonspeech "chirp" percept was thought to reveal the simultaneous operation of distinct phonetic and auditory modes of perception (Liberman et al., 1981; Repp, 1982).

In a recent article, Nusbaum, Schwab, and Sawusch (1983)--henceforth, NSS--proposed a new explanation. According to their "chirp identification hypothesis," the speech percept does not derive from fusion, but from phonetic identification of the chirp _without reference to the base_. NSS also reported two experiments whose results seem consistent with their hypothesis. Although counterevidence was published simultaneously by Repp, Milburn, and Ashkenas (1983), it was not accepted as such by NSS (see their Footnote 3). The purpose of this note is to examine the arguments and data presented by NSS and to expose their weaknesses. The conclusion will be that the chirp identification hypothesis is not a viable explanation of duplex speech perception and should be laid to rest.

Motivation for the Chirp Identification Hypothesis

From a brief review of some earlier research, NSS conclude that "taken together, the available evidence favors the dichotic integration explanation of duplex perception" (pp. 324-325). Nevertheless, to prepare the ground for their chirp identification hypothesis, NSS cite two findings that they consider to be at variance with the dichotic integration view.

One finding is Rand's (1974) observation that attenuation of second- and third-formant (F2 and F3) transitions in an intact syllable is more detrimental to phonetic perception than attenuation of the same transitions when they

---

*Also _Perception & Psychophysics_, in press.

Acknowledgment. Preparation of these comments was supported by NICHD Grant HD01994 to Haskins Laboratories. I am grateful to Alvin Liberman, Virginia Mann, Richard Pastore, Sigfrid Soli, and Michael Studdert-Kennedy for helpful comments on an earlier draft.

are removed from the base and presented to the opposite ear. NSS conclude that "this result demonstrates that the transitions are processed differently in an intact syllable and on the speech side of the duplex percept" (p. 325). They neglect the fact that Rand's (1974) and many subsequent split-formant studies (e.g., Danaher & Pickett, 1975; Hannley & Dorman, 1983; Nearey & Levitt, 1974; Perl & Haggard, 1974) were undertaken to investigate the effects of "upward spread of masking" due to the first formant (F1). Release from this form of masking consequent upon dichotic separation of formants is well documented. Within the framework of the dichotic integration hypothesis, then, there has been a widely accepted psychoacoustic explanation of the perceptual differences between intact and fused syllables, which does not imply that they are "processed differently."[1]

The second finding NSS cite as being incompatible with the dichotic integration hypothesis is Cutting's (1976) result that large differences in fundamental frequency do not substantially alter duplex perception. NSS argue that different fundamental frequencies signify different articulatory sources, and that the "phonetic processor" should not be able to integrate stimuli that appear to come from different sources. Several counterarguments may be offered, however: (1) The dynamic articulatory information conveyed by the time-varying properties of the chirp is likely to be much more important than that conveyed by fundamental frequency. (2) The chirp is not sufficiently speechlike to suggest any specific articulatory origin by itself. (3) Other forms of dichotic fusion are similarly unaffected by differences in fundamental frequency (Cutting, 1976; Repp, 1976a; Tartter & Blumstein, 1981).

Thus, contrary to NSS's arguments, there do not appear to be any serious problems for the dichotic integration explanation of duplex perception. The possibility remains that the chirp identification hypothesis might account equally well for the data in the literature. That it does not, however, is immediately evident from findings that NSS themselves cite as support for the dichotic integration hypothesis: How, for example, can the chirp identification hypothesis account for the fact that duplex speech identification deteriorates with increasing temporal asynchrony of chirp and base (Cutting, 1976)? Or for the finding that, with selective attention to the speech side of the duplex percept, the chirp receives a different perceptual interpretation depending on the base it is paired with (Liberman et al., 1981)? If there is no integration of chirp and base, it should not matter what the base is and when it occurs. NSS simply bypass these difficulties, which are painfully obvious.

The chirp identification hypothesis rests on three assumptions. The first one is reasonable: "With the appropriate instructions, subjects might at least be able to 'guess' from which consonant or place of articulation a chirp was derived" (p. 325). The second assumption, however, is bizarre: "When asked to identify the speech, subjects can no longer rely solely on the speech-like but phonetically constant base for responding. In order to avoid responding the same way on every trial, subjects must use the transitions (in some way) to produce a phonetic response" (p. 325). The base by itself sounds like a perfectly acceptable syllable (at least when the stimuli are derived from stop-consonant-vowel syllables), and if listeners could avoid fusing it with the chirp, they would surely respond to it the same way they identify it in isolation. Indeed, NSS's own data show that, when the base is presented repeatedly in isolation, subjects are not reluctant at all to give the same response over and over. The third assumption is that the speechlike character of the base leads listeners to "identify the phonetic response with the base instead of with the transition" (p. 324). However, an inability to attribute the response to its correct stimulus would be expected only in the case of fusion. Moreover, if there is no dichotic integration, as NSS maintain, listeners should be able to attend to the base and hear it the way it sounds in isolation. In other words, the chirp and the base should be perceived as separate and unrelated stimuli, which they most decidedly are not (e.g., Liberman et al., 1981; Repp et al., 1983).

In summary, it is evident that the chirp identification hypothesis is not only inconsistent with most data in the literature but also rests on extremely implausible assumptions.

The Nusbaum et al. (1983) Data

NSS's Experiment 1 confirmed the crucial prediction that isolated chirps can be identified consistently as phonetic segments. The stimuli were the synthetic two-formant syllables [ba] and [ga], which are distinguished by a rising vs. falling F2 transition. Repp et al. (1983) have pointed out that rising and falling F2 chirps bear an auditory resemblance to the glides, [w] and [j]. Thus, subjects may have arrived at their (surprisingly consistent) responses by perceiving the chirps not as [b]-like or [g]-like but as [w]-like or [j]-like, and by subsequently choosing the response category that most resembled the quasi-phonetic glide percept. Such a relatively straightforward association may not exist, however, for stimuli used by others in earlier duplex perception experiments. Perhaps unwittingly, NSS chose stimuli that *were uniquely suited to chirp identification.*

Even though the isolated chirps could be associated with phonetic labels, it by no means follows that the subjects of NSS also relied on chirp identification in the duplex condition of Experiment 1. The relative similarity of the overall response proportions for isolated chirps and duplex stimuli (shown in Figure 3 of NSS) is very weak evidence indeed; it not only amounts to accepting the null hypothesis but also merely reflects similar response consistency--not necessarily similar response strategies--in the two experimental conditions. In fact, it is not unlikely that whatever speechlike attributes chirps may possess in isolation (e.g., [w]-like, [j]-like) they lose in the duplex situation, due to competition from the fused speech percept. It is significant, in this connection, that NSS *never asked* their subjects to identify the chirps in the duplex condition while ignoring the bases (or, perhaps, some irrelevant syllables substituted for the bases). Without any demonstration that subjects actually can identify chirps phonetically in the presence of distracting contralateral speech stimuli, the results of Experiment 1 are totally inconclusive.

Experiment 2 was conducted to determine what NSS call the "labeling characteristics of the perceptual process (or processes)" (p. 328) used in the duplex paradigm. A six-member acoustic continuum from [ba] to [ga] was constructed by varying the onset frequency of the F2 transition in the presence of a constant F3 (with a rising transition, to inhibit [da] percepts). These stimuli were presented as full syllables, in a duplex condition, and in an isolated-chirp condition, where the isolated chirps included both the variable F2 and (for no apparent reason) the fixed F3 transition.

According to NSS, the dichotic integration hypothesis predicts that, "if the chirp and base are truly perceptually integrated in the duplex condition, this fused percept should be processed in the same manner as the intact syllables. Thus, the category boundaries should not differ in these two conditions" (p. 328). This prediction ignores once again the potential influence on the category boundary of release from masking due to F1 (as well as other possible psychoacoustic factors) in split-formant presentation (cf. Rand, 1974). While the direction of that influence is difficult to predict, there is no strong basis for expecting identical category boundaries in the two conditions. NSS further predict that, "since the isolated transitions must be processed differently from normal speech ..., the category boundary for isolated transitions should be different from the duplex and intact boundaries" (p. 328). This is simply a non sequitur. The boundaries on entirely unrelated continua may coincide, particularly when they fall near the center of the stimulus range. Unless an experiment is designed to permit the prediction of specific boundary locations (see Bailey, Summerfield, & Dorman, 1977), there is simply no logical connection between category boundaries and "manner" or mode of processing.

Although NSS do not state the predictions of the chirp identification hypothesis in detail, they apparently expected that the boundaries for isolated chirps and duplex stimuli would be the same, since both were thought to involve chirp identification, and different from the boundary for intact syllables because of the purported difference in "manner of processing." The results of Experiment 2 fit these predictions and thus were taken by NSS to support the chirp identification hypothesis. It should be clear from the foregoing discussion, however, that the results are just as compatible with the dichotic integration hypothesis, and that the experiment is logically flawed.

In their General Discussion, NSS make a surprising (and supremely confusing) turnabout by considering the possibility of dichotic fusion without abandoning the chirp identification hypothesis which, of course, postulates the absence of fusion. They suggest, however, that "this dichotic fusion might not occur prior to phonetic labeling. Rather, fusion should [sic!] occur after the phonetic features have been separately identified in the two ears" (p. 331). However, there is little evidence in favor of this new hypothesis. Since both the base and the chirp carry place-of-articulation and manner information, fusion after labeling would frequently result in the perception of two consonants--e.g., [bga] or [bja]--which never happens in duplex presentation. A weakened version of the hypothesis, which does not permit such double-consonant percepts, would be indistinguishable from the dichotic integration view.

NSS also suggest that duplex perception experiments should include an isolated-chirp control condition, to be able "to determine how much more information is contributed by hearing the acoustic attribute in the appropriate syllabic context" (p. 331). If this methodological recommendation were all that NSS wished to convey, there would be little to disagree with.

Clearly, despite the implausibility of the chirp identification hypothesis, there might be some value in demonstrating that chirp identification can not account for the results of a particular study. The experiments of NSS could then be accepted as carefully contrived situations in which it seems as if chirp identification had occurred in duplex perception. The problem with NSS's account, of course, is their insistence that chirp identification actually does occur. The correct conclusion should have been that there is no support for this hypothesis.

The Repp et al. (1983) Data

The data of Repp et al. (1983) were collected for the explicit purpose of refuting the chirp identification hypothesis, as described in an early version of the NSS paper (1981). In Experiment 1, stimuli from a [da]-[ga] continuum varying in the F3 transition were used in a design similar to Experiment 2 of NSS. All subjects but one were unable to label the isolated F3 transitions consistently, and that one subject consistently reversed the category assignment. All subjects, however, labeled the syllables accurately in the duplex condition. Thus, this study demonstrated that phonetic identifiability of isolated chirps is not a necessary condition for duplex speech perception. In Experiment 2 of Repp et al., an AXB syllable similarity judgment task was employed to facilitate selective attention to the ear receiving the base. Perception continued to be strongly influenced by the unattended contralateral chirp. This study disconfirmed a prediction that follows directly from the chirp identification hypothesis, viz., that subjects should be able to "recover" the base by selective attention to the ear receiving it.

In a footnote added in proof (Footnote 3, p. 332), NSS comment on Experiment 1 of Repp et al. Five points are made: (1) Instead of fusion of the chirp with the base, "it is possible that the context of the base in one ear facilitates the extraction of phonetic information from the chirp in the other ear." Note that this is yet another hypothesis, different from the chirp identification hypothesis that postulates that duplex speech identification proceeds without reference to the base. In fact, the only way in which this unannounced "facilitation hypothesis" seems to differ from the dichotic fusion hypothesis is that it predicts that selective attention to the base should be possible. However, Experiment 2 of Repp et al. (on which NSS do not comment) refutes that prediction. (2) NSS point out that the results of Repp et al. do not prove "that it is impossible for subjects to extract phonetic information from these isolated chirps." This is correct but irrelevant, for the point of the demonstration was that poorly identified chirps nevertheless lead to accurate consonant identification when paired with a base. (3) "Repp et al. did not establish the level at which this fusion occurs." Indeed, this was not the purpose of their study. (4) "According to the chirp-identification hypothesis, if fusion does occur, it should take place after some phonetic processing of the chirp." How can a prediction about fusion be derived from a hypothesis that explicitly postulates the nonoccurrence of fusion? (5) Finally, "although dichotic fusion may be a reasonable explanation of the results obtained by Repp et al., there is still no reason to assume that such fusion occurred when the chirps could be identified in isolation, as in the earlier duplex research." However, parsimony demands that a common account be provided for all duplex perception and split-formant experiments, and dichotic fusion is a highly satisfactory general explanation. Moreover, there is no evidence at all that the chirps in earlier duplex studies could be identified in isolation, since this was not tested and different types of stimuli were used. In summary, these comments of NSS do nothing to weaken the results of Repp et al., which clearly disconfirm the chirp identification hypothesis.[2]

CONCLUSION

To be sure, a lot more is to be learned about dichotic fusion and auditory segregation in speech stimuli. While fusion clearly takes place in duplex perception, we do not know at what level in the auditory system it occurs, what kinds of neural mechanisms it involves, and whether or not it is specific to phonetic perception. These interesting questions should be pursued without further distraction.

REFERENCES

Ainsworth, W. A. (1978). Perception of speech sounds with alternate formants presented to opposite ears. *Journal of the Acoustical Society of America*, *63*, 1528-1534.

Bailey, P. J., Summerfield, Q., & Dorman, M. (1977). On the identification of sine-wave analogues of certain speech sounds. *Haskins Laboratories Status Report on Speech Research*, *SR-51/52*, 1-25.

Bentin, S., & Mann, V. (1983). Using the duplex phenomenon to distinguish speech and nonspeech perception of second formant transitions. *Journal of the Acoustical Society of America*, *73*, S52-53. (Abstract)

Broadbent, D. E. (1955). A note on binaural fusion. *Quarterly Journal of Experimental Psychology*, *7*, 46-47.

Broadbent, D. E., & Ladefoged, P. (1957). On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America*, *29*, 708-710.

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

Danaher, E. M., & Pickett, J. M. (1975). Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss. *Journal of Speech and Hearing Research*, *18*, 261-271.

Darwin, C. J., Howell, P., & Brady, S. A. (1978). Laterality and localization: A "right ear advantage" for speech heard on the left. In J. Requin (Ed.), *Attention and Performance VII*. Hillsdale, NJ: Erlbaum.

Hannley, M., & Dorman, M. F. (1983). Susceptibility to intraspeech spread of masking in listeners with sensorineural hearing loss. *Journal of the Acoustical Society of America*, *74*, 40-51.

Isenberg, D. S., & Liberman, A. M. (1978). Speech and nonspeech percepts from the same sound. *Journal of the Acoustical Society of America*, *64* (Supplement No. 1), S20 (Abstract).

Jusczyk, P. W., Smith, L. B., & Murphy, C. (1981). The perceptual classification of speech. *Perception & Psychophysics*, *30*, 10-23.

Liberman, A. M. (1974). The specialization of the language hemisphere. In F. O. Schmidt & F. G. Worden (Eds.), *The neurosciences: Third study program*. Cambridge, MA: MIT Press.

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

Mann, V. A., & Liberman, A. M. (in press). Some differences between phonetic and auditory modes of perception. Cognition.

Nearey, T. M., & Levitt, A. G. (1974). Evidence for spectral fusion in dichotic release from upward spread of masking. Haskins Laboratories Status Report on Speech Research, SR-39/40, 81-89.

Nusbaum, H. C., Schwab, E. C., & Sawusch, J. R. (1981, April). Duplex perception: Dichotic integration or "chirp" identification? Paper presented at the 52nd meeting of the Eastern Psychological Association, New York.

Nusbaum, H. C., Schwab, E. C., & Sawusch, J. R. (1983). The role of "chirp" identification in duplex perception. Perception & Psychophysics, 33, 323-332.

Nye, P. W., Nearey, T. M., & Rand, T. C. (1974). Dichotic release from masking: Further results from studies with synthetic speech stimuli. Haskins Laboratories Status Report on Speech Research, SR-37/38, 123-137.

Pastore, R. E., Szczesiul, R., Rosenblum, L. D., & Schmuckler, M. A. (1982). When is a [p] a [t], and when is it not. Journal of the Acoustical Society of America, 72 (Supplement No. 1), S16 (Abstract).

Perl, N., & Haggard, M. (1974). Masking versus hemispheric sharing of processing for speech sounds. Speech Perception (Department of Psychology, The Queen's University of Belfast), 2:3, 47-54.

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

Repp, B. H. (1975). Dichotic masking of consonants by vowels. Journal of the Acoustical Society of America, 57, 724-735.

Repp, B. H. (1976a). Effects of fundamental frequency contrast on identification and discrimination of dichotic CV syllables at various temporal delays. Memory & Cognition, 4, 75-90.

Repp, B. H. (1976b). Dichotic "masking" of voice onset time. Journal of the Acoustical Society of America, 59, 183-194.

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

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

Schwab, E. C. (1981). Auditory and phonetic processing of tone analogs of speech. Unpublished doctoral dissertation, State University of New York at Buffalo.

Studdert-Kennedy, M. (1981). The emergence of phonetic structure. Cognition, 10, 301-306.

Tartter, V. C., & Blumstein, S. E. (1981). The effects of pitch and spectral differences on phonetic fusion in dichotic listening. Journal of Phonetics, 9, 251-259.

FOOTNOTES

[1] NSS later dismiss the possibility of (upward spread of) masking effects on the grounds that "this explanation cannot be invoked for the articulation-based dichotic integration hypothesis, since proponents of this position have explicitly stated that general auditory processes have no role in mediating phonetic perception (Liberman, 1974; Repp, 1982; Studdert-Kennedy, 1981)" (p. 330). This reflects a serious misunderstanding: By the same token, these

proponents would presumably have to argue that the intelligibility of speech should remain unimpaired in the presence of loud noise! Obviously, distortions due to interactions in the peripheral auditory system must precede any phonetic processing. The point of the authors cited by NSS was that phonetic classification cannot be explained by general auditory processes; however, perceptual changes may well result from factors that affect the internal spectro-temporal representation of speech signals. NSS also cite an unpublished dissertation by Schwab (1981) as showing that auditory masking is absent when stimuli are perceived as speech. While Schwab's results are intriguing, they are not directly applicable to the duplex situation because they did not rest on a comparison of monaural and dichotic presentation conditions. To conclude from Schwab's findings that auditory masking cannot occur in speech stimuli would be absurd.

[2] There are a variety of other observations that speak directly or indirectly against the chirp identification hypothesis. To mention only one particularly damaging result, both Rand (1974) and Cutting (1976) have found that duplex speech perception is resistant to severe attenuation of the chirp; in fact, Bentin and Mann (1983) recently demonstrated that speech identification is still good when chirp detection and discrimination scores are at chance. For other relevant results, see Ainsworth (1978), Bentin and Mann (1983), Broadbent (1955, 1957), Darwin, Howell, and Brady (1978), Isenberg and Liberman (1978), Jusczyk, Smith, and Murphy (1981), Mann and Liberman (in press), Nye, Nearey, and Rand (1974), Pastore, Szczesiul, Rosenblum, and Schmuckler (1982), and Repp (1975, 1976b).

FURTHER EVIDENCE FOR THE ROLE OF RELATIVE TIMING IN SPEECH: A REPLY TO BARRY*

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

Abstract. In an earlier paper (Tuller, Kelso, & Harris, 1982a) we suggested that the timing of consonant-related muscle activity was constrained relative to the period between onsets of muscle activity for successive vowels. Here we first reexamine those data based on reservations posed by Barry. Next, we present a kinematic study of articulation that extends, and strongly supports, our original observations. Finally, we very briefly survey some converging lines of evidence for a functionally significant vowel-to-vowel period in speech and how this may relate to the role of temporal invariance in motor skills in general.

In his review, Barry (1983) makes some well-reasoned comments that have given us further insight into our previously presented data and encouraged us to look at the results of a study we have just completed within a similar perspective. Barry's first point is that our results may be, in some sense, a statistical artifact. Just as most of the durational stretching and shrinking across rate and stress changes occurs in the vowel portion of the acoustic signal, the vowel-related electromyographic (EMG) activity is also the most elastic part of production. Changes in duration of consonant-related activity are smaller, though systematic (cf. Tuller, Harris, & Kelso, 1982). This alone--according to Barry--might account for the fact that the correlations we computed of the interval between the onsets of muscle activity specific to production of successive vowels and the timing of muscle activity for the intervening consonant (Barry's Figure 1a), are higher than correlations between the onsets of muscle activity for successive consonants and the timing of activity for the intervening vowel (Barry's Figure 1b). To explore this possibility, we followed Barry's suggestion and correlated the period between successive consonant onsets with the vowel onset-to-consonant onset interval. In all cases, this resulted in a lower correlation than our original measure. The shape of the histogram of correlations based on Barry's suggested analysis, presented in Figure 1a, is significantly different (Kolmogorov-

---

*Also Journal of Experimental Psychology: Human Perception and Performance, 1983, 9, 829-833.

+Also Cornell University Medical College

++Also University of Connecticut

+++Also City University of New York, Graduate Center

Acknowledgment. This work was supported by NINCDS grant NS-17778 to Cornell University Medical Center, and NINCDS grant NS-13617, U. S. Office of Naval Research Contract No. N00014-83-C-0083, and BRS Grant RR-05596 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-74/75 (1983)]

Smirnov for r > .8, p < .001) from the distribution arising from our original procedure, that is, by correlating the period between vowel onsets with the interval from vowel onset to consonant onset (see Figure 1b).

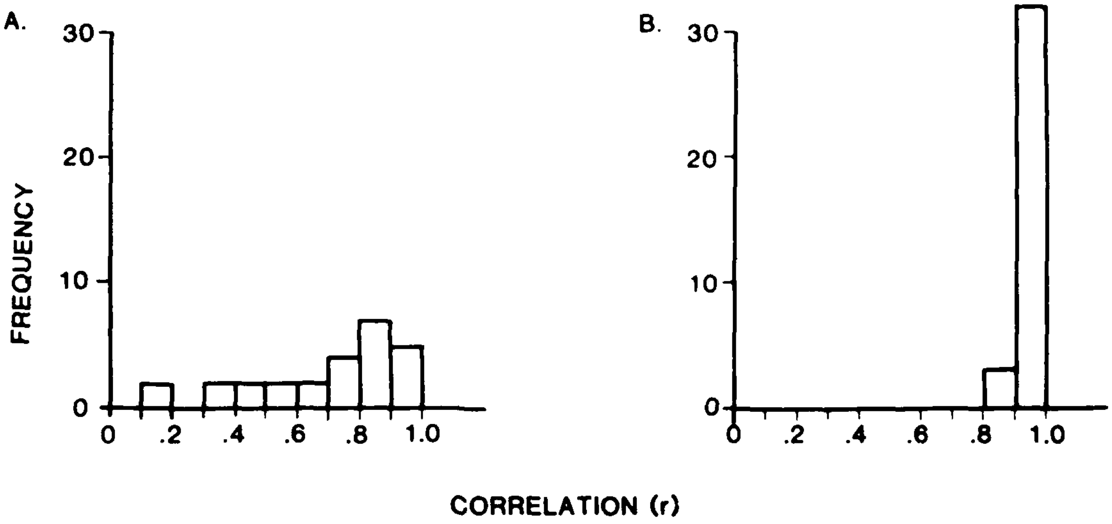


Figure 1. A) Distribution of correlations for the period between onsets of muscle activity for *successive consonants* and the latency of onset of vowel-related muscle activity. B) Distribution of correlations for the period between onsets of muscle activity for successive vowels and the latency of onset of consonant-related muscle activity.

Although this analysis shows that the correlation measure we used will give higher correlations than the one Barry suggested as a substitute, these results do not address a crucial point that underlies our argument, and is obliquely addressed by Barry. We believe that we obtain our correlation results because the small changes in duration of consonant-related activity are correlated with the relatively larger changes in duration of vowel-related activity, over the averaged effects of stress and speaking rate on an ensemble of tokens. If this is true in the average across stress and rate conditions, the same relations should hold for individual tokens within stress and rate conditions. As we pointed out in our original article (Tuller, Kelso, & Harris, 1982a, hereafter called our JEP article), there is no need to assume that changes in vowel- and consonant-related activity are ratiomorphic, and, indeed, neither we nor Barry believe they usually are. However, we cannot examine this point in detail using electromyographic data because it is not always possible to define onsets and offsets in individual repetition tokens of an utterance (see Baer, Bell-Berti, & Tuller, 1979, for a discussion of temporal measures of individual vs. averaged EMG records). For this reason, we will describe a more recent experiment in which we measured articulator movement trajectories, which can, of course, be analyzed on a token-by-token basis.

Since the publication of our JEP article, we have extended our observations to the kinematics of the jaw and lips during speech (Tuller, Kelso, & Harris, 1982b). Briefly, subjects produced utterances of the form /bVCab/ where V was either /a/ or /ae/ and C was from the set /p, b, v, w/. Each utterance was spoken with two stress patterns and at two self-selected speaking rates, conversational and relatively fast. In essence, the experimental design incorporated and extended the earlier design of our EMG study. Ten to twelve repetitions of each utterance type were produced. Articulatory movements in the up-down direction were monitored by an optoelectronic device that tracked the movement of lightweight, infrared, light-emitting diodes attached to the subject's lips and jaw. (Details of data collection and processing may be found in Tuller et al., 1982b.)

In order to examine more closely whether the high correlations obtained in the EMG experiment are a function of using means in the analyses, or perhaps are solely due to the effect of variations in vowel duration, we performed three analyses of /bapab/ (the one utterance common to both experiments) produced by the only subject who participated in both studies. First, we asked the original question about stress and rate variations: does the interval from vowel onset to consonant onset change systematically as a function of a vowel-to-vowel period? To this end, correlations were computed between the period from the onset of jaw lowering for the first vowel to the onset of jaw lowering for the second vowel and the interval between the onset of jaw lowering for the first vowel and the onset of consonant-specific movement (that is, a close movement analogue of our earlier EMG measure; Figure 2a). In separate analyses, the onset of movement for the medial labial consonant was defined either by the onset of upper lip lowering or by the onset of lower lip raising (independent of simultaneous jaw movements). Each correlation was based on 35 data points. The Pearson's product-moment correlations were .97 and .96 for the lower lip and upper lip, respectively (Figures 2b and 2c). These kinematic results, obtained from measures of individual repetitions of each utterance type, essentially mirror our earlier EMG findings, which were based on utterance ensemble averages.

In a second analysis, we examined the movement analogue of Barry's suggested analysis by correlating the interval between onsets of upper lip lowering (or lower lip raising) for successive consonants with the interval between vowel onset (as indexed by the onset of jaw lowering) and the following consonant onset. These correlations were significantly lower (using Fisher's r-to-z transform) than those obtained by our original definition of period and latency: when consonant production is indexed by upper lip movement, $r = .70$ versus .96, $t(32) = 3.704$, $p < .001$; when consonant production is indexed by lower lip movement, $r = .76$ versus .97, $t(32) = 4.384$, $p < .001$. Again, the variations in vowel duration alone cannot account for the systematic relationship between the timing of consonant articulation and the period between successive vowels.

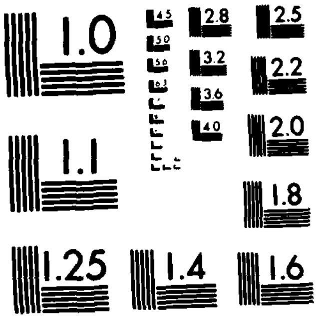
AD-A136 416 STATUS REPORT ON SPEECH RESEARCH A REPORT ON THE STATUS AND PROGRESS OF S..(U) HASKINS LABS INC NEW HAVEN CT A M LIBERMAN 30 SEP 83 SR-74/75(1983) N00014-83-C-0083

3/3

UNCLASSIFIED F/G 6/16 NL

END
DATE
FILMED
[illegible]-84
DTIC

END
DATE
FILMED
[illegible]-84
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

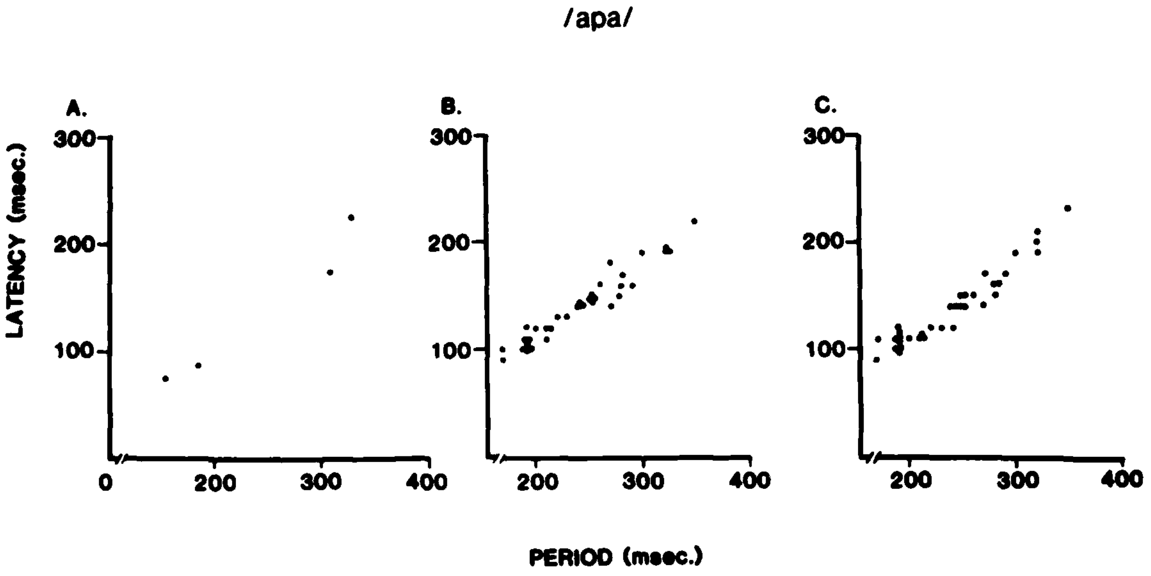


Figure 2. Timing of consonant articulation ("latency") as a function of the vowel-to-vowel period. Each graph contains data from two stress patterns and two rates produced by the same speaker. A) The onset of consonant-related activity in orbicularis oris is graphed relative to the interval between epochs of activity in anterior belly of digastric, y = .89x + 107, r = .89. Each point represents the mean of EMG data for 12 repetitions of "pa-pap." B) Timing of lower lip raising as a function of the vowel-to-vowel period indexed by jaw lowering movements. Each point represents one token of the utterance "ba-pab," y = .66x -18, r = .97. C) Same as (B), but with consonant articulation indexed by the onset of upper lip lowering, y = .7x -28, r = .96.

We undertook a final analysis to examine specifically whether the high correlations obtained are simply a function of the change in vowel duration contributing to both variables or whether they reflect some organizational attribute of each repetition's internal structure. To this end, we explored whether the small changes in duration of consonant-related articulatory movements were correlated with corresponding changes in vowel-related gestures (that is, Barry's suggested correlation of "period" and "period minus latency"). For all repetitions of /bapab/ at both stress and rate levels, we determined the duration of vowel-related movements, defined as the interval from the onset of jaw lowering for the first vowel to the onset of lip movement for the /p/, and the duration of movement specific to the consonant, defined as the interval from the onset of lip movement for the /p/ to the onset of jaw lowering for the second vowel. We then calculated the correlation between members of the pairs. If these correlations are significantly greater than zero, then the temporal relations between a vowel and its following consonant are not random and, although vowel duration does contribute to the high correlations, it is not the only significant factor. It was in fact the case that the durations of vowel and consonant movements were positively correlated: when consonant production was indexed by upper lip movement, $r = .74$, $t(32) = 5.37$, $\underline{p} < .001$; when consonant production was indexed by lower lip movement, $r = .72$, $t(32) = 5.14$, $\underline{p} < .001$. In conclusion, we believe that our results cannot be accounted for by vowel variation alone, but indicate that the timing of consonant articulation is constrained relative to the timing of articulation for the flanking vowels.

In order to unpack Barry's third point, we must return to consideration of the EMG data. Barry speculates on the interpretation of results reported in the JEP article relative to our own earlier findings that the temporal overlap of muscle activity for certain vowels and consonants altered little over marked changes in syllable duration (Tuller, Harris, & Kelso, 1981). Consider the schematic in Figure 3. The interval AC represents the duration of muscle activity specific to the first vowel, the interval BE represents the duration of activity in a different muscle for production of the consonant, and DF is the duration of muscle activity for the second vowel. The "overlap intervals" we have referred to are the time from the onset of consonant-related activity to the offset of activity specific to the preceding vowel (BC in Figure 3), and the time between the onset of activity for the second vowel and the offset of activity for the preceding consonant (DE). In our earlier work, we examined the duration of overlapping activity in a lip muscle (orbicularis oris), acting for production of the consonants "p" and "b," and a tongue muscle (genioglossus), acting for production of the vowels "ee" and "ay" in utterance such as in "pee-peep" and "pay-payp." The overlap intervals (BC and DE in Figure 3) remained remarkably constant across two stress patterns and two speaking rates. In a companion paper (Tuller, Kelso, & Harris, 1981), we extended these observations to the activity of various other articulator muscles--in fact, these were the same recordings analyzed for the JEP article. Although the relatively constant temporal overlap of activity in orbicularis oris and genioglossus again resulted, other muscle comparisons showed different patterns (e.g., for the production of "pa-pap" the temporal overlap of a jaw-lowering muscle, anterior belly of digastric, relative to a lip muscle, orbicularis oris, changed systematically as speaking rate increased). Our conclusion was that the temporal overlap of muscle activity in vowel-consonant and consonant-vowel pairs does not, as a rule, remain fixed over metrical variations in speaking rate and syllable stress.

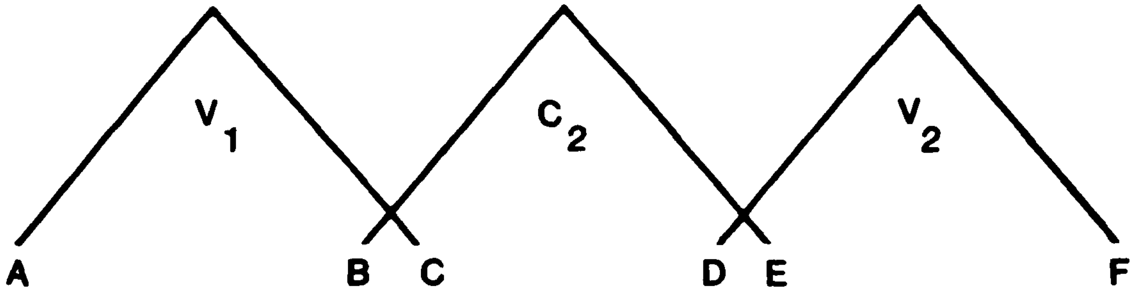


Figure 3. Schematic EMG activity for a vowel-consonant-vowel triad.

Following from this conclusion, we wish to point out that our thoughts have altered somewhat as to why the overlap interval between orbicularis oris and genioglossus remained unaltered in both experiments (see also Raphael, 1975). It may be that our assumption that the tongue is completely free to assume any position during production of /p/ is in fact incorrect (see also Alfonso & Baer, 1982; Bell-Berti, 1980; Harris & Bell-Berti, in press; Houde, 1967). Rather than conceiving different articulators as being either crucially involved, or uninvolved, in producing a given sound, we might do better to consider the entire vocal tract as involved in producing all sounds with only the relative importance of individual articulators shifting as the phonetic structure changes. Thus, the constant overlap of orbicularis oris and genioglossus may reflect the articulatory organization that in some way maximizes conditions for production of the bilabial stop consonant, and does not reflect feedback-dependent (or for that matter feedback-independent) control of the timing of successive segments.

In Barry's final comment, he expresses surprise that we find stable vowel-to-consonant timing relative to the interval between successive vowels even though the vowel and consonant are separated by a syllable boundary. He suggests that the subject was performing an articulatory syllabification different from that we have represented orthographically. Thus, perhaps the subject was saying something like "peep-eep" rather than "pee-peep." Apart from the fact that such a production strategy seems counterintuitive, we should remark that the intervocalic /p/ was aspirated, thus conforming to the conventional description of a syllable-initial form.

Leaving aside the question of articulatory strategies, an issue we have not addressed in any detail, we should remark that temporal and spatial coarticulatory effects are very well documented in the literature. These indicate that syllable boundaries do not necessarily disrupt acoustic or articulatory interactions between segments and, perhaps more to the point, that transsyllabic interactions may be stronger than intrasyllabic ones. For example, the measured acoustic duration of a vowel is strongly affected by the number of transsyllabic consonants that immediately follow it (Lindblom & Rapp, 1973). An effect on acoustic vowel duration of preceding, intrasyllabic consonants has not always been found (for review see Elert, 1964; see also Lindblom & Rapp, 1973). In addition, the acoustic duration of a vowel before a voiceless stop consonant (such as /p/) has long been known to be shorter than the same vowel occurring before a voiced stop consonant (such as /b/), both within ("rip" vs. "rib") and across ("rapid" vs. "rabid") syllables (House, 1961; Klatt, 1973; Petersen & Lehiste, 1960). Transsyllabic articulatory effects have also been documented. As a recent example, Harris and Bell-Berti (in press) report that in sequences such as [iʔi] and [uʔu] the glottal stop [ʔ] does not cause relaxation of the tongue for [i] sequences or the lips for [u] sequences. In other words, the syllable boundary between the first vowel and the stop does not seem to be articulatorily marked. More generally, there may not be any isomorphism between articulatory syllabification and syllabification as defined by linguists (that is, if linguists could agree on the rules for syllabification; cf. Bell, 1978).

In his comments, Barry agrees with us that it is at least "plausible" that vowel-to-vowel timing is important for rhythmic structuring. In fact, many pieces of evidence in the literature (in addition to the two papers Barry cites) suggest a functionally significant vowel-to-vowel period (and perhaps, by extension, that commonalities among segments are exploited in production; cf. Fowler, 1977). First, the description of English as being "stress-timed" is based on the perception that stressed vowels occur at approximately equal intervals. Although there is little support for a strict stress-timing hypothesis, there is evidence that speakers maintain at least a tendency toward stress-timing that may be more closely associated with the timing of the stressed vowels than with the accompanying consonants (for review, see Fowler, 1983).

A second source of evidence that a vowel-to-vowel articulatory period may be functionally significant is the literature on compensatory shortening and coarticulation. We have already mentioned that intervocalic consonants shorten the measured acoustic duration of the surrounding vowels. This may mean that all aspects of the articulation of vowels are produced in shorter time periods when consonants follow them. Alternatively, it may mean that the consonants and vowels are produced in concert, with the trailing edges of the vowels progressively "overlaid," as it were, by the consonants. In other words, consonants and consonant clusters might be produced on a background of continuous vowel articulation. An articulatory organization of this sort was first proposed by Öhman (1966), to explain the changes in formant transitions for intervocalic consonants as a function of the flanking vowels. More recent articulatory evidence that the influence of both preceding and following vowels is apparent throughout the intervocalic consonant (Barry & Kuenzel, 1975; Butcher & Weiher, 1976; Gay, 1977; Harris & Bell-Berti, in press; Sussman, MacNeilage, & Hanson, 1973) might also be interpreted as indicating a significant vowel-to-vowel articulatory period.

In conclusion, let us reiterate our previous conviction that the data reported here and in our JEP article are compatible with a style of motor organization in which the relative timing among individual electromyographic or kinematic events is preserved in the face of scalar changes in, for example, absolute duration and amplitude of EMG activity or articulator displacement and velocity (for reviews see Kelso, 1981; Kelso, Tuller & Harris, 1983). In fact we believe, with Bernstein (1967), that the cooperation observed among muscles and joints during coordinated activity is best described by a partitioning of variables into two classes; those that can effect scalar changes in a behavior and those that preserve its internal temporal "topology." Temporal invariance across scalar variation may be a design feature of all motor systems and may constitute one of Nature's solutions to the problem of coordinating complex systems, like speech, that possess many degrees of freedom.

REFERENCES

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

Baer, T., Bell-Berti, F., & Tuller, B. (1979). On determining EMG onset time. In J. J. Wolf & D. H. Klatt (Eds.), *Speech Communication Preprints, 97th Meeting of the Acoustical Society of America, Cambridge, Mass., 1979*. New York: Acoustical Society of America.

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

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

Bell, A. (1978). Segment organization phenomena and their explanations. In A. Bell & J. Hooper (Eds.), *Syllables and segments*. New York: North-Holland.

Bell-Berti, F. (1980). Velopharyngeal function: A spatial-temporal model. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 4, pp. 291-316). New York: Academic Press.

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

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

Elert, C. C. (1964). *Phonological studies of quantity in Swedish*. Uppsala: Almqvist & Wiksell.

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

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

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

Harris, K. S., & Bell-Berti, F. (in press). On consonants and syllable boundaries. In L. Raphael, C. Raphael, & M. Valdovinos (Eds.), Language and cognition: Essays in honor of Arthur J. Bronstein. New York: Plenum.

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

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

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

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

Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. Journal of the Acoustical Society of America, 54, 1102-1104.

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

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

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

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

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. Journal of Speech and Hearing Research, 16, 397-420.

Tuller, B., Harris, K. S., & Kelso, J. A. S. (1981). Articulatory motor events as a function of speaking rate and stress. Haskins Laboratories Status Report on Speech Research, SR-65, 33-62.

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

Tuller, B., Kelso, J. A. S., & Harris, K. (1981). Phase relationships among articulatory muscles as a function of speaking rate and stress. Haskins Laboratories Status Report on Speech Research, SR-65, 63-90.

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

Tuller, B., Kelso, J. A. S., & Harris, K. (1982b). On the kinematics of articulatory control as a function of stress and rate. Haskins Laboratories Status Reports on Speech Research, SR-71/72, 81-88.

REVIEW[*]

Phonological intervention: Concepts and procedures. Edited by Michael Crary (1982). San Diego, CA: College-Hill Press, Inc., 1982.

Katherine S. Harris[+]

One of the most important occupations of traditional American speech pathologists has been the provision of remediation services to misarticulating children. Out of this setting has come such classic work as Templin's *Certain language skills in children*, which has provided us with developmental norms for the various speech sounds of English, and a great deal of information on vocabulary development. While Templin's approach was essentially atheoretical, there is some underlying view that the speech sounds are learned one at a time, in an order that reflects articulatory ease. An entirely different tradition is represented by Jakobson's *Child language, aphasia and phonological universals*, which is, in some sense, an attempt to account for the acquisition of speech sounds against a background of taxonomic phonemics. Jakobson claimed that children learn contrasts, rather than individual sounds, and that the order of acquisition is set up so that maximal contrasts, presumably the easiest contrasts, are learned first. The specification for sounds in terms of features provides a matrix for degree of contrast between sound pair members. Another linguist with important insight into speech development has been Stampe, the originator of "natural phonology." Stampe's emphasis is on the dependence of the child's form on the adult's. The child is said to have innate processes that simplify his/her output production of a received adult model. Thus, the child begins with the easiest forms, those in which maximum simplification has been achieved, and gradually inhibits simplifying processes.

In the 1970's, linguistically based approaches of various kinds began to have a vogue in the traditional speech pathology setting. The book reviewed here represents this trend away from a focus on "articulation disorders" towards a focus on "phonological disorders." Each of the five chapter authors describes his/her interpretation of "phonological intervention" and goes on to discuss the nuts and bolts of diagnosis and remediation within that framework.

---

[*] Also *Applied Linguistics*, in press.

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

Acknowledgment. Preparation of this review was supported by NINCDS Grants NS-13617 and NS-13870 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-74/75 (1983)]

REVIEW*

Phonological intervention: Concepts and procedures. Edited by Michael Crary (1982). San Diego, CA: College-Hill Press, Inc., 1982.

Katherine S. Harris+

One of the most important occupations of traditional American speech pathologists has been the provision of remediation services to misarticulating children. Out of this setting has come such classic work as Templin's Certain language skills in children, which has provided us with developmental norms for the various speech sounds of English, and a great deal of information on vocabulary development. While Templin's approach was essentially atheoretical, there is some underlying view that the speech sounds are learned one at a time, in an order that reflects articulatory ease. An entirely different tradition is represented by Jakobson's Child language, aphasia and phonological universals, which is, in some sense, an attempt to account for the acquisition of speech sounds against a background of taxonomic phonemics. Jakobson claimed that children learn contrasts, rather than individual sounds, and that the order of acquisition is set up so that maximal contrasts, presumably the easiest contrasts, are learned first. The specification for sounds in terms of features provides a matrix for degree of contrast between sound pair members. Another linguist with important insight into speech development has been Stampe, the originator of "natural phonology." Stampe's emphasis is on the dependence of the child's form on the adult's. The child is said to have innate processes that simplify his/her output production of a received adult model. Thus, the child begins with the easiest forms, those in which maximum simplification has been achieved, and gradually inhibits simplifying processes.

In the 1970's, linguistically based approaches of various kinds began to have a vogue in the traditional speech pathology setting. The book reviewed here represents this trend away from a focus on "articulation disorders" towards a focus on "phonological disorders." Each of the five chapter authors describes his/her interpretation of "phonological intervention" and goes on to discuss the nuts and bolts of diagnosis and remediation within that framework.

---

*Also Applied Linguistics, in press.

+Also The Graduate School, City University of New York

Acknowledgment. Preparation of this review was supported by NINCDS Grants NS-13617 and NS-13870 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-74/75 (1983)]

Ingram, whose 1976 book Phonological disability in children provided the inspiration for the conference on which this volume is based, rediscusses and amplifies some of the practical problems in collecting data samples and inferring from them the natural simplification processes that form the basis of his approach to classification. Shriberg, whose theoretical stance is quite similar to Ingram's, presents a detailed scheme for diagnostic classification. He makes the interesting suggestion that, while some errors in the productions of a given child may arise from Stampe's "natural processes," operating on a developmentally delayed system, others may arise as a consequence of structural abnormalities, such as middle ear involvement. Fokes rediscusses some practical problems in making such an inventory, noting especially the difficulties posed for sampling by inherent variability, such as inconsistent productions or progressive idioms. Blache introduces an elaborate description of speech sounds in terms of what purports to be a distinctive feature analysis, uses this to hang a developmental analysis on, and uses this, in turn, as the basis of a diagnostic workup. Hodson simply describes the patterns of error in children of varying degrees of unintelligibility.

In spite of a difference in emphasis, there is a common theme. All the authors focus on the need to examine a sample of speech behavior that is sufficiently complete that each sound is assessed in a variety of contexts, along with the pattern of substitutions and the resulting neutralization of contrasts relative to the adult system. This emphasis, of course, results from an exposure to the phonologist's practice of writing rules to make conversion between one system and another or one level of representation and another. In the case of the misarticulating child, one might compare the child's system to that of the ambient community, or examine the operation of processes in remediation. However, both Shriberg and Ingram are quite cautious about the reality status of their inferred underlying phonological units, or the relationship of their analysis schemes to Stampe's natural phonology (Stampe, 1973). Shriberg also notes the possibility, raised by Dinnsen, Elbert, and Weismer (1980) and rediscussed in detail by Maxwell and Weismer (1982) that misarticulating children may differ among themselves in the relationship of their underlying phonological schemata to the adult model.

The authors do differ on substantive issues. Both Fokes and Blache advocate forms of discrimination training in remediation. Shriberg is very specific about his reasons for doubting its efficacy, and Ingram has been similarly skeptical in other writings. It should be noted that some disenchantment with discrimination training as a remediation technique has been voiced, as well, by speech pathologists who have not joined the "phonological intervention" camp (Shelton & McReynolds, 1979).

Another difference is that only one author, Blache, makes extensive use of feature notation. It should be said that, while his feature notation is rather vaguely attributed to Jakobson, the particular version used in this volume would not be recognized by its presumed originator, and the mode of presentation may confuse readers. However, trying to guess the possible reasons for the abandonment of feature notation by the other authors is a more interesting mission for a reviewer than disagreeing with the use of any particular form.

One can think of both structural and substantive reasons. As feature notation is commonly used in speech pathology, it does not represent any observation not present in the segmental notation; that is, the clinician, having written [b], for example, looks up the features of [b]:

$$\begin{bmatrix} -\text{vocalic} \\ -\text{consonantal} \\ -\text{high} \\ -\text{back} \\ +\text{anterior} \\ -\text{coronal} \\ +\text{voice} \\ -\text{continuant} \\ -\text{nasal} \\ -\text{strident} \end{bmatrix}$$

in Chomsky and Halle's (1968) notation, and inserts them in place of [b]. The fact that [b] is produced normally, at a phonetic level, without vocal fold vibration during closure in some environments is not relevant to the substitution, and no independent observation is made of voicing per se. Thus, the clinician has no greater contact with misarticulations in need of correction in the one notation than in the other. It should be pointed out that both Ingram and Shriberg have suggested use of narrow transcription, while they do not discuss a systematic use for it.

Another reason for the abandonment of feature notation is that, as Ingram has pointed out, the carefully collected data of the last decade (Yeni-Komshian, Kavanagh, & Ferguson, 1980) reveal the primacy of segment over feature in learning. Jakobson's predictions for a universal order of feature acquisition is not supported in detail. Furthermore, while an important early writer in the field, Compton (1970), has suggested that the correction of misarticulation of a feature in one segment may generalize to another segment, the empirical justification for such a view is not strong. Given these problems, a strong motivation for persuading speech pathologists to make the intellectual effort to translate from segments to features seems lacking, whatever the gain in elegance and simplicity this translation gives linguistic analysis.

Finally, it is impossible to leave this volume without remarking on one of its undiscussed premises, that the clinician's notational scheme, whether featural or segmental, adequately captures all the information needed in remediation. This may not be so. By its nature, transcription reduces the dynamic articulation process to a series of static symbols, thus minimizing the role of timing as a component of effective production. It has been shown (by Smith [1978], Kent & Forner [1980] and Bond & Wilson [1980], among others) that children develop adult temporal patterns only very slowly. It is not clear what effect various forms of timing pattern irregularity have on the transcription operation; neither is it clear what clinical significance temporal deviance might have. Hence, some of the information the clinician needs may be left outside transcriptional evaluation.

Furthermore, the assumption made throughout most of the book is that the child's errors are appropriately described as substitutions, that is, that they are produced as consistently as "correct" sounds. The assumption may be as much a reflection of the characteristics of the therapist's perception as of the child's productions. If the child produces a sound lying outside the clinician's native repertoire, the clinician may record it as a simple substitution of an item within his repertoire. It might be noted here that one old transcription category, the distortion, is missing. It seems at least plausible that some misarticulating children may produce sounds that no normal produces, with the consequence that the clinician has no appropriate model. Beyond that, the transcriptional scheme itself is not set up to capture differences in the variability of sounds produced, and variability information may be important in remediation.

Of course, one important reason for the use of transcription as the clinician's primary tool is that in most clinics, no other is available. Surely, then, it must be a goal of research effort to show the relationship of acoustic and transcriptional techniques in systematizing what competent clinicians know about the misarticulating child, and to investigate the relative utility of instrumental and non-instrumental approaches to speech production.

REFERENCES

Bond, L. S., & Wilson, H. F. (1980). Acquisiton of the voicing contrast by language-delayed and normally speaking children. Journal of Speech and Hearing Research, 23, 152-161.

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

Compton, A. J. (1970). Generative studies of children's phonological disorders. Journal of Speech and Hearing Disorders, 375, 315-339.

Dinnsen, D. A., Elbert, M., & Weismer, G. (1981). Some typological properties of functional misarticulation systems. In W. O. Dressler (Ed.), Phonologica 1980. Innsbruck: Innsbrucker Beitrage zur Sprachwissenschaft.

Ingram, D. (1976). Phonological disability in children. London: E. Arnold.

Jakobson, R. (1968). Child language, aphasia, and phonological universals. The Hague: Mouton.

Kent, R. D., & Forner, L. L. (1980). Speech segment durations in sentence recitations by children and adults. Journal of Phonetics, 8, 157-168.

Maxwell, E. M., & Weismer, G. (1982). The contribution of phonological, acoustic and perceptual techniques to a characterization of a misarticulating child's voice contrast for stops. Applied Psycholinguistics, 3, 29-43.

Shelton, R., & McReynolds, L. (1979). Functional articulation disorders: Preliminaries to treatment. In N. Lass (Ed.), Speech and language: Advances in basic research and practice. New York: Academic Press.

Smith, B. L. (1978). Temporal aspects of English speech production: A developmental study. Journal of Phonetics, 6, 37-67.

Stampe, D. (1973). A dissertation on natural phonology. Unpublished doctoral dissertation, University of Chicago.

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

Yeni-Komshian, G., Kavanagh, J., & Ferguson, C. (Eds.) (1980). Child phonology. Vol. 1: Production. New York: Academic Press.

REVIEW*

Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler. Edited by H. W. Dechert and M. Raupach (1980). The Hague: Mouton.

Michael Studdert-Kennedy+

How can we study the process of transforming thought into speech? Can we find in the temporal structure of speech overt, measurable indications of the speaker's underlying cognitive activity? Do pauses serve a communicative function in guiding the listener's segmentation of an utterance? Are there pausal patterns common to different languages or to speech in its many _genres_--reading, public speaking, telling a story, answering difficult questions, idly conversing, and so on? What light may be thrown by unusual patterns of pausing on the disordered origins of aphasic speech or on the difficulties of second language learners? These and related questions are the topics of this book.

The book comprises the revised versions of 34 presentations given at a workshop on "Pausological Implications of Speech Production," held in Kassel, West Germany in June, 1978. The workshop, jointly organized by psycholinguists at the Gesamthochschule Kassel and at St. Louis University, Missouri, was attended by 37 linguists and psychologists from Canada, France, the Netherlands, Norway, the United Kingdom, the United States, and West Germany. For publication the participants wisely agreed to abandon the pretentious neologism, _pausological_, in favor of a more accessible word, _temporal_, although, as I have indicated, the papers do not deal with timing as such, that is, with the origins and mechanisms of temporal order in speech. Rather the "temporal variables" of the title are simply the frequency, duration, and location of pauses in the speech flow from which underlying nontemporal processes (that take time to occur) may be inferred.

Having said this, I should add that many of the papers fall quite outside this rubric. In fact, the papers are extraordinarily heterogeneous, in both topic and quality. Most of them are short (6-10 pages), so that reading the book straight through makes for a bumpy ride, as we jounce from grand, speculative discussion of language and the brain (Karl Pribram) to an opaque pass at extending Thom's theory of catastrophes into the dynamics of verbal planning (Wolfgang Wildgen) to formal proposals for taxonomies of speech

---

*Also _Phonetica_, in press.

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

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

pauses and their role in grammar (Thomas Ballmer, Raimund Drommel) to the problems of the pause extraction in automatic speech recognition (Jens-Peter Köster, Hede Helfrich). Nonetheless, some coherence does emerge from the editors' grouping of the papers into the sections of the original workshop: general, syntactic, and structural, conversational, prosodic, and cross-linguistic aspects, and a final discussion.

The diversity of approaches evidently reflects some uncertainty among the participants as to what the object of study actually is. The St. Louis contingent (Daniel O'Connell and Sabine Kowal) seems to believe that the uncertainty might be resolved, if only the "field" could be granted a theoretical framework. Ballmer (p. 211) makes a valiant attempt to launch the needed theory with a taxonomy of pause types. He proposes a tripartite classification in terms of airflow intensity, controllability (unintentional vs. intentional) and the potential utility of pauses to speaker and hearer, listing under this last division some twenty-six types--and warning us that any particular pause may be classified under more than one type! The difficulty with such schemes, as Wallace Chafe points out in the final discussion (p. 327), is that interpretive (or functional) taxonomies invite disagreement. In Francois Grosjean's words: "There are maybe 40 or 50 different variables that can create a silence in speech. A silence may mark the end of a sentence, you can use it to breathe, you can use it to hesitate, there may be ten or fifteen different things happening during that silence" (p. 328). If this is so, there is more than enough room for disagreement on what the operative variables are. Nor are purely objective definitions of pause likely to be of greater use. For example, pause frequency and length may vary with speaker, social situation, speech rate, and a host of other contextual variables, many, if not all, of which are purely inferential. The prospect of filling the theoretical void in the face of this complexity and uncertainty is dim.

What seems to be needed is simplification: careful descriptive and experimental study with clearly defined variables. O'Connell and Kowal in their introductory "Prospectus for a science of pausology" evidently think the time for this is past: "If we are ever to transcend the trivialization which has beset modern psychology...we must find a way of engaging multilogic reality" (p. 9). For this we will find no better way than that of, say, James Joyce. In the meantime, there is science, and this calls for reliable data, systematically collected under well-controlled conditions.

An exemplary instance of an experimental approach is the work of Grosjean and his colleagues on the relations between syntactic structure and the distribution of pauses between words in a sentence. Grosjean reviews preliminary studies of spontaneous speech in interviews, showing that pauses tend to fall at major and minor constituent breaks. Later studies of oral reading showed that variations in syntactic complexity (measured, in one study, by subjects' parsings of the sentences) could account for as much as 56% of the variance in pause duration. Looking for other sources of variance, Grosjean and his colleagues noted that speakers tend to disregard syntactic breaks at certain points, so as to divide constituents into word groups of more or less equal length. They therefore worked up an elegant model to predict the distribution of pauses between words from a weighted index of syntactic complexity and constituent length. In a test of the model, they were able to account for 72% of the total variance in pause duration. Andrew Butcher (p. 90) reports a study of pauses in the reading of a German story in which the Grosjean model accounted for 86% of the variance.

Yet the matter is not simple. If a pause can be displaced from a syntactic break, it is evidently not a necessary consequence of the speaker's syntactic organization. Moreover, an inconsistent relation between pausing and syntax throws the communicative value of pauses as syntactic markers for the listener into question. Geoffrey Beattie (p. 131) addresses this issue in a study of spontaneous speech designed to assess whether pauses serve an encoding function for the speaker or a communicative function for the listener. He combined analysis of a speaker's speech into hesitant phases (high pause/phonation ratio) and fluent phases (low pause/phonation ratio) with an analysis of the speaker's gaze toward or away from his interlocutor. Beattie found that gaze aversion was very much more likely during hesitant phases than during fluent phases, and was significantly more probable at juncture pauses in a hesitant than in a fluent phase. If we assume that gaze aversion facilitates the self-absorption necessary for clausal planning, we may conclude that pauses, particularly during hesitant phases, may indeed reflect the encoding process. Beattie suggests, further, that "...juncture pauses in fluent phases, accompanied by speaker gaze at the listener, are presumably used to segment the speech for the decoder" (p. 139). However, this attempt to rescue a communicative function for juncture pauses by assigning them a dual function, depending on whether the speaker gazes at or away from the listener, strikes me as unduly tortuous.

The issue comes up again in a lucid and energetic paper by James Deese (p. 69), illustrating, among other things, the complexity of prosodic syntax markers in fluent speech. Deese reports selected analyses of substantial bodies of formal speech recorded at public hearings and committee meetings, at graduate seminars and in radio discussions. He analyzes pause structure in terms of short range grammatical relations within sentences and of long range relations in the structure of discourse. In the short range grammatical analysis, he makes several telling (if not always new) observations: (1) sentence boundaries are frequently (24% in one sample of 1043 randomly selected boundaries) marked neither by a rising or falling intonation contour nor by a break in acoustic energy (i.e., a pause); (2) where sentence boundaries are not marked by intonation or pause, they may often be marked by increased syllable rate on both sides of the boundary; (3) in tests with words excised from context listeners are most accurate in detecting a boundary when it is marked both by intonation contour and by a pause longer than 50 msec; (4) listeners judge a given pause as longer if it occurs at a clause break than if it occurs within a clause.

The burden of these observations is that the prosodic devices by which syntactic structure may be marked in fluent speech are far from simple. Moreover, the fact that listeners' judgments of pause length may be determined by the syntactic structure, rather than the reverse, suggests that other prosodic variables may be marking the syntax and may even be determining the pause structure.

Alan Henderson (p. 198) reports an ingenious study that speaks to this last point. Starting from the well-known click studies in which reaction time is elevated for a click placed in a syntactically marked, but prosodically unmarked, clause break, he asked whether he might not find a similar increase in reaction time for a tone placed in a syntactically unmarked, but prosodically marked, break. He measured English listeners' reaction times to a tone placed at the end of a word in each of six Czechslovakian sentences (none of the listeners knew, or recognized the language as, Czech). The sentences were manipulated so that the tone followed either an intonation fall and a pause, a fall alone, a pause alone, or neither. Reaction times to the tone were significantly longer in the three conditions where it followed a fall than in the other conditions. From this Henderson concludes that an intonation fall is a more salient cue to segmentation than a pause. Indeed, he turns the tables completely by suggesting that "...a break in signal energy is perceived as it is because of its context rather than being a cue to the structuring of the context" (p. 205). Certainly, as Henderson also sensibly suggests, a child (or an adult) learning a language is likely to find intonation a more reliable guide to syntactic structure than pauses--for which, the participants in this workshop unanimously agreed, the determinants are many and various.

If intonation is the principal cue to syntactic segmentation, might not the correlation of pause structure with syntax simply reflect a role of intonation in determining the location and, perhaps, length of pauses? Yet it cannot be the sole determinant, since the correlation between pausing and syntax would then be as high as between intonation and syntax. What then of rhythm and rate? Here the evidence is suggestive, though certainly not conclusive. Anne Cutler (p. 183) describes errors of syllable omission in spontaneous speech that have the effect of equalizing the number of syllables per foot, and thus making the speaker's output more isochronous. Interestingly, this may be just the effect of speakers' tendencies to bisect constituents, observed by Grosjean. Ballmer (p. 216) also remarks that pauses may serve to maintain the rhythmic pattern of an utterance. Finally, as far as rate is concerned, Grosjean (pp. 92-93) reports that pauses (both breathing and nonbreathing) tend to disappear, first from minor, then from major constituent breaks as rate is increased, until, at the highest rates (391 words per minute, in the study reported) only breathing pauses at some sentence boundaries remain.

What all this comes down to, then, is that pauses in fluent speech that seem to reflect the speaker's planning of syntactic structure, may be epiphenomenal consequences of other prosodic variables. As Butcher remarks: "...it would seem...neither feasible nor desirable to investigate pausing separately from certain other dependent variables, such as intonation, rhythm, and tempo" (p. 86). Butcher goes on to conjecture that: "...rather than all prosodic variation, including pausing, being determined by the syntactic structure, pausing is determined by intonation pattern, which in turn is normally coterminous with the syntactic pattern" (p. 90). If this proves to be so, we may conclude that syntax-marking pauses have little or no direct communicative function.

Let us consider now pauses to which we might be less inclined to assign intended communicative value: unfilled and filled pauses (that is, pauses containing hesitation sounds: uh...er, and the like) in which a speaker is quite evidently at a loss for a discourse plan, that is, for what to say. The central difficulty in studying the cognitive activity that underlies these hesitations is that, under normal circumstances, the investigator has even less idea of what speakers have in mind than the speakers themselves. One solution to the difficulty is to provide the speaker with a sort of open-ended script, a general discourse plan that the investigator knows, but that the speaker has to formulate. Thus, Wolfgang Klein (p. 159) induced much lengthy hesitation by asking people for route directions in a city. He could then compare the alternate routes, false starts, backtrackings, and roadblocks in the speakers' cognitive map, inferred from their utterances, with the clear "discourse plan" laid out in an actual map of the city.

Chafe (p. 169) offered his subjects a richer opportunity for self-revelation by asking them to tell what had happened in a 7-minute color movie (with sound effects, but no dialog) they had just seen. To introduce his analysis of the resulting spontaneous narratives, Chafe quotes William James on the stream of consciousness: "Like a bird's life, it seems to be made of an alternation of flights and perchings. The rhythm of language expresses this, where every thought is expressed in a sentence, and every sentence closed by a period" (James, 1890, p. 243). Chafe applies the metaphor to describe how someone tells a story, talking in spurts of a few seconds at a time, darting from one "focus of consciousness" to another. Foci, expressed in phrases or clauses with a rising pitch contour and a brief following pause, form "clusters" (or sentences) that end with a falling contour and a somewhat longer pause. Examining the content of foci within a cluster, we see how the speaker flits from point to point, capturing different aspects of a scene, or grouping a run of small events into a single purposive action. Long hesitations between clusters often reflect "time-consuming mental processing," as the speaker switches to a new time, place, actor, event, or scene.

Chafe argues that such "hesitation-ridden speech" should not be regarded as disfluent, even if technically ungrammatical, but rather "...should actually be highly valued as an accurate expression of a speaker's thoughts" (p. 180); he expects his mode of analysis to become "...an important and necessary aspect of hesitation research" (p. 180). Perhaps he is right, but I am not sure where it all leads. What he offers seems to be little more than a traditional *explication du texte*, extended from works of literature to the "creative act" (p. 170) of commonplace speech production.

Indeed, Chafe's chapter, like many others in this book, inadvertently draws attention to the contrast between pauses and errors as sources of inference about the cognitive processes of a speaker. Bernard Baars remarks in the general discussion, "...slips of the tongue are revealing in a way that pauses are not. Slips say something, and if you want to make inferences regarding deeper levels of control in speaking, you have more information to go on (p. 336)." In fact, the form of errors has already served to constrain our models of language processing, and their study is by no means exhausted. This point is well illustrated by two papers in the general introductory section of the book.

The first paper, by John Laver (p. 21), reports an experiment designed to induce errors by requiring subjects to speak pairs of vowels in a /pVp/ frame at increasingly fast rates. The hypothesis was that rapid, successive execution of vowel pairs drawing on relatively distinct neuromuscular systems (e.g., front-back, high-low) might invite competition between the two systems, leading to errors such as diphthongal glides, while different degrees of activation of roughly the same neuromuscular system (as in tense-lax front, or tense-lax back, vowel pairs) would preclude competition and so elicit few errors. This is precisely what was found. The experiment is modest and the report preliminary, but, as Laver points out, the principle of neuromuscular compatibility, illustrated in the pattern of errors, might be fruitfully applied in diverse areas of phonetic study, from the derivation of natural phonological classes to language acquisition and second language learning.

The second paper, by E. Keith Brown (p. 28), introduces the (for me) novel notion of "grammatical incoherence." An instance is the utterance of a young girl, stroking a moulting cat and holding up a hair: "How long do you suppose a life of a fur has?", spoken without hesitations and with apparent confidence that she had produced an intelligible utterance—as indeed she had. (For, as Brown remarks, listeners are far more tolerant of grammatical incoherence than of word distortion and such incoherence seldom impedes communication.) Brown uses this example to distinguish between two types of "blend," reflected in such incoherences. In a "cognitive blend" two related, but different cognitive structures with different surface realizations (fur, hair) compete, and the wrong one wins. Such errors may tell us about the organization of a speaker's lexicon and the processes of selection from it. In a "process blend," by contrast, "a single cognitive structure...may be realized by a number of surface forms and the resultant utterance is a blend of the processes that lead to these different forms" (p. 35). Thus, equivalent forms (e.g., How long a life does a hair have? How long a life has a hair? How long is the life of a hair?) may blend to produce "How long a life of a hair has?". Such errors may tell us about the processes of selecting from equivalence classes of syntactic forms. Of course, if a speaker avoids errors by pausing long enough to choose the right word or turn of phrase, we learn nothing: we detect his quandary, perhaps, but not its content. Brown's is an original and illuminating paper.

The final section of the book deals with cross-linguistic aspects. Here, it would seem, there might be an opportunity to dissociate general cognitive constraints due to syntax, tendencies toward stress or syllable timing and, perhaps, characteristic rates of speech. Thus, Grosjean, in a brief, but useful review paper (p. 307), reports that while pause time ratios in the spontaneous English and French of interviews are almost identical, they are arrived at in different ways: pauses are fewer, but longer in French, more frequent, but shorter in English. The constant ratios perhaps reflect breathing demands, common to all spoken languages; but the more frequent pauses of English reflect a tendency (syntactically governed, Grosjean implies, though it is not clear how) for speakers to insert pauses inside verb phrases, as they do not in French. On the other hand, a tendency, reported by Marc Faure (p. 287) for pauses in German to be most frequent before pronouns (as they are not in French) simply reflects a tendency, common to English, French, and German, to pause before the first or second word in a subordinate clause, of which the pronouns, in German, must be placed first.

Indeed, one may doubt the worth of including pause instruction in second language courses, recommended by Robert DiPietro (p. 320), for several reasons. First, the differences across the admittedly few languages that have been studied do not appear to be great. Alain Deschamps (p. 255) does report that French students tend to carry French patterns over into their English; but the most general effect among second language speakers, reported by both Deschamps and Manfred Raupach (p. 263) is, not surprisingly, an increase in the frequency (not the length) of pauses within sentences. Raupach reports, further, that many individuals have idiosyncratic pause patterns in their first language that they are likely to transfer into a second language. Finally, my overall impression, gathered from many papers in this book, is that pauses--other than those introduced for deliberate rhetorical effect--are largely automatic consequences of cognitive and physiological processes over which speakers have little control.

The last point emerges with particular cogency from studies, reviewed by Grosjean, comparing the pause structures of an oral language (English) with those of a manual-facial language (American Sign Language (ASL)). Freed from the demands of breathing, a sign language can reduce the amount of time spent in pausing: the pause time ratio for ASL is, in fact, less than half that of English. On the other hand, since a sign takes longer to form than a word, the overall rate of signs per minute is less than a third of the rate of English words per minute. Yet the proposition rates in the two languages are almost identical. The paradox is resolved by noting that, while the phonological and syntactic structures of a spoken language are largely due to sequential organization over time, a highly inflected signed language, such as ASL, can make extensive use of simultaneous manual, bodily, and facial gesture, distributed in space. Quite different means are thus used in the two languages to maintain what may be a natural rate of information flow common to all languages.

Despite these differences, the durations of ASL signs seem to be influenced by many of the factors that influence word duration, such as semantic novelty and position within a phrase. Moreover, the reduced pause time ratio of ASL is accomplished by shorter, not fewer pauses, so that its pause pattern can be quite similar to that of a spoken language. In fact the distribution of pauses between signs in "recited" sentences, like the distribution of pauses between words, reflects both constituent structure and the length of constituents: the model of Grosjean and his colleagues, discussed earlier, accounted for 72% of the variance in a study of ASL, as it had in a study of speech. Of course, the communicative function of pauses, no less than their possible determination by other prosodic variables, such as rhythm and rate, are even less understood for ASL than for spoken languages. Nonetheless, cross-modal comparison between signed and spoken languages promises to isolate universal cognitive and motoric constraints on language production.

In conclusion, we can be confident that universities will not now rush to establish Departments of "Pausology." On the contrary, the message of this interesting, if uneven, book is that the study of pauses in the speech flow will be advanced not by isolation, but by integration into other areas of phonetic and general psycholinguistic study.

REFERENCES

James, W. (1890). The principles of psychology. New York: Henry Holt. (Reprinted by Dover Publications, New York).

II. PUBLICATIONS

III. APPENDIX

IV. ERRATA

PUBLICATIONS

Beddor, P. S. (1983). *Phonological and phonetic effects of nasalization on vowel height*. Bloomington, IN: Indiana University Linguistics Club.

Bell-Berti, F., & Raphael, L. J. (1983). On identifying syllable boundaries. *Speech Communication*, *2*, 145-147.

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

Collier, R., Bell-Berti, F., & Raphael, L. J. (1982). Some acoustic and physiological observations on diphthongs. *Language and Speech*, *25*, 305-323.

Hanson, V. L. (1982). When a word is not the sum of its letters: Fingerspelling and spelling. In F. Caccamise, M. Garretson, & U. Bellugi (Eds.), *Teaching American Sign Language as a second/foreign language* (pp. 176-185). Silver Spring, MD: National Association of the Deaf.

Hanson, V. L. (1983). The reading process: Cognitive and psycholinguistic perspectives. *Contemporary Psychology*, *28*, 550-551.

Harris, K. S. (1983). Review (*The nature of stuttering*, edited by C. van Riper), *Applied Psycholinguistics*, *4*, 177-179.

Harris, K. S. (in press). Review (*Phonological intervention: Concepts and procedures*, edited by M. Crory). *Applied Psycholinguistics*.

Harris, K. S. (in press). Coarticulation as a component in articulatory descriptions. In R. Daniloff (Ed.), *Articulation assessment and treatment issues*. San Diego, CA: College Hill Press.

Katz, L., & Baldasare, J. (1983). Syllable coding in printed word recognition by children and adults. *Journal of Educational Psychology*, *75*, 245-256.

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

Kewley-Port, D., Pisoni, D. B., & Studdert-Kennedy, M. (1983). Perception of static and dynamic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, *73*, 1779-1793.

Liberman, A. M. (1983). What a perception-production link does for language. *The Behavioral and Brain Sciences*, *6*, 216.

Lindblom, B., MacNeilage, P. F., & Studdert-Kennedy, M. (in press). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Universals workshop*. The Hague: Mouton.

Lukatela, G., Moraca, J., Stojnov, D., Savić, M. D., Katz, L., & Turvey, M. T. (1982). Grammatical priming effects between pronouns and inflected verb forms. *Psychological Research*, *44*, 297-311.

Mann, V. A., & Liberman, A. M. (in press). Some differences between phonetic and auditory modes of perception. *Cognition*.

Mann, V. A., Shankweiler, D., & Smith, S. (in press). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. *Journal of Child Language*.

McGarr, N. S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research*, 134-141.

Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of /l/ and /w/. *Journal of the Acoustical Society of America*, *73*, 1751-1755.

Price, P. J., & Levitt, A. G. (1983). The relative roles of syntax and prosody in the perception of /š/-/č/ distinction. *Language and Speech*, *26*.

Recasens, D. (1983). Place cues for nasal consonants with special reference to Catalan. *Journal of the Acoustical Society of America*, *73*, 1346-1353.

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

Repp, B. H. (1983). Bidirectional contrast effects in the perception of VC-CV sequences. *Perception & Psychophysics*, *33*, 147-155.

Repp, B. H. (in press). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech Communication*.

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

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

Rubin, H., & Liberman, I. Y. (1983). Exploring the oral and within language errors made by language disabled children. *Annals of Dyslexia*.

Studdert-Kennedy, M. (Ed.). (1983). *Psychobiology of language*. Cambridge, *MA: MIT Press*.

Studdert-Kennedy, M. (1983). Limits on alternative auditory representations of speech. In C. W. Parkins & S. W. Anderson (Eds.), *Cochlear prostheses: An international symposium* (pp. 33-38). New York: New York Academy of Sciences.

Studdert-Kennedy, M. (in press). Perceiving phonetic events. In *W. H. Warren & R. E. Shaw* (Eds.), *Persistence and change: Proceedings of the First International Conference on Event Perception*. Hillsdale, NJ: Erlbaum.

Studdert-Kennedy, M. (in press). Review (*Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler*, edited by H. W. Dechert & M. Raupach).

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

APPENDIX

DTIC (Defense Technical Information Center) and ERIC (Educational Resources Information Center) numbers:

Status Report		DTIC	ERIC
SR-21/22	January - June 1970	AD 719382	ED 044-679
SR-23	July - September 1970	AD 723586	ED 052-654
SR-24	October - December 1970	AD 727616	ED 052-653
SR-25/26	January - June 1971	AD 730013	ED 056-560
SR-27	July - September 1971	AD 749339	ED 071-533
SR-28	October - December 1971	AD 742140	ED 061-837
SR-29/30	January - June 1972	AD 750001	ED 071-484
SR-31/32	July - December 1972	AD 757954	ED 077-285
SR-33	January - March 1973	AD 762373	ED 081-263
SR-34	April - June 1973	AD 766178	ED 081-295
SR-35/36	July - December 1973	AD 774799	ED 094-444
SR-37/38	January - June 1974	AD 783548	ED 094-445
SR-39/40	July - December 1974	AD A007342	ED 102-633
SR-41	January - March 1975	AD A013325	ED 109-722
SR-42/43	April - September 1975	AD A018369	ED 117-770
SR-44	October - December 1975	AD A023059	ED 119-273
SR-45/46	January - June 1976	AD A026196	ED 123-678
SR-47	July - September 1976	AD A031789	ED 128-870
SR-48	October - December 1976	AD A036735	ED 135-028
SR-49	January - March 1977	AD A041460	ED 141-864
SR-50	April - June 1977	AD A044820	ED 144-138
SR-51/52	July - December 1977	AD A049215	ED 147-892
SR-53	January - March 1978	AD A055853	ED 155-760
SR-54	April - June 1978	AD A067070	ED 161-096
SR-55/56	July - December 1978	AD A065575	ED 166-757
SR-57	January - March 1979	AD A083179	ED 170-823
SR-58	April - June 1979	AD A077663	ED 178-967
SR-59/60	July - December 1979	AD A082034	ED 181-525
SR-61	January - March 1980	AD A085320	ED 185-636
SR-62	April - June 1980	AD A095062	ED 196-099
SR-63/64	July - December 1980	AD A095860	ED 197-416
SR-65	January - March 1981	AD A099958	ED 201-022
SR-66	April - June 1981	AD A105090	ED 206-038
SR-67/68	July - December 1981	AD A111385	ED 212-010
SR-69	January - March 1982	AD A120819	ED 214-226
SR-70	April - June 1982	AD A119426	ED 219-834
SR-71/72	July - December 1982	AD A124596	ED 225-212
SR-73	January - March 1983	**	**

Information on ordering any of these issues may be found on the following page.

**DTIC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International
Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Language Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

Errata

Baer, T., & Alfonso, P. J. (1982). On simultaneous neuromuscular, movement, and acoustic measures of speech articulation. *Haskins Laboratories Status Report on Speech Research*, *SR-71/72*, 89-110.

p. 98: Figure 5. A corrected version of Figure 5 is provided below.

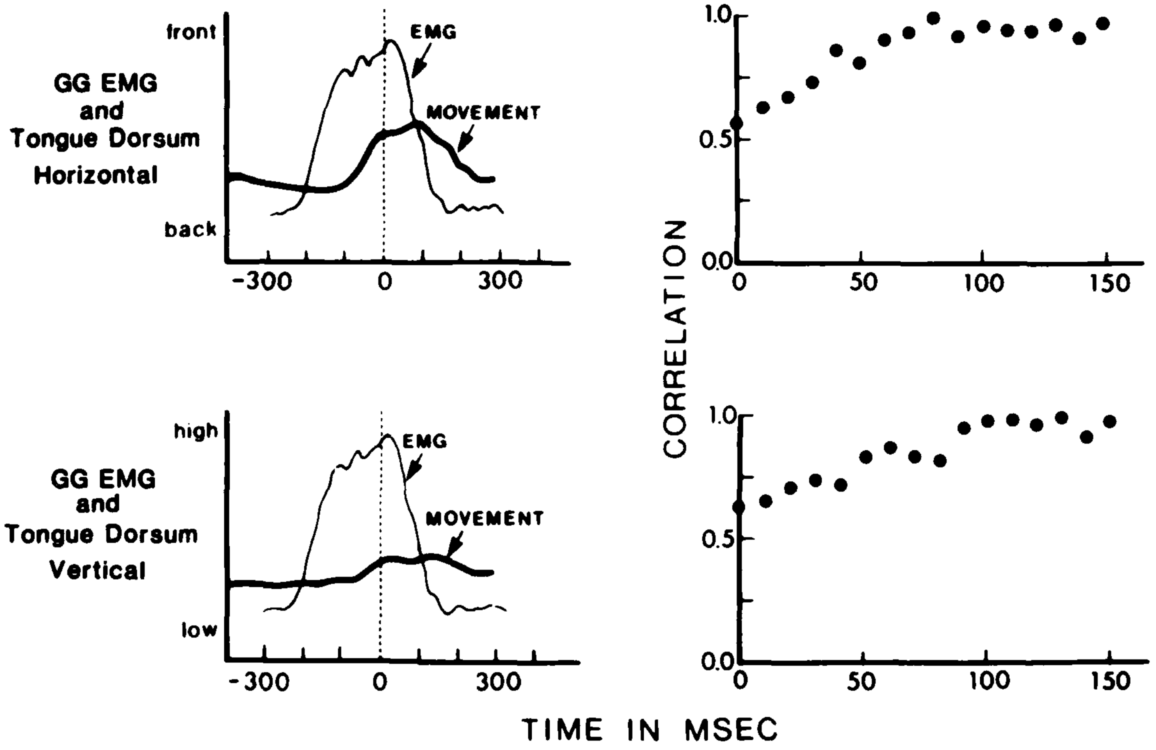


Figure 5. Genioglossus EMG activity with tongue dorsum horizontal movement (top left) and with tongue dorsum vertical movement (bottom left) during /i/. Correlation functions between the EMG curve and the respective movement curve are shown on the right.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY *(Corporate author)* Haskins Laboratories 270 Crown Street New Haven, CT 06510	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP N/A

3 REPORT TITLE

Haskins Laboratories Status Report on Speech Research, SR-74/75, April - September, 1983

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 September, 1983	7a. TOTAL NO. OF PAGES 228	7b. NO. OF REFS 467

8a. CONTRACT OR GRANT NO. HD-01994 NS13617 HD-16591 NS18010 N01-HD-1-2420 NS07237 RR-05596 NS07196 BNS-8111470 N00014-83-C-0083 NS13870	9a. ORIGINATOR'S REPORT NUMBER(S) SR-74/75 (1983) 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 April-30 September) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

-The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation;
-Phonetic coding and order memory in relation to reading proficiency: A comparison of short-term memory for temporal and spatial order information;
-Exploring the oral and written language errors made by language disabled children;
-Perceiving phonetic events;
-Converging evidence in support of common dynamical principles for speech and movement coordination,
-Phase transitions and critical behavior in human bimanual coordination,
-Timing and coarticulation for alveolo-palatals and sequences of alveolar +[j] in Catalan
-V-to-C coarticulation in Catalan VCV sequences: An articulatory and acoustical study;
-Prosody and the /š/-/č/ distinction;
-Intersections of tone and intonation in Thai
-Simultaneous measurements of vowels produced by a hearing-impaired speaker;
-Extending formant transitions may not improve aphasics' perception of stop consonant place of articulation;
-Against a role of "chirp" identification in duplex perception;
-Further evidence for the role of relative timing in speech: A reply to Barry
-Review (Phonological intervention: Concepts and procedures),
-Review (Temporal variables in speech).

DD FORM 1 NOV 65 1473 (PAGE 1)

UNCLASSIFIED

Security Classification

S/N 0101-807-6811 *This document contains no information not freely available to the general public. It is distributed primarily for library use.

A-31408

UNCLASSIFIED
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Perception: phonetic, events prosody, fricative tone, intonation, Thai formants, transitions, aphasics, stop consonants, place chirp, identification, duplex temporal variables, review Speech Articulation: dynamics, movement, coordination timing, coarticulation, alveolar, palatal, Catalan vowel, consonant, deafness timing, relative phonological intervention, review Reading: phonetic representation, speech comprehension phonetic coding, short term memory, temporal order, spatial order Motor Control: phase transitions, bimanual, coordination						

DD FORM 1 NOV 65 1473 (BACK)
S/N 0101-807-6821

UNCLASSIFIED
Security Classification

A-31409

LMED

8