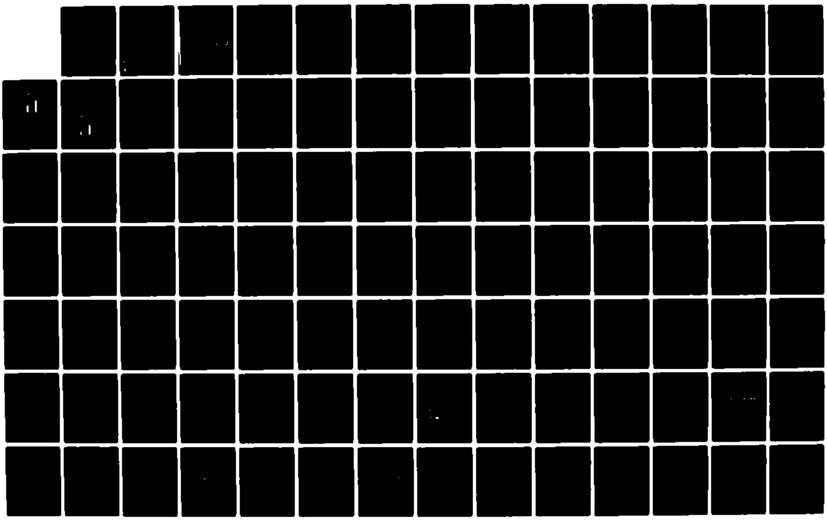
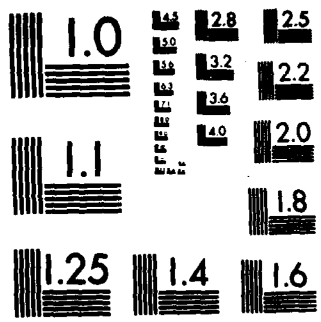
AD-A129 713 STATUS REPORT ON SPEECH RESEARCH A REPORT ON THE STATUS 1/4
AND PROGRESS OF S..(U) HASKINS LABS INC NEW HAVEN CT
M STUDDERT-KENNEDY ET AL. 1983 SR-73(1983)

UNCLASSIFIED N00014-83-C-0083 F/G 6/16 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

SR-73 (1983)

ADA 129713

Status Report on

SPEECH RESEARCH

January–March 1983

HASKINS LABORATORIES

SR-73 (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 January - 31 March 1983

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

(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.)

SR-73 (1983)
(January-March)

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 PRF-8006144
Grant BNS-8111470

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

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

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President and Research Director
Franklin S. Cooper,* Associate Research Director
Patrick W. Nye, Associate Research Director
Raymond C. Huey,* Treasurer
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*
Alice Healy*
Kiyoshi Honda[2]
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*
Michael Studdert-Kennedy*
Betty Tuller*
Michael T. Turvey*
Robert Verbrugge*
Douglas Whalen

Technical and Support Staff

Eric L. Andreasson
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
André Cooper
Tova Clayman
Patricia Ditunno
Steven Eady
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
+NIH Research Fellow

SR-73 (1983)
(January-March)

CONTENTS

I. Manuscripts and Extended Reports

I. MANUSCRIPTS AND EXTENDED REPORTS

THE INFLUENCE OF SUBCATEGORICAL MISMATCHES ON LEXICAL ACCESS

D. H. Whalen

Abstract. When the noise portion of an [s] or [š] is combined with vocalic formant transitions appropriate to the other fricative, the resulting consonantal percept is almost always that of the noise. Whalen (1982) has shown that the mismatch of transitions nonetheless slows the identification of that fricative. This result was extended to a lexical decision paradigm to answer two questions: Does the inappropriate transition slow down access of a word, or is the delay limited to tasks involving specifically phonetic judgments? Second, what could such a delay tell us about how the lexicon is searched? The stimuli were 48 English words and 48 phonotactically legal nonwords, each containing either [s] or [š]. Two versions of each stimulus occurred, one with the original vocalic portion, and one in which the vocalic formant transitions were inappropriate to the fricative. In a speeded lexical decision task, word judgments were slower when the transitions were inappropriate. A nonsignificant delay occurred in nonwords (as in a similar experiment by Streeter & Nigro, 1979). The implications for the logogen and cohort theories of lexical access are discussed. Lexical access is shown to be sensitive to fine phonetic detail.

INTRODUCTION

The noise portion of the alveolar and palatal voiceless fricatives in English is a powerful enough cue for place of articulation to override any place information in the vocalic formant transitions of accompanying vowels. Thus, if the vocalic segment from [sa] is excised and combined with the noise portion from [ša], the resulting percept is the syllable [ša]: The transitions seem to be ignored. Such an artificial mismatch, in which a cue is put in a new environment where its value is not sufficient to produce the appropriate percept, will be called a subcategorical phonetic mismatch; the cue that is overridden will be called a *mismatched cue*. The present experiment will determine whether such mismatched transitions affect decision time within a lexical decision task. The results will help us decide whether listeners make phonetic decisions based on isolated time slices of the acoustic stream, or rather integrate all the information they receive.

Acknowledgment. I would like to thank Louis Goldstein, Alvin M. Liberman, and Michael Studdert-Kennedy for helpful comments on this paper. This research formed part of a Yale University Ph.D. dissertation entitled Perceptual effects of phonetic mismatches. Support was provided by NICHD Grant HD-01994.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

Earlier experiments (Martin & Bunnell, 1982; Whalen, 1982) have shown that subcategorical mismatches, while not changing the phonetic percept, slow phonetic identification. When the transitions of fricative-vowel syllables are mismatched, phonetic identification of both the fricative and the vowel is slowed. Whalen (1982) argued that listeners attempt to integrate all cues available, even if the result of that attempt is not noticeable in the final phonetic judgment. Since those experiments elicited phonetic judgments, it is possible that the effect is limited to such rather unnatural tasks. The lexical decision task is more natural.

Subcategorical mismatches have been examined previously in a lexical decision task. Mismatched transitions into a medial stop resulted in slower times in a speeded lexical decision task (Streeter & Nigro, 1979). The effect only appeared for word judgments, but not for nonword judgments. Streeter and Nigro interpret this result in terms of an exhaustive lexical search, in which the physical nature of the nonword stimulus has no effect. There are other interpretations possible (one of which is given below in the Discussion section), and the effect itself needs replication. The present study uses the same lexical decision paradigm, and extends it.

One drawback to the Streeter and Nigro study was that the mismatched cue always preceded the overriding cue. Thus their results cannot distinguish between two inherently plausible explanations. One account would say that the subjects were slowed because they made a phonetic decision as the closure transitions were perceived and had to reverse that decision when the opening transitions were perceived. This account can be called "disposing," since each cue is dealt with in strict temporal order (cf. Whalen, 1982). The other account would assume that the subjects tried to integrate the information of each set of transitions and were slowed by the mismatch in its own right. This account can be called "integrating," since every cue over a (yet to be determined) time frame is examined in conjunction with the other cues. Only when the overriding cue comes first do these two accounts differ. The disposing account would then say that the mismatched cues should simply be ignored and thus not slow phonetic identification. The integrating account would say that the mismatched cues provide phonetic information, but if that information is to be overridden, the integration will take extra time no matter where the mismatch occurs. The present study will examine this question directly, by having the mismatched cue preceding the overriding cue in some cases, and following in others.

The phonetic experiments of Whalen (1982) have shown that mismatched cues that follow the overriding cue do slow judgments. This provided evidence against disposing theories (cf. Blumstein & Stevens, 1980; Cole & Scott, 1974; Klatt, 1979; Stevens, 1975). In a disposing theory, every time-slice of the acoustic stream is examined, without regard for context, for its phonetic contribution. Once this information is extracted, that time-slice is not considered further. The alternative, "integrating," theory (cf. Liberman, 1979; Liberman & Studdert-Kennedy, 1978; and Repp, 1982) was better able to account for the data of Whalen (1982). This account assumes that listeners deal with all phonetic information over a fairly large stretch of time, taking the overall acoustic context into account. Thus the mismatched cues that followed the overriding cues were just as disruptive as ones that preceded.

While the evidence from the phonetic experiments supported the integrating account, that account would lead us to expect mismatched cues in both words and nonwords to slow lexical decision. However, as already noted, Streeter and Nigro (1979) did not find an effect of mismatches in nonwords. If their finding is replicated, we would have to conclude either that the mismatch effect is limited to the strange combination of successful lexical access on the one hand and purely phonetic judgments on the other, or that the lack of an effect with the nonwords is an artifact of the lexical decision methodology. Finally, if we find no interaction between cue appropriateness and cue position, then the integrating account of speech perception will be further supported.

EXPERIMENTAL PROCEDURE

Materials

The test stimuli were 48 monosyllabic English words and 48 phonotactically possible, monosyllabic nonwords (see Appendix). Each contained either [s] or [š], in either initial or final position. All were chosen to be of relatively low frequency (less than 50 occurrences in the Kučera and Francis, 1967, corpus). For each word or nonword, there was another word or nonword that differed from it only in containing the other fricative. This matching made it possible to change only the transitions, leaving the vowel quality in the friction the same. Thus, for example, "soak" was matched with "shoak," "mess" with "mesh," and "sipe" with "shipe." The mean duration of test items was 569 msec. Words were slightly longer overall than nonwords (575 vs. 564 msec).

To avoid having fricatives in every word, two filler items were constructed for each test item. The fillers were all monosyllabic words or phonologically legal nonwords. The words were matched with the test words for frequency, and the distribution of phonemes in the nonwords approximated that of English words. The mean duration of filler items was 525 msec. Again, words were slightly longer than nonwords (532 vs. 518 msec).

A male native speaker of English recorded three tokens of each of the test and filler items. The stimuli were read in randomized order during a single recording session. Materials were low-pass filtered at 10 kHz and digitized at a sampling rate of 20 kHz. One token of each item was chosen for the experiment. Filler items were chosen for naturalness and clarity. Test items were chosen so that the friction and vocalic segment of the two corresponding items (such as "soak" and "shoak") were of equal duration. In this way, the two versions of each item (matched or mismatched transitions) were of equal duration.

Once the tokens had been selected, friction of each test item was combined with its corresponding vocalic segment. The resulting stimuli fell into four categories of interest: 1) The stimulus was a word containing vocalic formant transitions that matched the fricative percept generated by the noise ("appropriate transitions"); 2) The stimulus was a word, but the transitions were inappropriate; 3) The stimulus was a nonword, and the transitions were appropriate; and 4) The stimulus was a nonword, and the transitions were inappropriate. Note that every test item occurred with both appropriate and inappropriate transitions, and that, since friction always overrode transitions, both the matched and the mismatched versions of, for example, "soak" were identified as "soak."

The stimuli also varied systematically along other lines. There was an equal number of items with initial fricatives and items with final fricatives. This was varied to test the effect of mismatched cue position. In addition, there was an equal number of items whose lexical status changed from word to nonword or vice versa with the change of fricative (e.g., "soak," a word, and "shoak," a nonword) and items whose status remained the same with either fricative (e.g., the words "mess" and "mesh," and the nonwords "froose" and "froosh"). Thus in half the test items, the change from [s] to [š] would change the correct answer, and in half it would not.

Subjects

Two groups of subjects were tested, expert and naive. The exper listeners were 18 researchers at Haskins Laboratories, all of whom we phonetically trained and/or had extensive experience in phonetic researc Two were left-handed. The naive subjects were 18 volunteers, all nativ speakers of English, who were paid for their participation. One was left handed.

Apparatus

Subjects were seated in a quiet room and heard the stimuli over Telephonics TDH-39 headphones. They responded by pressing one of two buttons on a panel in front of them. The "yes" response was on the left and the "no" response on the right. During the test, if the answer was correct and within the stated time limit (longer than 100 msec and shorter than two seconds), a small light on the control box in front of them lit up. Their response time, answer, and the correctness of that answer went into a computer file after each trial.

Procedure

The subjects' task was to judge whether each item was an English word or not. They were told to hit the "yes" button if the item was a word and "no" if it was not. Examples of words and nonwords were given to the subjects. They were then instructed to judge the status of the item as quickly as possible. Subjects were told to expect a few mistakes, both because they could misperceive items and because they could press a button by accident. They were instructed to slow down if they made too many of the latter mistakes. It was explained that these were careful pronunciations, so that "toas" and "bline" were to be taken as nonwords, even if these pronunciations might occur instead of "toast" and "blind." Any word that was known only as a slang word was to be counted as a nonword. The feedback light was explained.

There were two conditions for the experiment. In the first, the subject heard all test items, half with appropriate transitions, half with inappropriate. Since there were two versions of each test item, only one could be presented to a subject in a standard lexical decision task (which requires each item to occur only once, in order to avoid priming effects). This forced the first analysis of the transition effect to be cross subject. In the second condition, the subject heard every test item again, but in its other version. The second condition thus resembled a lexical decision test in which each item has been primed by a repetition. The combination of the two conditions, while having the complication of speeded decisions on second presentation (cf. Dannenbring & Briand, 1982; Scarborough, Cortese, & Scarborough, 1977), allows the transition effect to be examined within subjects.

Two random sequences containing all the test and filler items were made. One version (with appropriate or inappropriate transitions) of each test item occurred in one sequence, with the other version occurring in the other. The assignment of subjects to one sequence or the other for the first condition was counterbalanced within groups.

A practice block, containing twenty words and twenty nonwords that did not occur in the test, was run to familiarize the subjects with the equipment and the task. After it was determined that no questions remained, the two test blocks of the first condition were run. A thirty second pause ocurred between blocks. Each block contained 144 trials, plus four "warm-up" stimuli at the beginning (which were not tallied in the results). After a short break, the two blocks of the second condition were run.

The stimuli were recorded on one channel of an audiotape while, on the other channel, a timing tone was recorded simultaneously with the onset of the stimulus. The inter-stimulus interval was three and a half seconds.

RESULTS

The results of the two conditions (first presentation of the test items vs. second presentation) and the two conditions together were analyzed similarly. An analysis of variance was performed on the mean reaction time with the following factors: Expert vs. naive subjects ("group"); vocalic formant transitions were appropriate to the fricative or not ("appropriate transitions"); word vs. nonword; and initial vs. final fricatives. A separate analysis was done for each condition, then a combined analysis with the added factor of condition.

Results for Condition 1

Only correct responses within the specified time limits (longer than 100 msec, shorter than 2 sec) were included in the analysis of the results. This gave an overall error rate of 8.6%. The rate was 10.7% for words and 6.4% for nonwords. One item effect showed up strongly in the errors: The word "deuce/douce" accounted for one out of seven errors on words. Errors occurred at approximately the same rate in the two versions of each word (8.7% for the original versions, 8.4% for the mismatched versions).

As can be seen from Figure 1, inappropriate transitions slowed lexical decision, $F(1,34) = 6.04$, $p < .02$. Subjects were 18 msec faster in their decisions when the transition was appropriate (means of 932 and 950 msec, respectively). It is also evident that nonwords took longer than words, $F(1,34) = 6.41$, $p < .02$. While inappropriate transitions delayed response for both words and nonwords, the effect was larger with the words, $F(1,34) = 4.16$,

$p < .05$. A separate analysis of variance of just the nonwords shows that the transition effect did not reach significance, $F(1,34) = 0.84$, n.s.

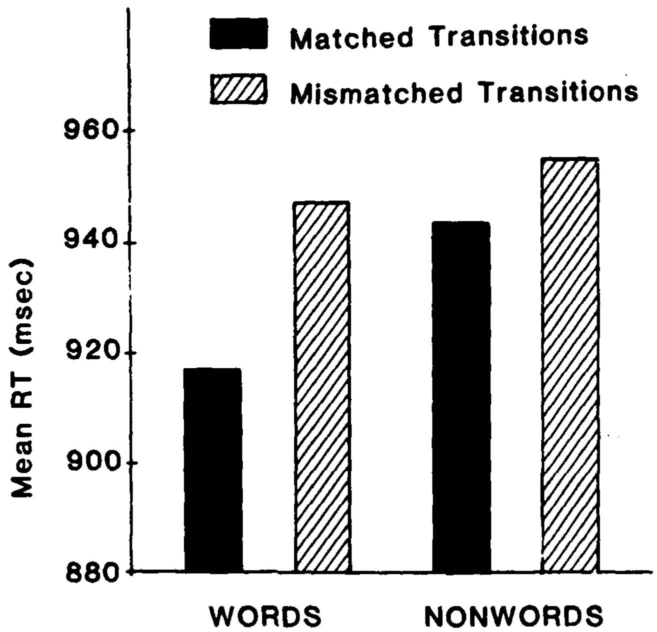


Figure 1. Lexical decision times for the first presentation of each item (Condition 1).

When the results were analyzed by item rather than subject, the transition effect did not reach significance, $F(1,92) = 2.17$, n.s. Since transition was a between-subject factor for the item analysis, and since the effect was of small magnitude, this outcome is not too surprising. However, it does mean that the results for the first presentation of an item alone do not allow us to conclude that the transition effect will hold for any word or nonword of English.

Items with initial fricatives (overriding cue preceding) took longer to identify than those with final fricatives (overriding cue following), $F(1,34) = 33.05$, $p < .001$ for the subject analysis, $F(1,92) = 6.06$, $p < .025$ for the item analysis. This occurred despite the greater average duration of the fricative-final items (583 msec for the final fricative items vs. 555 msec). This factor is not of great interest in itself. These groups necessarily contained different items. Thus the effect simply indicates that some items were reliably identified faster than others. However, there are many possible causes for such item effects, and we do not have the evidence for distinguishing among them. For present purposes, the initial/final factor is of interest only if it interacts with the appropriateness of transition factor, and this it did not do: The delay caused by inappropriate transitions was the same whether the friction came before the transitions or after: $F(1,34) = 1.56$, n.s., for the subject analysis, $F(1,92) = 1.37$, n.s., for the item analysis. Thus the effect was the same whether the overriding cue came first or not.

The experts were significantly faster than the naive subjects, $F(1,34) = 10.21$, $p < .01$. The means were 886 and 996 msec, respectively. One interaction involving this factor was significant. The inappropriate transitions slowed reaction times for both words and nonwords for both groups, but the difference for the word responses of the naive subjects was much larger than their nonword responses or the experts' response to either words or nonwords, $F(1,34) = 6.73$, $p < .02$. This could be a proportional effect due to the greater magnitude of their reaction times, since the transition effect was not significant for the nonwords for either group.

Results for Condition 2

The overall error rate for Condition 2 was 6.7%. The rate was 7.6% for words and 5.9% for nonwords. Errors occurred at roughly the same rate in the two versions of each word (7.2% for the original versions, 6.3% for the mismatched versions).

The results for this condition, as can be seen from Figure 2, are quite similar to those of the first condition. The effect of the appropriateness of

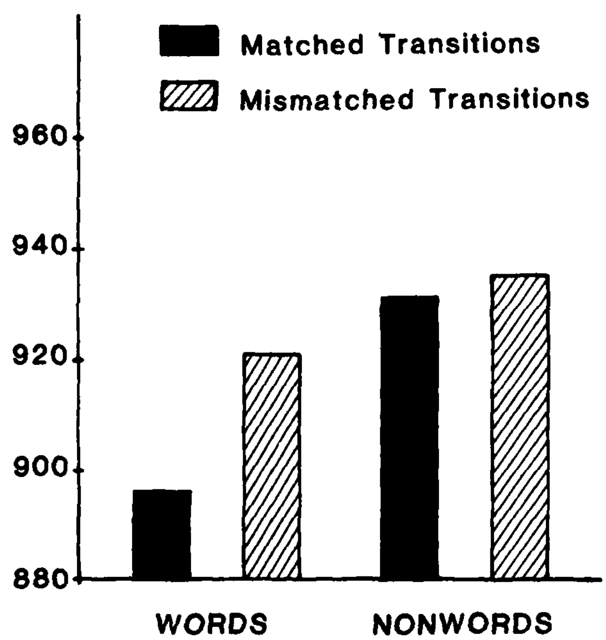


Figure 2. Lexical decision times for the second presentation of each item (Condition 2).

the transition was again significant, $F(1,34) = 5.64$, $p < .05$. Subjects were 14 msec faster in their decision when the transitions were appropriate. In the analysis by item, the transition effect again failed to reach significance, $F(1,92) = 2.28$, n.s., so that it still cannot, on these data, be reliably generalized to other items.

Decisions about words remained faster than about nonwords, $F(1,34) = 5.23$, $p < .05$. On average, subjects were 25 msec faster in their decision when the stimulus was a word (means of 908 msec vs. 933). In the analysis by item, this difference was not significant, $F(1,92) = 3.87$, n.s. Together, the analyses indicate that the word/nonword effect disappeared on the second presentation of these items (min $F'(1,112) = 2.22$, n.s.; cf. Clark, 1973). This occurred despite the larger difference in overall response time in comparison to the first condition (18 msec for Condition 1, 25 msec for Condition 2).

The interaction of word/nonword and appropriateness of transition did not reach significance, $F(1,34) = 3.28$, n.s., for the subject analysis, $F(1,92) = 1.38$, n.s., for the item analysis. However, since the first condition did have such an interaction, a separate analysis by subject of the nonword judgments was made. It showed that the transition effect was again not present for these subjects in the nonword judgments, $F(1,34) = 0.25$, n.s.

Items with initial fricatives were still identified more slowly than those with final fricatives, $F(1,34) = 31.16$, $p < .001$ for the subject analysis, $F(1,92) = 8.87$, $p < .05$ for the item analysis. The interaction of position of the fricative and appropriateness of transition was also not significant, $F(1,34) = 0.34$, n.s., for the subject analysis, $F(1,92) = 0.05$, n.s., for the item analysis. On second presentation of an item, then, inappropriate transitions again slowed the judgment whether they preceded or followed the friction.

The experts were again significantly faster than the naive subjects, $F(1,34) = 5.98$, $p < .02$. The means were 872 and 970 msec, respectively. No interactions with this factor were significant. Thus the effects of interest seem to be independent of linguistic sophistication.

Results for Conditions 1 and 2 Combined

When the results for first and second presentation of an item are considered together, the effect of the appropriateness of the transition was significant for the subject analysis, $F(1,34) = 15.26$, $p < .001$. The transition effect did not reach significance in the item analysis, $F(1,92) = 3.81$, $p = .054$, but the min F' did (min $F'(1,17) = 6.2$, $p < .025$). Decisions were 17 msec faster when the transition was appropriate (means of 922 msec for the appropriate and 939 for the inappropriate transitions). Since each subject's data now contain responses to both versions of each test item, intersubject variability is much reduced for the subject analysis. In the item analysis, each subject gave a response to each version of the item, so that the subject variability is much reduced there as well. The lack of an interaction between condition (i.e., first presentation vs. second presentation of each item) and appropriateness of transition, $F(1,34) = 0.08$, n.s., for the subject analysis, $F(1,92) = 0.36$, n.s., for the item analysis, indicates that the slowing effect of inappropriate transitions is the same for initial access of a word and for the second access.

Across the two conditions, the word/nonword factor interacted with the appropriateness of transition in the subject analysis, $F(1,34) = 6.68$, $p < .02$. The item analysis showed no interaction, $F(1,92) = 0.62$, n.s. While the decisions were slower to both words and nonwords when the transitions were inappropriate, the effect was much larger with words (28 msec vs. 8 msec). A separate analysis on just the nonwords showed that the delay with nonwords was again not significant in the subject analysis, $F(1,34) = 1.10$, n.s. The item analysis alone shows a significant transition effect for the nonwords, $F(1,92) = 4.46$, $p < .05$, but the two analyses together are not significant, min $F'(1,5) = 0.88$, n.s.

Decisions about words remained faster than about nonwords, $F(1,34) = 7.10$, $p < .025$ for the subject analysis, $F(1,92) = 5.59$, $p < .025$ for the item analysis. On average, subjects were 21 msec faster in their decision when the stimulus was a word (means of 920 msec vs. 941).

The initial/final factor was still extremely significant, $F(1,34) = 68.09$, $p < .001$ for the subject analysis, $F(1,92) = 24.31$, $p < .001$ for the item analysis. The items with initial fricatives took longer to decide upon (951 msec) than those with final fricatives (910 msec). However, in these combined results, there was still no interaction between initial vs. final fricative and the appropriateness of transition, $F(1,34) = 2.63$, n.s., for the subject analysis, $F(1,92) = 0.62$, n.s., for the item analysis.

The effect of hearing the item for the second time was one of shortening the decision time by an average of 20 msec, $F(1,34) = 4.70$, $p < .05$ for the subject analysis, $F(1,92) = 76.76$, $p < .001$ for the item analysis. This factor did not interact with either the word/nonword or the appropriateness of transitions factor, together or singly (the F value was less than 1 in most cases). That the speeding effect of repetition was present in the nonwords as well as the words is confirmed in the separate analysis of the nonword results. Responses to the second presentation of a nonword were, on average, 18 msec faster than to the first, $F(1,34) = 4.19$, $p < .05$ for the subject analysis, $F(1,92) = 42.90$, $p < .001$ for the item analysis.

The experts were significantly faster than the naive subjects, $F(1,34) = 8.25$, $p < .01$. The means were 879 and 983 msec, respectively. This factor was involved in three interactions. One involved only the location of the fricative (initial or final), which is not relevant to the present discussion except in its lack of an interaction with the transition factor. The two remaining interactions involved three and four other factors; no natural explanation for the interactions was apparent.

DISCUSSION AND CONCLUSION

The delay caused by inappropriate transition previously found in phonetic identification was found again in a more natural paradigm. A mismatch of fricative and transitions caused a delay in lexical access on both the first presentation and the second. Even when subjects are not paying attention specifically to the segmental phonetic structure of an item, a subcategorical phonetic mismatch slows the judgment. The effect failed to hold up in the nonword decisions. Since this result was obtained previously (Streeter & Nigro, 1979), it is not unexpected. However, the explanation given by those authors is not appealing. An alternative, that the lexical decision process itself is responsible for the disappearance of the effect, will be discussed below.

In the previous paragraph and in the discussion below, there is a benign ambiguity about the origin of the mismatch effect: We have assumed that mismatches slow phonetic analysis, but it is possible that the slower times simply reflect the subjects' lessened confidence in their judgments. In either event, the implications for the integration vs. disposal issue are equivalent. Experiments could be devised to choose between these alternatives, but the present study does not do so. The remainder of the discussion will argue the first interpretation only, although arguments for the second could be constructed with equal ease.

The lack of an interaction between the position of the transitions (whether the fricative was initial or final) and the appropriateness of transitions shows that listeners were attending to the mismatched transitions whether the overriding cue came before or after them. If the noise cue of fricative-initial stimuli were dealt with and disposed, then the place information of the transitions would not cause a delay even if it conflicted with the place information of the noise. Listeners do not "dispose" of each piece of the phonetic stream as it comes, but rather integrate over a larger stretch. The present stimuli do not help us decide just how large a stretch this integration covers.

Other considerations can be mentioned here (cf. Whalen, 1982). If each slice of the signal were treated as a cue to one or more phones independent of the rest of the signal, the phonetic construct would get out of hand. Each slice would give information about one particular phone, but there are often ten or more 25-msec slices in one fricative noise. Even if each slice is sufficient to identify the fricative, the phonetic construct does not have ten fricatives for each noise. In addition, some parts of the signal have a separate significance in isolation that would be misleading if each time slice were considered alone. For example, the transitions of the vocalic segment, if presented in isolation, give rise to a stop percept (cf. Whalen, 1982). There must be some way of telling that, with no silent closure, the transitions are not to be taken as constituting a stop. That is, the signal must be integrated over a larger piece of the signal. Thus, even a disposing account must make some use of integration.

Results similar to those obtained in this study were interpreted by Streeter and Nigro (1979) to support the notion that the mismatched cues are not dealt with in the construction of the phonetic percept, but rather are carried along in a "degraded" representation. Their claim relies on lack of a delaying effect of mismatches in nonwords. They assumed that the construction of the phonetic representation of an item's two versions would take the same amount of time but that the representation of the mismatched version would not be as well-constructed as that of the matched version. This difference was equated with the difference between a stimulus presented with and without added noise. The lack of an effect of mismatched cues in the nonwords would thus depend on there not being any entry in the lexicon to match, so that the quality of the stimulus would not affect the decision time. While no studies of lexical decision have used both auditory presentation and added noise, there have been visual analogs. Stanners, Jastrzembski, and Westbrook (1975), for example, found that a random dot pattern partially obscuring the words and nonwords slowed reaction times for both categories and in fact more so for the nonwords. Streeter and Nigro predict the opposite for auditory presentation. If they are wrong and nonwords in noise are classified as nonwords more slowly than those without added noise, then their proposal would be less than convincing. It seems more plausible that something in the nonword decision itself is responsible for the reduction in the mismatched cue effect.

One possible explanation attributes the reduced effect to an added step in the nonword decision. The extra time spent on nonword decisions may reflect phonetic reanalysis, in which even matched cues are treated as suspect: When a string is found to lack an entry in the lexicon, it may be rechecked for previously undetected phonetic ambiguities that might make it a word. If the original analysis is retained, the nonword decision is then made, but the process will have reduced the difference in response time between items with matched and mismatched transitions. If this account is correct, the delays found here and in Streeter and Nigro (1979) are inherent in the phonetic analysis; their disappearance in the nonwords is an artifact of the lexical decision methodology.

Some support for the added-step interpretation of the nonword data is contained in the data from the second condition. Previous results of repeated presentation are relevant here. Scarborough et al. (1977) demonstrated that repetition of items decreases reaction times even after a lag of 31 items. More importantly, they found that the effect of repetition on a well-known factor in lexical decision times (in this case, frequency of occurrence) varied across experiments. In some cases, the frequency effect disappeared, while in others it persisted.

With the present experiment, the effect of inappropriate transitions was the same on the first presentation as on the second. If anything, we might have expected the transition effect to weaken when the words were being heard for the second time, since the criterial levels for recognition would presumably be lowered. That did not happen. Thus the effect found seems to occur in both the initial access of a lexical item and on the second. The second presentation of words reduced the time required to respond to them, as would be expected (Forbach, Stanners, & Hochhaus, 1974). But repetition was equally effective in reducing the time required to judge nonwords (cf. Dannenbring & Briand, 1982). We would perhaps expect that all times would be reduced by practice, but that words, since they prime themselves (Scarborough et al., 1977), should show greater effects than nonwords. This was not the case. Streeter and Nigro (1979) and others assume that nonwords do not have an entry in the lexicon. An item without an entry in at least a temporary lexicon could not be self-priming. Any time gained in the nonword decisions, then, would be due to a faster search. This could be accomplished either through familiarity with the task or by searching a subset of the lexicon. Even if a subset of the lexicon is searched on the second presentation, the words should still have an added advantage from the priming. The evidence leads us to say that nonwords have lexical representations, at least within a test session.

Lexical decision judgments, then, are affected by subcategorical mismatches that do not result in overt ambiguities. Since most theories of lexical access are vague about the properties of the phonetic input, they can accommodate almost any result from experiments of the present sort. I will briefly discuss the treatment of the present results in two of them, the logogen theory (Morton, 1969, 1979) and the cohort theory (Marslen-Wilson, Note 1; Marslen-Wilson & Welsh, 1978).

The logogen theory assumes that words (or morphemes) are collections of phonetic, semantic, and other properties with an associated threshhold. If that threshhold is met, that word is accessed. Priming is a temporary lowering of the threshhold, while greater frequency within the language lowers the threshhold permanently. Logogens are completely passive.

The cohort theory asserts that words are organized by their initial sounds into groups or "cohorts." Once the initial sounds (probably a half syllable) are identified, all words in that cohort become candidates. These candidates are eliminated by further incoming data until only one word remains, or until none remains. Cohorts, then, are partially active.

One common feature of these two theories is a distinction between phonetic analysis and lexical access. Neither theory has much to say about the phonetic analysis, except that, if it occurs, it does so either before input to the logogens, or in step with cohort activity. The mismatches introduced into the present stimuli could have affected either process. If the phonetic analysis was slowed, the decision would be slowed for both words and nonwords. If the search was conducted on a degraded stimulus (as proposed by Streeter & Nigro, 1979), the decision for words would be slowed while that for nonwords might not be (see the discussion above). The two theories of lexical access are compatible with either interpretation.

The logogen theory is more easily made compatible with the delay in the phonetic analysis. In that event, the activation of logogens would be delayed until the phonetic analysis was completed, so the theory would not need to be modified to take account of these results. If the degraded stimulus version were correct, then a degraded stimulus would add less to the correct logogen's activation. Then the threshhold must be lowered over time or the activation increased for the word decision to be initiated.

The cohort theory is also compatible with both versions. The two versions look much more similar to each other with this theory. In both versions, early mismatches would slow the cohort's self-activation. If the selection of a cohort is delayed a few milliseconds because of a mismatched cue, then the final output of that cohort will be delayed. Later mismatches, occurring after the cohorts are active, would either be available to the cohort later, or would be more slowly utilized by the cohort. Since the lexical lookup stage in the cohort theory is interleaved with the phonetic decisions, the choice between the two explanations is of limited interest.

The two main theories of lexical access are thus unaffected by the choice between assigning the effect of the mismatched cues to the phonetic analysis or to the use of a degraded analysis in the search of the lexicon.

Note that both the logogen theory and the cohort theory are disposing. The logogen theory is obviously disposing, since each time-slice adds a certain amount to the relevant logogens' activation. Conflicting information would not lower that activation, but simply add to another logogen's activation. Thus the logogen theory has the same problem of explaining why it is that the transition cues for fricatives are not also treated as cues for stops. The cohort theory behaves similarly.

The proposal that nonword judgments require phonetic reanalysis would allow the cohort model to explain something it has had trouble explaining before, namely, the consistent finding that nonword decisions take longer than word decisions. When all words in a cohort are contradicted by the phonetic input, the nonword decision should be possible, thus giving faster reaction times for nonwords. If the cancellation of a cohort instead called for a phonetic reanalysis and check that the proper cohorts had been active, another step would be introduced and the effect would be explained. Shorter nonword decisions could be expected for items that eliminate all possibilities very early in the word. Since the present items were monosyllabic, they do not provide the best evidence for the cohort theory.

The phonetic reinterpretation proposal gives us an alternative proposal for another set of results as well. Phoneme monitoring has been shown to be speeded when the phoneme-bearing stimulus is a word as compared with a phonetically similar nonword (Rubin, Turvey, & Van Gelder, 1976). If subconscious lexical access is taking place, then subconscious failure of lexical access must be taking place as well. The theory proposed by Rubin et al. is that the phonological representation available to the words makes the phonemic judgment easier. It could also be that a phonetic reanalysis occurred with the nonwords (even though lexical status was not explicitly at issue), thus slowing the (equally well-supported) phoneme response.

The current results demonstrate that even in the paradigm of judging lexical status, subjects are sensitive to subcategorical phonetic mismatches. Since this effect occurs whether the mismatched cue precedes the overriding cue or follows it, we can conclude that listeners are attempting to attribute the proper value to every cue they receive, even if it seems redundant.

REFERENCE NOTE

1. Marslen-Wilson, W. D. Sequential decision processes during spoken word recognition. Paper presented at a meeting of the Psychonomic Society, San Antonio, Texas, November, 1978.

REFERENCES

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

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

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

Dannenbring, G. L., & Briand, K. Semantic priming and the word repetition effect in a lexical decision task. Canadian Journal of Psychology, 1982, 36, 435-444.

Forbach, G. B., Stanners, R. F., & Hochhaus, L. Repetition and practice effects in a lexical decision task. Memory & Cognition, 1974, 2, 337-339.

Klatt, D. H. Speech perception: A model of acoustic-phonetic analysis and lexical access. Journal of Phonetics, 1979, 7, 279-312.

Kučera, H., & Francis, W. N. Computational analysis of present-day American English. Providence, R.I.: Brown University Press, 1967.

Liberman, A. M. An ethological approach to language through the study of speech perception. In M. von Cranach, K. Poppa, W. Lepenies, & D. Ploog (Eds.), Human ethology. Cambridge: Cambridge Press, 1979, 682-704.

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

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

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

Morton, J. The interaction of information in word-recognition. Psychological Review, 1969, 76, 165-178.

Morton, J. Word recognition. In J. Morton & J. C. Marshall (Eds.), Psychologinguistics 2: Structure and processes. Cambridge, Mass.: MIT Press, 1979, 107-156.

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

Rubin, P., Turvey, M. T., & Van Gelder, P. Initial phonemes are detected faster in spoken words than in spoken non-words. Perception & Psychophysics, 1976, 19, 394-398.

Scarborough, D. L., Cortese, C., & Scarborough, H. Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 1-17.

Stanners, R. F., Jastrzembski, J. E., & Westbrook, A. Frequency and visual quality in a nonword classification task. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 259-264.

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

Streeter, L. A., & Nigro, G. N. The role of medial consonant transitions in word perception. Journal of the Acoustical Society of America, 1979, 65, 1533-1541.

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

Whalen, D. H. Subcategorical phonetic mismatches slow phonetic judgments. Haskins Laboratories Status Report on Speech Research, 1982, SR-71/72, 167-198.

Appendix--Stimuli for Lexical Decision Task

Numbers in parentheses are the frequencies from Kučera and Francis (1967).

			Initial s/š		Final s/š	
			word	nonword	word	nonword
s/š	s	1.	soak (7)	shoak	goose (4)	goosh
change		2.	sap (1)	shap	moss (9)	mosh
caused		3.	soup (16)	shoup	bus (34)	buhsh
change in					[buss (1)]	
word/		4.	soap (22)	shoap	toss (9)	tosh
nonword		5.	soy (1)	shoy	fleece (0)	fleesh
status		6.	silk (12)	shilk	fuss (4)	fush
	š	1.	shade (28)	sade	trash (2)	trass
		2.	shaft (11)	saft	cash (36)	kass
		3.	shout (9)	sout	gauche (1)	goass
		4.	shut (46)	sut	bush (14)	boos
		5.	shove (2)	suv	rash (1)	rass
		6.	chef (9)	sef	wash (37)	woss
s/š		1.	shoot (27)	shuke	mess (22)	giss
change			[shute (1)	chute (2)]		
caused			suit (48)	suke	mesh (4)	gish
no		2.	sift (0)	sipe	brass (19)	pless
change			[sifted (3), sifting (1)]			
in			shift (41)	shipe	brash (1)	plesh
word/		3.	sack (8)	sek	crass (2)	duss
nonword			shack (1)	shek	crash (20)	dush
status		4.	self (40)	sofe	lass (2)	koos
			shelf (12)	shofe	lash (6)	koosh
		5.	sock (4)	seeg	lease (10)	woas
			shock (31)	sheeg	leash (3)	woash
		6.	sake (41)	sud	douce (1)	froose
			shake (17)	shud	douche (0)	froosh

THE SERBO-CROATIAN ORTHOGRAPHY CONSTRAINS THE READER TO A PHONOLOGICALLY ANALYTIC STRATEGY*

M. T. Turvey,+ Laurie B. Feldman,++ and G. Lukatela+++

Abstract. The Serbo-Croatian language is written in two alphabets and its orthography is phonologically shallow: The grapheme to phoneme correspondences are simple and direct in both the Roman and Cyrillic alphabets. Results of a series of experiments that exploit the special properties of the Serbo-Croatian writing system indicate that in word recognition, skilled readers access the lexicon in a manner that must include an analysis of phonological components. This evidence for a phonological recognition strategy in Serbo-Croatian is not subject to the same criticisms as the evidence in English: 1) More consistent phonological effects have been demonstrated with words than with pseudowords; 2) The Cyrillic form of a word and the Roman form of that same word form the basis for comparison and these forms are necessarily equivalent both in terms of orthographic regularity and the reliability of grapheme-phoneme correspondences. In summary, interpretation of the data suggest that a phonological recognition strategy in Serbo-Croatian is not optional.

Among the Southern Slavic languages, there are two groups: an Eastern group from which Church Slavonic, Macedonian, and Bulgarian emerged, and a Western group from which Serbo-Croatian and Slovenian emerged. Old Church Slavonic was the literary language of Serbia (a republic of Yugoslavia) until the eighteenth century when it was replaced by Serbo-Croatian. Today, the Serbo-Croatian language includes three main dialects: a) štokavski, b) kajkavski, and c) čakavski. Within štokavski there are again three dialects and many of these variations (including some of a phonetic nature) are captured by the written language, for example, mliko, mleko, mlijeko (milk).

From the vantage point of the student of reading, the Serbo-Croatian orthography is of interest in two major respects. First, it bears a simple relation to the phonemics (as classically defined) of the language and introduces no special, rule-governed adjustments to preserve morphological

---

*To appear in L. Henderson (Ed.), Orthographies and reading. London: Lawrence Erlbaum, in press.

+Also University of Connecticut.
++Also Dartmouth College.
+++University of Belgrade.

Acknowledgment. Preparation of this paper and the research on which it is based were supported by NICHD Grant HD-01994 to Haskins Laboratories, NICHD Grant HD-08495 to the University of Belgrade, and NSF Grant BNS 7924409 to Laurie Feldman. We thank Carol Fowler for her help.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

relatedness. Moreover, it is a highly-inflected language. Indeed, the orthographic form of a root morpheme is sometimes varied in order to preserve a tight correspondence with the phonemes of the spoken language. For example, SNAH+A and SNAS+I are forms (nominative singular and dative singular, respectively) of the same word (daughter-in-law).[1] That the Serbo-Croatian orthography directly and consistently transcribes the phonemes of the language is due in large part to the deliberate alphabet reforms of the last century. The old Slavonic alphabet contained about 45 letters, some of which were not essential to Slavonic-Serbian, that is, the Serbo-Croatian language in use in Serbia in the second half of the eighteenth century. Although others preceded him, it was Vuk Karadžić (popularly referred to as Vuk) who systematically applied the principle of a strictly phonemic alphabet by deleting some characters and introducing new characters in place of compound letters. Karadžić adopted a simple principle: "Write as you speak and read as it is written." (Consequently, all written letters are pronounced and none are made silent by context.) Karadžić's work was controversial at the time, mainly because it reduced the similarity of Serbian and Russian Cyrillic script--it 'Latinized' the Serbian alphabet.

The second interesting aspect of the Serbo-Croatian orthography is that there are two alphabet versions--a Roman version and a Cyrillic version--as shown in Table 1 and Figure 1. Facility with both alphabets is commonplace among Yugoslavians although actual usage tends increasingly toward the Roman. Inspection of Figure 1 readily reveals that whereas there are letters unique to one or the other alphabet, some letters are shared. Of these shared letters, some (A, E, O, M, K, T, J) have a common phonemic interpretation; some (H, P, C, B) are ambiguous, receiving different phonemic interpretations depending on whether they are treated as Roman or as Cyrillic. From the perspective of the experimental investigation of processes underlying word recognition, this latter feature is especially useful, as will be evident below.

There has been much debate about whether fluent reading proceeds with reference to phonology.[2] Negative arguments usually predominate when the departure point is a consideration of the English orthography, which represents the phonology of the language in a complex fashion. It is felt that the internal processing costs of referencing the phonology are prohibitive and the benefits nonexistent. Not surprisingly the argument is more positive when the point of departure is a consideration of the Serbo-Croatian writing system. Experimentally, the debate has come to ground as the issue of phonological influences on lexical decision: Do phonological variables affect the speed of distinguishing letter strings that are words from letter strings that are not words? The research reviewed here has shown that for native Serbo-Croatian readers and written Serbo-Croatian material the answer is "Yes." On the basis of this research it can be argued that visual word-recognition in Serbo-Croatian proceeds with reference to the phonology.

When discussing phonological involvement in word recognition, it is important to distinguish between the notions of (i) a phonologically analytic strategy that precedes lexical access and (ii) a phonological representation that is arrived at only subsequent to lexical access. Continuous with the latter notion is the often made claim that, in reading, the lexicon is accessed via visual aspects of the printed word. A phonologically analytic strategy, on the other hand, is continuous with the claim that in reading, the

TABLE 1

SERBO-CROATIAN				
ROMAN		CYRILLIC		
PRINTED UPPER CASE	PRINTED LOWER CASE	PRINTED UPPER CASE	PRINTED LOWER CASE	LETTER NAME IN I.P.A.
A	a	А	а	a
B	b	Б	б	bə
C	c	Ц	ц	tsə
Č	č	Ч	ч	tʃə
Ć	ć	Ћ	ћ	tʃjə
D	d	Д	д	də
Đ	đ	Ђ	ђ	dʒjə
DŽ	dž	Џ	џ	dʒə
E	e	Е	е	ɛ
F	f	Ф	ф	fə
G	g	Г	г	gə
H	h	Х	х	xə
I	i	И	и	i
J	j	Ј	ј	jə
K	k	К	к	kə
L	l	Л	л	lə
LJ	lj	Љ	љ	ljə
M	m	М	м	mə
N	n	Н	н	nə
NJ	nj	Њ	њ	njə
O	o	О	о	ɔ
P	p	П	п	pə
R	r	Р	р	rə
S	s	С	с	sə
Š	š	Ш	ш	ʃə
T	t	Т	т	tə
U	u	У	у	u
V	v	В	в	və
Z	z	З	з	zə
Ž	ž	Ж	ж	ʒə

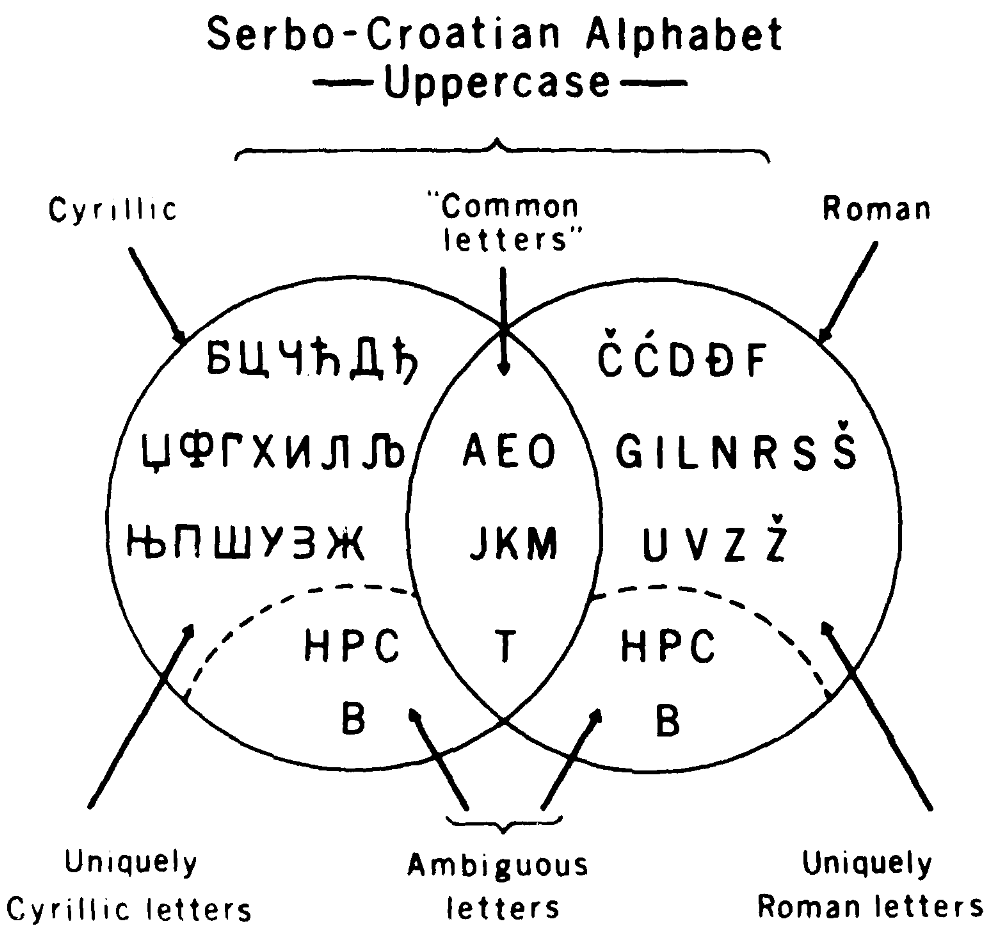


Figure 1. Letters of the Roman and Cyrillic alphabets.

lexicon is accessed via phonological aspects of the printed word that are specified in the details of the orthographic structure. In recognizing a word, the word's morphophonological structure must be determined and lexical access is a process that arrives at the morphophonological representation of the word from the details of its orthographic specification. The argument that lexical decision proceeds by reference to the phonology is intended to be an argument for a phonologically analytic access strategy. Given the nature of the Serbo-Croatian orthography (i.e., morphophonemes map relatively simply to classical phonemes as well as to orthography), a phonologically analytic strategy is the most simple and the most efficient.3

Before reviewing the Serbo-Croatian experiments we should note two kinds of data from lexical decision research that are interpreted as evidence for phonological involvement in the accessing of English lexical items. First, rejecting a pseudoword (e.g, BRANE) that sounds exactly like a real word (e.g, BRAIN) is more difficult (that is, associated with slower latencies) than rejecting a pseudoword that does not sound like any word (Coltheart, Davelaar, Jonasson, & Besner, 1977). An analogous observation on homophonous words is tenuous, holding only when the pseudoword foils do not sound identical to lexical items (Davelaar, Coltheart, Besner, & Jonasson, 1978).

We cannot take too seriously an argument for phonological involvement in lexical access that is based solely on the results obtained with pseudowords homophonous with words. Ignoring discussion as to whether or not the pseudoword homophone effect can be attributed to visual similarity (contrast Martin, 1982, with McQuade, 1980), the argument rests on the truth of the assertion that a pseudoword like BRANE is responded to comparatively slowly because it is phonologically identical to BRAIN. But letter strings that sound alike when spoken aloud may not be identical in terms of the phonological description that governs lexical decision; formally it is appreciated that the phonetic representation of an English word is distinct from its morphophonological representation. In sum, the comparative slowness of BRANE cannot be attributed unequivocally to phonological factors, viz., a morphophonological representation in common with that of an actual word.

Second, English words that are "regular," in the sense of complying with grapheme-phoneme correspondence rules such as Venezky's (Venezky, 1970) are accepted faster than English words that are "exceptions" to these rules. Results are inconsistent, however (compare Coltheart, Besner, Jonasson, & Davelaar, 1979, with Bauer & Stanovich, 1980, and Parkin, 1982). In part, the controversy may reflect a difficulty in defining regular and irregular correspondences for graphemic units of English (see Parkin, 1982); the difficulty may be with respect to regularity (Bauer & Stanovich, 1980; Glushko, 1979) or with respect to letters which comprise a unit (Venezky, 1970).

The preceding discussion of the situation in English is intended to highlight the fact that hard evidence for a phonologically based lexical decision process is difficult to come by with English. As we will attempt to show, such evidence is easy to come by with Serbo-Croatian.

Roughly, the basic experimental procedure has been to compare the lexical decision time to a letter string that is written in a mix of <u>unique</u> and <u>common</u>

letters with the lexical decision time to a letter string written in a mix of ambiguous and common letters. A letter string of the former kind can be read in only one way and has a single morphophonological representation. In contrast, a letter string of the latter kind can be read in two ways because it is written in the letters shared by the two alphabets, some of which are phonemically equivocal; a letter string of this kind has two distinct morphophonological representations. If lexical decision proceeds with reference to the phonology, then a morphophonologically ambiguous letter string might be expected to extend decision time relative to a letter string that receives a unique morphophonological representation. This hypothesis has been evaluated in two ways: via a comparison of different letter strings (Lukatela, Popadić, Ognjenović, & Turvey, 1980; Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978) and via a comparison of different versions (Roman and Cyrillic) of the same letter string (Feldman, 1981; Feldman, Kostić, Lukatela, & Turvey, 1981).

Consider the experiment by Lukatela et al. (1980). The participants in this experiment (and the other experiments) were students from the University of Belgrade who were facile with both alphabets. They were presented with 144 letter strings, one half of which were words and one half of which were pseudowords. Of the word stimuli, 36 could be read in only one way and 36 could be read in two ways.[4] Of the pseudowords, 54 were associated with a single reading and 18 with a double reading. The task of a participant in the experiment was simply to identify, by a key press, whether or not a letter string, be it Cyrillic or Roman, represented a word in the Serbo-Croatian language, and to do so as quickly as possible. The results were straightforward: Lexical decision times were significantly slower for letter strings that were phonologically ambiguous and the decision time difference, between phonologically ambiguous and phonologically univocal letter strings, was more pronounced for words than for pseudowords. Phonological ambiguity is more detrimental to words than to pseudowords.

When different words are compared in a lexical decision experiment for the purpose of evaluating phonological factors, problems arise of matching the words on frequency of occurrence in the language, richness of meaning, length, number of syllables, etc. These problems can be virtually eliminated by taking advantage of the fact that some words can be transcribed in the Roman and Cyrillic alphabets such that in one alphabet the reading is phonologically ambiguous whereas in the other alphabet the reading is phonologically unique. To evaluate the phonological contribution to lexical access, the bi-alphabetical nature of Serbo-Croatian permits a comparison of a written word with itself. Table 2 gives several examples of words and pseudowords that are phonologically ambiguous or not depending on the alphabet in which they are transcribed.

In an experiment by Feldman (1981), bi-alphabetical readers made rapid lexical decisions about words and pseudowords including tokens of the types shown in Table 2. Consider the Serbo-Croatian word meaning savanna. This word is phonologically ambiguous when transcribed in Cyrillic (CABAHA) and phonologically unequivocal when transcribed in Roman (SAVANA). A number of words and pseudowords exhibiting the contrast exemplified by CABAHA and SAVANA were among the items presented to the subjects. The principal expectation was that decisions on letter strings like CABAHA would be significantly slower

Table 2

Types of Letter Strings and Their Lexical Status

Composition of Letter String	Phonemic Interpretation	Meaning
AMBIGUOUS and COMMON		
CABAHA*	Cyrillic /savana/	savanna
	Roman /tsabaxa/	nonsense
KOBAC	Cyrillic /kovas/	nonsense
	Roman /kobats/	hawk
KACA	Cyrillic /kasa/	safe
	Roman /katsa/	pot
HEPETAC*	Cyrillic /neretas/	nonsense
	Roman /xepetats/	nonsense
COMMON		
JAJE	Cyrillic /jaje/	egg
	Roman /jaje/	egg
TAKA	Cyrillic /taka/	nonsense
	Roman /taka/	nonsense
UNIQUE and COMMON		
SAVANA*	Cyrillic impossible	
	Roman /savana/	savanna
NERETAS*	Cyrillic impossible	
	Roman /neretas/	nonsense
КОБАЦ	Cyrillic /kobats/	hawk
	Roman impossible	
ПУДАЛ	Cyrillic /pudal/	nonsense
	Roman impossible	

(*indicates those letter string types included in the present experiment) than decisions on letter strings like SAVANA. Underscoring again the fact that the letter strings exemplified by CABAHA and SAVANA are the same word and, therefore, are identical in all respects but one, viz., the number of morphophonological representations, it is a noteworthy empirical observation that their associated decision times differed by more than 300 msec. (Similar magnitudes of difference were observed by Feldman et al., 1981.)

Clearly, with native Serbo-Croatian readers and written Serbo-Croatian material, lexical decision is intimately connected with the phonological level of the language. It is sometimes said that for native English readers and written English material, phonological access is an option that is taken or not depending on the conditions of the lexical decision task (Davelaar et al., 1978) and, further, that the more general preference of English readers is for a faster, visual strategy. In sharp contrast, referencing the phonology appears to be mandatory and not optional for the Serbo-Croatian reader. And if there is, in addition, a visual strategy at the disposal of the Serbo-Croatian reader, it is neither preferred nor faster. The impact of these results lies with the observation that phonological ambiguity retards lexical decision even when experimental conditions and instructions discourage the participant from making reference to the phonology. In one experiment (Lukatela et al., 1978) both the design of the experiments and the instructions to the subject attempted to constrain the reader to a Roman reading. Nevertheless, subjects were not able to eliminate the Cyrillic interpretation. With regard to a potentially preferred visual strategy that takes advantage of familiar visual form it should be noted that there is evidence that mixed alphabet letter strings (that do not include phonologically ambiguous characters) do not yield consistently slower lexical decision times than letter strings appearing in their natural visual format (Katz & Feldman, 1981). Also, the naming of mixed alphabet letter strings (with no ambiguous characters) is not slowed in comparison to naming the same letter strings in their strictly Roman transcription (Feldman & Kostic, 1981).

It remains for us to make a few remarks highlighting the analytic nature of the processes underlying lexical decision in Serbo-Croatian. Feldman (1981) and Feldman and Turvey (1983) showed that, with the number of syllables containing ambiguous characters held constant, the greater the number of ambiguous characters in a letter string the slower the lexical decision time. Further, Feldman (1981) observed that with the number of ambiguous characters controlled, clustering two ambiguous characters within one syllable retarded lexical decision more than having the two ambiguous characters appearing in different syllables. Most evidently, in the process of deciding on the lexical status of a letter string the native reader of Serbo-Croatian pays close heed to its internal phonologic structure.

To conclude, the Serbo-Croatian orthography is phonologically very regular (permitting a valid prediction of how a word is spoken solely on the basis of the letters comprising the word) and as such encourages neither the development of options for accessing the lexicon nor, relatedly, a sensitivity to the linguistic situations in which one option fares better than another. In this important respect it is very different from the phonologically deep English orthography that encourages (and, perhaps, demands) flexibility. For the beginning reader and for the fluent reader of Serbo-Croatian there are few enticements to try any strategy other than one that is phonologically analytic. Such a strategy is efficient, economical, and most befitting the Serbo-Croatian orthography.

REFERENCES

Bauer, D., & Stanovich, K. Lexical access and the spelling-to-sound regularity effect. Memory & Cognition, 1980, 8, 424-432.

Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E. Phonological encoding in the lexical decision task. Quarterly Journal of Experimental Psychology, 1979, 31, 489-502.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. Access to the internal lexicon. In S. Dornic (Ed.), Attention and performance VI. London: Academic Press, 1977.

Davelaar, E., Coltheart, M., Besner, D., & Jonasson, J. Phonological recoding and lexical access. Memory & Cognition, 1978, 6, 391-402.

Feldman, L. B. Word recognition in Serbo-Croatian is primarily phonological. Haskins Laboratories Status Report on Speech Research, 1981, SR-66, 167-201.

Feldman, L. B., & Kostić, A. Word recognition with mixed-alphabet forms. Haskins Laboratories Status Report on Speech Research, 1981, SR-66, 203-211.

Feldman, L. B., Kostić, A., Lukatela, G., & Turvey, M. T. An evaluation of the "Basic Orthographic Syllabic Structure" in a phonologically shallow orthography. Haskins Laboratories Status Report on Speech Research, 1981, SR-66, 273-291.

Feldman, L. B., & Turvey, M. T. Word recognition in Serbo-Croatian is phonologically analytic. Journal of Experimental Psychology: Human Perception and Performance, 1983, 9, 288-298.

Glushko, R. J. The organization and activation of orthographic knowledge in reading aloud. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 674-691.

Katz, L., & Feldman, L. B. Linguistic coding in word recognition: Comparisons between a deep and a shallow orthography. In A. Lesgold & C. A. Perfetti (Eds.), Interactive processes in reading. Hillsdale, N.J.: Erlbaum, 1981.

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

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. Lexical decision in a phonologically shallow orthography. Memory & Cognition, 1980, 8, 124-132.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. Bi-alphabetical lexical decision. Language and Speech, 1978, 21, 142-165.

Martin, R. C. The pseudohomophone effect: The role of visual similarity in non-word decision. Quarterly Journal of Experimental Psychology, 1982, 34A, 395-409.

McQuade, D. V. Variable reliance on phonological information. Language and Speech, 1980, 24, 99-100.

Parkin, A. Phonological recoding in lexical decision: Effects of spelling-to-sound regularity depend on how regularity is defined. Memory & Cognition, 1982, 10, 43-53.

Venezky, R. L. The structure of English orthography. The Hague: Mouton, 1970.

FOOTNOTES

[1]The + designates the boundary between base morpheme and inflectional affix. The h→s alternation is representative of a class of lawful variations.

[2]There is some ambiguity about the term "phonology" according to whether one assumes a descriptive linguistic or a Chomskyan perspective. By the former, "phonological" usually means classical phonemic as distinct from morphophonemic. By the latter, "phonological" refers to systematic phonemic and thus, is closer to morphophonological in the terminology of descriptive linguistics. Our meaning of "with reference to phonology" can be interpreted as lexical access, mediated by a phonetic/surface phonemic reading.

[3]As a consequence of its inflectional morphology, the skilled reader of Serbo-Croatian is also analytic at the level of constituent morphemes. We see phonological analysis and morphological analysis as two aspects of the same skill in that they focus on the internal structure of the word.

[4]Of the phonologically ambiguous words, one third were different words by their Roman and Cyrillic alphabet readings, e.g., KACA. One third were words by their Roman reading and nonsense by their Cyrillic reading, e.g., KOBAC. Finally, one third were words by their Cyrillic reading and nonsense by their Roman reading, e.g., CABAHA. (The examples come from Table 2 and do not necessarily represent words that were actually presented in this experiment.) Results for the three kinds of ambiguous words were not significantly different.

GRAMMATICAL PRIMING EFFECTS BETWEEN PRONOUNS AND INFLECTED VERB FORMS

G. Lukatela,+ Jelena Moraca,+ D. Stojnov,+ M. D. Savić,+ L. Katz,++ and M. T. Turvey++

Abstract. It is well known that deciding on the lexical status of a word can be facilitated by a preceding, semantically related word. Three experiments are reported demonstrating a different kind of facilitation due to the grammatical relation between function words and content words in Serbo-Croatian. A pronoun facilitated or inhibited the lexical decision made to a following verb depending on whether the person of the verb, as represented by its inflected ending, agreed with the person of the pronoun. Also, verbs primed subsequent pronouns, but the pattern of results for priming of pronouns by verbs was markedly different from that for priming of verbs by pronouns. The results suggest that the organization of the internal lexicon is sensitive to grammatical as well as semantical relations between words.

The facilitation of the perception of one word by the perception of another has been the subject of much recent experimental inquiry. Facilitation effects have been demonstrated largely, but not exclusively, in the context of word lists and primarily, but not exclusively, with words that are either associatively or semantically related. Almost without exception, however, these effects have been demonstrated in the lexical decision task where the subject is asked to decide, as rapidly as possible, whether or not a given letter string is a word. Thus, the standard demonstration of facilitation effects is of the following form: Given two words, simultaneously or successively, the lexical decision latency for the pair (are they both words?) or just to the second of the two can be shown to depend on the semantic relation that exists between them (e.g., Fischler, 1977; Meyer, Schvanaveldt, & Ruddy, 1975; Neely, 1977).

Recently, evidence was provided of a different facilitation effect, one that would appear to deserve the epithet "grammatical" rather than "semantic" (Lukatela, Kostić, Feldman, & Turvey, 1983) because the formal relation between prime and target words depends on the target's grammatical inflection. Inflection is the major grammatical device of Serbo-Croatian, Yugoslavia's principal language. Nouns are declined with the individual grammatical cases formed by adding a suffix to a (quasi) root morpheme. In normal linguistic

+University of Belgrade.
++Also University of Connecticut.
Acknowledgment. This research was supported by NIH Grants HD-08495 to the University of Belgrade, and HD-01994 to Haskins Laboratories. We gratefully acknowledge the help of M. Gurjanov.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

usage, a noun is often preceded by a governing preposition that requires the noun to be in a particular grammatical case (or, for some prepositions, one of two grammatical cases). This redundancy makes clear the noun's function in the sentence. The lexical decision task was adapted to the question of whether the processing of an inflected noun is facilitated by the prior presentation of a grammatically consistent preposition. The answer was positive: Lexical decision times to nouns were faster when the preceding preposition was appropriate to the case of the noun than when it was either inappropriate or simply a nonsense syllable. The present paper pursues a further potential instance of grammatical facilitation, one that is defined over the relation of pronoun to verb. The person of a Serbo-Croatian verb is specified by the suffix of the verb and by a preceding or following pronoun (or noun) that is the subject of the verb. Insofar as a given pronoun and a given inflected form of the verb co-occur consistently in normal linguistic usage, the perception of the one may facilitate the perception of the other. In particular, a prior pronoun might facilitate lexical decision on a subsequent verb with which it is grammatically consistent, and vice versa.

The types of facilitation under consideration here--that of noun by preposition and of verb by pronoun--may not be open to the kind of interpretation applied to the more familiar instances of facilitation between semantically similar items. The notion of an automatic spread of activation, originally described by Quillian (1969) and elaborated recently (for example, Anderson, 1976; Collins & Loftus, 1975; Neely, 1977; Posner & Snyder, 1975), refers ultimately to a specific linkage between particular representations of particular words. The idea that there is a specific linkage between (certain) internal word-representations, so that the direct stimulation of one representation mechanically leads to the (indirect) stimulation of others, identifies a medium for the automatic accessing of word meaning in long-term memory. Such automaticity is useful--it prunes degrees of freedom in the search process. Thus, glass leads mechanically and eventually to ice, cave to mine, nurse to wife, and so on (from the appendix of Fischler, 1977).

There is, therefore, a certain intuitive appeal to the notion of automatic spreading activation. However, the relation of preposition to inflected noun in Serbo-Croatian cannot be sensibly portrayed as a linkage between particular internal word-representations. English is sufficient to make this point: What could possibly motivate or rationalize specific linkages between the lexical representations of in and wall, from and chalk, below and jogger? A potentially more sensible portrayal follows from the suggestion that morphemes rather than words are specifically linked. Thus, spreading activation might be defined over connections between the small set of Serbo-Croatian prepositions and the small set of inflected endings of Serbo-Croatian nouns. The prepositional priming of lexical decision on an inflected noun could then be said to rest on the partial activation of the noun, namely, of its inflected ending (compare with Stanners, Neiser, Hernon, & Hall, 1979). Against this interpretation, however, is (i) evidence that the inflected Serbo-Croatian nouns are represented in the internal lexicon as singular units rather than as morphological concatenates (Lukatela, Gligorijević, Kostić, & Turvey, 1980); (ii) evidence that priming or facilitation does not occur between two semantically unrelated nouns that are in the same grammatical case (Lukatela & Popadić, Note 1); and (iii) the argument that the evidence for morphological decomposition reported for English materials (e.g., Stanners, Neiser, & Painton, 1979; Taft & Forster, 1975) may be an artifact of overrepresenting multimorphemic stimuli in the experimental design (Rubin, Becker, & Freeman, 1979).

We have belabored the problem of applying an interpretation of semantic facilitation to grammatical facilitation in order to underscore that an explanation that addresses relations among some word types may not address relations among all word types. For example, how relations are effected among words of the open class (e.g., adjectives, verbs, and nouns) may not be how relations are effected among words of the closed class (e.g., pronouns, prepositions, determiners, auxiliaries), nor how relations are effected across the two classes--such as the facilitation of an inflected noun by a grammatically consistent preposition. The distinction of open and closed classes is not just a formal distinction--readers of English relate to the two vocabulary types in qualitatively different ways suggesting, among other things, largely distinct recognition procedures (Bradley, 1978; Friederici & Schoenle, 1980; Garrett, 1978; Zurif, 1980). This division of the lexicon into two categories not only militates against a single account of facilitation effects, but also argues, more generally and most obviously, against a unitary view of the lexicon; on a pluralistic view, words would be expected to differ widely in the manner of their lexical organization and the means by which they are accessed. For example, it seems unlikely that, within the open class, nouns and verbs should be organized and retrieved along identical lines. The characterization of nouns as clusters of correlated attributes in a hierarchical organization contrasts with the characterization of verbs as clusters of uncorrelated attributes in a matrix-like organization (Huttenlocher & Lui, 1979; Kintsch, 1972; Miller & Johnson-Laird, 1976). With regard to the inflected nouns of Serbo-Croatian, it appears that the grammatical cases of any given noun comprise a system of words with the more frequent nominative singular form as the nucleus around which the oblique case forms cluster uniformly (Lukatela et al., 1980). Preliminary work on how the various forms of inflected Serbo-Croatian verbs relate among themselves suggests, however, no prominent member in the verb system that is comparable to the nominative singular in a noun system even though there are large differences among the verb forms in their individual frequencies of usage (Mandić & Ognjenović, Note 2).

The upshot of the foregoing is that semantic facilitation and grammatical facilitation are probably best understood not as expressions of a single mechanism, but rather as expressions of different mechanisms that stand in a complementary relation; it should not be surprising to find different species of facilitation if, as can be supposed, the organization of the lexicon is pluralistic rather than unitary.

EXPERIMENT 1

In Serbo-Croatian the inflectional forms of the verb identify voice (active or passive), mood, tense, number, and person; a pronoun subject agrees--in normal usage--with the inflectional form in number and person. When a pronoun occurs, it most often precedes the inflected verb form; sometimes the verb precedes the pronoun. The first experiment examined the effect of a preceding appropriate, inappropriate, or nonsense pronoun on a subsequent lexical decision made to a Serbo-Croatian verb. Two inflectional forms were used: the first person singular present and second person singular present. Our expectation was that when the pronoun agreed with the inflected verb form, lexical decision time for the verb would be shorter than when the pronoun did not agree with the inflected form, or when the 'pronoun' was, in fact, a nonsense syllable.

Method

Subjects. Sixty-four students from the Department of Psychology, University of Belgrade, received academic credit for participation in the experiment. A subject was assigned to one of four subgroups, for a total of sixteen subjects per subgroup.

Materials. Letter strings, each consisting of five or six upper-case letters, were typed and used to prepare black-on-white slides.

Two kinds of slides were constructed. In one kind, the letter string was arranged horizontally in the upper half of a 35 mm slide and, in the other, the letter string was arranged horizontally in the lower half of a 35 mm slide. Letter strings in the first type of slide were always pronouns (or their pseudoword analogues) and letter strings in the second type of slide were always inflected verbs (or pseudoword analogues). Altogether, there were 640 slides; 320 "pronoun" slides and 320 "verb" slides with each set evenly divided into 160 words and 160 pseudowords. The 160 verb slides that were real words consisted of two sets of 80, representing the same 80 verbs in the first person singular present tense, and in the second person singular present tense. These 80 verbs were selected from the middle frequency range of a corpus of one million Serbo-Croatian words (Kostić, Note 3). A different set of 80 verbs of the same frequency and in the same person and the same tense was used to generate the pseudowords. This was done by simply changing one letter in the root morpheme of the verb, leaving the inflected ending unchanged. The replacement was an orthotactically and phonotactically legal letter. Then, a second set of 80 pseudowords was created where the words differed from those in the first set in their inflections for person, that is, first person became second person, and vice versa.

As an illustration of how the verb and pseudoverb slides were prepared, consider a typical mini-list of Serbo-Croatian verbs presented in Table 1. All these verbs are from the mid-frequency range and display the three possible endings in the first person (-IM, -AM, -EM) and in the second person (-IS, -AS, -ES) of the present tense. From the list of 160 verbs exemplified by the mini-list in Table 1, one half were used to produce the verb slides. The other half were transformed into pseudoverbs by changing the initial or the second consonant. In this manner, the letter strings in Table 2 were obtained from the mini-list of Table 1 although, as stated, a unique set of real verbs was actively used to generate the pseudowords. To reiterate, in deriving a pseudoverb from a verb, the final syllable was never changed, and the final syllables (-IM, -AM, -EM, -IS, -AS, -ES) were balanced across all verbs and pseudoverbs.

Table 1

Examples of Serbo-Croatian Verbs[1]

Infinitive form	First person present tense	Second person present tense
RADI-TI (to work)	RADI-M	RADI-Š
ČITA-TI (to read)	ČITA-M	ČITA-Š
PISA-TI (to write)	PISE-M	PISE-Š
PUŠI-TI (to smoke)	PUŠI-M	PUŠI-Š
PEVA-TI (to sing)	PEVA-M	PEVA-Š
PI-TI (to drink)	PIJE-M	PIJE-Š

[1]The hyphens have been added to emphasize the inflections.

The slides were grouped into pronoun-verb pairs such that (1) the inflected verb slides contained a word in one half of the pairs and a pseudoword in the other half, and (2) the pronoun slides contained the first-person singular pronoun JA, or the second person singular pronoun TI, or a monosyllabic pseudoword (a pseudopronoun). Six monosyllabic pseudowords--JO, VA, DA, TR, ZI, KI--were derived from the pronouns JA and TI by changing the initial or final letter. Forty monosyllabic pseudoword slides were prepared with the letter string JO, twenty slides with VA, twenty slides with DA, forty slides with TR (R can function as a vowel in the language), twenty slides with ZI, and twenty slides with KI.

Table 2

Pseudoverbs Derived from the Verbs in Table 1

Infinitive form	First person present tense	Second person present tense
KUSI-TI	KUSI-M	KUSI-Š
JEVA-TI	JEVA-M	JEVA-Š
DI-TI	DIJE-M	DIJE-Š

[1]The hyphens have been added to emphasize the inflections.

In total, there were 640 different pairs of slides of which a given subject saw 160 pairs. Forty other different pairs of slides were used for the preliminary training of subjects.

Design. As remarked, each verb and pseudoverb appeared in two persons. A constraint on the design of the experiment was that a given subject never saw a given verb or pseudoverb--in either inflected form--more than once. In one half of the 160 trials the second stimulus in a pair was a verb and in the other half the second stimulus was a pseudoverb. The set of 80 verbs that was presented to a subject consisted of 40 verbs in first person singular and 40 other verbs in second person singular. Similarly, the set of 80 pseudoverbs that was presented to a given subject consisted of 40 pseudoverbs in the first person singular and 40 other pseudoverbs in the second person singular.

The two groups of verbs and the two groups of pseudoverbs were each further divided into four subgroups of ten. Items in these four subgroups, two of verbs and two of pseudoverbs, were preceded by the nominative first person pronoun JA. Four other subgroups, two of verbs and two of pseudoverbs, were preceded by the nominative second person pronoun TI. With respect to the pseudopronouns, two groups of verbs and two groups of pseudoverbs were preceded by the pseudopronouns JO, VA, or DA. The other two groups of verbs and pseudoverbs were preceded by the pseudopronouns TR, ZI, or KI.

There were four groups of 16 subjects each. All received the same experimental manipulation and differed only with regard to the particular stimuli they were presented. Each subject in each group of 16 subjects saw each pronoun-verb, pseudopronoun-verb, pronoun-pseudoverb, and pseudopronoun-pseudoverb combination. Put differently, each subject saw the same verbs and pseudoverbs as every other subject, but not necessarily in the same person nor necessarily preceded by the same pronoun or pseudopronoun type.

Procedure. On each trial, two slides were presented. Each slide was exposed in one channel of a three-channel tachistoscope (Scientific Prototype, Model GB) illuminated at 10.3 cd/m2. The subject's task was to decide as rapidly as possible whether the letter string contained in a slide was a word. Both hands were used in responding to the stimuli. Both thumbs were placed on a telegraph key button close to the subject and both forefingers on another telegraph key button two inches further away. The closer button was depressed for a "No" response (the string of letters was not a word), and the further button was depressed for a "Yes" response (the string of letters was a word).

Latency was measured from the onset of a slide. The subject's response to the first slide terminated its presentation and initiated the second slide, unless the latency exceeded 1300 msec, in which case the second slide was initiated automatically. The presentation of the second slide, unlike that of the first, was fixed at 1300 msec.

Results

Analyses were performed only on those latencies to the second slide for which responses were correct and which were less than 1300 msec. Total error rate was 1.3 percent. Mean lexical decision reaction times for verb and pseudoverb trials are presented in Table 3.

An analysis of variance was performed on each subject's mean reaction times in each combination of prime lexicality (pronoun vs. pseudopronoun), target lexicality (verb vs. pseudoverb), and person (first vs. second). Because, for this and for subsequent analyses, results were essentially similar for both persons, the presentation and interpretation of the results have been simplified. When the person of the prime and target were the same, the combination has been labeled "appropriate"; when different, the combination has been labeled "inappropriate." Thus, for Table 3, data for both the first and second persons have been combined to give a mean for "appropriate" priming of real verbs of 652 msec. Similarly, the mean of the "inappropriate" cell, 780 msec, is a combination of data for two conditions: first person pronouns preceding second person verbs and second person pronouns preceding first person verbs.

Table 3

Experiment 1: Mean reaction time in milliseconds to verbs and pseudoverbs when primed by grammatically appropriate or inappropriate pronouns or by pseudopronouns.

	Target	
Prime	Verbs	Pseudoverbs
Appropriate pronoun	652	758
Inappropriate pronoun	780	731
Pseudopronoun	726	794

The analysis of word data showed that there were no significant differences between groups of subjects, $F(3,60) = .93$, $MSe = 34418$, $p > .50$. Also, the average latency of a verb preceded by a pronoun did not differ from the average latency of a verb preceded by a pseudopronoun, $F(1,60) = 2.91$, $MSe = 4026$, $p > .10$. However, the interaction of verb ending with pronoun person was significant, $F(1,60) = 118.91$, $MSe = 4086$, $p < .001$, accounting for the nonsignificant main effect of pronoun versus pseudopronoun. Further, inflected verb ending, pronoun person, and pronoun lexical status (real or pseudo) formed a three-way interaction: $F(1,60) = 137.79$, $MSe = 3993$, $p < .001$. This is to say that latencies to inflected verb forms varied as a function of whether (i) the prime was a pronoun or a pseudopronoun; and (ii) the pronoun was appropriate or inappropriate. Inspection of Table 3 reveals that the decision time for verbs was shorter when the pronoun was grammatically appropriate.

The analysis of variance on pseudoverb data showed no main effect due to subject group, $F(3,60) = .44$, $MSe = 47985$, $p > .50$. However, there was a significant main effect of the pronoun's lexical status, $F(1,60) = 54.48$, $MSe = 5267$, $p < .001$, such that pronouns (relative to pseudopronouns) reduced reaction times to pseudoverbs. There was a significant two-way interaction of verb ending with pronoun person, $F(1.60) = 13.42$, $MSe = 1168$, $p < .001$, which must be interpreted relative to a three-way interaction of verb ending, pronoun person, and pronoun vs. pseudopronoun, $F(1,60) = 21.14$, $MSe = 1061$, $p < .001$. This suggests that it was more difficult to reject pseudoverbs that were preceded by an appropriate pronoun than to reject the same inflected pseudoverbs preceded by an inappropriate pronoun. Finally, when a pseudoverb was preceded by a pseudopronoun, there were no significant differences among the inflected forms of the pseudoverb. In sum, pseudoverb rejection latencies were faster when the preceding item was a pronoun than a pseudopronoun but, for these faster latencies, an appropriate pronoun slowed pseudoverb rejection more than an inappropriate pronoun.

Discussion

Facilitation of lexical decision by a preceding item is generally said to occur either by means of a process that is automatic or by a process that is conscious and attentional (Neely, 1977; Posner & Snyder, 1975). As an example of the latter, lexical decision on inflected verbs that were preceded by a grammatically appropriate pronoun may have been facilitated by the subjects' consciously expecting to see the inflected ending specific to the pronoun before the verb was displayed. If such was the case--that the facilitation we observed was due entirely to the allocation of selective attention--then there would be little reason to believe that the observed facilitation is characteristic of the process of lexical access during natural discourse. It is well known that attentional priming is slow relative to automatic priming (e.g., Stanovich & West, 1981) and it is unlikely that attentional priming could play a useful role in the lexical access of verbs, given the normally close temporal contiguity between pronoun and verb.

First consider the pseudoverb results, which are consistent with the notion of automatic processing. To begin with, there was no general inhibition effect. Compared to pseudopronouns, inappropriate as well as appropriate pronouns expedited negative decisions on pseudoverbs. The overall reduction in rejection latencies induced by a preceding pronoun suggests that pronouns and verbs may stand in a special relation. One speculation is that pronouns trigger a verb processing mechanism that operates on the morphological structure of verbs. The pseudoverb data are consistent with the notion that verb processing begins with a decomposition of the verb into stem and suffix and that a preceding pronoun primes the mechanism that performs this morphological parsing.

Assuming, therefore, that a pronoun quickened the decomposition of a following verb, argument can be given that this effect occurred automatically. Consider the contrary possibility, that the effect was due to an attentional mechanism. If the pseudopronoun-pseudoverb sequence is regarded as an instance of neutral priming, then the pronoun-pseudoverb sequence can be regarded as an instance of negative priming, misleading the subject to consciously expect a verb. Because of a pronoun, an attentional expectation of a verb is formed directing processing capacity to the verb region of the lexicon and reducing the processing capacity for the pseudoverb that follows.

If the latter were the case, then pseudoverb decision times should have been slowed by a pronoun relative to the pseudoverb decision times associated with a pseudopronoun. The fact that the opposite outcome was observed suggests that the grammatical relation between pronoun and verb facilitated rejection of the pseudoverb automatically rather than attentionally.

A further observation on pseudoverbs suggests the involvement of post-lexical processes. Reaction time to a pseudoverb preceded by a pronoun appropriate to its inflected ending was slower than reaction time to a pseudoverb preceded by a pronoun inappropriate to its inflected ending (see Table 3). The congruency between a morpheme currently being processed (the inflected ending of the pseudoverb) and a recently processed pronoun may retard the decision to reject the rest of the target item--the pseudoverb stem--as nonsense.

In contrast to the pseudoverb data, the verb data are not consistent with the notion of automatic processing. The latencies to verbs preceded by inappropriate pronouns were slower than the latencies to verbs preceded by pseudopronouns. This fact is easy to understand in terms of attentional facilitation and difficult to understand in terms of automatic facilitation. Selective attention (but not automatic priming) uses conscious processing capacity and when it is directed to the wrong target (for example, by an inappropriate pronoun), the subject has fewer resources to use in processing the actual target that is displayed.

Attentive rather than automatic processing is said to dominate at longer temporal separations between the priming stimulus and the target stimulus. With short temporal separations, inhibition effects are negligible, becoming increasingly more substantial as the separation is lengthened (Neely, 1977). If the effects of pronouns on verbs are mediated by attentive processing, then the latency of accepting as a word a verb that follows an inappropriate pronoun should be greater when the verb is separated from the pronoun by a long interval than when the separation interval is short. This hypothesis is evaluated in the second experiment, which, in addition, seeks to replicate the pattern of results obtained in the first experiment.

EXPERIMENT 2

The design of Experiment 2 permitted a systematic examination of the automaticity hypothesis by studying the effect of the length of time permitted for pronoun processing before the appearance of the verb. Two stimulus onset asynchronies were used, 300 msec and 800 msec. These intervals bracket the average intervals subjects produced themselves in Experiment 1. In contrast to the first experiment, subjects in Experiment 2 were required to make a lexical decision only to the second stimulus (the verb or pseudoverb target). In further contrast, the first stimuli in the second experiment were always pronouns; there were no pseudopronouns. In all other respects the design and the stimuli were the same as Experiment 1. Verb and pseudoverb targets were preceded by pronouns that were either appropriately or inappropriately matched to the targets' inflectional suffixes.

Method

Subjects. Eighty students from the Department of Psychology, University of Belgrade, received academic credit for participation in the experiment. None of the subjects previously took part in Experiment 1.

Materials. The stimuli were the same as in Experiment 1 with the exception of the pseudopronoun stimuli, which were not used. In total there were 160 different pronoun-verb pairs and 160 pronoun-pseudoverb pairs.

Design. A subject was assigned to one of eight groups, with ten subjects per group. Each subject saw 80 pairs of stimuli. The first stimulus in each pair was a pronoun. In half of the 80 trials, the second stimulus in a pair was a verb and in the other half, the second stimulus was a pseudoverb. Each subject in each odd-numbered group of 10 subjects (i.e., in Groups 1, 3, 5, 7) saw 40 different stimulus pairs in the pronoun-verb combination and 40 other different stimulus pairs in the pronoun-pseudoverb combination. Within each combination, the pronoun, verb, or pseudoverb appeared equally often in the first and the second person. The onset-onset interval between prime and target in these groups was 300 msec. Similarly, each subject in each even-numbered group of 10 subjects (i.e., in Groups 2, 4, 6, and 8) saw the same stimuli pairs as his/her counterpart in the odd-numbered groups. The onset-onset interval for these groups was 800 msec.

Procedure. The procedure was similar to that in Experiment 1 except that the subject gave a response only to the second stimulus in each trial. The first stimulus in each trial was always presented for 300 msec; the second stimulus was presented with no delay (for half the subjects) or with delay of 500 msec (to the other half).

Latency was measured from the onset of the second slide. Display of the second slide was terminated by a key press.

Results and Discussion

An analysis of variance was performed on each subject's mean reaction time computed on all correct responses out of the ten trials in each experimental situation. All latencies shorter than 300 msec and longer than 1300 msec were considered as errors. The total error rate was 1.7%.

Table 4 presents the mean reaction time data for verb targets primed by appropriate or inappropriate pronouns at stimulus onset asynchronies of 300 msec or 800 msec. Inspection of the results for real verbs suggests that appropriate pronouns facilitated verb recognition relative to inappropriate pronouns. There is also the suggestion that the relative priming facilitation increased as the interval between prime and target onsets increased. Inspection of the pseudoverb results suggests that the four pseudoverb conditions that were preceded by pronouns did not differ.

Table 4

Experiment 2: Reaction time in milliseconds to verbs and pseudoverbs when primed by appropriate or inappropriate pronouns at 300 or 800 millisecond stimulus onset asynchronies.

	Target			
	Verbs		Pseudoverbs	
SOA	300 msec	800 msec	300 msec	800 msec
Prime				
Appropriate pronoun	666	643	731	722
Inappropriate pronoun	729	739	717	714

Analyses supported these suggestions. First, an analysis of variance was performed on the average verb and pseudoverb latencies in each experimental condition for each subject. There were several interactions that reflected effects due to counterbalancing the assignment of specific verbs and pseudoverbs to the various conditions. For example, the five-way interaction for counterbalanced subject groups with stimulus onset asynchrony, verb/pseudoverb, first person/second person pronoun, and appropriate/inappropriate suffix was significant, $\underline{F}(3,72) = 3.39$, MSe = 1259.9, $\underline{p} < .03$. Inspection of this and other interactions involving groups indicated that the trends in the data were similar for all groups; the ordinal relationships in the data discussed below were true for all groups although the sizes of the differences changed.

The interaction of verb/pseudoverb by appropriate/inappropriate inflection by stimulus onset asynchrony was significant, $\underline{F}(1,72) = 6.01$, MSe = 1777.4, $\underline{p} < .02$. This three-way interaction was studied further by performing two analyses of variance, separately, on verbs and pseudoverbs. As Table 4 suggests, the two-way interaction between appropriate/inappropriate inflection and stimulus onset asynchrony was significant, $\underline{F}(1,72) = 10.45$, MSe = 1915.1, $p < .002$. Inspection of the table shows that the large difference between appropriately and inappropriately primed verbs at the short 300 msec asynchrony (666 and 729 msec, respectively) is somewhat larger at the 800 msec asynchrony (643 and 739 msec, respectively). Thus, the increasing onset asynchrony between prime and target was effective in increasing the differential between appropriate and inappropriate primes. It is clear that there is a strong main effect for appropriateness over and above its interaction with onset asynchrony; the latency difference between verbs with inflected endings appropriate to the pronoun and verbs with inflected endings inappropriate to the pronoun was highly significant, $\underline{F}(1,72) = 262.6$, MSe = 1915.1, $\underline{p} < .001$. This main effect of appropriateness was the most striking result of the verb analysis, confirming the large effect that was found in Experiment 1. There were also reliable effects due to the person of the pronoun (not shown in Table 4); verb reaction times were faster following a first person pronoun prime than a second person pronoun prime, $F(1,72) = 1601$, MSe = 949.8, $p < .001$.

A different picture emerged from the analysis of pseudoverbs. There, the two-way interaction between appropriate/inappropriate inflection and onset asynchrony was not significant and, in fact, its mean square was small, $F(1,72) = .76$, MSe = 507.7. However, the main effect of appropriateness, although small, was very reliable, $F(1,72) = 16.1$, MSe = 655.9, $p < .001$. As Table 4 indicates, the pseudoverbs with inflected endings that were appropriate to the preceding pronoun were rejected as words more slowly than inappropriate pseudoverbs. Finally, although not indicated in Table 4, the person of the preceding pronoun was again significant. The first person pronoun facilitated subsequent lexical decisions more than the second, $F(1,72) = 15.3$, MSe = 1017.6, $p < .001$.

Thus, the pattern that was observed in Experiment 1 was replicated under the conditions of Experiment 2. Verb lexical decision was faster and pseudoverb lexical decision was slower in the presence of a grammatically appropriate pronoun relative to an inappropriate pronoun. Additional results from the present experiment suggested that the relative facilitation of verbs and inhibition of pseudoverbs was largely completed within the 300 msec onset asynchrony; only small increases occurred when the pronoun was displayed for 800 msec before the verb came on.

Although the significant interaction between appropriateness and temporal separation for the verbs is in accordance with the attentional hypothesis, the fact that the effect of appropriateness was largely established by the 300 msec interval implies that the pronominal influence is principally automatic and not attentional. And, as in Experiment 1, the data for pseudoverbs lend no support to an attentional source of the priming effect. When the latter result is considered together with the grammatical influence on verbs at a 300 msec separation of pronoun and verb, an automatic view of the pronominal influence on verbs emerges as the most parsimonious.

EXPERIMENT 3

Verbs and pronouns are open and closed word classes, respectively. There is evidence, as noted in the Introduction, that words of an open class and words of a closed class may not be processed in the same manner. It might also be the case that the effects on the processing of items of one class induced by items of the other class are not symmetrical. In particular, pronominal influences on verbs may not be identical to verbal influences on pronouns. A third experiment was conducted that was similar to the first experiment in all respects except for a reversal of the order of stimuli within each pair--the prime was a verb (or pseudoverb) and the target was a pronoun (or pseudopronoun).

Twenty-five students from the Department of Psychology, University of Belgrade, participated in the experiment. None of them had participated in the first or second experiments.

Results and Discussion

Mean decision times for the pronoun and pseudopronoun targets are presented in Table 5. Mean acceptance latency for pronouns was faster when preceded by grammatically appropriate verbs than by inappropriate verbs. Slowest were pronouns preceded by pseudoverbs. In contrast, mean rejection

Table 5

Experiment 3: Reaction time in milliseconds to pronouns and pseudopronouns when primed by appropriate and inappropriate verbs.

	Target	
Prime	Pronouns	Pseudopronouns
Appropriate verb	550	645
Inappropriate verb	575	645
Pseudoverb	613	656

latencies for pseudopronouns were approximately equal whether preceded by appropriate verbs, inappropriate verbs, or pseudoverbs. With regard to the verb and pseudoverb targets that appeared as first stimuli in each trial, the average acceptance latencies for verbs in first and second person in the present tense were 735 msec, and 752 msec, respectively, whereas the mean rejection latencies for pseudoverbs in first and second person were 771 msec, and 774 msec, respectively. The total error rate (wrong responses and slow responses) on first and second stimuli was 1.8% and 2.0%, respectively.

The suggestions that the decision time to a pronoun was shorter when the pronoun was preceded by a verb as opposed to a pseudoverb and that the latency to an appropriately primed pronoun was shorter than to an inappropriately primed pronoun were substantiated by the statistical analyses. An analysis of variance revealed that the legality of the prime (verb vs. pseudoverb) was significant, $F(1,24) = 48.33$, $MSe = 1925$, $p < .001$. Grammatical person of the pronoun target (first vs. second) was not significant, but a three-way interaction among legality of prime (verb or pseudoverb), inflected ending of prime (appropriate or inappropriate), and the person of the pronoun was significant, $F(1,24) = 5.54$, $MSe = 634$, $p < .05$. This significant interaction means that grammatical consistency between the inflected ending of the preceding verb or pseudoverb and the pronoun was an important factor only when the preceding item was a verb. With regard to pseudopronouns, inspection of Table 5 suggests that in all combinations, the rejection latencies were about the same, a suggestion that was supported by the analysis of variance.

The average acceptance latency for a pronoun was shorter when it was preceded by a verb than when it was preceded by a pseudoverb. Importantly, this reduction occurred whether or not the ending of a priming verb was grammatically appropriate to the person of the pronoun. Clearly, the obtained data cannot be explained in terms of priming the pronoun by the verb ending, since all the pseudoverbs that were used in this experiment had the same endings as the verbs (m, s) yet the lexical decision on pronouns was indifferent to the pseudoverbs that preceded them. The acceptance latencies to pronouns in the grammatical and non-grammatical pseudoverb-pronoun combinations were virtually identical.

A closer examination of verb-pronoun combinations reveals that the average decision latency for pronouns was statistically faster when the verb ending was appropriate to the pronoun than when it was not appropriate. This observation suggests that an appropriate inflected ending was able to enhance lexical decision on a pronoun over and above the enhancement produced by a preceding verb. Importantly, a differential effect of the appropriateness of the inflected ending to the pronoun was not found with pseudoverbs.

An interpretation of these data is that a verb preceding a pronoun primes the (small) set of pronouns, a pseudoverb does not. In addition, the verb primes the particular member in the pronoun set that is congruent with the verb's inflected ending. This priming would appear to be automatic. Inhibition effects were absent and the presence of a verb significantly affected the latencies for accepting pronouns as words even though throughout the experiment subjects could rely on the fact that only pronouns and pronoun analogues would appear as second stimuli.

In summary, the most noticeable commonality between the first two experiments and the third is that the shortest acceptance latency for a word target was in the condition in which the word pair was grammatical. In short, pronouns and verbs are mutually facilitating. The most noticeable difference between the first two experiments and the third is that the data of the third experiment display no inhibition effect (pronouns preceded by grammatically inappropriate verbs were responded to faster, not slower, than pronouns preceded by pseudoverbs) and exhibit no differentiation within the group of decision latencies on pseudopronouns. In short, verbs affect the pronouns they precede differently from the way that pronouns affect the verbs that follow them.

Taken together, the results of the three experiments suggest that pronouns can automatically facilitate verbs and that verbs can automatically facilitate pronouns, but that the mechanism of facilitation is not the same in the two cases.

REFERENCE NOTES

1. Lukatela, G., & Popadić, D. Priming nouns by nouns. Unpublished manuscript (in Serbo-Croatian). Laboratory of Experimental Psychology, University of Belgrade, 1979.
2. Mandić, G., & Ognjenović, P. Representation of verbs in the lexicon. Unpublished manuscript (in Serbo-Croatian). Laboratory of Experimental Psychology, University of Belgrade, 1980.

3. Kostić, Dj. Frequency of occurrence of words in Serbo-Croatian. Unpublished report. Institute of Experimental Phonetics and Speech Pathology, Belgrade, Yugoslavia, 1965.

REFERENCES

Anderson, J. Language, memory, and thought. Hillsdale, N.J.: Lawrence Erlbaum, 1976.

Bradley, D. Computational distinction of vocabulary type. Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1978.

Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. Psychological Review, 1975, 82, 407-428.

Fischler, I. Semantic facilitation without association in a lexical decision task. Memory & Cognition, 1977, 5, 335-339.

Friederici, A. D., & Schoenle, P. W. Computational dissociation of two vocabulary types: Evidence from aphasia. Neuropsychologia, 1980, 18, 11-20.

Garrett, M. F. Word and sentence perception. In R. Held, H. W. Liebowitz, & H. L. Teuber (Eds.), Handbook of sensory physiology, (Vol. 8): Perception. Berlin: Springer-Verlag, 1978.

Huttenlocher, J., & Lui, F. The semantic organization of some simple nouns and verbs. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 141-162.

Kintsch, W. Notes on the structure of semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Lukatela, G., Gligorijević, B., Kostić, A., & Turvey, M. T. Representation of inflected nouns in the internal lexicon. Memory & Cognition, 1980, 8, 415-423.

Lukatela, G., Kostić, A., Feldman, L. B., & Turvey, M. T. Grammatical priming of inflected nouns. Memory & Cognition, 1983, 11, 59-63.

Meyer, D. E., Schvanaveldt, R. W., & Ruddy, M. G. Loci of contextual effects on visual word-recognition. In P. M. A. Rabitt & S. Dornic (Eds.), Attention and performance V. New York: Academic Press, 1975.

Miller, G. A., & Johnson-Laird, P. N. Language and perception. Cambridge, Mass.: Harvard University Press, 1976.

Neely, J. H. Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. Journal of Experimental Psychology: General, 1977, 106, 226-254.

Posner, M. I., & Snyder, C. R. Facilitation and inhibition in the processing of signals. In P. M. A. Rabitt & S. Dornic (Eds.), Attention and performance V. New York: Academic Press, 1975.

Quillian, M. R. The teachable language comprehender. Communications of the ACM, 1969, 12 459-476.

Rubin, G. S., Becker, C. A., & Freeman, R. H. Morphological structure and its effect on visual word recognition. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 757-767.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. Memory representation for morphologically related words. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 399-412.

Stanners, R. F., Neiser, J. J., & Painton, S. Memory representation for prefixed words. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 733-743.

Stanovich, K. E., & West, R. F. The effect of sentence context on ongoing word recognition: Tests of a two-process theory. Journal of Experimental Psychology: Human Perception & Performance, 1981, 3, 658-672.

Taft, M., & Forster, K. I. Lexical storage and retrieval of prefixed words. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 638-647.

Zurif, E. B. Language mechanisms: A neuropsycholinguistic perspective. American Scientist, 1980, 68, 305-311.

MISREADINGS BY BEGINNING READERS OF SERBO-CROATIAN*

Vesna Ognjenović+, G. Lukatela,+ Laurie B. Feldman,++ and M. T. Turvey+++

Abstract. Errors in reading aloud by the beginning reader have been interpreted as reflecting the difficulty and the importance of phonemic segmentation for the acquisition of reading skills. Results from previous studies on English words patterned as consonant-vowel-consonants showed: 1) more errors on vowels than on consonants; 2) more errors on word final consonants than on word initial consonants; and suggested that 3) consonant errors were based on phonetic confusions while vowel errors were not. In contrast to their English counterparts, the beginning readers of Serbo-Croatian tested in the present study committed proportionally fewer errors on their reading of vowels than of consonants but in common with their English counterparts, their reading of final consonants was more vulnerable to error than their reading of initial consonants. This pattern of errors was found for both word and pseudoword consonant-vowel-consonant structures and the pattern of vowel confusions, like the pattern of consonant confusions, was rationalized by speech-related factors. The differences between the patterns of confusions for Serbo-Croatian and for English could be due to the difference between the two orthographies in the precision with which they represent the phonology or to the fact that the vowels of English are qualitatively less distinct phonologically than the vowels of Serbo-Croatian.

For any alphabetic orthography the highly encoded nature of phonemes in the spoken language bears significantly on the task of learning to read analytically--that is, learning to relate to letter strings in a way that efficiently exploits the specification of a letter string's pronunciation by its spelling. The significance of speech encodedness to reading has been extensively discussed by Gleitman and Rozin (1977) and it has shaped the orientation of the Haskins Laboratories group to the task that befalls the

---

*Also *Quarterly Journal of Experimental Psychology*, 1983, 35A, 1-13.
+University of Belgrade.
++Also Dartmouth College.
+++Also University of Connecticut.
Acknowledgment. Special thanks to Carol Fowler, Isabelle Liberman and Donald Shankweiler for reviewing several drafts of this manuscript. Thanks also to Manojlo Gurjancv and Milan Savić for their computer expertise. This research was supported in part by NICHD Grant HD 08495 to the University of Belgrade and in part by NICHD Grant HD 01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

beginning reader (Liberman, I. Y., Note 1; Liberman, I. Y., Shankweiler, Orlando, Harris, & Bell-Berti, 1971; Fowler, Liberman, & Shankweiler, 1977; Mattingly, 1972; Shankweiler & Liberman, 1972; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). To read analytically the child must explicitly realize that continuous speech is divisible into phonemes and that each word is decomposable into a specific number of phonemes ordered in a specific way. This explicit realization--"linguistic awareness" (Mattingly, 1972)--is made difficult, it is argued, by the fact that the phonemes are not represented in the speech stream as discrete, isolable entities but rather they are encoded into the structure of the syllable (Liberman, A. M., Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, A. M., Mattingly, & Turvey, 1972). In contrast to speech perception, reading entails a more deliberate appreciation of component structure. The word "bat" is comprised of three phonetic segments, yet acoustically there are no distinct segments. The child's putative difficulty, it should be emphasized, is not with differentiating minimally contrastive word pairs--such as bad and bat--but rather with appreciating that each word is decomposable into three segments, the first two of which are shared by the two words and the third of which distinguishes them (Liberman, I. Y., Shankweiler, Liberman, A. M., Fowler, & Fischer, 1977).

There is considerable evidence that young children have difficulty segmenting the spoken word (see Gleitman & Rozin, 1977; and Liberman & Shankweiler, 1979, for a review). It has been proposed that this difficulty is reflected in the pattern of errors a child produces in reading. Shankweiler and Liberman (1972) had third grade American children read aloud consonant-vowel-consonant letter strings, all of which were words. They observed that errors on the final consonants were far more numerous than errors on the initial consonants; in addition, they observed that errors on the medial vowels far exceeded those on consonants in both final and initial position. Similar error patterns had been noted in earlier reports (Daniels & Diack, 1956; Venezky, 1968; Wheeler, 1970). Shankweiler and Liberman (1972) proposed two interpretations. According to the first interpretation, the error pattern reflects the beginning reader's differential difficulty segmenting sounds occurring in the initial, medial, and final positions in the syllable. That is to say, the error difference between the initial consonants, medial vowels, and final consonants is attributed to the relative positions within the syllable occupied by the different types of sound and not to differences among the sound-types themselves. According to this first interpretation, the higher error rate for the medial vowels than for the initial and final consonants is because the individual vowel is spread throughout the syllable. Other speech-related arguments for the greater susceptibility of medial vowels to be read incorrectly can be cited. Generally, there is reason to suppose that the properties of vowels in speech as distinguished from the properties of consonants may have perceptual consequences (Liberman, A. M. et al., 1967; Liberman, A. M. et al., 1972). The categorical perception that marks the (stop) consonants is less obviously characteristic of vowel perception. In addition, the contribution of the consonants to the phonological message is not matched by the vowels. On the other hand, the vowels as the nuclei of syllables support prosodic characteristics and provide the major medium for individual and regional variations in the spoken language.

In sum, the higher error on vowels might be related to the embedded and context-sensitive status of vowels in speech. Let us refer to this as the "universal" interpretation, for it emphasizes aspects common to all languages. This universal interpretation can be contrasted with one that might be termed "particular," so-called because it emphasizes the particularities of the English orthography. On this second interpretation of Shankweiler and Liberman's (1972), the higher error rate for medial vowels might be due to the fact that many of the complexities of English spelling are concentrated on the vowels--there are many possible pronunciations for most of the vowel graphemes and each vowel phoneme can be transcribed by one of several graphemes (Dewey, 1970). (For example, /u/ is represented by a number of different letters or digraphs: u, o, oo, ew, etc.) Relevant to the particular interpretation of the magnitude of vowel errors is Shankweiler and Liberman's (1972) report that the error rate on the individual medial vowels was related to their orthographic complexity, that is, to the number of graphemes and digraphs by which they are represented in the orthography.

Evidence bearing on the foregoing interpretations of the differential rate of errors on medial vowels and initial and final consonants is to be found in a further study (Fowler et al., 1977) that was motivated in part by a concern for the difference between the consonant sets used for the syllable-initial and syllable-final consonants of the original experiment. This difference in consonant sets raised the question of whether the pattern of errors might in fact be due to the difference in the phonologic (or orthographic) properties of the consonants occupying the final and initial positions rather than to the positions themselves. With the consonant sets equated, the later experiment (Fowler et al., 1977) replicated the position-dependency of consonant errors. As before, final consonant errors exceeded the errors on initial consonants by a margin of 2:1. Moreover it was shown that many phonetic features of the presented consonant were shared by the nature of the incorrect consonant that was given in its place. With regard to vowels, however, Fowler et al. reported that whether they placed vowels in initial, medial, or final syllabic positions, errors did not vary systematically with positions in the word. Further, the substituted (incorrect) vowels were not phonetically related to target words. Finally, there is evidence (Fowler, Shankweiler, & Liberman, 1979) that learning to read entails a progressive appreciation of the different phonemic values that a vowel grapheme can assume and the orthographic contexts in which particular spelling-sound correspondences can apply.

It can be claimed, therefore, that the errors on vowels and consonants by beginning readers of English differ in nontrivial ways and mimic, in reading, an opposition between these phonemic categories that is universal in speech. It can also be claimed, however, that with respect to the vowels, the child's misreadings do not primarily reflect difficulties in phonological segmentation. The speech-related factors that account for consonant errors do not account for vowel errors. Fowler et al. (1979) and Liberman, I. Y. et al. (1977) suggest, therefore, that the vowel errors are probably due to the complexity and variability of the spelling-to-sound correspondences in English. In brief, they suggest the language-particular interpretation of vowel errors rather than the universal interpretation. In the experiment reported here (which replicates with Yugoslav readers the conditions of the Shankweiler and Liberman et al. experiments) it is also the particular interpretation that is favored although the dissociation in reading, of vowels and consonants, is not strictly upheld. For beginning readers of Serbo-Croatian vowel errors like consonant errors are owing largely and equally to speech-related factors.

THE SERBO-CROATIAN WRITING SYSTEM

The English and Serbo-Croatian languages differ in the depth of their alphabetic orthographies. As a consequence, the simple letter-sound correspondences of English are significantly more variable than the correspondences of Serbo-Croatian. Where the English writing system is both morphemic and phonemic in its reference, the Serbo-Croatian alphabet demonstrates a clear priority for the phonemic.

This simple correspondence between letter and sound reflects the deliberate alphabet reforms introduced into Serbo-Croatian by Vuk Stefanović Karadžić and by Ljudevit Gaj in the 19th century. In this respect, the Serbo-Croatian orthography--which takes two forms, the Cyrillic and the Roman (see Lukatela, Savić, Ognjenović, & Turvey, 1978)--might be regarded as a nearly ideal medium of instruction by advocates of a purely phonetic writing system for the initial teaching of reading: Each phoneme is transcribed by only one letter or letter pair and each letter or letter pair is always pronounced.[1] (In the Cyrillic version there are only single letters.)

Does the fact that the grapheme-phoneme correspondencies of Serbo-Croatian are direct and consistent facilitate their acquisition? If it does, then the beginning readers of Serbo-Croatian may be less subject to errors in their reading of vowels and consonants. It is our intention to compare the two classes of phonemes within and between the orthographies of English and Serbo-Croatian. To this end we give due consideration, in what immediately follows, to the different accents that the five vowels of Serbo-Croatian may assume, suggestive as they are of a violation of the claimed-for spelling-to-sound regularity.

There are four variants of accent that can appear in syllables of Serbo-Croatian (see Figure 1). There is both a falling and a rising voice, each of which can occur in both a short and in a long form. These variations in accent can uniquely distinguish among different words (e.g., SEDI, see Footnote 2) but they are not specified by the script. The possible accents for any particular vowel are constrained by the position of the syllable within the word: Polysyllabic words may have any of the four accents on the penultimate syllable but the last syllable is usually unaccented. For monosyllabic words--the kind used in the present experiment--only long or short (falling) accents are possible.

As mentioned above, the Serbo-Croatian vowel set contains only five members. In terms of the F_1-F_2 vowel space, these vowels are qualitatively distinct as no region is shared by two different identities. One could claim that the four accents for each vowel introduce complexity into the simple and systematic relation between grapheme and phoneme as there are sometimes four possible interpretations for a particular Serbo-Croatian vowel grapheme. An inspection of acoustic parameters, however, suggests that the determiners of accent are basically independent of the particular vowel--that vowel identity, at least as it is defined by formant structure in some restricted phonemic

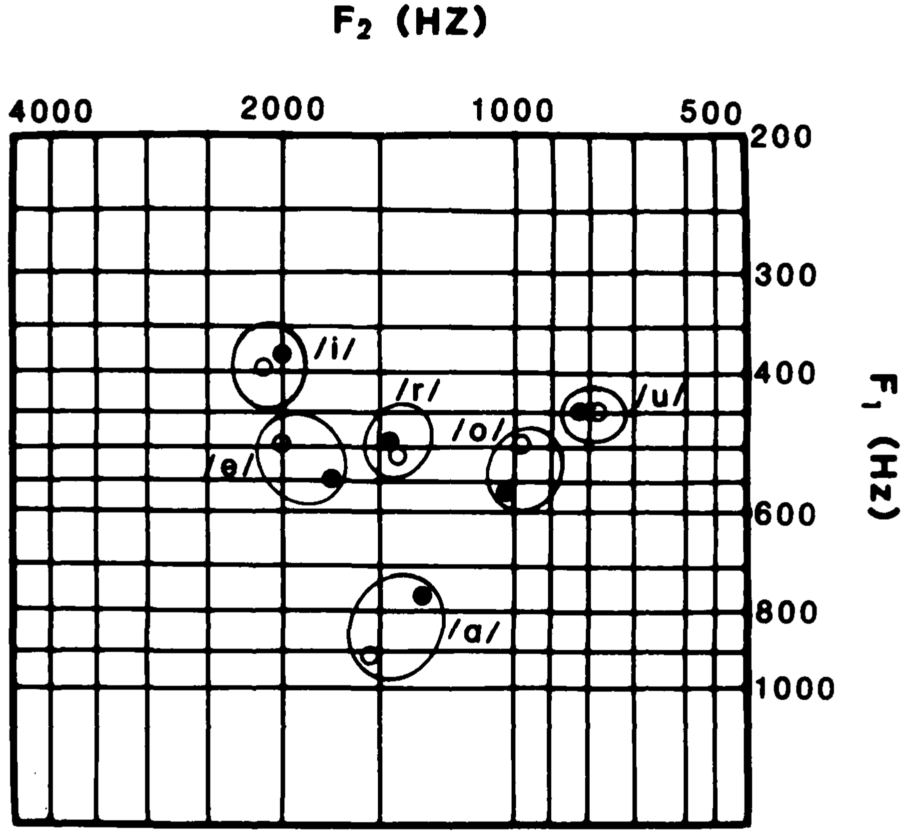


Figure 1. Acoustic vowel diagram of accented syllable nuclei occurring in approximately 400 Serbo-Croatian words produced by one speaker. Filled dots represent syllable nuclei bearing the short falling accent; Circles represent syllable nuclei with the long falling accent. (Modified from Lehiste & Ivić, 1963, p. 84.)

environments (Kalić, 1964), is not disturbed by variations in accent. These accent options for Serbo-Croatian vowels are to be contrasted with the complexities that characterize the pronunciation and the acoustics of English vowels. Of potential significance is the claim (Magner & Matejka, 1971) that the ideal accentual system as presented in Serbo-Croatian grammars "has little or no relationship with the accentual system(s) employed in many urban areas" [p. 189]. Speakers in the Magner and Matejka (1971) study could not always differentiate the four accentual variants. Discrimination between short rising and short falling forms was particularly vulnerable to error although contrasts between long rising and long falling accents were also commonly missed.

The implication of the foregoing is that the accent imposed on a particular vowel does not seem to influence its identification relative to other vowel options. So, for the child learning to read in Serbo-Croatian, the orthography will respect a simple, relatively context-free mapping between grapheme and phoneme for both vowels and consonants relative to the English orthography where the relationship for vowels is substantially more complex than the relationship for consonants. It is important to underscore that the orthographically distinct vowels of Serbo-Croatian are also phonetically distinct, in terms of the formant defined vowel space. It will not be possible, therefore, to distinguish orthographic from phonetic effects among Serbo-Croatian vowels.

METHOD

Subjects

Sixty-five first grade students at an elementary school in Belgrade participated in this study. Their ages ranged from 6.5 to 7.5 years and all had I.Q.'s within the normal range. At the time of testing, they had completed their first semester of school and had an active knowledge of the Cyrillic alphabet.

Materials and Design

Two hundred monosyllabic letter strings patterned as consonant-vowel-consonant (CVC) were constructed. One half of these CVCs were words and one half were pseudowords. All words were familiar to first graders as determined by Lukić (1970) and by consultation with the childrens' teachers. Following Fowler et al. (1977), in both the word and pseudoword lists, the twenty-five Serbo-Croatian consonant phonemes (which can occur in both the initial and in the final positions of a word) appeared twice in each position. In the majority of the trigrams, the medial letter was one of the five Serbo-Croatian vowels (/i/, /e/, /a/, /o/, /u/). In some trigrams, however, the medial letter was the semi-vowel /r/. In Serbo-Croatian, monosyllabic words of this type--consonant-/r/-consonant--occur relatively frequently. Of the one hundred words, twenty-five could be reversed to produce other words. For example the word "BOR" (pine) if read from right to left becomes "ROB" (slave).

Each string of three uppercase Cyrillic characters was arranged horizontally at the center of a 3" x 5" white card. These stimuli were printed in Cyrillic such that individual letter shapes were similar to the form generally used by the classroom teacher. The cards were placed face down in front of the child and were turned over one by one by the examiner. Each child was asked to read each letter string aloud as it was presented. Responses were written down by the examiner and were recorded simultaneously on magnetic tape.

Each child participated in two sessions. As in the procedure adopted by Fowler et al. (1979) words and pseudowords were blocked into separate lists and one list was presented in each session. Children who read the word list in the first session read the pseudowords list in the second and vice versa. The order of presentation was balanced across children.

Results

The responses to the stimuli revealed several types of errors: (a) reversal of sequence in which a letter string or a part of it was read from right to left, (b) omission, (c) addition, (d) substitution. Single letter orientation errors did not occur because the Cyrillic upper case letters did not provide opportunity for reversing letter orientation.

Sequence reversals. The analysis of errors showed that sequence reversals accounted for only a small proportion of the total of misread letters even though the lists were constructed to provide ample opportunity for the complete reversal of the sequences. (As noted, 25 percent of the words were "reversible;" and 13 percent of the pseudowords were words if read from right to left, for example the pseudoword NIS would become SIN, meaning "son").

The complete sequence reversals are distinguished from the partial and the total reversal scores for words and pseudowords and given in Table 1. Proportions of opportunity for error (in percentages) are presented within brackets. Sequence reversals were rare.

Table 1

Errors of sequence reversals (and proportion of opportunities, based on number of reversible letter strings)

	Complete sequence reversal	Partial sequence reversal	Total
Words	17 (1.1%)	6 (0.0%)	23
Pseudowords	21 (2.5%)	13 (1.5%)	34

Omissions. Single letter omission errors were also quite rare. Their distribution on initial and final consonants and on medial vowel/semivowel is presented in Table 2. Omissions of the final consonant in words seem to be more frequent than in pseudowords, but the respective proportions of opportunity are too small to allow any reliable conclusion on their distribution.

Table 2

Omission errors

	Initial consonant	Medial vowel	Final consonant	Total
Words	1	4	11 (0.2%)	16
Pseudowords	4	3	3	10

Additions. Errors of addition were distributed in a nonrandom manner (see Table 3). Additions of a single phoneme were more frequent before the final consonant (FC_1) than after the final consonant (FC_2), other types of additions being relatively infrequent.

Table 3

Errors of addition of a single phoneme

	Initial consonant	Medial vowel	Before final consonant FC_1	After final consonant FC_2	Total
Words	6	10	52	12	80
Pseudowords	1	9	52	25	87

In words and pseudowords where the medial letter was R (the semivowel /r/), additions of a single phoneme in front of the final consonant and after the semi-vowel were the most frequent. For example, the word GRB was often misread as /grab/, /grub/, or /grob/. In four words (GRB, VRH, TRG, TRN) there were 45 single vowel additions and in four pseudowords (BRS, DRN, KRP, PRK) there were 47 single vowel additions of FC_1 type. (Although all letter strings were printed in Cyrillic script, the Roman equivalents are presented here.) The proportion of opportunity for this particular error expressed as a percentage was 17 in the four words and 18 in the four pseudowords. This is a notable result. Apparently, in order to facilitate the phonetic representation of the letter string the child inserted a vowel between the medial semivowel and final consonant.

Substitutions. Substitutions of single phonemes were the major sources of error. The distribution of substitution errors on the initial and final consonant and on the medial vowel/semivowel for both words and pseudowords is presented in Table 4, which gives the raw error scores and the respective percentage (within brackets).

Table 4

Single phoneme substitution errors

	Initial consonant	Medial vowel	Final consonant	Total
Words	172 (2.6%)	93 (1.4%)	264 (4.1%)	529
Pseudowords	213 (3.3%)	113 (1.7%)	368 (5.7%)	693

An analysis of variance on total errors revealed that the word-pseudoword or lexicality contrast was not a significant source of variance, $F(1,198) = 3.51$, $MS_e = 43.74$, $p < .10$; neither was the interaction between lexicality and position within the syllable, $F(2,396) = .93$, $MS_e = 10.69$, $p > 1$. On the other hand, the position of a letter in a syllable was a highly significant contributor to the overall variance, $F(2,198) = 21.5$, $MS_e = 10.69$, $p < .001$. A protected t-test confirmed the previously-reported inferiority of performance on the final consonant relative to performance on the initial consonant in the present data, $t(99) = 268$, $p < .001$. However, it is plainly the case that performance on the vowels was inferior to performance on neither the initial or final consonants. In fact, protected t-tests reveal that performance on vowels was superior to performance on both initial and final consonants, $t(99) = 196$, $p < .001$ and $t(99) = 463$, $p < .001$, respectively. This is contrary to the findings in English.

Closer inspection of the children's response protocols revealed that syllables that included the character Ч, Џ, ћ, or Ђ symbolizing, respectively, the affricates /tʃ/, /dʒ/, /tʃj/, /dʒj/ were disproportionately subject to error. The affricates are notoriously more difficult to distinguish by ear and to produce distinctively than other sounds of Serbo-Croatian. Excluding those syllables (seventeen words and seventeen pseudowords) in which affricates occurred in either initial or final position substantially reduced the overall errors and eliminated the absolute difference between the initial consonant errors and the medial vowel errors as can be seen in Table 5.

Table 5

Errors when the affricates

(Ч, Џ, Ћ, and Ђ) were excluded*

	Initial consonant	Medial vowel	Final consonant
Words	44	51	124
Pseudowords	104	97	258

*Total errors with 17 word stimuli and 17 pseudoword stimuli excluded.

Relation between errors and target consonants. A matrix of confusions between stimulus letter and substituted response was constructed separately for initial position and for final position errors. A correlation of the two matrices yielded a value of $r = .73$; which means that 53 percent of the variance in the patterns of errors for initial and final consonants was common.

A correlation was then computed between the number of shared phonetic features and the frequency of error. (Only those target-error combinations were included in which a subject actually produced an error.) Using Jakobson's (1962) feature matrix for Serbo-Croatian and including the feature values for those features that need not be specified in order to capture only the minimal distinctive contrasts of the Serbo-Croatian phonology, two new matrices of shared features were created--one for target vowels (including /r/) with error vowels and one for consonants (including /r/) with consonants. Here, shared features can assume seven values. For word-initial consonants, the relation between common features and frequency of errors among presented-substituted letter pairs was $r = .23$ $N = 200$, $p < .01$. For word-final consonants, the relation was $r = .30$ $N = 200$, $p < .01$. In both cases, the frequency of confusions and number of shared phonetic features do correlate. We can interpret this to mean that phonetic similarity does account significantly for some portion of the variance in the pattern of confusions among presented and substituted consonant pairs. This finding is consistent with the pattern of errors derived from studies of English consonants (Fowler et al., 1977).

Relation between errors and target vowels. Unlike the English vowel findings, however, the vowel confusions in Serbo-Croatian can also be related to the degree of phonetic contrast. The proportion of error confusions is given in Table 6. The correlation between number of shared features and frequency of each presented-substituted letter pair confusion was $r = .52$ $N = 30$, $p < .001$. This value of r is particularly high given the restricted range (vowels share between 3 and 6 features) and the relatively small N (there are 30 possible confusions). It suggests that the vowel substitutions of Serbo-Croatian, like the consonant substitutions of Serbo-Croatian and unlike the vowel substitutions of English are, at least in part, phonetically governed.

Table 6

Percent of total errors. Rows represent presented vowel. Columns represent incorrect substitution. (r was never substituted for another vowel.)

	a	e	i	o	u	r
a		9	2	10	3	
e	9		5	2	1	
i	2	5		1	4	
o	10	4	1		3	
u	2	1	1	8		
r	8	1	1	3	3	

Discussion

The two major contrasts between the present data for beginning readers of Serbo-Croatian and those previously reported for beginning readers of English are that: (1) vowels in the medial position of a written consonant-vowel-consonant syllable are no more likely to be read incorrectly--indeed are less likely to be given an incorrect reading--than the initial and final consonants; and (2) vowel errors are no less likely to be rationalized by phonetic feature considerations than are consonant errors. Let us consider each contrast in turn.

As noted above, the Serbo-Croatian vowel set is numerically smaller than its English counterpart (the Serbo-Croatian vowels are only five in number) and qualitatively better defined (the Serbo-Croatian vowels are non-overlapping in the F_1-F_2 space regardless of accent). Is the fact that the Serbo-Croatian vowel set is smaller--and, therefore, that the likelihood of correctly reading a member of the set by chance is greater--reason enough for the proportionately smaller number of errors on Serbo-Croatian vowels? A guessing explanation is worthy of consideration if there is a good reason to believe that a random guessing strategy was being used. There were, in all, 13,000 opportunities for vowel errors in the present experiment (200 syllables, 65 subjects). As is evident from Table 4, the number of actual vowel errors totaled 205, which is far below the number of errors to be expected if the children were merely guessing at the vowels. (Since the guessing probability for consonants is trivially low, it would not alter the actual error rate and is not discussed.) Clearly a general guessing strategy has to be ruled out, which does not, of course, rule out guessing as a back-up strategy when all else fails. The 205 errors, therefore, might be interpreted as representing those occasions on which the children were forced to guess and guessed wrongly. Which is to say that 205 represents four-fifths of all those occasions when the children guessed because on one-fifth of these occasions they guessed correctly. By this reasoning, therefore, the number of times the children were forced to guess amounted to about 256 so that even disallowing correct guessing would not elevate the vowel errors above the consonant errors (see Table 7). In short, the fact that the vowels were not the major source of errors for beginning readers of Serbo-Croatian, as they were for beginning readers of English, is probably not attributable--at least, not in full--to the smaller size of the Serbo-Croatian vowel set; that it might be attributable, in larger part, to the greater distinctiveness of members of Serbo-Croatian vowel set is considered below.

Table 7

Total number of errors including all CVC strings

(i.e. 100 word stimuli, and 100 pseudoword stimuli)

	Initial consonant	Medial vowel	Final consonant
Words	172	93	264
Pseudowords	213	112	368

Let us now turn to the observation that beginning readers of Serbo-Croatian produced vowel errors that were, like consonant errors, rationalized by the degree of phonetic contrast. Recall that the observation for beginning readers of English was that vowel errors, unlike consonant errors, did not bear a feature-based relation to their target sounds (Fowler et al., 1979). This contrast might index a significant difference between the two orthographies and the challenge they pose to the neophyte reader. However, attempts to cash this promissory note must be prefaced by a necessary caveat: That the aforementioned contrast could be illusory, a trivial consequence of whether one has hit upon the propriety feature set for defining vowels. Possibly, a feature matrix for English vowels other than that used by Fowler et al. (1979) would capture a more pronounced phonemic basis for the vowel errors of their beginning readers.

Assuming that this possibility is not correct, we can raise two questions concerning the contrast currently under consideration: (1) Why should the errors in reading Serbo-Croatian vowels be speech-related when the errors in reading English vowels are not?; (2) What are the consequences for (beginning) reading of this conformity of vowels and consonants in Serbo-Croatian and this dissociation of vowels and consonants in English? As noted in the introduction, the Serbo-Croatian orthography is phonographic in a way that the English orthography is not, viz., that totally reliable guides to the pronunciation of a word occur even at the orthographic grain-size of the single letter. English orthography, being simultaneously but complexly a representation of morphology and phonology--where these representations are mixed fairly inconsistently from word to word (Gleitman & Rozin, 1977)--mandates that often the only reliable guides to pronunciation are to be found at an orthographic grain size that sometimes encompasses several letters and very often encompasses entire words. Put differently, English orthography is partly morphemic. Thus, the beginning reader of Serbo-Croatian can relate to the orthography as simply a phonological representation and derive the pronunciations of the 'consonantal' and 'vocalic' constituents of a word purely on phonological grounds. In comparison, the beginning reader of English must relate to the orthography as both a phonological representation and a morphological representation and may not necessarily be able to derive the pronunciations of the 'consonantal' and 'vocalic' constituents of a word in precisely the same way as the beginning reader of Serbo-Croatian.

Consider now a theory of initial reading acquisition that follows from the notions of linguistic awareness and encodedness (Mattingly, 1972). A fairly standard scenario is one in which the visual form of a word seen by the child co-occurs with the acoustic form produced by the 'teacher.' Now it must be assumed that the child's internal lexicon already represents familiar words in a way sufficient for the purposes of saying them and recognizing them when heard. These representations have been established largely on tacit grounds as the inevitable consequence of a decoding device that condenses out discrete phonemes from the continuous speech stream. In learning to read analytically, however, that which is normally done tacitly must now be done explicitly: The heard word produced by the 'teacher' must be explicitly decomposed into its constituents in order to effect a mapping between its structure and the constituent structure of the seen symbol string.

Somehow, the child must actively fashion either a special lexicon, one to which visually encountered words can be referred, or a new (orthographic) way of accessing the already-existing (phonologically accessible) lexicon. In either case, the facility with which the child can internally represent written words as ordered linguistic segments abstractly consonant with the ordered visual segments depends on the child's linguistic awareness, the awareness that speech is divisible into those phonological segments that the letters represent (Liberman, I. Y., Liberman, A. M., Mattingly, & Shankweiler, 1980). If a special lexicon is fashioned, then it should be referred to as an explicit lexicon (to distinguish it from the lexicon fashioned on mainly tacit grounds that supports speech perception and speech production). This explicit lexicon will be fallible and, similarly, the fashioning of a new mode of lexical access will be difficult, to the degree that the encodedness of speech obscures for the individual listener the phonemic composition of heard words.

We return at this juncture to a focal question: Is an appeal to encodedness sufficient to account for the difference in the relative magnitudes of vowel errors between beginning readers of English and beginning readers of Serbo-Croatian? It would seem not. The degree to which words resist explicit decomposition into their constituent phonemes should be more or less the same for both languages. However, the non-overlapping nature of the Serbo-Croatian vowel space would guarantee greater consistency in the assignment of internal descriptors to the vowels in the formation of an internal representation. And in this regard the fact that, for spoken Serbo-Croatian, any one point in the F_1-F_2 space is associated with only one vowel (or no vowel at all) is buttressed by the fact that, for written Serbo-Croatian, any one vowel character in the alphabet is associated with only one vowel phoneme. It can be argued, therefore, on two counts, that the pronunciation of a Serbo-Croatian vowel (by a beginning reader) is more likely to be correct, *ceteris paribus*, than the pronunciation of an English vowel (by a beginning reader). However, it remains equivocal whether the truth of this argument is grounded in the orthography or the phonology of Serbo-Croatian vowels.

REFERENCE NOTE

1. Liberman, I. Y. *Segmentation of the spoken word and reading acquisition*. Paper presented to the Society for Research in Child Development, Philadelphia, PA, 1973.

REFERENCES

Daniels, J. C., & Diack, H. *Progress in reading*. Nottingham: University of Nottingham, Institute of Education, 1956.

Dewey, G. *Relative frequency of English spellings*. New York: Teachers College Press, 1970.

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

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

Gleitman, L. R., & Rozin, P. The structure and acquisition of reading: Relation between orthographic and the structure of language. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading: The proceeding of the CUNY Conferences*. Hillsdale, N.J.: Erlbaum, 1977.

Jakobson, R. *Roman Jakobson selected readings*. *I Phonological studies*. The Hague: Mouton, 1962.

Kalić, D. *Akusticka analiza vokala u ekavskom govoru Srpskohrvatskog jezika: I Tehnicke mogucnosti prepoznavanja pomocu filtara*. (Acoustic analysis of vowels in Ekavian dialect of Serbo-Croatian: And technical possibilities of recognition by filters.) Unpublished doctoral dissertation, Faculty of Electrical Engineering, University of Belgrade, 1964.

Lehiste, I., & Ivić, P. Accent in Serbo-Croatian: An experimental study. *Michigan Slavic Materials 4*. Ann Arbor, Mich.: University of Michigan, 1963.

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

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

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

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. 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, N.J.: Erlbaum, 1977, 207-225.

Liberman, I. Y., Shankweiler, D., Orlando, C., Harris, K., & Bell-Berti, F. Letter confusions and reversals of sequence in the beginning reader: Implications for Orton's theory of developmental dyslexia. Cortex, 1971, 7, 127-142.

Lukatela, G., Savić, M., Ognjenović, P., & Turvey, M. T. On the relation between processing the Roman and Cyrillic alphabets: A preliminary analysis with bi-alphabetic readers. Language and Speech, 1978, 21, 113-141.

Lukić, V. Active written vocabulary of pupils at the elementary school age (in Serbo-Croatian). Belgrade: Zavod za Izdavanje Udzbenika SR Srbije, 1970.

Magner, T. F., & Matejka, L. Word accent in Serbo-Croatian. University Park, Pa.: Pennsylvania State University Press, 1971.

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

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

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

Venezky, R. L. In M. Kavanagh (Ed.), Communicating by language: The reading process. Bethesda, Md., 1968.

Wheeler, D. D. Processes in word recognition. Cognitive Psychology, 1970, 1, 59-85.

FOOTNOTES

[1]There are exceptions to this characterization: For example, the first "d" in "predsednik" is generally interpreted as /t/. The number of violations is small, however.

[2]"Sedi," with differing accents, can mean grey as an adjective, a man with grey hair, the third person singular of the verb "to grey" or the third person singular of the verb "to sit."

BI-ALPHABETISM AND WORD RECOGNITION*

Laurie B. Feldman+

THE LINGUISTIC ENVIRONMENT OF YUGOSLAVIA

The linguistic environment in Yugoslavia allows investigation of the interrelation among various symbolic systems. Several Slavic languages are spoken within the boundaries of one relatively small country. This contact among languages permits a variety of bilingual environments to develop and allows for the study of the symmetric and nonsymmetric influences in the acquisition and mastery of two languages. In addition, and more to the focus of the present work, among people whose first spoken language is Serbo-Croatian, which is the official language of Yugoslavia, a large portion learns to read and write that language completely in two different alphabets—Roman and Cyrillic. This reflects, in part, an educational requirement that both alphabets be taught within the first two grades. (The Roman alphabet is taught first in the western part of Yugoslavia and the Cyrillic alphabet is taught first in the eastern part of thhe country.) This bi-alphabetic environment invites study of the cognitive relation between two alphabetic symbol systems. In my report, I summarize results of a series of experiments that explored how visually presented letter strings are recognized by readers who command two alphabetic systems. Then I discuss implications of these findings with respect to the interrelation between the two visual alphabetic systems of Serbo-Croatian. Before I review these results, however, some special properties of Serbo-Croatian and its writing systems need to be described.

The Serbo-Croatian language is written in two different alphabets, Roman and Cyrillic. The two alphabets transcribe one language and their graphemes map simply and directly onto the same set of phonemes. These two sets of graphemes are, with certain exceptions, mutually exclusive (see Table 1). Most of the Roman and Cyrillic letters are unique to their respective alphabets. There are, however, certain letters that the two alphabets have in common. In some cases, the phonemic interpretation of a shared letter is the same whether it is read as Cyrillic or as Roman; these are referred to as

---

*A portion of this paper was presented at the NATO Conference on the Acquisition of Symbolic Skills, University of Keele, England, July 1982.

+Also Dartmouth College.

Acknowledgment. The research was supported in part by NICHD Grant 08495 to the University of Belgrade, in part by NICHD Grant HD 01994 to Haskins Laboratories, and in part by NSF dissertation fellowship BNS 7924409 to Laurie Feldman. In addition, funds were contributed by the Faculties of Philosophy and Electrical Engineering at the University of Belgrade and by the University of Connecticut.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

TABLE 1

SERBO-CROATIAN				
ROMAN		CYRILLIC		
PRINTED		PRINTED		
UPPER CASE	LOWER CASE	UPPER CASE	LOWER CASE	LETTER NAME IN I.P.A.
A	a	А	а	a
B	b	Б	б	bə
C	c	Ц	ц	tsə
Č	č	Ч	ч	tʃə
Ć	ć	Ћ	ћ	tʃjə
D	d	Д	д	də
Đ	đ	Ђ	ђ	dʒjə
DŽ	dž	Џ	џ	dʒə
E	e	Е	е	ɛ
F	f	Ф	ф	fə
G	g	Г	г	gə
H	h	Х	х	xə
I	i	И	и	i
J	j	Ј	ј	jə
K	k	К	к	kə
L	l	Л	л	lə
LJ	lj	Љ	љ	ljə
M	m	М	м	mə
N	n	Н	н	nə
NJ	nj	Њ	њ	njə
O	o	О	о	ɔ
P	p	П	п	pə
R	r	Р	р	rə
S	s	С	с	sə
Š	š	Ш	ш	ʃə
T	t	Т	т	tə
U	u	У	у	u
V	v	В	в	və
Z	z	З	з	zə
Ž	ž	Ж	ж	ʒə

common letters. In other cases, a shared letter has two phonemic interpretations, one in the Roman reading and one in the Cyrillic reading; these are referred to as ambiguous letters (see Figure 1).

Whatever their category, the individual letters of the two alphabets have phonemic interpretations (classically defined) that are virtually invariant over letter contexts. (This reflects the phonologically shallow nature of the Serbo-Croatian orthography.) Moreover, all the individual letters in a string of letters, be it a word or nonsense, are pronounced—there are no letters made silent by context. Finally, Serbo-Croatian is a highly inflected language. Many aspects of the syntax are marked by appending a suffix, commonly composed of a vowel, or a vowel and a consonant, to some base form.

Given the relation between the two Serbo-Croatian alphabets, it is possible to construct a variety of types of letter strings. A letter string composed of uniquely Roman and common letters (e.g., FABRIKA) or of uniquely Cyrillic and common letters (e.g.,ФАБРИКА) would be read in only one way and could be either a real word or a nonsense word. A letter string composed entirely of the common and ambiguous letters (e.g., EKCEP) is bivalent. That is, it could be pronounced in one way if read as Roman and pronounced in a distinctly different way if read as Cyrillic; moreover, it could be a word in one alphabet and nonsense in the other or it could represent two different words, one in one alphabet and one in the other, or finally, it could be nonsense in both alphabets (see Table 2).

The present research focused on the detriment to performance incurred with phonologically bivalent letter strings in both skilled and beginning readers. These effects are interpreted as evidence of the influence of phonological decoding on visual word recognition (i.e., lexical decision and naming). To anticipate, results of the adult studies indicate that the effect of phonological bivalence is evidence of a mandatory phonological analysis in word recognition among skilled readers, an analysis that would not be predicted by any conventional (visual) lexical account. Results of the children's study show that reliance on a phonological recognition strategy varies with reading skill and suggest that the successive acquisition of two alphabetic systems by the beginning reader may increase the demands of decoding phonology.

LEXICAL DECISION AND NAMING PERFORMANCE IN BI-ALPHABETIC ADULT READERS

When bi-alphabetic adult readers of Serbo-Croatian performed a lexical decision task, letter strings composed of ambiguous and common characters (i.e., those letter strings that could be assigned both a Roman and a Cyrillic alphabet reading, e.g., CABAHA) incurred longer latencies than the unique alphabet transcription of the same word (e.g., SAVANA) (Feldman, 1981). This effect of phonological ambiguity was significant both for ambiguous words and pseudowords, but it was more consistent for words (see Figure 2). In an analogous naming task where subjects were instructed to read each letter string by its word reading when that option existed (Feldman, 1981), the same basic pattern of results occurred (see Figure 3).[1] Correlations between tasks were computed by taking the mean reaction time for individual words and pseudowords in the lexical decision and naming tasks. When the ambiguous and unique alphabet transcriptions were considered separately, both correlations

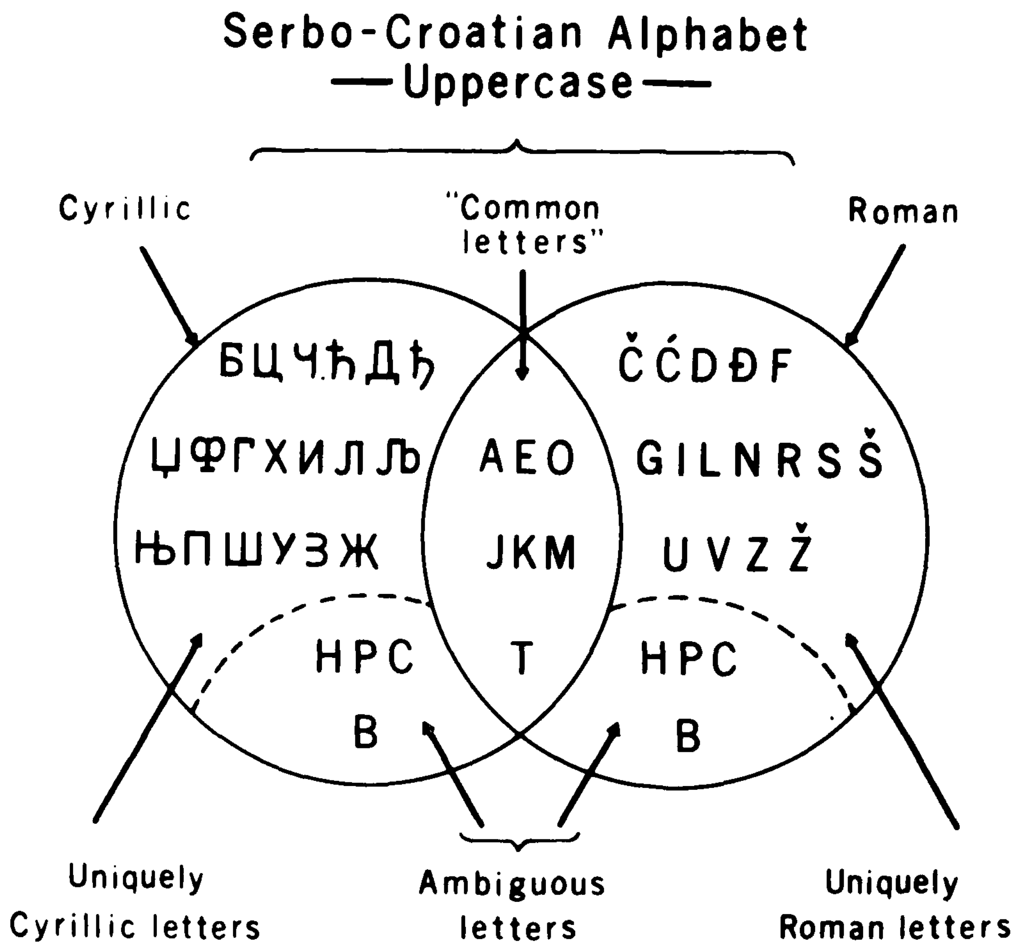


Figure 1. Letters of the Roman and Cyrillic alphabets.

Table 2

Types of Letter Strings and Their Lexical Status

Composition of Letter String	Phonemic Interpretation	Meaning
AMBIGUOUS and COMMON		
EKCEP*	Cyrillic /ekser/	nail
	Roman /ektsep/	nonsense
PATAK*	Cyrillic /ratak/	nonsense
	Roman /patak/	duck
KACA	Cyrillic /kasa/	safe
	Roman /katsa/	pot
HABOT*	Cyrillic /navot/	nonsense
	Roman /habot/	nonsense
COMMON		
JAJE	Cyrillic /jaje/	egg
	Roman /jaje/	egg
TAKA	Cyrillic /taka/	nonsense
	Roman /taka/	nonsense
UNIQUE and COMMON		
EKSER*	Cyrillic impossible	
	Roman /ekser/	nail
NAVOT*	Cyrillic impossible	
	Roman /navot/	nonsense
ПАТАК*	Cyrillic /patak/	duck
	Roman impossible	
ХАБОТ*	Cyrillic /habot/	nonsense
	Roman impossible	

(*indicates those letter string types included in the children's experiment)

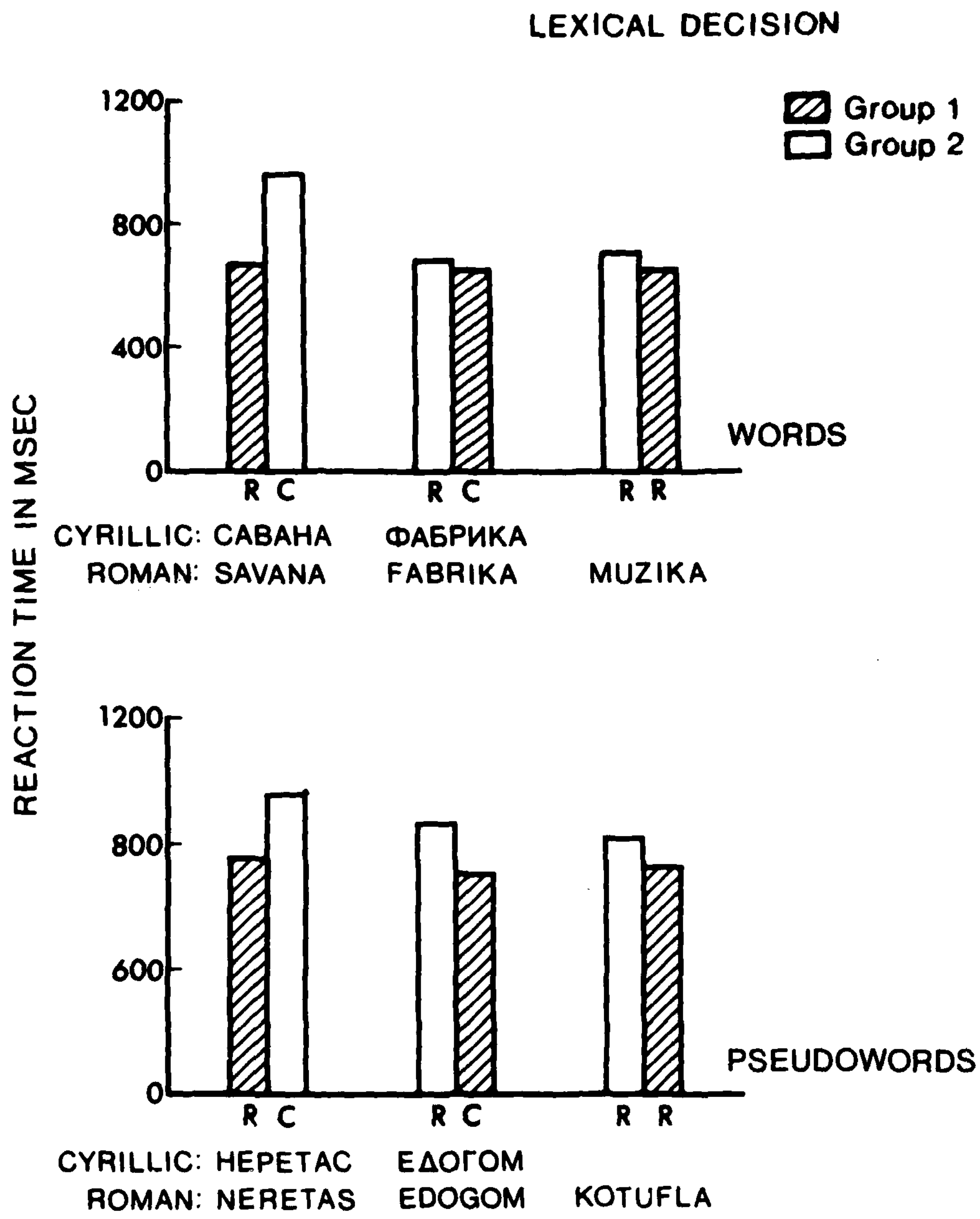


Figure 2. Mean reaction time for lexical decision on AMBIGUOUS (CABAHA) and UNAMBIGUOUS (FABRIKA, MUZIKA) words and pseudowords (in their Roman and Cyrillic transcriptions).

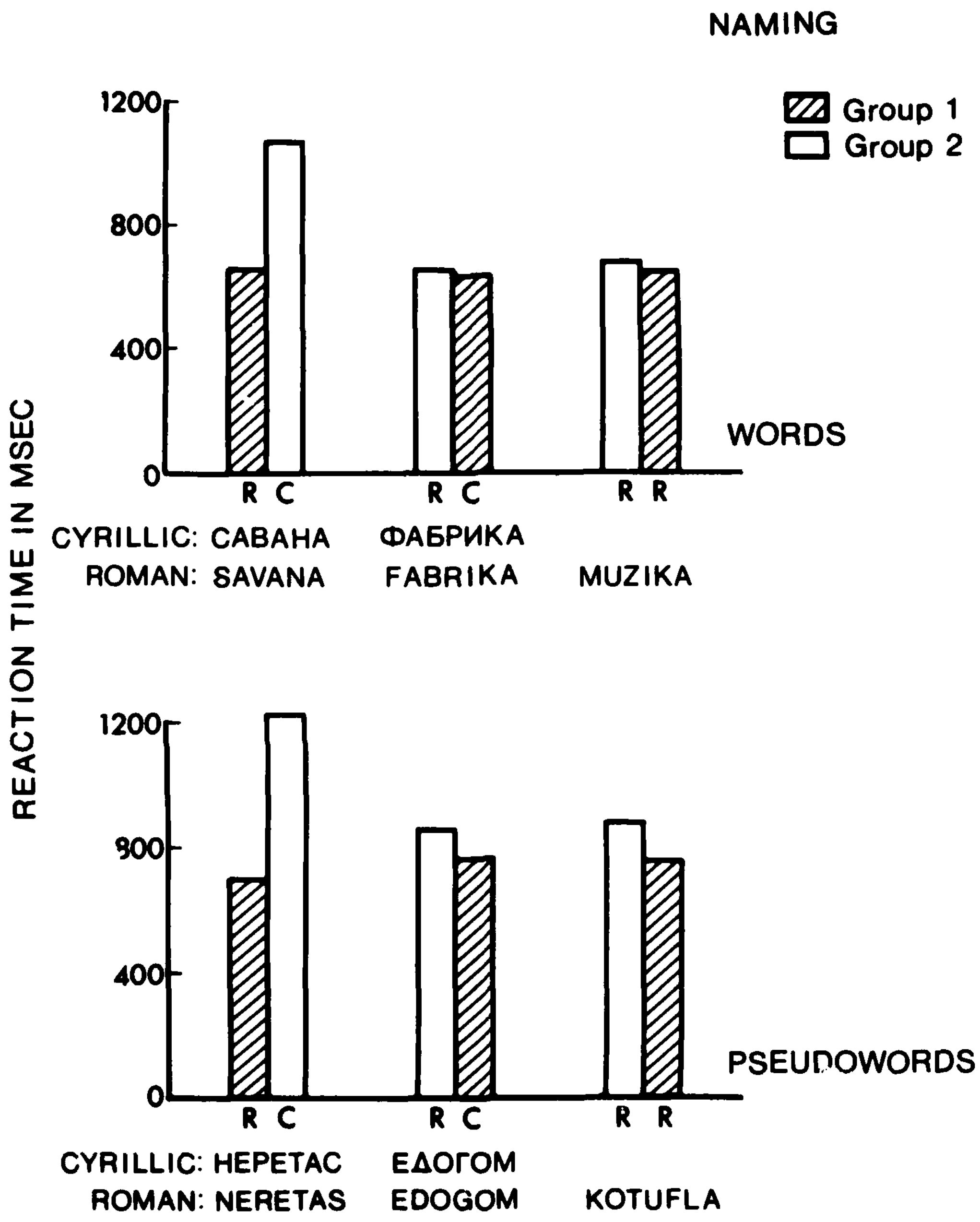


Figure 3. Mean reaction time to name AMBIGUOUS (CABAHA) and UNAMBIGUOUS (FABRIKA, MUZIKA) words and pseudowords (in their Roman and Cyrillic transcriptions).

between tasks were significant: For ambiguous letter strings, r = .48; for the unique alphabet transcriptions, r = .34. When means for all word and pseudoword forms within a condition were included (and the correlation between tasks was averaged over experimental conditions), the overall correlation between lexical decision and naming was even stronger, r = .66. This correlation, supported by the similarity of the figures for lexical decision and naming, implicates similar processes in both tasks. In the adult experiments, words were selected so as to include a varied distribution in the number and position of the ambiguous characters within the letter string (see Table 3). Results indicated that all letter strings that could be assigned both a Roman and a Cyrillic reading incurred longer latencies than the unique alphabet transcription of the same word and that the magnitude of the difference between the ambiguous form of a word and its unique alphabet control depended on the number and distribution of ambiguous characters in the ambiguous letter string (see Tables 4 and 5). These results with phonologically bivalent letter strings were interpreted as evidence that both lexical decision and naming in Serbo-Croatian necessarily involve an analysis that is sensitive to phonology and component orthographic structure. Moreover, skilled readers were not able to suppress the phonological analysis even though it was detrimental to performance.

In those experiments, all phonologically ambiguous letter strings that were words, were words by their Cyrillic interpretation. But the unique alphabet words and pseudoword strings included both Roman letter strings and Cyrillic letter strings. That is, by the design of the experiment, in performing the lexical decision or naming task, skilled readers were obliged to switch between alphabets in order to consider both a Roman and Cyrillic interpretation.

Results of earlier lexical decision experiments (Lukatela, Popadić, Ognjenović, & Turvey, 1980; Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978) have shown that the large decrement to performance incurred when Serbo-Croatian letter strings are associated with two phonological interpretations is not easily explained in terms of an account based on problems of letter identification due to interference between alphabets, however. In the earlier bi-alphabetic lexical decision experiments by Lukatela and his colleagues (Lukatela et al., 1978), both the design of the experiment and the instructions to the subject were intended to restrict subjects to the Roman reading: There were no uniquely Cyrillic characters presented anywhere during the experimental session and subjects were asked to interpret letter strings by their Roman reading. Nevertheless, in a pure Roman context, positive decision times to ambiguous Roman words were significantly slowed and more prone to error relative to decision times to (other) unambiguous Roman words. An unpublished study by the present author (Feldman, Note 1) supports this finding. In that experiment, all letter strings composed of ambiguous and common characters that were words, were words by their Roman interpretation and all other letter strings contained unique Roman and common characters (but no unique Cyrillic characters). Performance on ambiguous letter strings was again significantly more prone to error than on the unique alphabet transcription of the same letter strings (and a trend in the reaction time data, although it missed significance, suggested the same results). To summarize, lexical decision latencies to letter strings composed of ambiguous and common letters were slowed relative to their appropriate controls, both in a mixed

Table 3

Distribution of Ambiguous Letters and Pronunciation for AMBIGUOUS Cyrillic Letter Strings.

Three Syllable Letter Strings	Possible Pronunciations	Meaning	Number of Ambiguous Letters	Number of Ambiguous Syllables
CABAHA	Cyrillic /savana/	savanna	3	3
	Roman /tsabaxa/	nonsense		
KAPABAH	Cyrillic /karavan/	caravan	3	2
	Roman /kapabax/	nonsense		
OCTABKA	Cyrillic /ostavka/	resignation	2	2
	Roman /otstabka/	nonsense		
Two Syllable Letter Strings				
OPMAH	Cyrillic /orman/	cabinet	2	2
	Roman /opmax/	nonsense		
CAHTA	Cyrillic /santa/	iceburg	2	1
	Roman /tsaxta/	nonsense		
KOTBA	Cyrillic /kotva/	anchor	1	1
	Roman /kotba/	nonsense		

Table 4

Mean Reaction Time for Lexical Decision on AMBIGUOUS Cyrillic/Unique Roman Words.

Three Syllable Letter Strings	Number of Ambiguous Letters	Number of Ambiguous Syllables	Cyrillic Reaction Time	Roman Reaction Time	Difference between Cyrillic and Roman
CABAHA	3	3	960	676	284
KAPABAH	3	2	1038	646	392
OCTABKA	2	2	894	710	184
Two Syllable Letter Strings					
OPMAH	2	2	927	655	272
CAHTA	2	1	1001	617	384
KOTBA	1	1	880	625	255

Table 5

Mean Reaction Time to Name AMBIGUOUS Cyrillic/Unique Roman Words.

Three Syllable Letter Strings	Number of Ambiguous Letters	Number of Ambiguous Syllables	Cyrillic Reaction Time	Roman Reaction Time	Difference between Cyrillic and Roman
CABAHA	3	3	1049	661	388
KAPABAH	3	2	1047	609	438
OCTABKA	2	2	933	594	339
Two Syllable Letter Strings					
OPMAH	2	2	1125	703	422
CAHTA	2	1	1201	687	514
KOTBA	1	1	1071	667	404

alphabet and in a pure alphabet context. Together, these results invalidate an account of bivalence that depends exclusively on a strategy-based conflict or interference between the two alphabet modes.

Other variations of the bi-alphabetic lexical decision task invalidate a decision process account of the detriment due to bivalence that posits (post-lexical) interference between conflicting lexical judgments. Lexical decision latencies to letter strings composed entirely of ambiguous and common letters were always slowed, whether 1) both the Cyrillic interpretation and the Roman interpretation yielded a positive response (Lukatela et al., 1980; Feldman, Note 1); 2) both the Cyrillic interpretation and the Roman interpretation yielded a negative response (Feldman, 1981; Lukatela et al., 1978, 1980); or 3) the Cyrillic interpretation and the Roman interpretation yielded one positive response and one negative response (Feldman, 1981; Lukatela et al., 1978, 1980). Although methodological considerations make it impossible to compare these three results directly, it is evident that the effect of bivalence is not confined to instances in which the Roman and Cyrillic interpretation produce conflicting lexicality judgments.

Two other aspects of bi-alphabetic lexical decision need to be remarked upon. First, words composed entirely of common letters (with no ambiguous or unique letters), e.g., JAJE, were accepted (as words) no more slowly than letter strings that included common and unique letters. Likewise, pseudowords composed entirely of common letters, e.g., TAKA were rejected (as words) no more slowly than letter strings that included common and unique letters. Because the distinction between common letters and ambiguous letters is based on their phonemic interpretation, this result suggests that it is phonological bivalence rather than a visually-based alphabetic bivalence that governs the effect (see Lukatela et al., 1978, 1980, for a complete discussion).

Finally, the effects of bivalence did not occur if a letter string composed predominantly of ambiguous and common characters contained even one unique character. Specifically, the presence of one unique letter that occurs as an inflectional suffix on a singular noun, is sufficient to cancel any effect of bivalence in lexical decision (Feldman, Kostić, Lukatela, & Turvey, 1981). It seems that while the presence of ambiguous and common letters is a necessary condition for phonological bivalence and the size of the effect depends on the number of such ambiguous letters, nevertheless any effect can be cancelled by the presence of even a single character that uniquely specifies alphabet.

At this point it is tempting to conclude that skilled readers of Serbo-Croatian, when performing the lexical decision (and naming) task, are always sensitive to the presence of ambiguous and unique characters. However, results of two experiments suggest that there is need for further qualification. Given the availability of two alphabets for Serbo-Croatian, it is possible to create a novel but interpretable string by mixing characters from the Roman and Cyrillic alphabets. When words were selected so as not to include any potentially ambiguous characters in their mixed alphabet form, lexical decision judgment times for words (Katz & Feldman, 1981) and naming times for words (Feldman & Kostić, 1981) were no slower for mixed alphabet forms (e.g., ФLAШA) than for pure alphabet forms of the same letter strings (e.g., FLAŠA). Evidently, skilled readers can perform both lexical decision and naming in a phonologically analytic manner that is indifferent to mixed alphabet distortions to visual form. In conclusion, under the special conditions of bi-alphabetically induced phonological ambiguity, attention to some visual characteristics of letter strings is manifest only when it serves to disambiguate alphabet.

NAMING PERFORMANCE FOR BI-ALPHABETIC BEGINNING READERS

When beginning readers of Serbo-Croatian performed a naming task, letter strings composed of ambiguous and common characters were named more slowly than the unique alphabet transcription of the same word (Feldman, Note 2). In that experiment, half the letter strings were ambiguous and half were unique to one alphabet. Among the ambiguous letter strings, half were words by their Cyrillic reading (and pseudowords by their Roman reading) and half were words by their Roman reading (and pseudowords by their Cyrillic reading). Further, among those letter strings that contained unique and common letters, half were unequivocally Cyrillic and half were unequivocally Roman. Finally, within both ambiguous and unique letter strings, half were words by one of their readings and half were always pseudowords. Subsequent to the bi-alphabetic naming task, each subject named a list of pseudowords, all of which were written in an unequivocally Cyrillic transcription. Third- and fifth-grade students, all of whom had learned Cyrillic print in first grade and Roman print in second grade, served as subjects.

Results indicated that overall, naming was slower for third-graders than for fifth-graders and that both third and fifth graders were slowed more when naming phonologically bivalent letter strings than when naming unique alphabet controls. This result occurred with ambiguous words (both Roman and Cyrillic) and with ambiguous pseudowords. Thus, the effect of bivalence is consistent with the naming data in adults reported above. The design of this experiment also permitted a comparison of bivalence across alphabets. For third-graders, the degree of impairment was greater when the ambiguous letter string is a word by its Roman reading (and a pseudoword by its Cyrillic reading), e.g., BATAK, than when it is a word by its Cyrillic reading (and a pseudoword by its Roman reading), e.g., EKCEP. For fifth-graders, however, there was no such interaction (see Figure 4). The asymmetric interference of first-learned and second-learned alphabet in naming ambiguous letter strings for younger readers but not for older readers suggests that the asymmetry is only temporary and that it may be equalized through experience.

In subsequent analyses, mean pseudoword naming time was used as a measure of reading skill for each child; the difference between each subject's latency to name all unique words and his or her latency to name all ambiguous words served as a measure of the impairment due to phonological bivalence. The correlation computed between pseudoword naming time and impairment due to phonological bivalence was significant and negative, $r = -.33$, $t = 2.80$, $p < .05$. That is, those readers who were fastest at decoding pseudowords were most slowed with bivalent letter strings.

In summary, results for naming ambiguous letter strings in both skilled and less-skilled beginning readers revealed a significant effect of phonological ambiguity on naming time. In addition, the phonological analysis required to recognize a phonologically bivalent letter string may be more vulnerable to

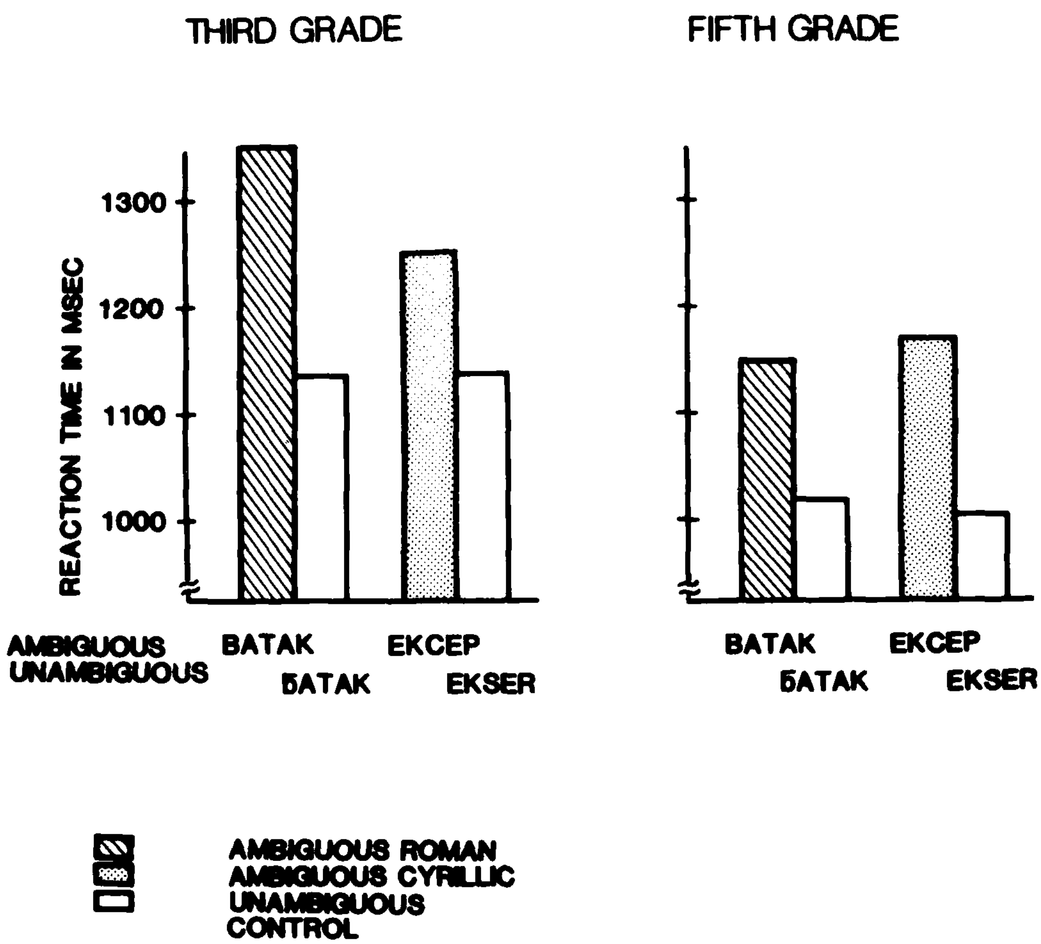


Figure 4. Mean reaction time for third- and fifth-graders to name AMBIGUOUS (Roman and Cyrillic) words and the UNAMBIGUOUS alphabet transcription of the same words.

disruption when that letter string is a word by the second-learned alphabet reading than when it is a word by the first-learned alphabet reading.[2] Finally, using pseudoword naming speed as an index of reading skill, the detriment to performance caused by reliance on a phonologically analytic recognition strategy when naming ambiguous letter strings was greater in skilled beginning readers than in less-skilled beginning readers.

THE COMMAND OF TWO SYMBOL SYSTEMS

The above results provide the following characterization of bi-alphabetism: 1a) When confronted with a letter string composed entirely of ambiguous and common letters, readers are slowed relative to their performance on an alternative transcription of the same word that is comprised of characters that are unique to one alphabet. However, with a letter string composed exclusively of common letters, readers are no slower than with a letter string that includes at least one unique letter. 1b) The magnitude of the difference between the ambiguous transcription of a letter string and the unique alphabet transcription of that same letter string increases as the number of ambiguous characters increases. 2) The presence of a single unique letter is sufficient to neutralize any effect of ambiguous letters. 3) When one word contains a mix of unique letters from both the Roman and Cyrillic alphabets, readers are not slowed relative to the performance on the same letter string transcribed in purely Roman or purely Cyrillic script. 4) Appreciation of bivalent phonology with a subsequent impairment to performance is enhanced as the efficacy of phonological decoding skill increases.

In summary, the findings on phonological ambiguity imply that in the act of reading, full command of the alphabets of Serbo-Croatian does not entail two functionally independent symbol systems. There are experimental circumstances in which violations to alphabetic integrity have <u>no</u> detrimental effect. These include: 1) distortions of surface orthographic form in the case where unique characters from both alphabets are merged together in one letter string or 2) mixed contexts in which some words are printed in Roman and other words are printed in Cyrillic. In other cases, inability to differentiate between alphabets impairs performance. Skilled readers are not able to restrict themselves deliberately to the Roman alphabet when the alphabetic context of the experiment and/or the instructions to the subject would invite an exclusively Roman mode. Moreover, readers of Serbo-Croatian proceed in a phonologically analytic manner: The extent of the detriment produced by ambiguous letter strings depends on the number and distribution of characters that occur in both alphabets, provided that those characters engender a different phonemic interpretation in each. It is also the case, however, that command of two alphabetic symbol systems allows the skilled reader to designate which alphabetic interpretation to apply by scanning the entire letter string for a unique character, a process that occurs independently of performing a phonological analysis. That is, in a fully ambiguous bi-alphabetic context, skilled readers are not indifferent to components of orthographic structure: The presence of a unique character may constrain the reader by specifiying one particular alphabet. Collectively, the results of experiments on the two alphabetic systems of Serbo-Croatian suggest that skilled readers typically do not separate these two symbol systems: Command of the two symbol systems of Serbo-Croatian does not mean two autonomous alphabetic systems.

REFERENCE NOTES

1. Feldman, L. B. *The influence of lexical status on a phonological strategy for word recognition*. Unpublished manuscript, 1981.
2. Feldman, L. B. *Phonological bivalence and the beginning reader*. Manuscript in preparation, 1982.

REFERENCES

Feldman, L. B. Word recognition in Serbo-Croatian is primarily phonological. *Haskins Laboratories Status Report on Speech Research*, 1981, *SR-66*, 167-201.

Feldman, L. B., & Kostić, A. Word recognition with mixed-alphabet forms. *Haskins Laboratories Status Report on Speech Research*, 1981, *SR-66*, 203-211.

Feldman, L. B., Kostić, A., Lukatela, G., & Turvey, M. T. An evaluation of the "Basic Orthographic Syllabic Structure" in a phonologically shallow orthography. *Haskins Laboratories Status Report on Speech Research*, 1981, *SR-66*, 273-291.

Katz, L., & Feldman, L. B. Linguistic coding in word recognition: Comparisons between a deep and a shallow orthography. In C. Lesgold & C. Perfetti (Eds.), *Interactive processes in reading*. Hillsdale, N.J.: Erlbaum, 1981.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. Lexical decision in a phonologically shallow orthography. *Memory & Cognition*, 1980, *8*, 124-132.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. Bi-alphabetical lexical decision. *Language and Speech*, 1978, *21*, 142-165.

FOOTNOTES

[1]In the naming task, a correct reading of an ambiguous pseudoword permitted two options. In analyzing the pseudoword data, either interpretation was accepted. For the word data, there was only one correct interpretation.

[2]In this interpretation, I am assuming that there is no intrinsic difference between alphabet and that analogous results would be obtained in a Roman first, Cyrillic second environment. This outcome has not been tested, however,

ORTHOGRAPHIC AND PHONEMIC CODING FOR WORD IDENTIFICATION: EVIDENCE FROM HEBREW

Shlomo Bentin,+ Neta Bargai,+ Amiram Carmon,+ and Leonard Katz++

Abstract. In Hebrew script, vowels are represented by small dots that are added to the consonants. In most printed material the dots are omitted, so that the reader sees only consonant strings. Because several different words (with different vowel structures) can share the same consonant string, a unique pronunciation for such a string is determined by the syntactic and semantic contexts. The purpose of this study was to investigate the influence of this phonemically ambiguous script on the reader's use of phonemic information for printed word recognition. In the first experiment, subjects were asked to name, as fast as possible, isolated words presented as consonant strings without vowels. Naming was faster when a single lexically valid pronunciation was possible than when the stimulus could be pronounced in several ways. In contrast, in the second experiment, the same phonemic ambiguity did not interfere with lexical decision, suggesting that phonemic codes were not used for printed word recognition. This suggestion was further investigated in a subsequent lexical decision task in which all consonant strings (words and nonwords) were presented with the vowel dots. There were three groups of nonwords: (1) the nonwords were homophonic to real words but, because of one different consonant, looked different; (2) the nonwords were made up of the same consonants as real words (orthographically similar) but, because of different vowels, sounded different; (3) the nonwords were neither phonemically nor orthographically similar to real words. Response time was fastest for the totally dissimilar nonwords and longest for the orthographically similar nonwords. Presumably, graphemic information provided by the print was more important than phonemic information in partially activating real word lexical entries and, thereby, slowing rejection of the orthographically similar nonwords. In contrast, those real words that had been primed by phonemically or orthographically similar nonwords were facilitated equally by both. This equality suggests that the priming effect had been mediated by those same real words that had been activated in the lexicon by the similar nonword primes. Several implications for models of printed word recognition are discussed.

+Aranne Laboratory of Human Psychology, Hadassah Hospital, and Department of Psychology, Hebrew University, Jerusalem.
++Also University of Connecticut.
Acknowledgment. This work was supported in part by NICHD Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

The present study was concerned with the process of printed word recognition and with the way in which print is related to the representation of words in the internal lexicon. A close relationship should exist between the nature of the phonological information provided by an orthography and the way the print maps onto the internal lexicon. For example, the Serbo-Croatian spelling system keeps an isomorphic relationship between letters and phonemes; letter-to-phoneme translation is therefore straightforward and requires minimal contextual linguistic information. It might seem reasonable to suggest, therefore, that phonemic codes mediate between print and the lexical item it represents. On the other hand, English spelling most often represents the morphophonemic level rather than the phonemic; the invariance in meaning between words is represented by an invariant spelling in spite of changes in phonemics (as in "heal-health" and "decagram-decimal"). This makes the rules for letter-to-phoneme translation more complex and indirect, suggesting that phonemic codes may be less often used by the skilled English reader. It seems plausible that skilled reading in English and Serbo-Croatian are efficient processes because the behavior is a well-exercised one. However, what is efficient for one orthography is not necessarily efficient for the other.

Differences in the reading process between Serbo-Croatian and English may be particularly strong in the subprocess that is involved in word identification, because it is here that the two orthographies differ most. Word identification is most often studied in the laboratory by means of the lexical decision task. It has been suggested that the major factor determining a skilled reader's use of phonemic recoding in making a lexical decision is the directness with which the reader's orthography maps onto the phonemic space of his/her language (Feldman & Turvey, in press; Katz & Feldman, 1982). Indeed, the evidence presented by Feldman and Turvey (in press), strongly supports the notion that printed word identification in Serbo-Croatian depends heavily on a phonemically derived code, while in English, most evidence presented so far suggests that phonemic codes are less often used (Coltheart, Davelaar, Jonasson, & Besner, 1977; Forster & Chambers, 1973; Frederiksen & Kroll, 1976). Katz and Feldman (1983) support this suggestion with data that directly compare Serbo-Croatian and English readers.

The present study extends the consideration of the relation between orthography and the process of printed word identification to Hebrew. The Hebrew orthography offers a unique opportunity for studying a reader's dependence on phonemic codes, because it allows manipulation of the phonological information carried by a single string of letters. Hebrew has an unusual system for representing vowels in print: small graphic symbols (dots) that are appended to the consonants, but cannot stand by themselves. The full writing system (consonants and dots) is initially taught in the first grades of elementary school, but the adult reader sees it only infrequently outside of prayer books and poetry. In all other printed material the vowel dots are omitted. This produces a situation where many (but not all) Hebrew words with the same sequence of consonant characters can be pronounced in several ways, each one a different legal Hebrew word (Figure 1). In order to pronounce the word, the reader must assign one of these alternatives to the character string on the basis of the context.

The Hebrew orthography can be considered, therefore, to represent phonemic information even more indirectly than English. While, in English, vowel symbols are always present but may represent alternative phonemic representations, such vowel symbols are totally absent in normally printed Hebrew, and the phonemic representation of a string of letters becomes correspondingly more ambiguous. Importantly, the missing information is vowel information, so that no articulation of the remaining consonants is specified in the print; only abstract consonantal phonemic information remains. Given this lack of specificity in the phonemic realization of the word, it would seem to be likely that printed words in Hebrew map directly to more abstract morphophonological representations.

Examples of single and multiple pronuncable Hebrew consonant strings

The word as seen in print ספר

The different pronunciations (with vowel dots)

Hebrew words	סֵפֶר	סַפָּר	סִפֵּר	סָפַר	סְפֹר	סֻפַּר	סַפֵּר
Phonetic representation	sefer	sapar	siper	safar	spor	supar	saper
English translation	book	barber	(he) told (he) cuts	he counted	count	was told was cut	tell

The word as seen in print כסף

The single pronounciation kesef כֶּסֶף

English translation money

Figure 1. Examples of single and multiple pronounciation Hebrew words.

The present study was designed to test this hypothesis, that is, to determine the extent to which phonemic information is relied on for word identification in Hebrew. If the relation between the directness of an orthography and phonemic coding that we have described above is true, then we should find little dependence on phonemic coding. Lexical decision in Hebrew should be less dependent on phonemic translation of the print. Nevertheless, the suggestion has been made (Navon & Shimron, 1981) that the skilled Hebrew reader uses phonemic information in general and vowel information in particular in accessing the mental lexicon, that is, that printed word identification depends on a phonemic code. Navon and Shimron base this proposal on experiments in which subjects named Hebrew words that had been printed either with or without vowels. Naming was found to be faster for words with vowels. Furthermore, substituting a graphemically different but allophonically identical vowel for the correctly spelled one did not slow the response, that is, graphemic dissimilarity did not disrupt naming. But the authors' proposal that lexical access is dependent on a phonemic code was an extrapolation from their naming experiments; no lexical decision experiments had been run. In naming, both prelexical and postlexical factors influence the performance. On the other hand, because naming necessarily involves the use of phonemic codes, it is a task in which their possible effect on performance can be investigated. The experiments reported here use both naming and lexical decision paradigms in a complementing manner to study the use of phonemic coding of print.

In our first two experiments, subjects were presented with strings of consonants without the vowel dots. Response times to two types of strings were compared: strings that represent one and only one word uniquely (single-word strings) and strings that represent more than one word depending on the vowels (multiple-word strings) (Figure 1 gives an example of each type). As in the example, each multiple-word consonant string represents several real words, each of which would display a different set of vowels if the vowels were printed. Thus, multiple-word letter strings are phonemically and morphophonologically more ambiguous than those strings that can be related to only one lexically valid phonemic representation. An initial experiment was required in order to demonstrate that, in performing a task in which phonemic codes are used, multiple pronunciations interfere with the response. A word naming task was used for this purpose. Although a naming response can, in theory, be generated lexically, without a letter-level grapheme-to-phoneme process, it appears that the phonemic code is, in fact, characteristically used for naming printed words (Navon & Shimron, 1981). A phonemic ambiguity effect was, in fact, obtained in our naming experiment; the same stimuli were then used to assess the use of a phonemic code in lexical decision. If indeed a complete phonemic code (consonants plus vowels) is necessary for a lexical decision, response time should be delayed for multiple word strings relative to uniquely pronounceable letter strings. On the other hand, if no retardation is found, it could be because no phonemic analysis occurred, or only a partial analysis occurred that took only consonants into account. The process of word recognition was further investigated in a third experiment in which all stimuli were presented with vowel dots so that each had a unique pronunciation. The use of phonemic coding was assessed by comparing the response times to nonwords that were either phonemically or orthographically similar to real words, and to the real words that had been primed by these similar nonwords. It was expected that phonemic similarity would be less effective than orthographic similarity.

EXPERIMENT 1

Before multiple-word and single-word consonant strings could be compared in a lexical decision paradigm, we had to establish the validity of the manipulation. That is, we had to determine first that multiple-word strings were in fact more ambiguous than single-word strings when a complete phonemic code had to be utilized by the subject. Therefore, a naming paradigm was chosen; the requirement to pronounce the stimulus consonant string ensured that the correct vowels as well as the consonants would be coded at some point in the process. If multiple-word strings failed to be pronounced more slowly than single-word strings, the same comparison would be of no value in a lexical decision paradigm. On the other hand, a positive result would allow further exploration of this ambiguity effect.

Method

Subjects. Eight male and eight female undergraduate students participated as part of the requirements of an introductory psychology course. They were all native speakers of Hebrew with normal or corrected-to-normal vision, and were naive with regard to the experimental hypothesis.

Stimuli and apparatus. Three hundred words, printed as consonant strings without vowels, were presented to 15 judges who classified each as high, medium, or low frequency. All words consisted of three letters and were two syllables in length. Since some of the characters in Hebrew may be given a vowel sound in addition to their customary consonant reading, only words that are spelled with pure consonants were selected. Those words that were classified by at least 13 of the 15 judges in one of the two extreme frequency groups were considered for inclusion in the set of experimental words. From each of the two frequency groups, 12 nouns with only one legal pronunciation each and 12 words with at least three legal pronunciations each (one of which was a noun) were selected, making a total of 48 stimuli in all.

All of the stimuli were generated by a computer to appear in the center of a cathode ray tube. The size of each letter was 1 cm x 1 cm and the length of the whole word was 5 cm, subtending a visual angle of approximately 4.1 degrees.

The subject's verbal response was recorded by a Mura DX-118 microphone, which was connected to a voice key. The reaction time was measured by the computer from stimulus onset.

Procedure. The experiment took place in a semi-darkened soundproof room. Subjects sat approximately 70 cm from the screen. They were instructed to name, as fast as possible, individual words that appeared on the screen at a rate of one every two seconds. Stimulus duration was terminated by the subject's response. (There were no failures to respond within two seconds.) The verbal response given by the subject was recorded by the experimenter in order to detect reading errors and pronunciation preferences, if any. All 48 words were presented in one session that was preceded by 5 training trials.

Results

Reaction times were averaged for each subject over the 12 words in each combination of frequency (high/low) and number of pronunciations (single/multiple). The reliability of these means was assessed by calculating a coefficient of variation (the ratio of standard deviation over mean). All the coefficients were lower than 0.2, suggesting that the means were reliable estimates for the individual distributions.

Inspection of Figure 2 suggests that there were effects of both frequency and phonemic ambiguity. This was supported by an analysis of variance that revealed that both the frequency and phonemic ambiguity factors were significant: Response times to high frequency words were faster than to low frequency words, $F(1, 15) = 48.99$, MSe = 2543, $p < .001$. With both high and low frequency words, the response to strings that were phonemically ambiguous was delayed relative to those strings that had only one legal pronunciation, $F(1, 15) = 31.94$, MSe = 5728, $p < .001$. The interaction was not significant.

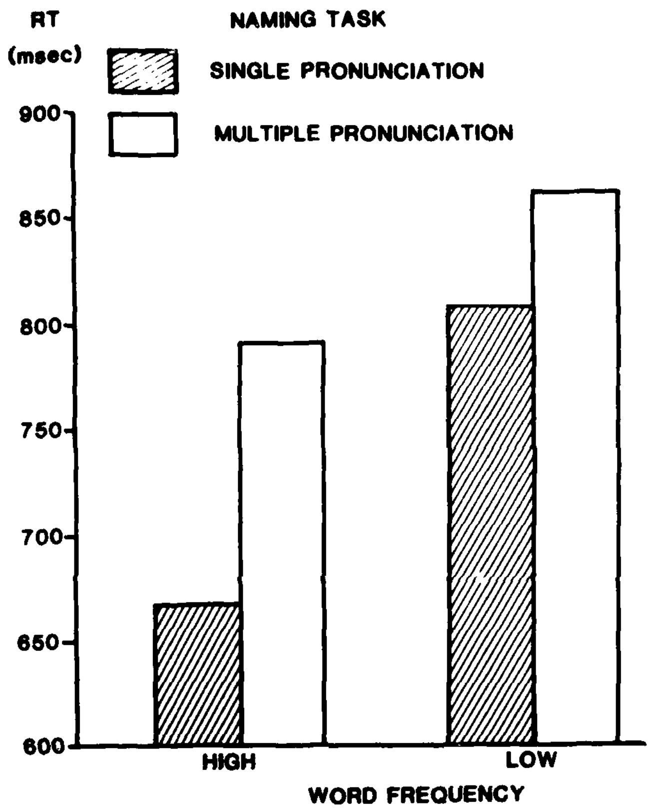


Figure 2. Naming time for single and multiple pronounciation, low and high frequency words.

Analyses of the specific pronunciations produced for multiple-word items by each subject showed that all words were given a legal pronunciation. However, there was variability in the specific word that subjects chose to assign to a given consonant string. For the set of 24 multiple-word items, the range of the number of subjects giving identical responses was 5 to 15 (out of a total of 16 subjects) with a median number of 7.

Discussion

Multiple-word consonant strings were named more slowly than single-word strings. It is clear, therefore, that in naming, subjects could not ignore the multiple phonemic (or semantic) representations of the ambiguous string.

However, the results were equivocal with regard to the locus of the effect; both prelexical and postlexical explanations remained viable for the naming task. Nevertheless, the outcome of this experiment placed constraints on the interpretations of possible outcomes for a lexical decision experiment. The absence of an ambiguity effect in a lexical decision experiment could only indicate that the source of the effect in Experiment 1 was postlexical in nature and that phonemic ambiguity (and, therefore, a phonemic code) has no effect on lexical access.

EXPERIMENT 2

Multiple-word and single-word consonant strings without vowels were presented in a lexical decision paradigm. If multiple-word strings are recognized by means of a phonemic code, then the ambiguity in the transform from print to phonemics should delay the decision to those strings relative to single-word strings. On the other hand, if no effect of ambiguity is found, this result, together with the outcome of Experiment 1, will suggest that a phonemic transform of print does not play an important role in word recognition in Hebrew.

Method

Subjects. Eight male and eight female undergraduate students participated as part of the requirements for an introductory psychology course. They were native Hebrew speakers and were about the same age as the subjects in Experiment 1.

Stimuli and apparatus. The same 48 words used for naming in Experiment 1 were used for lexical decisions in this experiment: 24 high frequency and 24 low frequency words. In each frequency group, half of the consonant strings could take only one legal pronunciation, while the others could be pronounced in at least three different ways. Forty-eight nonwords were added; they were formed by permuting the order of the consonants of the real words so that the result had no possible pronunciation that would form a legal word. Since the vowels were not printed, all the nonwords could be pronounced by arbitrarily assigning vowels. All 96 stimuli were presented with a different randomization for each subject.

Procedure. The conditions of Experiment 1 were repeated in this experiment. In addition, the subjects were instructed to press one of two alternative microswitch buttons, according to whether the stimulus on the screen was or was not a legal Hebrew word. The dominant hand was always used for "Yes" (i.e., "word") responses and the contralateral hand for the "No" responses.

Following the instructions, ten training trials (5 words and 5 nonwords) were presented. Then, 96 test trials were given in two blocks of 48 trials each. A ready signal preceded each block. The subject started the test stimulus sequence in each block by pressing a start button that cleared the screen. The interstimulus interval was 2 sec. The interblock time interval was between 3 and 5 minutes.

Results

The reaction times for correct "Yes" (i.e., "word") responses were averaged for each subject over the twelve words in each combination of high and low frequency and single and multiple pronunciation. These averages were tested for reliability by computing a coefficient of variation. All coefficients of variation were smaller than 0.2.

Responses to high frequency words were significantly faster than responses to low frequency words, $F(1, 15) = 57.21$, $MSe = 3171$, $p < .001$. In addition, a significant interaction was found between frequency and phonemic ambiguity, $F(1, 15) = 10.37$, $MSe = 1204$, $p < .001$. Examination of the means revealed an unexpected result. Although Fisher's protected t-tests indicated that reaction times to single-word and multiple-word stimulus strings were not different for high frequency words, there were differences for low frequency words. In contrast to the delayed response to multiple-word strings that was found in Experiment 1, the lexical decisions for low frequency, multiple-word strings were faster than for low frequency, single-word strings, $t(15) = 3.18$, $p < .01$. These results are presented in Figure 3.

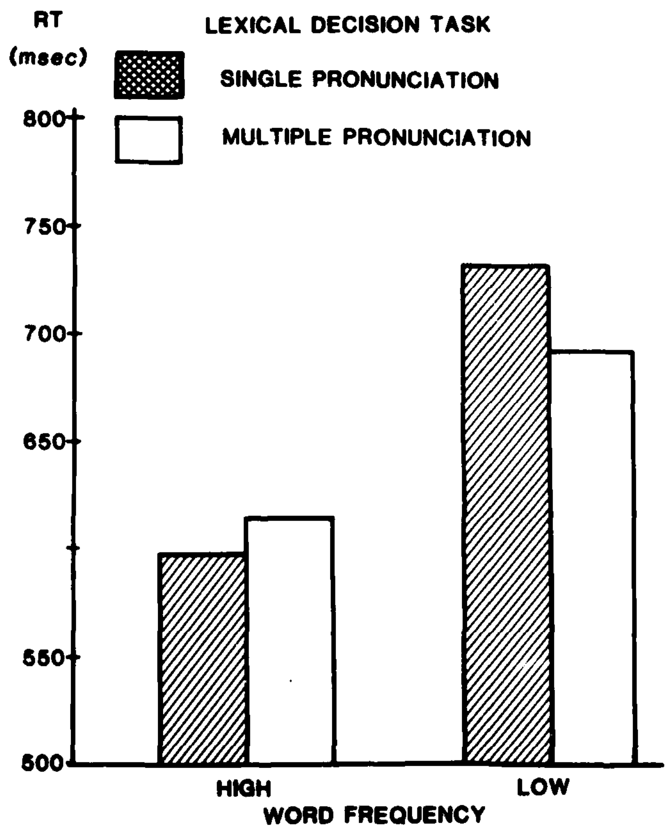


Figure 3. Lexical decision time for single and multiple pronounciation low and high frequency words.

Error percentages are presented in Table 1. An analysis of variance on the percentage of errors in each group revealed that there were significantly more errors for low frequency words than for high frequency words, $F(1,15) = 14.99$, $p < .001$. No other effects were found.

Comparison of the response times in Experiments 1 and 2 revealed that it took significantly longer to name the words than to recognize them in the lexical decision task, $t(28) = 3.11$, $p < 0.004$.

Table 1

Percentage of Incorrect Responses to Single and Multiple Pronunciation High and Low Frequency Words in a Lexical Decision Task.

	Pronunciation	
	Single	Multiple
Frequency		
High	7.29%	5.21%
Low	1.56%	1.56%

Discussion

In contrast to the effects found for naming, lexical decision time was not slower for multiple-word strings. On the contrary, multiple-word strings were recognized even faster than single-word strings, for low frequency words (but not for high frequency words). There are two alternative explanations for these results.

The first explanation is based on the assumption that in Hebrew the phonemic code plays only a minor role in lexical access. Consequently, phonemic ambiguity should have no effect on the response time when overt naming is not required. The delayed response for multiple word strings in naming would be, then, the result of a postlexical interference such as the requirement for response selection.

This hypothesis would predict no phonemic ambiguity effects for both high and low frequency words. However, if the frequency of the letter strings is considered (by a cumulative frequency of all the possible phonological realizations of the same consonant string), the response facilitation for low frequency multiple word strings might be the result of an artifact of the procedure used to select high and low frequency word stimuli for the experiment. Frequency was determined by means of ratings obtained from judges, but the judges may have systematically underestimated the true frequency of the multiple word consonant strings. This would have happened if the judges considered only one (probably the most frequent) meaning, of the several belonging to each string, ignoring the additional phonological realizations that were possible. Our introspections suggest that this certainly could have happened. The underestimation would affect the low frequency strings more, since the frequency added to a given string by each phonologic alternative is relatively higher. Thus, the apparent facilitation of the low frequency multiple-word strings would be accounted for as a simple frequency effect; the multiple-word strings we used may have been more frequent than the single word strings. Unfortunately there is no reliable source of word frequency data in Hebrew; therefore, this hypothesis could not be verified.

A *second way of accounting* for the absence of an interference effect due to phonemic ambiguity is based on the assumption that a multiple-word string activates its several different phonemic codes, which activate different entries in the lexicon simultaneously. The facilitation might be accounted for as an interaction among phonemic representations. Then, the interference effect in naming associated with phonemic ambiguity must be accounted for as the net result of a tradeoff between a process of rapid parallel lexical access and interference among the resultant phonemically coded words that compete for articulation. However, this hypothesis does not explain the interaction between the frequency and the number of phonemic realizations.

We favor the first explanation, in which a direct mapping of the print to abstract morphophonological representations is suggested. Support for this explanation is provided by other data that indicate that, when multiple phoneme codes are used for lexical access, the result is an inhibitition, *rather than a facilitation*, of word recognition. The data are from experiments in the Serbo-Croatian language. As we stated above, printed words in Serbo-Croatian have unique pronunciations. However, printed material can be produced in either of two different alphabets, the Cyrillic and the Roman. Although the two alphabets consist of distinct graphemes, for the most part, there are some graphemes common to both alphabets, and some of these have different pronunciations in the two alphabets. That is, there are some letters that look identical but sound different. A string that is made up of these phonemically ambiguous letters will have two pronunciations, one in each alphabet, either or both of which may be a real word. Both alphabets are taught to all children in elementary school and native speakers typically become facile at reading in either. Experiments by Feldman and Turvey (1982), and by Lukatela, Popadić, Ognejenović, and Turvey (1980) have demonstrated that subjects are slower in recognizing phonemically ambiguous words in lexical decision and naming tasks and that the inhibition is due to the ambiguity of the *phonemics* and *not* to the duality of meaning. In contrast, in English, it has been shown that multiple meanings speed lexical decisions rather than inhibit them (Forster & Bednall, 1973; Jastrzembski & Stanners, 1975). Therefore, in the present experiment, it seems unlikely that the phonemic ambiguity of the Hebrew multiple-word strings would be the source of facilitation in lexical decision, a result that would be inconsistent with the findings in Serbo-Croatian. Rather, consistent with the findings for English, the facilitating effect on multiple-word strings is more likely to be due to causes unrelated to phonemic coding.

EXPERIMENT 3

Although the evidence in Experiment 2 suggests that full phonemic coding does not precede lexical access, the results were not unequivocal. Therefore, a third experiment was run. A lexical decision priming paradigm was used in which all stimuli, both targets and primes, were printed with full notation, that is, including vowels. The critical target words were preceded by nonwords that were either orthographically similar to the target or were phonemically similar. The two members of an orthographically similar prime-target pair were spelled with identical consonants but with different vowels, so that the pronunciation of the prime resulted in a nonword. A phonemically similar pair contained members that were pronounced identically but were spelled differently, by using one different, but allophonic, consonant between the two strings. Examples are given in Figure 4.

Examples of phonetic and orthographic priming

FREQUENCY	TYPE OF PRIMING	PRIMES (nonwords)		TARGETS (words)		
		STIMULUS	PHONEME	STIMULUS	PHONEME	ENGLISH TRANSLATION
HIGH	ORTHO-GRAPHIC	אָבֵן	aven	אֶבֶן	even	stone
	PHONETIC	קֶסֶף	kesef	כֶּסֶף	kesef	money
LOW	ORTHO-GRAPHIC	נְקִיב	nekiv	נֶקֶב	nekev	hole
	PHONETIC	אָתָב	atav	אָטָב	atav	clothes pin

Figure 4. Examples of orthographically and phonemically similar nonwords.

The implications of this manipulation are straightforward: If one type of priming facilitates and the other does not, the dominant code type is the one that is important for word recognition.

A second effect is to be expected, one that does not involve priming but concerns nonwords alone. The critical (similar) nonwords, due to their construction, either look like real words (because of their consonant pattern) but do not sound like real words (because of their vowel pattern) or, conversely, sound like real words but do not look like them. Therefore, correct responses to these nonwords should be delayed if a search of the lexicon discovers real words that are similar. Again, the implication is clear: Either phonemically similar or orthographically similar nonwords will be slower, whichever is closer to the primary lexical code.

Method

Subjects. Eight male and eight female students who had not participated in any of the previous experiments took part in this experiment as a requirement of an introductory psychology course.

Stimuli and design. The stimuli were 48 words and 48 nonwords, all printed with the vowel dots, giving each stimulus a unique pronunciation. Twenty-four high frequency and 24 low frequency words were selected from the 300 three-consonant words as described in Experiment 1. Twelve out of the 24 words in each frequency group were selected to be targets to priming and preceded by a trial in which a nonword was presented. Twelve out of these 24 nonword primes (six for each word frequency group) were designed to produce a primarily phonemic facilitation in recognizing the following words by being identical homophones. The substitution of one letter with an allophone made them orthographically nonwords (Figure 4). The other 12 nonword primes (six in each word frequency category) had consonant strings identical to their following words, but different vowel dots made them sound like nonwords. They were expected to have a primarily orthographic priming effect (Figure 4). The other 24 words were not specifically primed.

The 24 nonwords that were not used for priming (nonsimilar nonwords), were strings of 3 consonant characters plus vowels that were obtained by recombining the consonant characters in the 24 unprimed words. Twelve additional nonwords were presented but were not considered for analysis: These nonwords were similar to words (six orthographically and six phonemically), but they were not followed by any real word counterparts. These 12 words were presented in order to discourage the subjects from predicting the occurrence of a word on the trial following a nonword that was similar to a real word. Different quasi-randomizations were used for each subject. The only constraint on the randomization was to keep together the pairs of priming nonwords and priming words. All stimuli were generated on a CRT in the same way as for Experiments 1 and 2.

Procedure. The procedure was similar to that followed in Experiment 2. The subjects were instructed to press the appropriate button as fast as possible. They were told that both the spelling and the sound of a stimulus counted for the decision. Ten training trials (5 words and 5 nonwords) preceded the first experimental trial block.

Results

Both reaction times and error percentages were averaged over the words within conditions for each subject. Errors were few (from zero to a maximum of three errors per condition). Analyses of variance on errors produced no significant results.

Inspection of Figure 5 shows that for reaction times, both graphemic and phonemic similarity interfered with correct nonword responses. However, graphemic similarity delayed the correct "No" responses significantly more than the phonemical similarity. This suggestion was supported by a one-way analysis of variance on the correct "NO" responses, which revealed that nonwords that were not similar to words were the easiest for the subject to reject as words. Response time was fastest for the dissimilar nonwords and longest for those that were similar graphemically, $F(2,30) = 9.87$, $MSe = 3421$, $p < 0.001$. Two-way analysis of variance on the correct responses for only those critical nonwords that were similar to words revealed that it took significantly longer to reject the nonwords that were graphemically similar to words than the nonwords whose similarity was mainly phonemic, $F(1,15) = 5.45$, $MSe = 16136$, $p < .04$. Also, it was found that nonwords that were similar to high frequency words were rejected faster than those nonwords that were similar to low frequency words, $F(1,15) = 11.44$, $MSe = 2632$, $p < .004$. There was no significant interaction.

Words that were preceded by similar nonwords were responded to faster than words that were preceded by unrelated nonwords or by unrelated words (Figure 6). However, the facilitation effect of both graphemic and phonemic similarity did not differ significantly. An analysis of Frequency (High/Low) by Priming (Primed/Unprimed) for reaction times on correct word responses revealed that primed words were, in fact, responded to faster than unprimed words, $F(1,15) = 58.7$, $MSe = 6057$, $p < .001$. Also the reaction times to high frequency words were faster than to low frequency words, $F(1,15) = 27.72$, $MSe = 5613$, $p < .001$. A second analysis of variance of Frequency (High/Low) by Priming Mode (Graphemic/Phonemic) on the reaction times to primed words revealed that even within the group of primed words, the high frequency words were responded to faster than the low frequency words, $F(1,15) = 8.06$, $MSe = 9630$, $p < .01$. The reaction times to the graphemically primed words appeared to be faster than to phonemically primed words, but this difference failed to reach statistical significance. Also, the Frequency and Priming Mode factors did not interact significantly.

The response time to the unprimed words and the nonsimilar nonwords in Experiment 3 was compared with the response time to words with single pronunciation and nonwords in Experiment 2. Two factors analysis of variance revealed that the response times were faster in Experiment 2, $F(1,30) = 41.95$; $MSe = 24292$, $p < 0.0001$. Also, the response time to words was faster than to nonwords in Experiment 3, but slower in Experiment 2. This interaction was supported by the analysis of variance, $F(1,30) = 11.07$, $MSe = 4430$, $p < 0.002$.

Discussion

Those nonwords that were misspelled but phonemically similar to words were rejected faster than those that were similar in print but differently pronounced. In addition, the responses to both of these groups of nonwords were delayed relative to responses to regular nonwords (i.e., nonwords that neither look nor sound like real words).

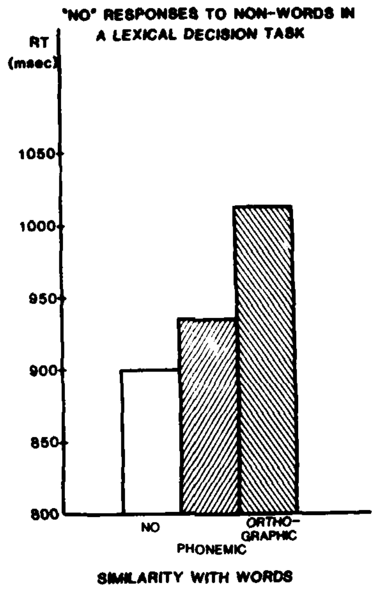


Figure 5. The reaction time to nonwords that were similar or nonsimilar to real words in a lexical decision task.

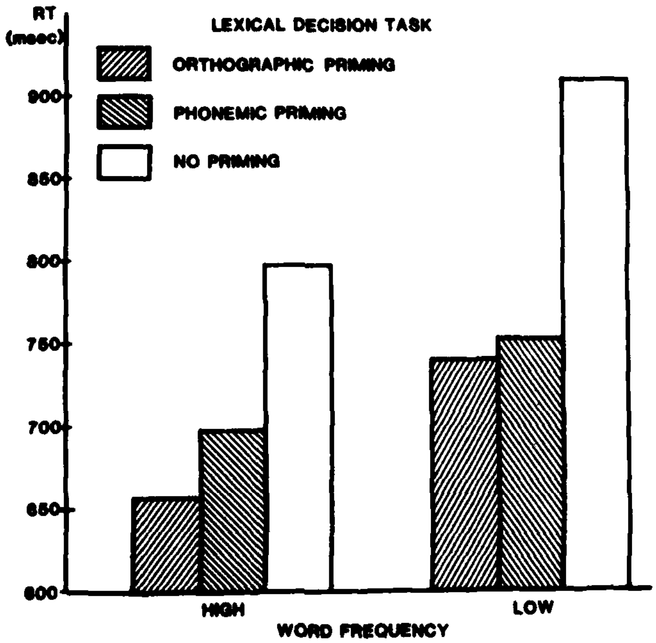


Figure 6. The reaction time to primed and unprimed words in a lexical decision task.

Other investigators have also demonstrated that certain classes of nonwords are harder to reject as real words. For example, it has been reported that nonlegal nonwords were responded to faster than legal nonwords (Stanners & Forbach, 1973). In a different study, Coltheart et al. (1977) assigned nonwords an index "N," where "N" was the number of different English words that could be produced by changing just one of the letters in the string to another letter, preserving letter positions. Nonwords with higher "N"s were responded to slower than nonwords with lower "N"s. These results suggest that the more similar a nonword is to a real word, the longer is the lexical decision time required to reject it. It seems, therefore, that in the present study the orthographically similar nonwords were associated with the real words more closely than were the phonemically similar nonwords. Of course, both groups of similar nonwords shared both phonemic and orthographic information with real words. It was reported, however, that the rejection of pseudohomophones is interfered with by their visual rather than phonemic similarity to words (Martin, 1982).

A correct "No" response to the orthographically similar nonwords must have been based on reading the vowel dots in addition to the consonants. In contrast, correct rejection of the phonemically similar nonwords could be made by considering only the consonantal letters alone. Since the adult Hebrew reader does not habitually read the vowels, it could be argued that this might, by itself, explain the precedence given to consonants and, thus, the difference observed between the two nonword categories. This explanation assumes that identification of printed words in Hebrew is primarly based on the consonant configuration that contains only partial information about a word's phonemics. Thus, this implication is in complete agreement with the hypothesis raised in this study, that the process of printed word recognition in Hebrew is based mainly on the orthographic information provided by the consonant letters.

The interference with correct "No" responses found in this study can be explained within the context of the logogen theory suggested initially by Morton (1969, 1970), and later expanded to explain nonword responses by Coltheart et al. (1977). According to this model, lexical memory includes a set of evidence-collecting devices--the logogens. These logogens serve as an interface between the sensory system and the cognitive lexical memory. Each word in memory has its own logogen. Logogens are activated by stimuli that are physically similar to the words to which the specific logogens are related. There is a positive correlation between the amount of similarity and the level of the logogen excitation. Logogens have thresholds that are inversely related to word frequency. Whenever a logogen is excited beyond its threshold, the access to the word in the cognitive lexicon is acheived and the "Yes" response is generated. However, if no logogen was excited beyond its threshold within a given time limit, a "No" response is generated. This time limit is dynamically adjusted up and down during processing. Stimuli that are similar to words represented in the lexicon tend to excite the logogen system more rapidly. As a consequence, the probability that the stimulus is indeed a word is high, and the time limit for a "No" response is increased. Within this conceptual frame, the nonwords in this experiment that were similar in print to real words would have excited the logogen system more rapidly, and to a greater extent than those whose similarity was mainly phonemical. We may conclude then that the orthographic analysis of the stimuli was completed first, while the phonemic analysis was only secondary.

Words are responded to faster if they are repeated within an experiment (Scarborough, Cortese, & Scarborough, 1977), or when preceded by semantic associates (Meyer, Schvaneveldt, & Ruddy, 1975). This effect is explained by the logogen theory as a "temporal summation" effect: When a logogen is fired, its threshold is reduced, and returns to baseline very slowly (Morton, 1979). Although not specified by Morton, this effect may not need to depend on above threshold preactivation of the logogen. Even limited arousal of a logogen might increase its baseline arousal level for a limited time period. Within this time period, less analysis would be required to fire this logogen, therefore faster response times would be measured (compare with the graded postsynaptic potentials and temporal summation of neurons). The priority of the letter analysis in the word identification process that was indicated by the correct "No" responses to nonwords suggests that real words that immediately follow orthographically similar nonwords should be responded to faster than those words that are preceded by the phonemically similar nonwords. However, the results failed to support this prediction. The facilitation effect of both the phonemically and the orthographically similar nonwords on the following real words was significant, but the amount of priming was not significantly different for the two conditions. One way to explain this incongruity between the similarity effects on "Yes" and "No" responses would be to assume that in the process of printed stimulus analysis, lexical activation of related items occurs. In this experiment, although the correct "No" response was generated by the logogen system in a nonword trial, the lexical memory could have been accessed either by a post decision analysis or through a verification process involved in the decision process itself (Becker, 1979; Becker & Killion, 1977). If the lexical entry of a real word that was suggested by the nonword was indeed accessed, the priming could be explained by a feedback from the cognitive system to the logogens in the same way this model would explain contextual priming effects (Besner & Swan, 1982). In this account the similarity of the nonwords would not have affected the thresholds of the real words directly, but rather, indirectly through an abstract, conceptual mediator, which once accessed, had lost the orthographic or phonemic specificity.

GENERAL DISCUSSION

The question investigated in this study was to what extent identification of printed words involves the use of phonemic codes on the letters. The results suggested that, in Hebrew, printed word recognition is not primarily mediated by a phonemic code. Phonemic ambiguity, which did interfere with the naming of words, did not interfere with their silent identification as words (i.e., in lexical decisions). Furthermore, subjects found it more difficult to reject a nonword that looked like a real word but sounded differently, than to reject a nonword that sounded like a real word but was orthographically different; orthographic information appeared to fit more closely to the code used by the reader for word identification than did phonemic information. The data suggest that, at least in Hebrew, a direct mapping exists from the print to a representation in the lexicon more abstract than the phoneme. These representations may be morphophonological in nature, consisting, for example, of the consonantal root from which the several inflectionally and derivationally related versions are eventually formed. However, there were only a few orthographically similar nonwords that were mistaken for words, indicating that phonemic information (as vowel information) must also have been used at some point. An alternative explanation is that the incorrect vowel dots altered the orthographic representation of the stimulus. This seems implausible, because a reader's lexical representations are unlikely to include orthographically represented vowels (Navon & Shimron, 1981). Therefore, the printed vowel information would almost certainly be used as cues for articulation by producing explicit phonemic rather than orthographic information. This phonemic encoding may have been used to disambiguate the orthographically similar nonwords; such a "verification" process is described below.

Several studies have suggested that the use of a phonemic code is optional and task dependent. Subjects will employ this strategy depending on the advantages and the disadvantages of its use in a particular task (Coltheart, 1978; Davelaar, Coltheart, Besner, & Jonasson, 1978; Stanovich & Bauer, 1978). Our results support this hypothesis. As a rule, the response times to comparable stimulus groups were longer in Experiment 3 where the vowel dots were added to the consonant strings, than in Experiment 2 where the vowel dots were not included. The response time to unprimed words in Experiment 3 was longer than the response time to the words in Experiment 2. Similarly, the response time to regular nonwords in Experiment 3 was longer than the response time to nonwords in Experiment 2. The presence of the additional phonemic information (i.e., inclusion of vowel dots) in Experiment 3 was not ignored by the subjects, who probably used it for further stimulus verification. The need for phonemic verification may have been increased in Experiment 3 by the presence of the orthographically similar words. In a previous study (Bentin & Carmon, Note 1), we have found that when words were presented with vowel dots, the nature of the nonwords determined the amount of phonemic verification. High and low frequency words with similar consonants were not responded to differently when the nonwords were meaningless permutations of the same letters. In contrast, the expected frequency effect was found when the nonwords were the same consonants with different vowel dots. We suggest that, in Hebrew, phonemic translation of the print is normally not necessary for word identification, and is employed only when the phonemic code is the single discriminative factor between words and nonwords.

The nature of the code used by subjects for word recognition does not depend only on the nature of the task. The complexity of the mapping rules from the orthographic to phonemic sets is probably a more basic and important factor. It has been demonstrated that in languages in which the mapping function is a simple isomorphism, such as in Serbo-Croatian, printed word recognition usually includes letter to phoneme transformation (Feldman & Turvey, in press). The language factor probably explains also the longer response times found in this study for lexical decisions (in Experiment 2) relative to naming. Forster and Chambers (1973) reported longer response times for lexical decisions than for naming in English. This relationship was replicated in Serbo-Croatian, but not in English (Katz & Feldman, in press). In the latter study, it was reported that semantic priming facilitates lexical decisions in both languages, whereas naming is facilitated only in English. It was suggested that in the shallow orthography of Serbo-Croatian, naming might be a direct mapping of phonemic information extracted from the script, to the articulatory system. In Hebrew, in contrast, print does not normally provide sufficient phonemic information, and therefore, naming must be mediated by the internal lexicon. This additional step slows down naming relative to lexical decision.

The mediation of the internal lexicon probably explains the similar priming effects of the orthographically and phonemically similar nonwords. This mediation suggests that the lexicon had been accessed by the nonwords that were similar to words. Since correct "No" responses were given to those nonwords, this lexical access could have happened either before a final verification was performed, or following the correct "No" response. Both alternatives have interesting implications for models of word recognition and reading. Lexical access preceding final verification implies that lexical access does not automatically elicit a "Yes" response in a lexical decision task. On the other hand, access to the internal lexicon following the response would imply that, for the literate adult, strings of letters trigger an automatic process of word recognition that is terminated only when a complete exhaustive linguistic analysis is achieved. Further investigation is necessary to determine whether either of the two alternatives, or both, are valid.

REFERENCE NOTE

1. Bentin, S., & Carmon, A. Hebrew word recognition with and without vowels: A lexical decision task. Manuscript in preparation.

REFERENCES

Becker, C. A. Semantic context and frequency effects in visual word recognition. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 252-259.

Becker, C. A., & Killion, T. H. Interaction of visual and cognitive effects in word recognition. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 389-401.

Besner, D., & Swan, M. Models of lexical access in visual word recognition. Quarterly Journal of Psychology, 1982, 34, 313-325.

Coltheart, M. Lexical access in simple reading tasks. In G. Underwood (Ed.), Strategies in information processing. London: Academic Press, 1978.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. Access to the internal lexicon. In S. Dornic (Ed.), Attention and performance VI. Hillsdale, N.J.: Erlbaum, 1977.

Davelaar, C., Coltheart, M., Besner, D., & Jonasson, J. Phonological recoding and lexical access. Memory & Cognition, 1978, 6, 391-402.

Feldman, L. B., & Turvey, M. T. Visual word recognition in Serbo-Croatian is primarily phonological. Journal of Experimental Psychology: Human Perception and Performance, in press.

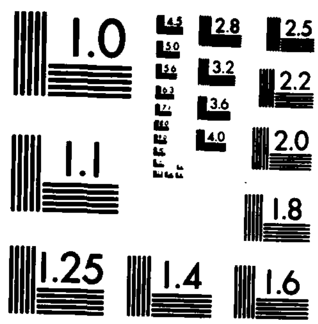
Forster, K. I., & Bednall, E. S. Terminating an exhaustive search in lexical access. Memory & Cognition, 1973, 4, 53-61.

Forster, K. L., & Chambers, S. M. Lexical access and naming time. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 627-635.

Frederiksen, J. R., & Kroll, J. F. Spelling and sound: Approaches to the internal lexicon. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 361-379.

AD-A129 713 STATUS REPORT ON SPEECH RESEARCH A REPORT ON THE STATUS AND PROGRESS OF S..(U) HASKINS LABS INC NEW HAVEN CT M STUDDERT-KENNEDY ET AL. 1983 SR-73(1983) 2/4

UNCLASSIFIED N00014-83-C-0083 F/G 6/16 NL



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

Jastrzembski, J., & Stanners, R. Multiple word meanings and lexical search speed. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 534-547.

Katz, L., & Feldman, L. B. Linguistic coding in word recognition: Comparisons between a deep and a shallow orthography. In A. Lesgold & C. A. Perfetti (Eds.), Interactive processes in reading. Hillsdale, N.J.: Erlbaum, 1981.

Katz, L., & Feldman, L. B. Relation between pronunciation and recognition of printed words in deep and shallow orthography. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1983, 9, 157-166.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. Lexical decision in a phonologically shallow orthography. Memory & Cognition, 1980, 8, 124-132.

Martin, C. R. The pseudohomophone effect: The role of visual similarity in nonword decisions. Quarterly Journal of Experimental Psychology, 1982, 34, 395-409.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Loci of contextual effects in visual word recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), Attention and performance V. London: Academic Press, 1975

Morton, J. Interaction of information in word recognition. Psychological Review, 1969, 76, 165-178.

Morton, J. A functional model for memory. In D. A. Norman (Ed.), Models of human memory. New York: Academic Press, 1970.

Morton, J. Facilitation in word recognition: Experiments causing change in the logogen model. In P. Kolers, M. Wrolstad, & H. Bouma (Eds.), Processing of visible language. New York: Plenum, 1979.

Navon, D., & Shimron, J. Does word naming involve grapheme-to-phoneme translation? Evidence from Hebrew. Journal of Verbal Learning and Verbal Behavior, 1981, 20, 97-109.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 1-17.

Stanners, R. F., & Forbach, G. B. Analysis of letter strings in word recognition. Journal of Experimental Psychology, 1973, 98, 31-35.

Stanovich, K., & Bauer, D. Experiments on the spelling-to-sound regularity effect in word recognition. Memory & Cognition, 1978, 6, 410-415.

STRESS AND VOWEL DURATION EFFECTS ON SYLLABLE RECOGNITION*

Charles W. Marshall+ and Patrick W. Nye

Abstract. Systems designed to recognize continuous speech must be able to adapt to many types of acoustic variation, including variations in stress. A speaker-dependent recognition study was conducted on a group of stressed and destressed syllables. These syllables, some containing the short vowel /ɪ/ and others the long vowel /æ/, were excised from continuous speech and transformed into arrays of cepstral coefficients at two levels of precision. From these data, four types of template dictionaries varying in size and stress composition were formed by a time-warping procedure. Recognition performance data were gathered from listeners and from a computer recognition algorithm that also employed warping. It was found that for a significant portion of the data base, stressed and destressed versions of the same syllable are sufficiently different from one another to justify the use of separate dictionary templates. Second, destressed syllables exhibit roughly the same acoustic variance as their stressed counterparts. Third, long vowels tend to be involved in proportionally fewer cross-vowel errors, but tend to diminish the warping algorithm's ability to discriminate consonantal information. Finally, the pattern of consonant errors that listeners make as a function of vowel length shows significant differences from that produced by the computer.

INTRODUCTION

To keep the analysis task within practical bounds, some form of segmentation of the acoustic signal into analyzable units is an intrinsic feature of

---

*A revised version of an oral paper presented at the 102nd Meeting of the Acoustical Society of America, 30 November - 4 December 1981, Miami Beach, Florida (Journal of the Acoustical Society of America, 1981, 70, Suppl. 1, S61).

+Present address: Department of Computer Science, Yale University, New Haven, Connecticut 06520.

Acknowledgment. This research was supported by NSF grant MCS 79-16177 and BRS grant RR-05596 to Haskins Laboratories. The authors wish to thank Franklin Cooper and Paul Mermelstein for much useful advice given during the planning of these experiments. We also wish to express our appreciation to Franklin Cooper, Louis Goldstein, Ignatius Mattingly, Paul Mermelstein, and Bruno Repp for the helpful comments they offered on various drafts of this paper.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

all current computer-based speech recognition methods. The choice of segments actually employed in recognition algorithms and in recognition studies has encompassed a wide variation in duration. This has ranged, for example, from centisecond units (Bahl, Baker, Cohen, Cole, Jelinek, Lewis, & Mercer, 1978) to phonemic segments (Klatt, 1978) to demisyllables (Dixon & Silverman, 1977; Rosenberg, Rabiner, Levinson & Wilpon, 1981) and beyond to syllables (Fujimura, 1975) and to words (Rabiner & Wilpon, 1979). Moreover, among these different choices, syllables and syllable-sized units have been lately receiving increasing attention.

There are several important features that qualify the syllable as a recognition unit. First, one must acknowledge the evidence that both speakers and listeners are aware of the existence of syllables and that they are usually in good agreement as to the number present in a given utterance. Second, syllables are the smallest units that can be uttered in isolation and for which, in many instances, it can be claimed that they are produced by completely executed articulatory gestures (roughly defined as maneuvers involving a single opening and closing of the vocal tract that in turn, cause transient increases in the acoustic energy contour). Third, further merit stems from the fact that, especially for closed syllables (CVCs), the coarticulation effects between the phones within the syllable can be assumed (on average) to be stronger than they are across syllable boundaries. Hence, in principle, the selection of the syllable as a recognition unit should present a simpler segmentation task because the boundaries are located in the less strongly coarticulated regions of the signal (Fujimura, 1975).[1] Fourth, syllables may also be said to hold a strong claim to being the authentic building blocks of speech because they constitute many common words in their entirety and can be combined in appropriate sequences to form all the multisyllabic words as well. And finally, syllables provide the basis for an important feature of word and sentence patterning whereby, through the exercise of selective syllable emphasis (stressing) and lack of emphasis (destressing), information about the syntactic structure and semantic content of a sentence is encoded in the acoustic signal.

However, variations in syllable stress bring about significant changes in the acoustic duration and spectral composition of most syllables. The magnitude of these changes can vary considerably with speaking rate, syntactic role and phonetic context. Thus, the effects of stress variation are an inherent feature of speech acoustics--a feature that must be accommodated by all recognition systems. Included among these systems are, of course, those that seek to identify linguistically relevant entities such as syllables, usually by matching acoustic segments to a dictionary of templates. Proposals for countering acoustic variation have generally taken one of two extreme positions, which can be referred to as collection versus computation. These positions hold that the template dictionary should either include (1) a collection of all the allophonic variants of each syllable to be recognized, or (2) only canonical, or stressed, examples from which all the expected variants are computed by an algorithm. The former approach carries the requirement of a large memory capacity, while the latter one promises a significantly lower memory cost that has to be traded against a somewhat increased computation cost and is consequently of practical as well as theoretical interest.

In this paper, we report on a preliminary investigation into the problem of linguistic variation and dictionary composition and describe data that have a bearing on the collection versus computation issue. Using selected sets of syllable-sized segments--some stressed and some destressed--taken from continuously spoken speech, we examined the recognition performance of a computer algorithm and compared it with that of human listeners. For computer recognition purposes, we used a syllable recognition algorithm prepared by Mermelstein (1978). Because it was expected that the severity of stress effects might vary as a function of phonological vowel length, two groups of syllables were employed, one incorporating the short vowel /ɪ/ and the other the long vowel /æ/. The study obtained empirical estimates of the error rates that occur during the recognition of stressed and destressed syllables (1) as a function of vowel length and (2) for dictionaries containing different combinations of stressed and destressed syllables. A study of the cluster structures produced by stressed and destressed syllables in a cepstral distance space was also undertaken.

METHOD

Selection of Syllables

Twenty-three pairs of vocabulary words were employed from a set of twenty-four pairs that had been originally selected. (The twenty-fourth pair was eliminated after a preliminary examination of the acoustic data.) Twelve pairs contained CVC syllables with an /ɪ/ vowel nucleus while the remainder contained similar syllables incorporating the vowel /æ/. One word of each pair (e.g., tidbit) contained the target syllable [tɪd] in stressed form while another word (e.g., wanted) contained its destressed counterpart. When choosing the words containing destressed examples of each syllable, a deliberate attempt was made to select only those in which, in the judgment of our linguist colleagues, the color of the nuclear vowels, when spoken by eastern American speakers, would not be likely to go to schwa when destressed.[2] Table 1 contains the vocabulary items that were included in a total of 58 sentences. The sentences were structured in such a way that the contrast between stressed and destressed syllables was retained and the placement of any of the vocabulary words in sentence-final position was carefully avoided.[3] For example, one of the sentences was "Old Bagdad on the Tigris offered an array of fantastic delights," which contained the syllables [dæd] and [fæn]. The sentences were composed in a variety of syntactic forms to induce the production of different speaking rhythms and to offset any reader tendency to adopt a sing-song or monotonous delivery. Each vocabulary word occupied at least two different contexts in the sentence set. However, four syllables were inadvertently included three times. They were the stressed syllables [læm], [tɪd] and [mæn] and the destressed syllable [dɪg].

Table 1

Syllables employed in recognition study.

Syllables containing /ɪ/		Syllables containing /æ/	
Stressed	Destressed	Stressed	Destressed
Rigmarole	Outrigger	Catalog	Catastrophic
Dignification	Indignation	Tactics	Tictactoe
Indigenous	Indigestion	Lambfaced	Lambaste
Filtrate	Infiltrate	Fatuous	Arafat
Simple	Simplicity	Tangent	Tangerine
Permissable	Premise	Fantail	Fantastic
Distant	Distinguish	Daddy	Bagdad
Tidbit	Wanted	Automatic	Automat
Litmus	Starlit	Hapless	Mishap
Bin	Coal-bin	Manic	Bagman
History	Historic	Bagman	Grab-bag
Sister	Catharsis		

Speaker Characteristics

Two male speakers (DZ and LL) were employed to allow speaker-dependent effects to emerge. Both were natives of the eastern United States, and had accents typical of that region. Each speaker read the list of sentences under instructions to imagine himself in circumstances in which each of the sentences might have been spoken and to reproduce them in an extemporaneous manner. During a preliminary examination of their speech data, it was found that one of the originally selected syllables failed to retain its vowel color when destressed and, therefore, it was eliminated from the study, leaving a total of 23 syllables. Four recording sessions were scheduled for each speaker at minimum intervals of about two weeks. Two recordings of the sentences were made at each recording session. Thus, the speakers provided eight different readings of each sentence and at least 16 examples of each syllable-pair (the four syllables noted above each yielded 24 examples). Therefore, in total, the data base contained 1,536 examples of the chosen syllables.

Parametric Conversion Procedures

After low-pass filtering at 4.9 kHz, the speech material was digitized at a 10 kHz rate and stored. A phonetician then isolated the target syllables by examining a display of the digitized waveform, adjusting a pair of cursors to mark the head and tail of each syllable at a zero crossing point in the waveform, and verifying the identity of the segment by listening to its output reproduced through a digital-to-analog converter and loudspeaker. The phonetician also made vowel duration measurements on a portion of the speech data from both speakers. Segmentation by visual inspection was preferred over automatic segmentation in order to keep the number of segmentation errors to an absolute minimum. Earlier work with an automatic segmentation algorithm (Mermelstein, 1975) has revealed the types of segmentation errors that automatic processing tends to introduce.[4]

Having isolated all of the syllables by hand, their sampled representations were converted into sequences of cepstral coefficient vectors at two levels of precision. For the first precision level (PL1) spectral values were obtained by FFT analysis of the digitized segments at a frame interval of 128 samples; for the second precision level (PL2) the interval was set at the higher resolution level of 64 samples per frame interval. In both cases, a frame consisted of 256 samples weighted by a Hamming window. Then, to shape the spectral energy content of the data so that it more closely resembled the frequency response of the human ear, the logarithms of the spectral amplitudes were weighted by a group of 20 triangular filters located at equal intervals along the mel-scale of frequency. This was done to gain the enhanced performance achieved previously with this transform (Davis, 1979; Davis & Mermelstein, 1980). Next, vector arrays of six cepstral coefficients were computed at PL1 and ten coefficients at PL2 for successive time-frame intervals (the gain-dependent zeroth coefficient was omitted from these arrays). Therefore, for any given syllable, the number of PL2 coefficients exceeded the number of PL1 coefficients by a factor of 3.3.

<u>Template Construction and Distance Measurement</u>

The procedure for creating syllable templates from the available tokens employed a dynamic programming algorithm described by Mermelstein (1976, 1978). This algorithm was based on principles employed in earlier work (Bridle & Brown, 1974; Itakura, 1975; Velichko & Zagoruyko, 1970), but differed from that work in some important details.

Each syllable was represented by a temporal sequence of mel-scale cepstral coefficient vectors. These vectors formed a matrix with the *n*th row representing the feature vector for the *n*th time frame. The non-linear warping consisted of selectively repeating or deleting rows in pairs of matrices.

Before warping any pair of syllables together to form a template, an initial optimum alignment was found by adding to each end of the shorter syllable an amount of silence equivalent to the difference in duration. Then this syllable, plus its silent attachments, was shifted with respect to the longer syllable until an interim minimum in the distance between the syllables (i.e., a minimum in the summed squares of the cepstral differences of corresponding time frames) had been found. At this point, the excess silence at the edges of the shorter syllable was pruned away so that the two matrices contained the same number of rows.

Following this length equalization and alignment, the non-linear warping algorithm was used to form the pattern of repetitions and deletions of rows dynamically from each matrix that gave the best match between them. The procedure involved the warping of both matrices onto a third time sequence (Sakoe & Chiba, 1978) and the derivation of a symmetric distance measure based on the sum of the squares of corresponding vector elements. The possible warps were constrained in such a way that the ends of the matrices always remained aligned together. Out of the warping procedure, the optimum path and its associated minimum distance were obtained. The optimum path was used to specify the corresponding time frames that were subsequently averaged together during template construction and, during recognition, the inverse of the computed distance was employed as a measure of the likelihood that a token represented the same syllable as a given template.

Having averaged two tokens together to form the first interim version of a template, this template was then warped together with a new token and the average of the resulting pair of matrices was computed by a procedure that weighted the matrix representing the interim template in proportion to the number of tokens it already contained. This process was repeated until the supply of tokens was exhausted--usually after the fourth or eighth warp.

The tokens used to construct templates were warped together in a fixed order but, to minimize possible order effects, four groups of dictionaries (one from each of the four speaking sessions) were formed and distance measurements were computed between each of these dictionaries and tokens drawn from one or more of the other three sessions. Thus, tokens to be recognized were never components of the template sets (dictionaries) against which they were matched; they were, however, drawn from the same words and sentence contexts as the templates, and they were spoken by the same speaker but at a different session. The pattern of comparisons is shown in Table 2.

Table 2

The speaking sessions that served as tokens and templates.

Run No.	Tokens		Templates
1	Session 1	tested against	Session 2
2	Session 1	tested against	Session 3
3	Session 2	tested against	Session 4
4	Session 2	tested against	Session 3
5	Session 3	tested against	Session 4
6	Session 4	tested against	Session 1

Composition of the Dictionaries

The four groups of syllable tokens produced by each of the two speakers (one group per speaking session) were converted into parametric form at both levels of precision. Following conversion, the tokens of each group were warped together by the dynamic programming technique (Mermelstein, 1978; Rabiner, Rosenberg, & Levinson, 1978) to give three classes of templates from which four dictionaries per speaker were derived (see the flowchart shown in Figure 1).

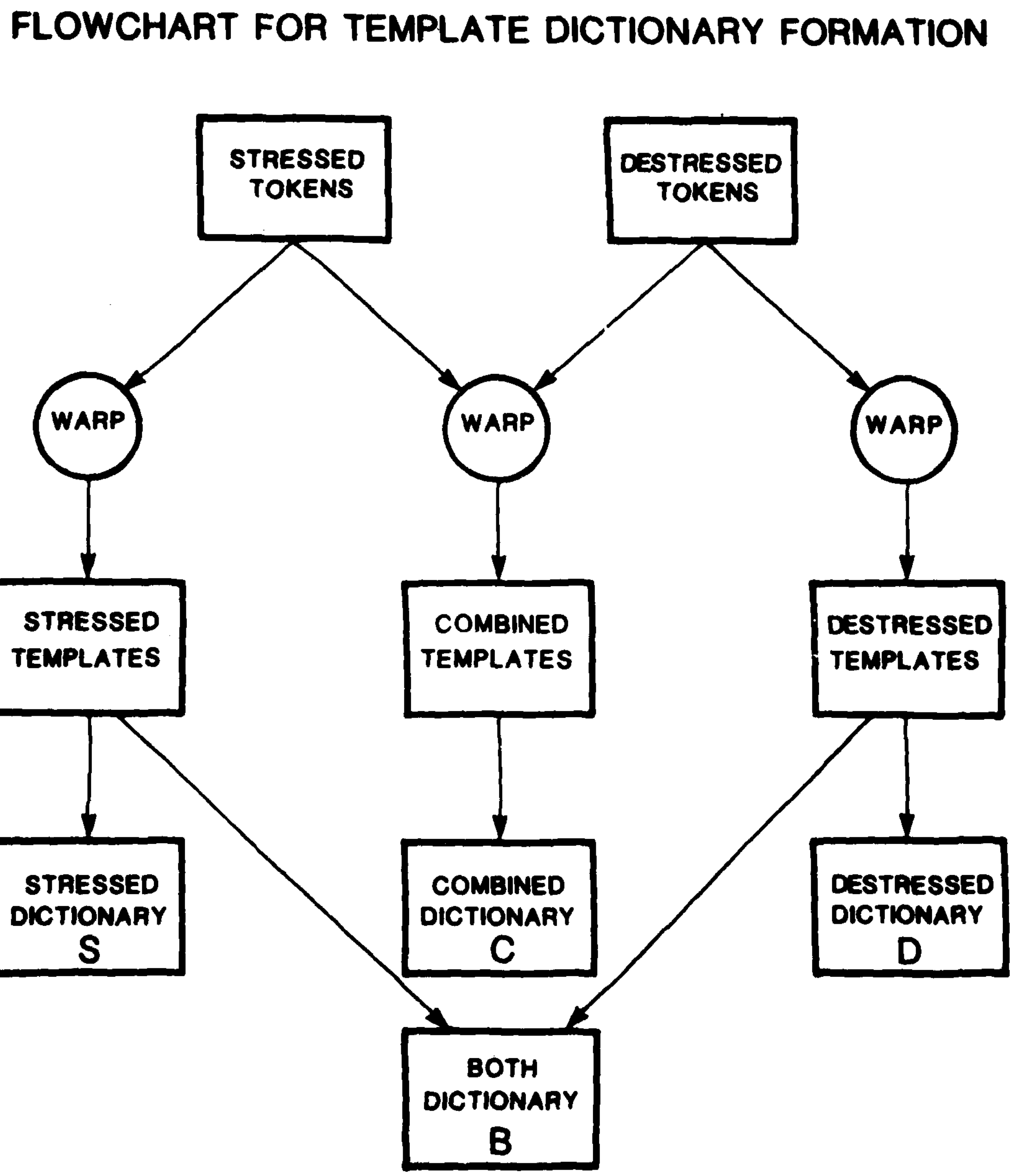


Figure 1. Flowchart illustrates the production of four types of dictionaries labeled B, C, S and D. For each such dictionary, the source data were stressed and destressed tokens extracted from a single speaking session.

The "stressed" (S) dictionaries contained templates formed by warping together only stressed tokens, while the "destressed" (D) dictionaries contained templates formed exclusively from destressed tokens. Consequently, each of these dictionaries contained 23 entries. The "combined" (C) dictionaries were formed by warping the stressed and destressed occurrences of each syllable token together and, therefore, also numbered 23 entries. The "both" (B) dictionaries contained the union of the stressed and destressed templates formed from a given speaking session (i.e., dictionaries S plus D); hence, they were twice the size of the other dictionaries and contained a total of 46 templates. As already noted, one dictionary was formed from each speaking session. Therefore, the total number of dictionaries produced amounted to 32 (four sessions x two speakers x four dictionary types).

During the recognition procedure, a warping was performed for each token with each of the templates in the appropriate dictionary (see Table 2) and the "recognized" syllable was identified as the top member of the list of hypothesized candidate syllables ranked in order of increasing token-template distance. These lists were employed in later studies that examined, in cases where the top candidate was in error, the frequency with which the correct choice appeared later in the list.

Collection of Data from Listeners

To establish a baseline from which to assess and, perhaps, to gain further insights into the performance of the computer recognition algorithm, a recognition test using the same isolated speech segments was presented to a group of 10 listeners. These listeners consisted of colleagues and their graduate students. All had taken part in many previous experiments of a similar nature and were fully familiar with the phonetic alphabet. They were given a list of the 23 syllables in phonetic transcription, informed that each presentation would be drawn from that list, and instructed to record each identification (or guess if necessary) by placing a check in a column below the appropriate entry in the list. The listeners were not asked to record stress levels. The syllables were delivered to the listeners at 5-second intervals via TDH-39 earphones from a tape recording of the computer output. Five seconds between each stimulus provided sufficient time for the listeners to make their responses. However, to ensure the detection and avoidance of missed responses, an 8-second interval was inserted after each group of five syllables and a 10-second interval after every twentieth syllable. The listeners heard (in random order) all of the target syllables produced by both speakers. Each one was repeated four times. Four of the syllables, as noted earlier, were inadvertently repeated six times. Hence, each subject heard 192 syllable-presentations from each speaker. The subjects' identification data were then entered into the computer and stimulus/response matrices for both the stressed and destressed syllables of each speaker were constructed.

RESULTS

Introduction

The results were examined from several points of view. To verify that our speech data did actually contain the expected durational variations, vowel duration and syllable duration measurements were examined. Then, the computer recognition errors were sorted and analyzed by precision level, vowel type, dictionary type, and stress. The data gathered from human listeners were, where possible, sorted and analyzed in similar fashion and compared with the computer results. Finally, the acoustic parameters were examined by means of a multi-dimensional scaling technique to reveal the clustering structures of stressed and destressed syllables.

Phonological vs. Physical Vowel Durations

Phoneticians have long believed that the vowel /æ/ has a longer duration in American English speech than the vowel /ɪ/. The classic experimental support for this assertion was provided by Peterson and Lehiste (1960), who showed that the intrinsic durations of /æ/ and /ɪ/ as syllabic nuclei in American English averaged 330 and 180 msec. However, they also observed that the length of a syllabic nucleus varied according to whether it was followed by a voiced or voiceless consonant. Since the final consonants of the CVC syllables employed in this study were drawn from both voiced and voiceless classes without regard to ensuring equal representation, it was necessary to verify empirically that a significant difference in duration was retained for the syllables we had chosen. To do this, it was deemed sufficient to perform vowel duration measurements on a representative portion of the data base and, for this purpose, data from one session by each speaker were selected. In contrast with the measurement procedure adopted by Peterson and Lehiste, which tended to include a large portion of the consonantal transition as a part of the vowel, the vowel durations measured in this study were confined to so-called steady-state regions of the syllables. These regions were defined as those portions of the syllables in which the cepstral frequencies did not deviate by more than 10 percent from their central values. Average overall durations of the syllables containing /æ/ and /ɪ/ were computed from the total numbers of samples stored per syllable.

The results of the vowel duration measurements are shown in Figure 2. The four distributions represent /æ/ stressed, /æ/ destressed, /ɪ/ stressed and /ɪ/ destressed. It can be seen that, on average, the durations obtained from speaker DZ were just a few percent shorter than those obtained from speaker LL. (The difference between the speakers in overall syllable duration was, however, considerably larger--about 35 percent.) The difference in median duration between stressed and destressed productions of the vowel /ɪ/ are shown in the figure to be 9 msec in the case of LL and 11 msec for DZ. Smaller reductions are apparent for the vowel /æ/. (A difference of the same sign was also evident in the overall syllable durations.) Thus, the syllables incorporating long vowels tended to retain the property of vowel length, while those incorporating short vowels were found to exhibit even further shortening in their destressed forms. In addition, it was found that destressing caused the consonantal regions of the syllables to be reduced in amplitude and overall spectral definition.

Overall Errors in Computer Syllable Recognition

The overall effects of stress on the performance of the recognition algorithm are best summarized in terms of the average error per syllable. Figure 3 shows the percentages of recognition errors made per syllable on the

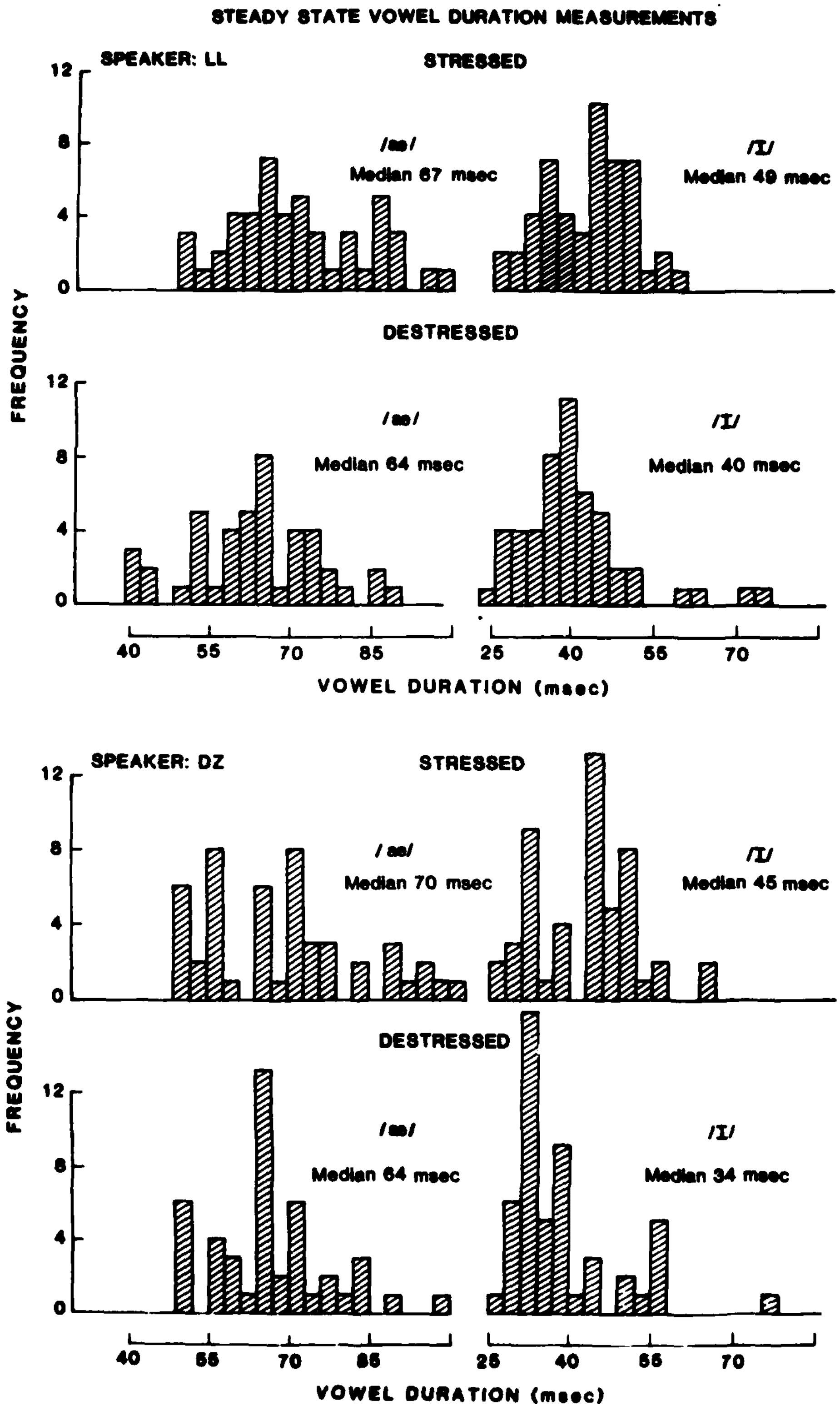


Figure 2. Frequency distributions of vowel durations measured for speakers DZ and LL from data collected at a single session. For speaker DZ, stress reduction results in a median reduction of 6 msec for /æ/ and 11 msec for /ɪ/. Corresponding reductions for speaker LL are 3 msec for /æ/ and 9 msec for /ɪ/.

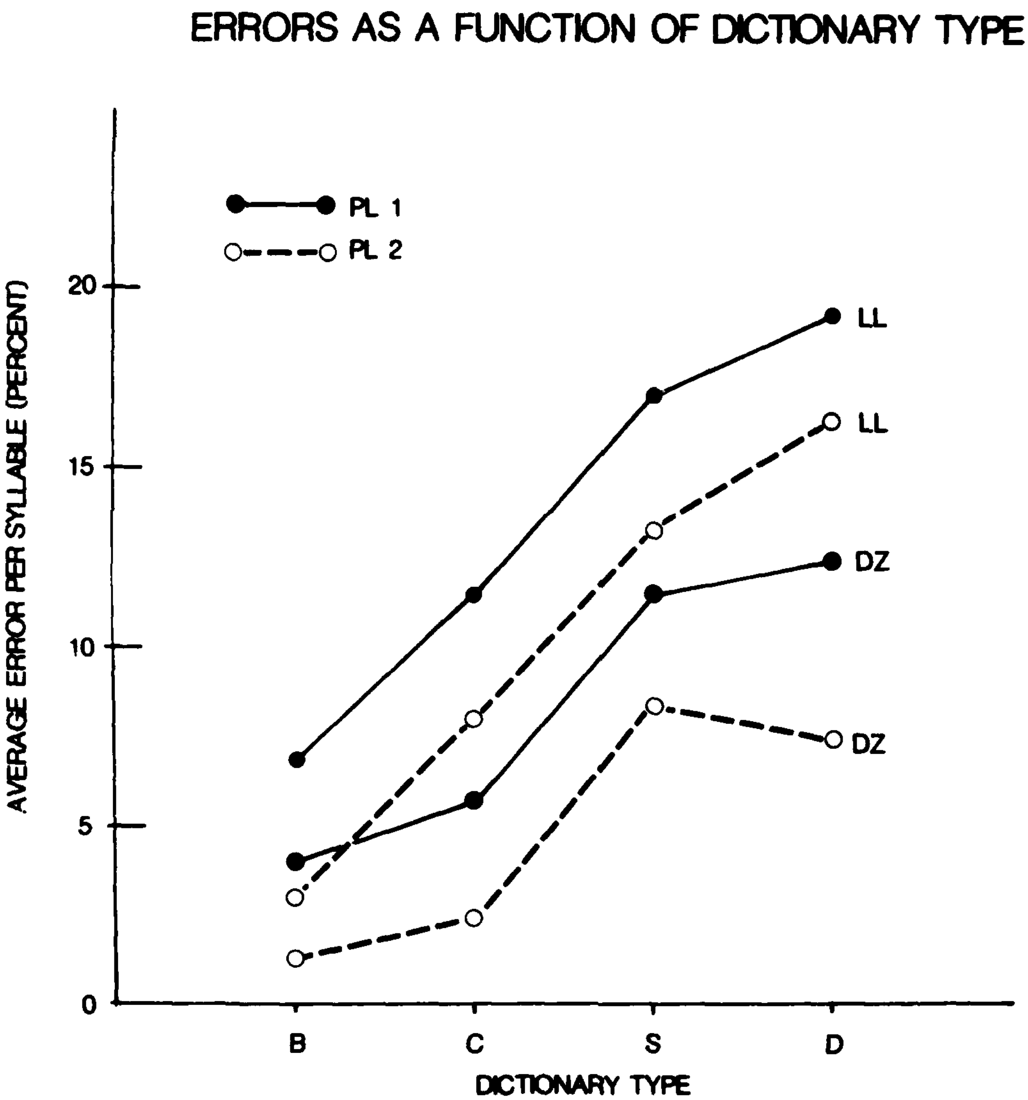


Figure 3. The average error per syllable plotted against dictionary type for two speakers (LL and DZ) and at two precision levels (PL1 and PL2). At PL1 spectral values were computed at a frame interval of 128 samples and at PL2 the frame interval was set at 64 samples. Window size remained fixed at 256 samples.

speech of LL and DZ as a function of the dictionary type and precision level. The data were obtained by averaging over six recognition runs. Each run was "open" and speaker-dependent and compared all 192 tokens from one session with each of the four dictionary types (containing twenty-three or forty-six templates). The syllables obtained from each recording session were employed once as the raw material for a group of dictionaries and one or more times as the unknowns (see Table 2). The error data for dictionary B in Figure 3 neglected errors in stress assignment.

The unknown tokens comprised equal numbers of stressed and destressed syllables whereas the dictionaries, except for B, contained only one template per syllable. Hence, recognition by the algorithm was considered correct when the syllable identity of the token (without regard to its stress) agreed with that of the template. Only for dictionary B was it possible to get separate estimates for errors of identity and of stress level. Confusion matrices for each of the individual recognition runs were formed and these were later summed together to create a single matrix from which were calculated the average error for each dictionary type, precision level, and speaker.

Four principal findings emerge from these data. The first is that the B dictionary gives the best overall performance. Second, the C dictionary is superior to both the S and D dictionaries. Third, the performance for the higher precision level (PL2) is significantly better than those for the lower precision level (PL1). Finally, all these features are apparent in the data of both speakers.

These results clearly show that the degree to which stress variation is included in syllable template formation is reflected in subsequent performance. For both speakers, the best recognition performance occurred when using the B dictionaries that contained both stressed and destressed templates and employed the higher precision spectral coefficients.

The next best performance emerged when the C dictionaries were used. Here the results show that, although occupying half of the storage space employed by the B dictionaries and the same space as the S and D dictionaries, the C dictionaries sucessfully embodied a high proportion of the variation due to stress--sufficient indeed to outperform the S and D dictionaries easily. Moreover, since the average error rate obtained with the C dictionaries was less than twice that of the B dictionaries, this suggests that, in principle, it should be possible to replace the least reliable C templates by separate stressed and destressed templates. This procedure would thereby create hybrid dictionaries that perform as well as B dictionaries but occupy less storage space than B dictionaries demand.

Figure 3 also shows a systematic speaker difference, with the speech of DZ yielding lower error rates than the speech of LL under the same conditions. This difference is comparable to the difference introduced by variations in dictionary type and is larger than the difference brought about by a change in precision level. It is of interest to note that the same speakers were employed in an earlier study that compared the effects on recognition performance arising from the use of different types of acoustic coefficients (Davis & Mermelstein, 1980). In that study, a similar speaker difference was found with each type of coefficient.

Furthermore, Figure 3 indicates that between dictionaries B and C and for a given error rate, there exists the opportunity to trade dictionary type (structure) against coefficient resolution. However, since computational complexity varies as the square of the number of coefficients involved, it is apparent that if the coefficient resolution were doubled for dictionary C, twice as many computational operations would be necessary to recognize a token using C as would be necessary to perform a recognition using dictionary B following a doubling of the number of templates in the dictionary. Hence, a greater increase in recognition accuracy per datum (bit) can be achieved by carefully increasing the number of templates than by using a larger number of higher-resolution coefficients per template. Also, once a lower bound has been reached for errors through improvements achieved by increasing coefficient resolution, it is apparent that further improvements may still be achieved by increasing the number of allophonic variants represented in template form to a point where a balance is found between the benefits of error reduction and an increasing computational cost.

Errors Classified by Vowel Identity

The computer recognition errors classified as a function of dictionary type and vowel identity are shown in the upper half of Table 3. In all four types of dictionary, more recognition errors occurred between syllable-tokens and templates incorporating the same vowel nucleus than occurred between syllables having different vowel nuclei. Moreover, a larger number of syllable identity errors was associated with the longer of the two vowels. This evidence strongly suggests that the errors arose because the vowel /æ/, constituting a substantial portion of the syllable, made a larger contribution to the distance measurement than did the flanking consonants. In other words, the presence of long vowels tended to "dilute" the consonant discriminability.

Table 3 also shows that if the cross-vowel errors involving /æ/ are expressed as a proportion (P%) of all errors involving /æ/, this proportion is smaller than the corresponding proportion for the vowel /ɪ/. This is true for both subjects and all dictionaries with the exception of B where, against the background of a small total number of cross-vowel errors involving /ɪ/, the proportions (P%) exhibit the opposite relationship because this total is exceeded by an isolated set of confusions peculiar to the speech of LL. Thus, taken as a whole, the number of errors involving long vowels tends not to include a substantial proportion of cross-vowel errors. Since long vowels constitute a prominent proportion of the syllables they occupy, they offer more information about their spectral structure and, hence, provide greater inherent protection against cross-vowel error.

Finally, Table 3 prompts the observation that if cross-vowel errors from dictionaries B, C, and S only are considered in that order, the number of those errors involving the vowels /æ/ and /ɪ/ increases at a roughly equal rate despite the differences in vowel duration. The major reason for this result probably stems from the properties of the dynamic warping algorithm whose nonlinear adjustment of the time axis has a tendency to provide some compensation for differences in vowel duration.

Table 3

Syllable errors classified by dictionary type and vowel.

Recognition by Computer

(summed over speaker, stress and precision level)

	Dictionary B			Dictionary C			Dictionary S			Dictionary D		
	/æ/	/ɪ/	P%	/æ/	/ɪ/	P%	/æ/	/ɪ/	P%	/æ/	/ɪ/	P%
/æ/	86	14	14.0	170	21	11.0	295	33	10.1	308	53	14.7
/ɪ/	9	67	11.8	19	100	16.0	40	220	15.3	122	166	42.7
	Total 176			Total 310			Total 588			Total 649		

Recognition by Listeners

(summed over speaker)

	Stressed			Destressed			Totals		
	/æ/	/ɪ/	P%	/æ/	/ɪ/	P%	/æ/	/ɪ/	P%
/æ/	42	6	12.5	50	5	9.1	92	11	10.7
/ɪ/	8	15	34.8	21	136	13.3	29	151	16.1
	Total 71			Total 212			Total 283		

Key: Symbols /æ/ and /ɪ/ at the left of the table refer to vowel nuclei of misidentified syllable tokens while the same symbols located at column heads refer to the nuclei of syllable templates that were mistakenly selected.

P% refers to the proportion of cross-vowel errors expressed as a percentage of all errors involving that vowel.

Comparison with Human Listeners

The lower section of Table 3 shows the listeners' data classified by vowel and stress level. An examination of the cross-vowel errors shows agreement with the bulk of the computer error data (upper section) inasmuch as the largest proportion of the human errors also involved /ɪ/ as compared with /æ/. The result suggests, of course, that the listeners were also able to make good use of the greater amount of vowel information available in the stimuli containing long vowels. The closest agreement with the listeners' overall performance is offered by dictionary C; here, both the proportion of cross-vowel errors (P%) and the total number of errors are of similar magnitude (listeners, 283; dictionary C, 310). However, the listeners' data differ from the computer results by posting a higher total of errors involving the vowel /ɪ/ (i.e., listeners, 180 vs. dictionary C, 119). Hence, the data provide evidence that the listeners' abilities to recognize the consonants of a syllable were not impaired by the presence of a long vowel and suggest that the recognition processes in the two cases are quite different. This conclusion is further supported by a comparison of listener and computer data in respect to the ten most frequently-made consonant errors. These data reveal that virtually no consonant confusions were shared in common. Furthermore, a classification of these errors in terms of voicing, manner, and place of articulation (occurring either alone or in combination) showed no systematic differences--they appeared in both groups of data with roughly equal frequency.

Further results from the listening experiment are given in Table 4, classified by speaker. The table shows that the syllables produced by LL were more accurately recognized by listeners than those produced by DZ--a result that is again at variance with that obtained by computer. In addition, for both speakers, and contrary to our expectations, the error percentages indicate that the overall human recognition performance was somewhat worse than the best computer performance (i.e., at PL2).

Table 4

Syllable identification errors classified by speaker.

Comparison of Listener Recognition and Computer Recognition

Recognition Method	Speaker	Percent Error
Listening	DZ	10.0
Listening	LL	7.5
Computer	DZ	1.4
Computer	LL	3.0

Computer data obtained using parameters at PL2 and dictionary B.

Errors Classified by Stress

A more revealing comparison of the listeners' recognition results with the computer results, and of the effects of dictionary type on computer performance, can be obtained if the errors are separately calculated for stressed and destressed tokens. Turning first to the computer data, Figure 4 indicates that the difference between stressed and destressed error rates was smallest when the B and C dictionaries were in use--notwithstanding the relatively larger difference that emerged from the speech of LL. A comparison of the listeners' recognition data with the computer data also reveals some marked speaker-dependent effects. While the listeners' error rate for stressed-token recognition of LL's speech is closely comparable to the error rate turned in by the computer, their corresponding error rate on DZ's stressed speech shows a three-fold increase over the computer error rate. A reason for this difference was revealed by a detailed examination of the listeners' errors on stressed tokens. This showed that 38 percent of the errors could be accounted for by two confusions, namely, those between DZ's articulation of [mæn] versus [mæt] and [hɪs] versus [dɪs]. In the destressed syllable data, however, no similar pair of confusions accounted for a comparably large proportion of the errors and the listeners' overall error rate consistently exceeded that delivered by the computer. Thus, in summary, there was evidence that on the stressed tokens, the listeners tended to perform only slightly worse than the computer, while on destressed tokens their performance was considerably below the computer using dictionary B.

A review of the composition of the four dictionaries can assist in explaining a substantial proportion of the error-rate differences appearing in Figure 4. In the case of the B and C dictionaries, the computer error rates for stressed and destressed tokens differed from one another by small amounts relative to the corresponding differences for dictionaries S and D, with the B dictionary evidencing a lower error rate on both stress types. Since only the B and C dictionaries contained both stressed and destressed information, their overall superiority was certainly to be expected. Meanwhile, using the S dictionary, the error rate for stressed tokens emerged as being nearly identical with that obtained when using the B dictionary. Destressed tokens, on the other hand, fared about four times worse when using dictionary S than when using dictionary B, a direct consequence of the lack of destressed information in S dictionaries. Conversely, when dictionary D was in use, errors involving destressed tokens occurred at roughly the same frequency as they did when using dictionary B, while the stressed tokens submitted to dictionary D yielded, as expected, an extremely high error rate.

The foregoing analysis ignored stress assignment as long as a syllable's identity was found correctly. Dictionary B provides the only opportunity to analyze stress-only-errors and Table 5 presents these data. The results show that, summed across both speakers and precision levels, errors in stress assignment occurred with 3.7 times greater frequency than did errors in syllable identity (cf., column B=649 and column C=176).

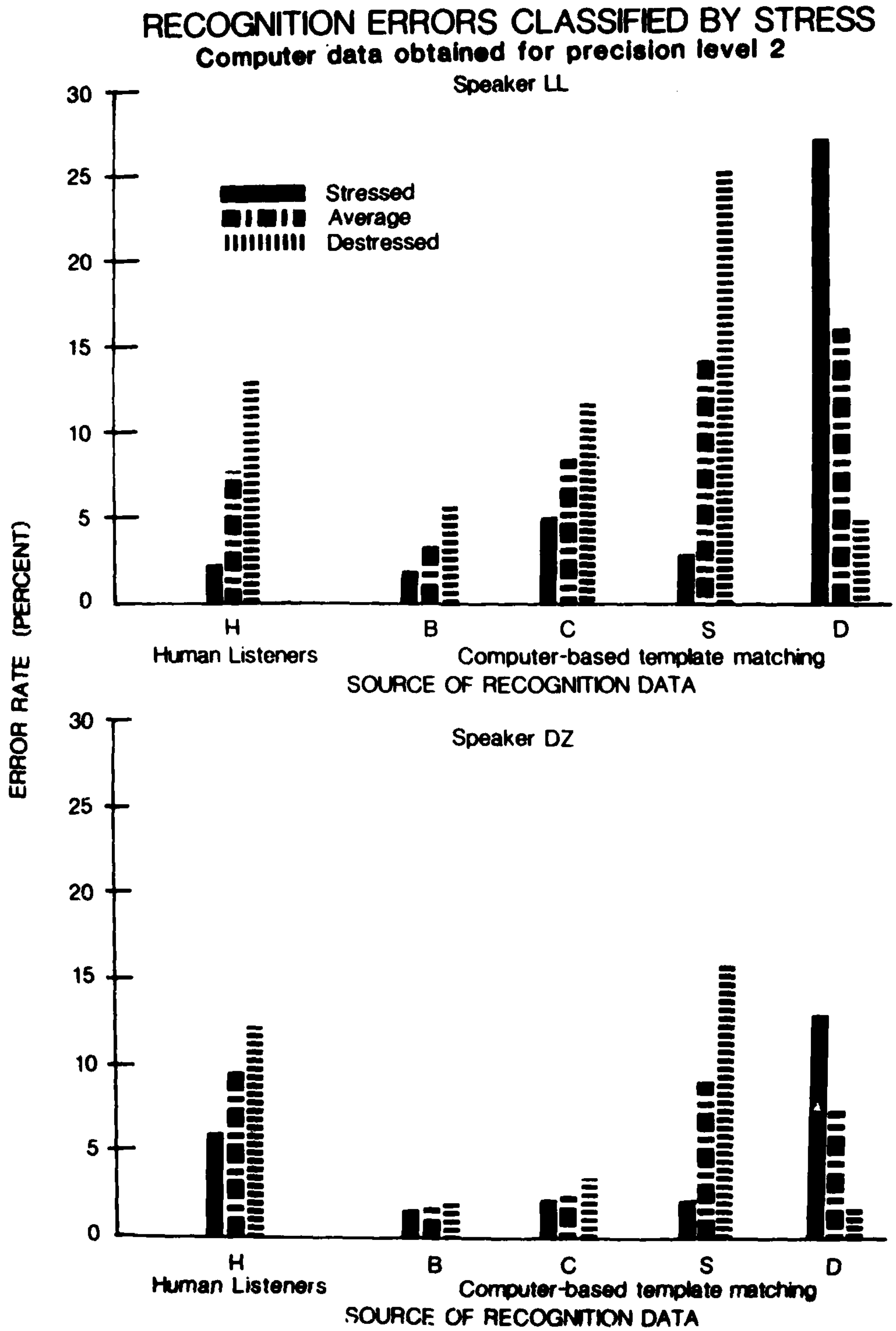


Figure 4. Comparison of error rates for human and computer recognition of syllables supplied by speakers LL and DZ. Results labeled H were obtained from listeners. Labels B, C, S and D refer to the four types of computer dictionary formed from coefficient data computed at PL2 (see text for explanation). The computer employed a dynamic warping and recognition algorithm with each dictionary in turn to recognize a closed set of unknown tokens.

Table 5

Recognition scores using dictionary B.

Classified by speaker, precision level and stress of token

Speaker		Token	A	B	C	Totals
DZ	PL1	Stressed	495	68	25	588
		Destressed	439	105	21	564
	PL2	Stressed	521	61	7	588
		Destressed	473	82	9	564
LL	PL1	Stressed	445	110	33	588
		Destressed	442	75	47	564
	PL2	Stressed	499	79	10	588
		Destressed	470	69	25	564
	Totals		3784	649	176	4608

Key:
A - Correct syllable identity and stress.
B - Correct syllable identity but incorrect stress.
C - Incorrect syllable identity.

Examination of Recognition Rank

An analysis was made of the number of times that the correct syllable appeared in second, third, fourth, and fifth positions in the rank of ordered distance measures obtained during the recognition computations. The results showed that about 70 percent of the syllables that failed to occupy the first rank (and, therefore, be "recognized") appeared in the second rank. Overall, the third rank captured about 18 percent of the unrecognized syllables and the fourth rank accounted for a further 5 percent. Speaker differences were another major feature of these data. In the case of LL, the proportions of syllables appearing in the various ranks did not vary significantly as a function of precision level. Speech data from DZ, on the other hand, showed higher proportions of unrecognized syllables entering the second rank in runs employing PL2. The magnitude of this shift was particularly prominent in the data for dictionary B, which indicates that this effect was related to the lower number of errors arising under PL2 conditions.

Geometry of the Stress Distance Space

The more significant features of the results just described can be explained by reference to the concept of a syllable distance space. Within this space, five possible configurations of the stressed and destressed tokens can be intuitively expected. Four of these are shown in Figure 5. The fifth configuration (Asymmetric Clusters; Equal Discriminability) shares features illustrated by configuration types (II) and (IV) and has been omitted. In

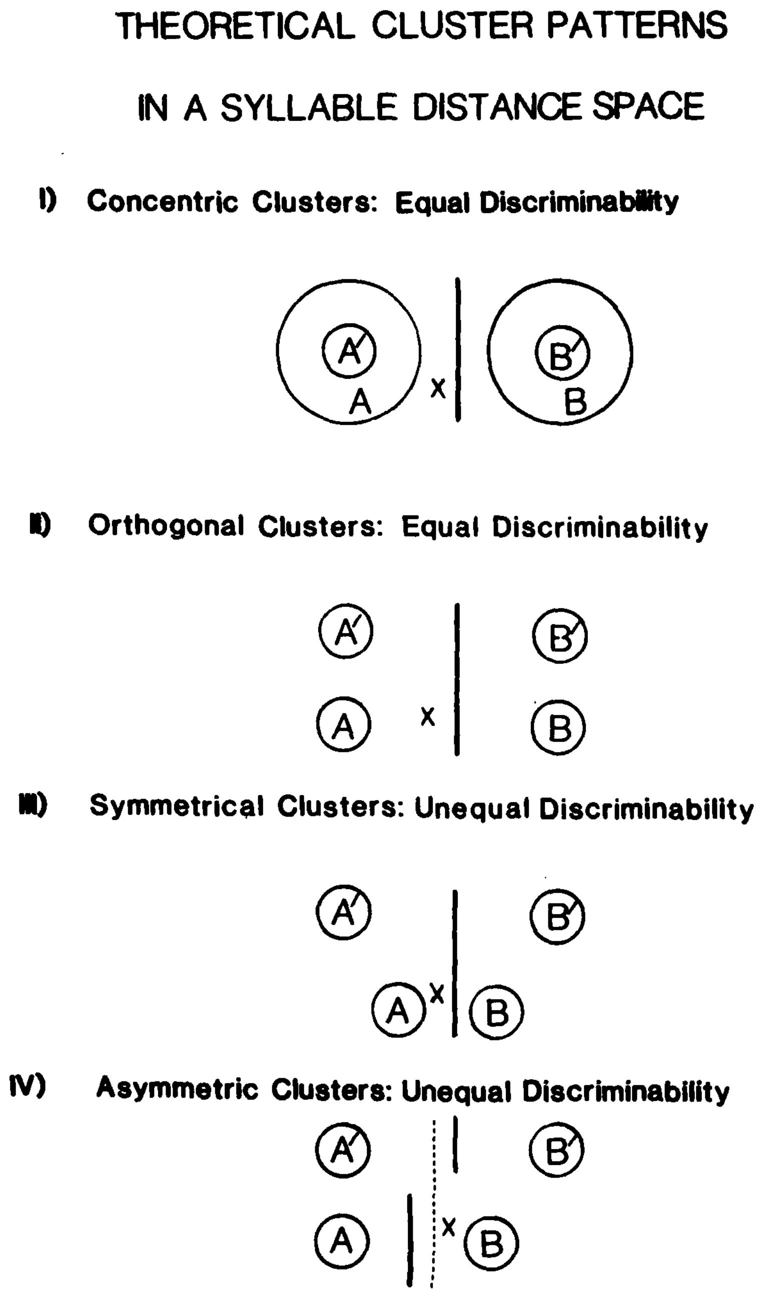


Figure 5. The symbol (X) represents the spatial location of an unknown token. Four types of cluster patterns for A and B are shown. Types (I), (II) and (III) are so distributed that a single decision boundary would serve for recognition of both stressed and destressed syllables and would lead to the classification of (X) as a member of the class "destressed A." For type (IV), different boundaries are required for unbiased decisions between stressed and destressed A and B. Hence, the token (X) is potentially classifiable as a "destressed B" or "stressed A."

each case, Figure 5 shows the theoretical relationship of two phonetically close syllables A and B occurring in both stressed (A') (B') and destressed (A) (B) forms. The heavy vertical bar that bisects an imaginary line linking the mid-points of the A and B distributions marks the position of the decision boundary between distributions A and B, which are assumed to be of similar size and conformation. (X) represents an unknown token. The first case, type (I), assumes that destressed syllables have the same central tendency as stressed syllables and form a large (noisy) cluster surrounding a smaller, more dense cluster of stressed tokens. This pattern would predict that a dictionary of stressed syllables (S) should serve well with both syllable types and, therefore, outperform all the other dictionaries. However, the data we have reported do not fit this prediction. Types (II) through (IV) postulate different formations of separate clusters for stresssed and destressed syllables. Type (II), consisting of four symmetric and orthogonal clusters, would suggest that stressed and destressed syllables should be found to be equally discriminable. Type (III) might arise when the discriminability of destressed pairs is less than that of stressed pairs but a single decision boundary can still serve to determine whether token (X) belongs to A or B. The fourth cluster configuration, type (IV), also gives rise to unequal discrimination but additionally requires the adoption of a second decision boundary to ensure the proper classification of the unknown (X).

To determine which of these theoretical models best fits the data, the distances obtained during recognition calculations were assembled in matrix form and input to the multidimensional scaling program KYST (Kruskal, Young, & Seery, undated). This program enabled us to generate graphic displays of the actual cluster structures of stressed and destressed syllables under a variety of dimensional constraints. The first observation to note is that, viewed overall, the clusters of destressed tokens consistently appeared to be only slightly less compact than the clusters of stressed tokens and, therefore, to possess a different but almost equally distinct acoustic form. In the two-dimensional case, the results contained examples of clusters that fitted each of the last three cases shown in Figure 5. For example, Figure 6 shows some actual distributions for both speakers obtained from data accumulated over all their speaking sessions. The spatial distributions are for the syllables [dɪg], [dɪj] and [dɪs], chosen because they represent minimal pairs (i.e., pairs of syllables that differ by a single phoneme). For the speaker LL, the upper half of the figure provides an example of orthogonal clusters resembling type (II) of Figure 5, while below is shown an equivalent group of clusters for the speaker DZ. In the latter case, the clusters tend to be asymmetrical and to resemble type (IV). In fact, by far the largest proportion of examples studied could be classified as type (IV). Thus, overall, the fourth case emerged as the best general model for the recognition data.

The type (IV) configuration (Figure 5) illustrates that, if a destressed token (X) is submitted to an S dictionary, the difference in location of the stressed decision boundary (upper vertical bar) will result in (X) being recognized as belonging to the class A. This makes it clear why poor recognition performances were obtained when the tokens were of different stress than the available templates. The diagram can also offer an explanation as to why a dictionary containing the combined templates was found to give better results and why better performance will always be achieved by using both stressed and destressed templates. To follow the explanation

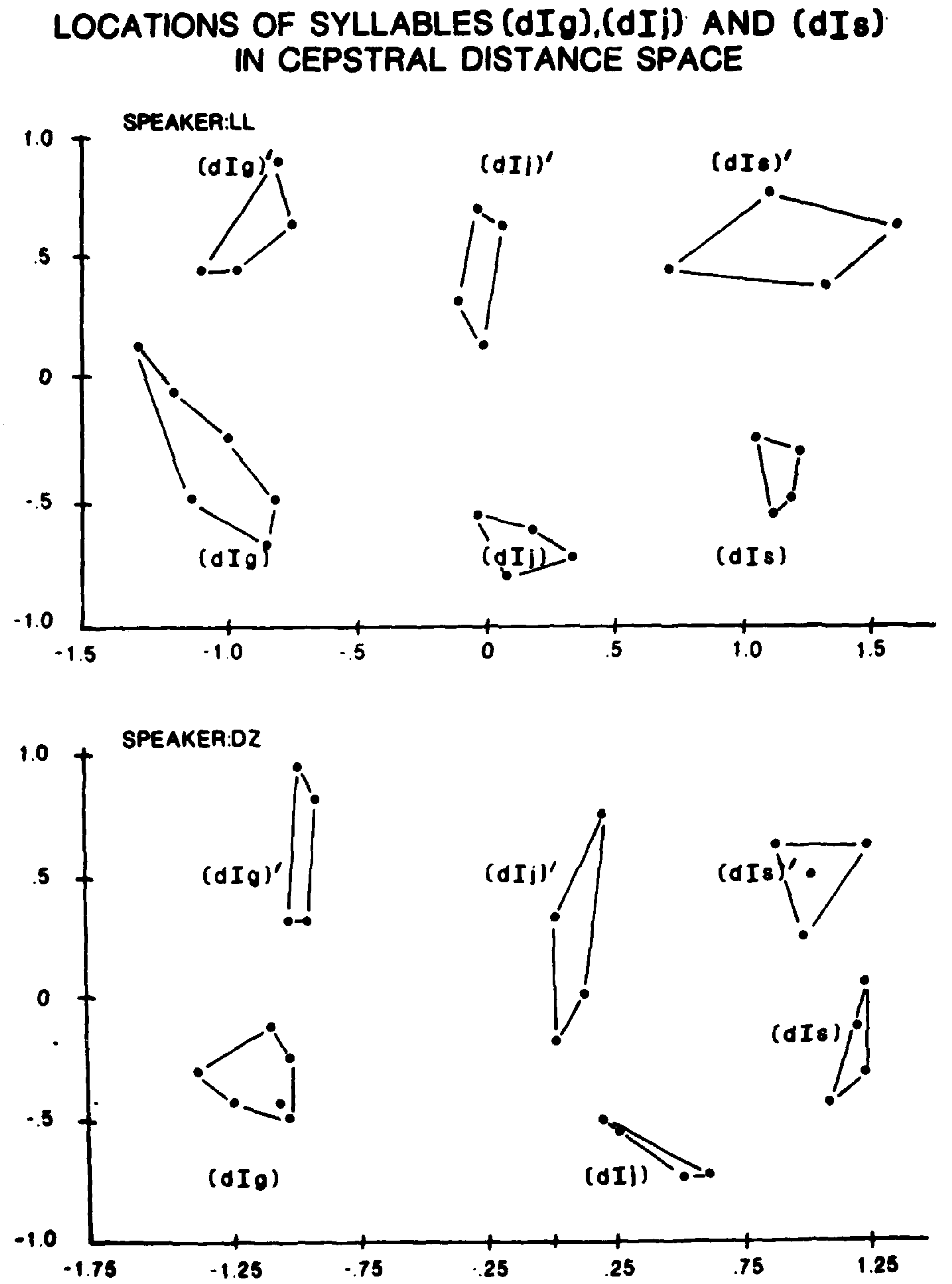


Figure 6. Cluster configurations for the syllables [dɪg], [dɪj] and [dɪs] obtained by analysis performed by the multidimensional plotting program KYST. Primes indicate stressed syllables. Syllable data were extracted from single sessions delivered by speakers LL and DZ.

offered in this case, we must assume that the clusters representing the combined templates for A and B will lie midway along the axis joining the centers of the stressed and destressed distributions. Therefore, the decision boundary (dotted vertical line) will now move to a point midway between the original stressed and destressed boundaries and (X) referred to this new boundary would now be correctly classified as belonging to class B.

CONCLUSIONS

Summary and Comments

We have conducted an investigation into the effects of stress and vowel duration on the performance of a recognition algorithm and we have compared some aspects of this performance with data gathered from listeners. In an effort to gain better control over our speech data, we chose to examine a form of stress variation that, while present in continuous speech, was sufficiently constrained that it could not be claimed to be representative of the more extreme forms that stress reduction can take. We deliberately omitted those types of stress reduction that result in (1) the syllabic vowel being pronounced as a schwa and, (2) the consonantal features being severely attenuated. Nevertheless, despite the relatively modest amount of stress variation present, its effects on recognition performance were quite large.

Our results showed that recognition accuracy for stressed and destressed syllables can be improved in three ways; these are, in increasing order of effect, (1) by increasing the resolution of the acoustic parameters as exemplified by exchanging PL1 for PL2, (2) by <u>combining</u> the acoustic features of stressed and destressed syllables into a single template dictionary or, (3) by doubling the size of the dictionary to include templates for <u>both</u> stressed and destressed syllables. Moreover, the results indicate that when computational economy is at issue, the nature of the trade-off between parameter resolution and dictionary size promises greater gains in recognition accuracy per bit of information from dictionary enlargement (the inclusion of individual stressed and destressed templates) than from increases in parameter precision.

We also examined the cluster structure adopted by pairs of linguistically different stressed and destressed syllables and found that the bulk of them can be classified as asymmetric distributions offering unequal discriminability for stressed and destressed forms. Moreover, we found that destressed tokens form clusters that are only marginally less compact than their stressed counterparts. This observation was confirmed by the fact that the overall recognition rate for destressed tokens submitted to a dictionary of destressed templates was very similar to the rate observed for stressed tokens matched against a dictionary of stressed templates.

The reason for the unusual compactness of the destressed tokens must almost certainly be sought in the environments in which these syllables were produced. The restrictions that were placed on the amount of stress reduction we wished to permit imposed strict limitations on the number of syllables and lexical environments that were available. Thus, the fact that any given word containing a target syllable appeared in only two different sentence environments provided little opportunity for a variety of coarticulation effects to extend from neighboring phones to the target syllables. Moreover, the experimental conditions fostered the likelihood that the magnitude of any coarticulatory interaction would vary according to a target syllable's position within a word. For example, as seen in Table 1, destressed syllables occupied word-initial, mid-word, and word-final positions on a roughly equal basis, whereas stressed syllables appeared prominently in word-initial position. Hence, to the extent that the strongest coarticulatory influence was likely to occur between target syllables and immediately adjacent phones, it may be assumed that, by virtue of the constancy of their immediate environment, approximately one third of the destressed syllables were produced with substantially the same coarticulation.

In addition, we confirmed that the phonologically short vowels were, according to our measurement criteria, shorter than phonologically long vowels. We also found that the shortening of vowel length and syllable length that accompanies stress reduction is greater in the case of the shorter vowel. Since some degree of time normalization is an intrinsic feature of the warping algorithm, one might expect that any bias in favor of longer vowels would be offset. Certainly this is suggested by the fact that cross-vowel error rates for short and long vowels increase at an approximately equal rate across dictionaries B, C, and S. However, the study also indicates that long vowels have two important advantages and suffer one disadvantage when subjected to warping and recognition procedures. First, among the advantages is the fact that identification errors involving long vowels tend to include a smaller proportion of cross-vowel errors than is found to be included among the identification errors involving short vowels. Second, long vowels tend to be associated with lower vowel-error rates than short vowels. The disadvantage that long vowels face is due to the preponderant contribution they make to the distance measure. This contribution is so large that it masks or "dilutes" consonant information to such a degree that syllable identity errors increase. We must therefore conclude that in future attempts to develop improved distance metrics, an effort directed at enhancing the contribution made by consonants should be given priority.

Another group of observations made in this study centered on the similarities and differences between recognition performances delivered by listeners and those produced by the computer. Evidence indicated that listeners could achieve a recognition accuracy on stressed tokens that is roughly comparable with that achieved by computer. On the other hand, computer recognition rates for destressed syllables under the most favorable conditions are found to be superior to the rates achieved by listeners. One tentative explanation for this possibly surprising observation rests on the notion that the listeners tend to be biased (or pre-primed) for stressed-item recognition by the phonetically-spelled syllable transcriptions displayed on their response forms. Yet another explanation acknowledges the fact that listeners must carry in their heads many more syllable templates than were listed on the response form. Given this fact, an unknown destressed token X may not be directly identified with the nearest syllable (A') listed on the response form but can be identified instead with template (C), not included in the response list, because distance $D(X,A')>D(X,C)$. Subsequently, X having lost its own acoustic identity (by decay of short-term memory) and assumed that of C, a search for the nearest template identified in the response list leads to the incorrect selection of template (B') because $D(C,B')<D(C,A')$.

Finally, it might be noted that the recognition of stressed syllables is a highly practiced task, whereas the recognition of destressed syllables is not. This is because, in continuous speech, destressed syllables are normally recognized with the aid of their context. To recognize them in isolation is a relatively unfamiliar task and consequently poorer performance is to be expected. Of course, the present data provide no opportunities to examine these alternative hypotheses properly. In the final analysis, it has to be conceded that the behavior of listeners and the behavior of the computer algorithm are so different as to make it obvious that the recognition principles employed by both are quite different.

Our finding that the spatial distributions of our stressed and destressed syllables do not greatly differ in size suggests that it might be possible to derive the acoustic properties of each destressed syllable by applying a warp in both the time and frequency domains to its appropriate stressed counterpart. Moreover, if warps of this kind proved to have properties that were common to a large class of syllables, say all CVCs of a given vowel type, this would be of considerable help in controlling the rate of dictionary growth. One way of applying such a warp would be by means of a matrix that would provide the opportunity to compute a composite or standard warp for a given syllable class by averaging together the warps obtained from many CVCs.

Stress effects are among the most difficult of the many obstacles that lie in the path of achieving a practical continuous speech recognition capability. In this study, we have begun a systematic approach to this problem by attempting to generate controlled, yet realistic, data and to observe their interaction with recognition variables such as dictionary composition, parameter precision and widely used recognition techniques such as dynamic pattern matching. We have succeeded in identifying many of the interactions that take place and in several cases have been able to point out their boundary conditions. Future work on the problem of stress variation should involve the gradual relaxation of some of the input constraints adopted here, the collection of additional observations, and the development of new and better algorithms.

REFERENCES

Bahl, L. R., Baker, J. K., Cohen, P. S., Cole, A. G., Jelinek, F., Lewis, B. L., & Mercer, R. L. Automatic recognition of continuously spoken sentences from a finite state grammar. IEEE International Conference on Acoustics, Speech and Signal Processing Record, 1978, 418-421.

Bridle, J. S., & Brown, M. D. An experimental automatic word recognition system. JSRU Report (Joint Speech Research Unit, Ruislip, England), 1974, No. 1003.

Davis, S. B. Order dependence in templates for monosyllabic word identification. IEEE International Conference on Acoustics, Speech and Signal Processing Record, 1979, 570-573.

Davis, S. B., & Mermelstein, P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech and Signal Processing, 1980, ASSP-28, 357-366.

Dixon, N. R., & Silverman, H. F. The 1976 Modular Acoustic Processor (MAP). IEEE Transactions on Acoustics, Speech and Signal Processing, 1977, ASSP-25, 367-379.

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

Fujimura, O. The syllable as a unit of speech recognition. IEEE Transactions on Acoustics, Speech and Signal Processing, 1975, ASSP-23, 79-82.

Itakura, F. Minimum prediction residual principle applied to speech recognition. IEEE Transactions on Acoustics, Speech and Signal Processing, 1975, ASSP-23, 67-72.

Klatt, D. H. SCRIBER and LAFS: Two new approaches to speech analysis. In W. A. Lea (Ed.), Trends in speech recognition. Englewood Cliffs, N.J.: Prentice Hall, 1978.

Kruskal, J. B., Young, F. W., & Seery, J. W. How to use KYST-II; A very flexible program for multidimensional scaling and unfolding. Computing Information Service, Bell Laboratories, undated.

Mermelstein, P. Automatic segmentation of speech into syllabic units. Journal of the Acoustic Society of America, 1975, 58, 880-883.

Mermelstein, P. Distance metrics for speech recognition--psychological and instrumental. In C. H. Chen (Ed.), Pattern recognition and artificial intelligence. New York: Academic Press, 1976.

Mermelstein, P. Recognition of monosyllabic words in continuous sentences using composite word templates. IEEE Transactions on Acoustics, Speech and Signal Processing, 1978, 708-711.

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

Pike, K. Intonation of American English. Ann Arbor, Mich.: University of Michigan Press, 1945.

Rabiner, L. R., Rosenberg, A. E., & Levinson, S. E. Considerations in dynamic time warping algorithms for discrete word recognition. Journal of the Acoustical Society of America, 1978, 63, S79.

Rabiner, L. R., & Wilpon, J. G. Speaker-independent isolated word recognition for a moderate size (54 word) vocabulary. IEEE Transactions on Acoustics, Speech and Signal Processing, 1979, ASSP-27, 583-585.

Rosenberg, A. E., Rabiner, L. R., Levinson, S. E., & Wilpon, J. G. A preliminary study on the use of demisyllables in automatic speech recognition. IEEE International Conference on Acoustics, Speech and Signal Processing Record, 1981, 967-970.

Sakoe, H., & Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech and Signal Processing, 1978, ASSP-26, 583-586.

Trager, G., & Smith, H. An outline of English structure. Studies in Linguistics: Occasional Papers 3. Norman, Ok.: Battenburg Press, 1951.

Velichko, V. M., & Zagoruyko, N. G. Automatic recognition of 200 words. International Journal of Man Machine Studies, 1970, 2, 223-234.

FOOTNOTES

[1]It is the complex nature of the coarticulatory interaction between phones (particularly within syllables) that has proved to make segmentation strategies based on phonemic units so difficult to develop.

[2]Many linguists (Pike, 1945; Trager & Smith, 1951) have drawn attention to the fact that English speech has more than two levels of stress. Furthermore, the comments of our colleagues and reviewers have made it obvious that there is insufficient agreement on a terminology for stress designation

to permit us to use the words "stressed" and "destressed" without the following explanatory remarks: The syllables employed in this study were obtained from words in which they customarily receive contrasting degrees of lexical stress. These stress contrasts were potentially subject to enhancement or reduction by the sentential context although the most obvious syntactic influences such as word-final lengthening were avoided. Therefore, a syllable labeled as "stressed" did not necessarily bear the primary or highest sentential stress. Syllables labeled as "destressed," on the other hand, always bore less stress than their stressed counterparts but were never so severely reduced as to cause the nuclear vowel to be produced as a schwa. In general, experience leads us to expect that the stress reduction exhibited by syllables incorporating /ɪ/ to be greater than the reduction for syllables incorporating /æ/.

[3]Because syllables in phrase-final position tend to undergo lengthening and because syllable lengthening is one of the principal correlates of stress (Fry, 1955), it was particularly necessary to avoid the interaction of such position effects with the syllables chosen for this study.

[4]Errors occurred primarily in the syllable-duration category and were due to a failure of the segmentation algorithm to include released bursts in final position as an integral part of the preceding syllable. A secondary problem was the occasional omission of destressed syllables. Such errors were not acceptable for the purposes of the present study.

PHONETIC AND AUDITORY TRADING RELATIONS BETWEEN ACOUSTIC CUES IN SPEECH PERCEPTION: FURTHER RESULTS*

Bruno H. Repp

Abstract. The series of studies begun by Repp (1981), with the purpose of examining whether trading relations between acoustic cues are obtained within phonetic categories, is continued with three experiments. Despite some unexpected complexities, the results tend to support the hypothesis that the trading relations studied are a consequence of phonetic categorization.

Whenever two or more acoustic cues contribute to the perception of a phonetic distinction, a trading relation among the cues can be demonstrated in categorization, given that the speech stimuli are phonetically ambiguous. That is, a change in one cue can be compensated for by a change in another cue, so as to maintain the same degree of perceptual ambiguity. In a previous paper (Repp, 1981) I asked whether cues would continue to engage in trading relations when the stimuli are phonetically unambiguous. An affirmative answer to this question would mean that the trading relation examined is either psychoacoustic in origin or that it derives from a phonetic mode of processing that extends beyond the mere assignment of category labels. A negative answer, on the other hand, would imply that the trading relation is either tied to phonetic categorization or that it is a psychoacoustic phenomenon specifically limited to the phonetic boundary region. Thus, while these answers do not distinguish between all possible hypotheses, they usefully restrict the set of alternatives. Further arguments and experimental evidence may then be adduced to arrive at the most likely explanation for a given trading relation.

Phonetic classification of unambiguous stimuli evidently does not yield the kind of information sought. In my earlier experiments, I employed instead a fixed-standard same-different discrimination paradigm with stimuli that either straddled a phonetic category boundary or came from within a phonetic category. Four different trading relations were examined. One of them, suspected to be of psychoacoustic origin, held up regardless of phonetic ambiguity; two others, suspected to be byproducts of phonetic categorization, disappeared for within-category stimulus comparisons; the results of the fourth experiment were inconclusive. The three experiments to be reported in the present paper supplement and extend my earlier research using exactly the same methodology.

Acknowledgment. This research was supported by NICHD Grant HD01994 and BRS Grant RR05596. I am grateful to Sharon Manuel for assistance in running the experiments.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

GENERAL METHOD

A graphic illustration of the paradigm in the form of a geometric analogy is provided in Figure 1. The two acoustic cues whose trade-off is to be investigated are depicted here as the height and width of rectangles. The dimension resulting from the perceptual integration of the two cues, analogous to the phonetic percept (though without any clearly defined category boundary), is the area of the rectangles, a measure of which (in arbitrary units) is given by the numbers in Figure 1. The subjects' task is to discriminate a standard, which occurs first in each stimulus pair, from a limited set of alternative stimuli. A series of practice trials is presented first, with subjects having foreknowledge of the correct responses. Half the stimulus pairs are "same" trials in which the standard is paired with itself; the other half are "different" trials in which the standard is followed by a stimulus that differs in one (the "primary") cue dimension (height in Figure 1) by a fairly large amount. Three blocks of test trials follow. In each of these, there are three types of trials occurring with equal frequency: "same" trials, 1-cue "different" trials in which the difference is only in the primary cue, and 2-cue "different" trials in which the comparison stimulus differs from the standard on both cue dimensions. The difference in the second (the "secondary") cue dimension (width in Figure 1) is fairly small and chosen so as to counteract the difference in the primary cue with respect to the integrated percept; thus, in Figure 1, increased height is coupled with reduced width. The size of the primary cue difference (height) decreases across the three test blocks, whereas the secondary cue difference (width) remains constant.

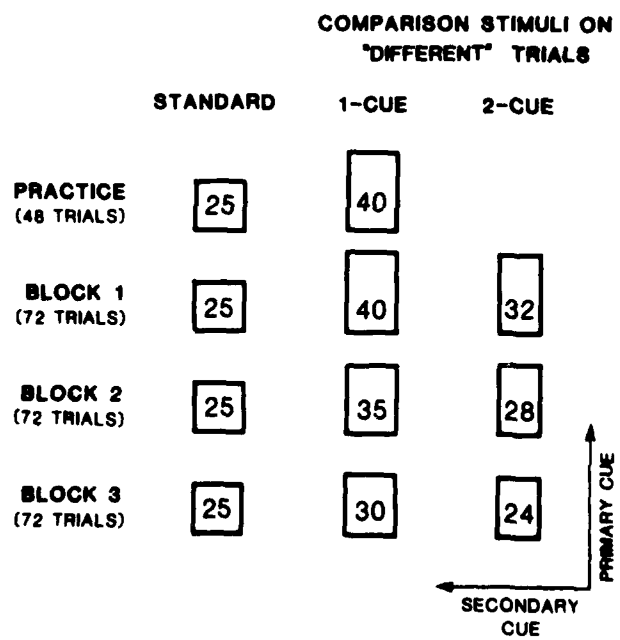


Figure 1. Schematic diagram of the experimental paradigm.

If listeners discriminate the stimuli on the basis of an integrated property derived from both cues (area), then the prediction is that, paradoxically, 1-cue differences should be easier to detect than 2-cue differences: In Figure 1, the standard-comparison difference in area is larger on 1-cue than on 2-cue trials. If, however, subjects do not integrate the two cues and instead either focus on a single cue or divide attention between two separable cue dimensions, then there should either be no difference between 1-cue and 2-cue trials (if only the primary cue is attended to), or 2-cue trials should yield higher detection scores than 1-cue trials. In the latter case, a divided-attention strategy may be distinguished from a secondary-cue focus by gauging the extent of the advantage for 2-cue trials and the extent of the decline in 2-cue discrimination performance over test blocks.

Each experiment has two conditions, a between-category (Between) and a within-category (Within) condition. Each condition includes the complete paradigm shown in Figure 1; the difference lies solely in the values chosen for the primary cue dimension. In the Between condition, they are chosen so that the standard stimulus is close to a phonetic boundary and the comparison stimuli tend to fall even closer to, or on the opposite side of, the boundary. This enables listeners to make use of phonetic category distinctions and thus encourages the phonetic strategy of deriving a single integrated percept from the two cue dimensions and of basing same-different judgments on a comparison of these percepts (i.e., categorical perception). This condition should yield the expected phonetic trading relation (revealed as a superiority of 1-cue over 2-cue trials) and thus serves as a control. In the Within condition, the primary cue values are chosen so that all stimuli fall well within a phonetic category. Here, listeners presumably can no longer make phonetic distinctions and have to rely on perceived auditory differences between the stimuli. The critical result is the relative performance on 1-cue and 2-cue trials. If this relation is significantly different from that observed in the Between condition, the conclusion is warranted that a different (presumably nonphonetic) perceptual strategy was used in within-category discrimination. It should be noted that, although the clearest result would be 1-cue superiority in the Between condition and 2-cue superiority in the Within condition, a significant change in the 1-cue versus 2-cue relation across conditions (i.e., a significant Cues by Conditions interaction in an analysis of variance) is sufficient to permit conclusions about differing perceptual strategies. The results may not always be ideal because, as in many other tasks concerned with categorical perception of speech, phonetic and auditory strategies may be used simultaneously in varying degrees, particularly in "between-category" discrimination. (See Repp, 1981, for presumed instances in the present paradigm.)

The experimental setup in the present experiments differed from that of my earlier studies in several minor respects. First, the number of test trials was increased by one-sixth to 84 per block. Second, the number of practice trials was reduced to 28, and instead of following a random sequence, they alternated between "same" and "different." As before, during practice the subjects checked off the correct responses printed on the answer sheet. Third, a change in the direction of primary-cue differences was introduced in parts of Experiments 2 and 3 and is described later. Fourth, more extensive identification data were collected than in the earlier studies. These data were always obtained after the discrimination tasks or in a separate session (or from different subjects altogether), to avoid biasing the listeners too strongly toward use of a phonetic strategy. On the other hand, the Between condition always preceded the more difficult Within condition, to permit subjects to get used to the stimuli and to the task. This, finally, constituted another change from my earlier studies, in which the Within condition was presented twice, both before and after the Between condition. Since there were no significant differences between these two presentations in any of the four previous experiments, the present use of a single run following the Between condition was fully justified, even though the total number of responses obtained was thereby reduced.

EXPERIMENT 1: "SAY"-"STAY"

The purpose of this study was to supplement my earlier Experiment 1, which was concerned with the stop manner distinction in "say" versus "stay." This distinction is of special interest because Best, Morrongiello, and Robson (1981) have reported results that suggest a phonetic basis for the trading relation between the two cues of silent closure duration and first-formant (F1) onset frequency. In my earlier study, I employed stimuli composed of a natural-speech "s" noise followed by a variable amount of silence (the primary cue) and one of two synthetic vocalic portions differing in F1 onset (the secondary cue). The results were encouraging but statistically weak, due to high variability (an aspect of the data that was also encountered in the present experiments, unfortunately). Although the expected trading relation was apparent both in the Between condition (as 1-cue superiority) and in the post-discrimination labeling data, it did not reach significance in either set of data. However, there was a significant 1-cue superiority in the Within condition, and a significant Cues-by-Conditions interaction confirming the reversal. Clearly, then, the phonetic trading relation was absent when the subjects could not draw any category distinctions, which supported the conclusion of Best et al. (1981) that the trading relation may be specific to phonetic perception.

The weakness of the phonetic trading relation in the earlier Between condition may have been due to a mixture of phonetic and auditory strategies in discrimination; however, the similar weakness in the labeling data cannot be so explained. Rather, it suggests that the stimulus materials were not optimal. The original purpose of the present study was to provide a replication with improved stimuli. All-natural stimuli were envisioned for that purpose. Since F1 onset frequency is difficult to manipulate directly in natural speech, it was planned to take vocalic portions from utterances of "say" and "stay," which were thought to contain the required difference in F1. Pilot tests (of a limited nature, to be sure) suggested, however, that the two vocalic portions--the particular tokens used, in any case--had no differential effect in perception and did not generate any trading relation. Although I could have extended my efforts at finding stimuli that "worked," I decided instead to vary a different, but equally relevant, secondary cue: the release burst that occurs immediately following the closure in "stay" but is absent in "say."

Method

The utterances "say" and "stay" were recorded by a female speaker and were digitized at 20 kHz. In order not to bias perception too strongly toward "stay," the fricative noise portion of "say" was employed in the experimental stimuli. However, to counteract a possible bias in the opposite direction, the final low-amplitude portion was trimmed off, leaving a noise waveform of 157 msec duration. The experimental stimuli were created by following this noise with a variable silent interval and one of two waveforms derived from the 400-msec post-closure portion of the "stay" utterance. Originally, this "day" portion began with a powerful release burst of approximately 25 msec duration, more than sufficient to cue perception of "stay" even when immediately preceded by an "s" noise without closure silence (Repp, 1982). To obtain stimuli that would permit perception of "say" in the same situation, the onset of the "day" portion was cut back by 20 and 29 msec, respectively, resulting in stimuli that, in analogy to Best et al. (1981), may be called strong "day" and weak "day" (relatively speaking). The strong "day" retained the last 4 msec of the release burst, which were of rather low amplitude. In the weak "day," this residual burst was eliminated together with the first 5-msec pitch period, which was of very low amplitude and was overlaid with some aspiration noise. Essentially, then, the strong and weak "day" differed in the presence versus absence of a residual release burst at onset.

In the Between condition, the fixed standard consisted of the "s" noise immediately followed by the strong "day"--a stimulus expected to be perceived as "say." The comparison stimuli in the three test blocks had silent closure intervals of 40, 30, and 20 msec, respectively. In the Within condition, the standard, which again contained the strong "day" portion, had a closure interval of 40 msec (expected to lead to the perception of "stay"), and the comparison stimuli had silences of 100, 80, and 60 msec. A separate identification tape contained ten random sequences of 14 stimuli generated by following the "s" noise with either the strong or the weak "day," separated by silent intervals ranging from 0 to 60 msec in 10-msec steps. The subjects were nine paid volunteers, mostly Yale undergraduates. For details of method not mentioned here, the reader is referred to Repp (1981).

Results and Discussion

Figure 2 displays the average post-discrimination identification results. Percent "stay" responses is shown as a function of silence duration. It is evident that the stimuli containing the strong "day" portion generated an orderly labeling function, with the category boundary at 25 msec of silence. The stimuli that served as standards in the discrimination task, with 0 and 40 msec of silence, received 2 and 91 percent "stay" responses, respectively, which confirms that they had been appropriately chosen as instances of "say" and "stay." The labeling function for the stimuli containing the weak "day," however, was unexpectedly gradual, reaching not even 50 percent "stay" responses at the longest silence. (Only two of the nine subjects reached 100 percent "stay" responses.) This was surprising, for exactly the same stimuli had been used in another study (Repp, 1982) where many more "stay" responses were obtained. The resulting exaggerated trading relation (if it still can be called that) between the silence and release burst cues has implications for the discrimination tasks: On one hand, an especially clear trading relation should emerge in the Between condition; on the other hand, the failure of the weak "day" stimuli to reach 100 percent "stay" responses (presumably even at silences longer than 60 msec, judging from Figure 2) gave subjects an unexpected opportunity to detect phonetic distinctions in the Within condition

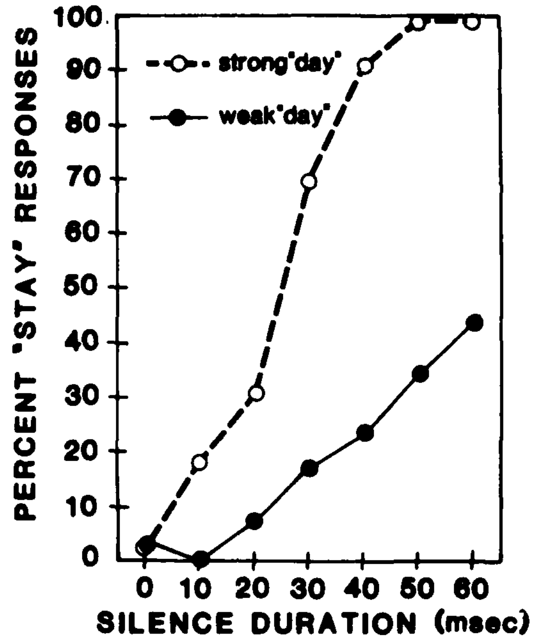


Figure 2. Identification results of Experiment 1.

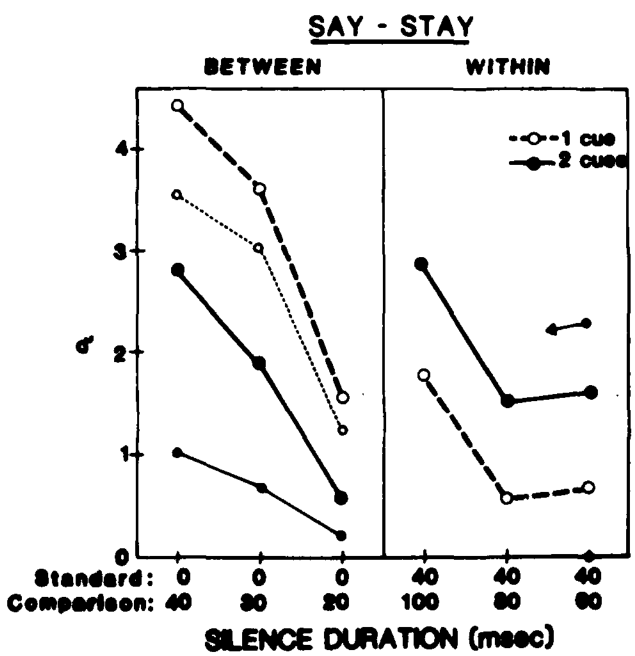


Figure 3. Discrimination results of Experiment 1.

as well. Here, however, a phonetic strategy should lead to higher scores on 2-cue than on 1-cue trials. (Consider the 40-msec strong "day" standard and the two 60-msec comparison stimuli in Figure 2.) Therefore, a reversal in the relation between 1-cue and 2-cue discrimination scores is predicted on phonetic grounds alone, which complicates (but still permits) an interpretation of the discrimination results.

These results are shown in Figure 3 as d' scores (heavy lines). The pattern is very clear: In the Between condition there is a large advantage for 1-cue trials, $F(1,8) = 31.7$, $p < .001$, while, in the Within condition, there is a strong trend in the opposite direction that, however, failed to reach significance, $F(1,8) = 3.3$. The Cues-by-Conditions interaction is highly significant, $F(1,8) = 25.2$, $p < .002$. In addition, performance declined across test blocks, $F(2,16) = 24.5$, $p < .001$, except for blocks 2 and 3 in the Within condition, where scores remained constant.

The results of the Between condition confirm the expected trading relation and bolster the somewhat weak results obtained in the same condition of the earlier "say"-"stay" study (Repp, 1981). The thin lines in Figure 3 indicate the results expected if subjects had relied on phonetic labels alone. These expected d' values were derived after predicting individual hit and false alarm rates according to the classic "Haskins model" of categorical perception. It can be seen that performance was a good deal better than predicted; this may be attributed to anchoring or contrast effects due to the fixed standard (Repp, Healy, & Crowder, 1979). The smaller gain for 1-cue trials may be attributed to a ceiling effect ($d'max = 4.64$). Thus, the data are consistent with the hypothesis that, in the Between condition, subjects relied primarily on phonetic labels in discriminating the stimuli. They are also consistent, however, with the alternate hypothesis that a psychoacoustic trading relation localized in the phonetic boundary region is responsible for the effects seen.

The results of the Within condition are less straightforward. Predicted d' values were computed for the last test block and are shown in Figure 3. It can be seen that performance on 1-cue trials was better than predicted (predicted d' was near zero) while performance on 2-cue trials was worse than predicted. As a result, the obtained difference between 1-cue and 2-cue discrimination was smaller than predicted. If the assumption is accepted that subjects used primarily a phonetic strategy even in the Within condition, the depressed scores on 2-cue trials may indicate that a psychoacoustic trading relation favoring 1-cue trials (as in the Between condition) counteracted the trends generated by the phonetic strategy. That purely auditory discrimination played an additional role is clear, at the very least, from the elevated scores in the first test block; note that the predicted scores must be lower in the first than in the last test block, as indicated by the arrow in Figure 3. (This can easily be verified with the aid of Figure 2.)

In the hope of clarifying the situation, the Within-condition results of individual subjects were inspected. All of the five subjects who gave very few "stay" responses to the weak "day" stimuli showed the predicted 2-cue superiority. So did, however, one of the two subjects whose "stay" responses reached 100 percent at or before the 60-msec silence duration (and whose predicted scores were, therefore, zero throughout) and one of two subjects whose labeling results indicated that 100 percent "stay" responses might have been reached somewhere beyond 60 msec. These results suggest the use of an auditory strategy favoring 2-cue trials, which implies that there was no psychoacoustic trading relation favoring 1-cue trials. On the other hand, one of the two subjects with reasonable labeling scores showed (as the only subject) a substantial advantage for 1-cue trials in the Within condition. The other one of the two subjects with excellent labeling scores performed near chance throughout (as predicted), which suggests that he was a strictly categorical perceiver and failed to make any use of auditory information.

In summary, the results of the present study, while not crystal-clear, do lend some support to the phonetic/localized-psychoacoustic pair of hypotheses; they tend not to favor the generalized-phonetic/psychoacoustic pair. Within the favored pair, the distinction rests on whether the postulated psychoacoustic interaction and its specific location can be supported by independent arguments or evidence. At present, such evidence is in short supply; however, some negative arguments will be presented in the General Discussion.

EXPERIMENT 2: "SLIT"-"SPLIT"

All the experiments up to now (including the four studies in Repp, 1981, and the present Experiment 1) had in common that the primary cue was temporal in nature, and that the Within condition used longer values on that temporal dimension than the Between condition. This was so out of necessity, since the category boundaries were located at relatively short durations of the temporal cue and did not leave sufficient "room" for a full discrimination paradigm (Figure 1) at the short end of the continuum. Also, to the extent that the boundary coincided with a psychoacoustic threshold of some sort (cf. Miller, Wier, Pastore, Kelly, & Dooling, 1976; Pastore, Ahroon, Baffuto, Friedman, Puleo, & Fink, 1977; Pisoni, 1977), one might have expected discrimination to be at chance below that threshold, i.e., at the very short end of the continuum. Nevertheless, it became increasingly evident that an application of the present paradigm to the short end of a temporal dimension might be a desirable strategy to pursue. After all, few psychoacousticians would be surprised by the finding that an interaction between cues occurring in the vicinity of some hypothesized threshold disappeared at long temporal separations of signal components: Temporal proximity may be a prerequisite for the interactions (be they masking or integration) that are thought to underly a trading relation. If so, however, then the psychoacoustic interaction should become even stronger when temporal separation is further reduced. On the other hand, if the stimuli with these short temporal values all fall in the same phonetic category, then the phonetic hypothesis would predict a disappearance of the trading relation. Moreover, finding that subjects can discriminate these stimuli at all would cast doubt on the hypothesis equating category boundaries with auditory thresholds.

To pursue this possibility, it is necessary to find a stimulus continuum on which the boundary is at somewhat longer durations of a temporal cue. The "slit"-"split" distinction seems to fit the bill. In a recent study by Fitch, Halwes, Erickson, and Liberman (1980), the average boundary on a continuum of varying silent closure durations was somewhere between 50 and 80 msec, depending on the precise characteristics of the stimuli. This gives rise to the hope of obtaining above-chance discrimination scores strictly within the "slit" category.

Experiment 2 was conducted in two parts. Part a included the Between condition and the Within ("slit") condition just described. Part b included the same conditions but with a different choice of standards, as described below, plus a second Within ("split") condition using long values of the temporal cue dimension.

Method

The stimuli were created in a similar way as those of Experiment 1. A female speaker recorded the utterance "split," which was digitized at 20 kHz. The pre-closure "s" noise, 141 msec in duration, was separated from the post-closure "blit" portion, which consisted of an initial 15-msec low-amplitude release burst followed by a 230-msec voiced portion, a 137-msec "t" closure, and a final "t" release burst. Two versions were derived from this portion by waveform editing: a strong "blit" that retained the final 12 msec of the release burst, and a weak "blit" that had no release burst left.

In the Between condition of Part a, the standard had a closure silence of 40 msec preceding the strong "blit." The comparison stimuli in successive test blocks had silences of 80, 70, and 60 msec. In the Within ("slit") condition of Part a, the standard had no silence preceding the strong "blit," while the comparisons had silences of 40, 30, and 20 msec. In the Within ("split") condition of Part b, the standard had 140 msec of silence preceding the strong "blit," while the comparisons had silences of 200, 180, and 160 msec. The Between and Within ("slit") conditions of Part b essentially reversed the standard and comparison stimuli of the corresponding conditions in Part a. In the Between condition, the standard initially had 80 msec of silence followed by the weak "blit," and the comparisons had 40 msec of silence. Over successive test blocks, the silence of the standard decreased from 80 to 70 to 60 msec, while that of the comparison remained constant. In the Within ("slit") condition, the silence in the standard decreased from 40 to 30 to 20 msec (followed by the weak "blit"), while that of the comparison remained fixed at 0 msec. The reason for these changes will become apparent below.

The identification test included ten random sequences of 20 stimuli. Silences ranged from 30 to 120 msec in 10-msec steps; stimuli included either the weak or the strong "blit." The identification test was taken by nine subjects, only four of whom also took Part a of the discrimination tests. Eight of the nine paid volunteers in Part a had also been subjects in Experiment 1. Seven new subjects were run in Part b. The subjects in Part b listened to the Within ("split") tape at the end of the session.

Results and Discussion

The average results of the identification test are shown in Figure 4. They proved to be very orderly. The category boundaries were at 49 and 70 msec for the strong and weak "blit," respectively. Note that the standards used in the Within "slit" (Part a) and "split" (Part b) conditions, with

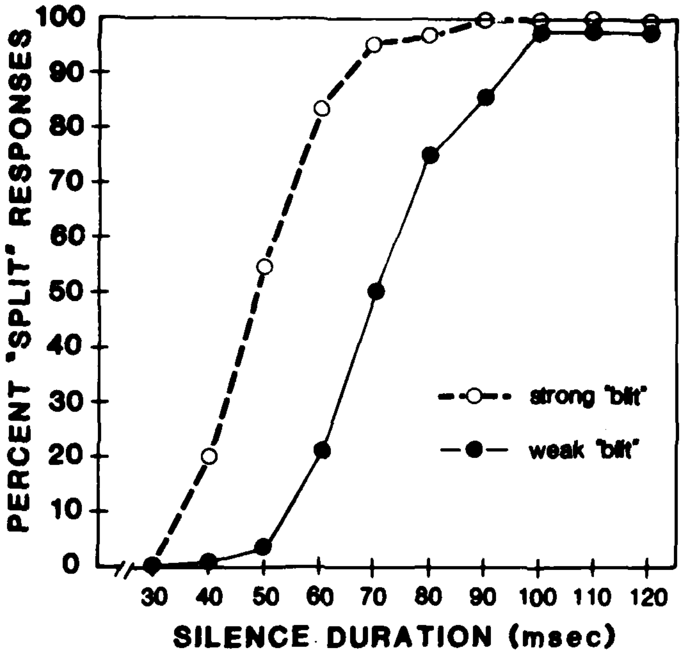


Figure 4. Identification results of Experiment 2.

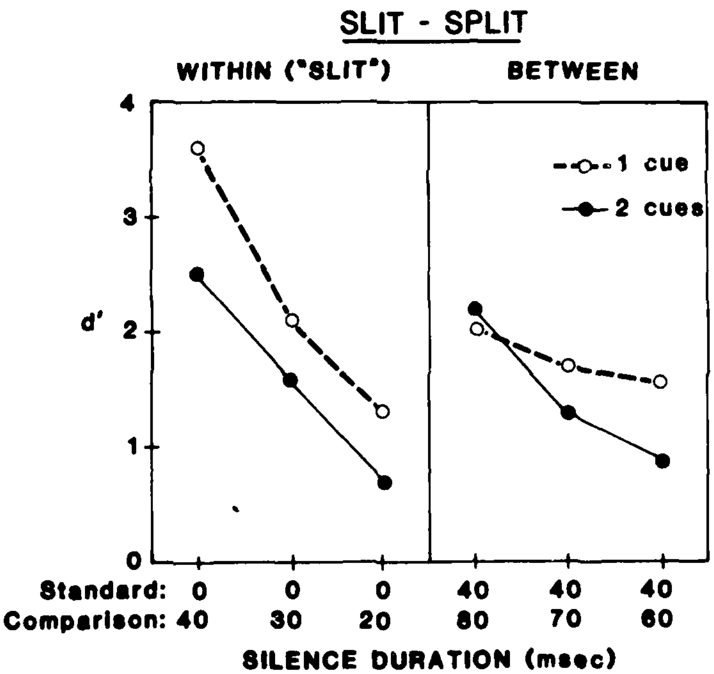


Figure 5. Discrimination results of Experiment 2, Part a.

silences of 0 and 140 msec, were unambiguous instances of "slit" and "split," respectively, as intended.

The average results of the discrimination tests of Part a are shown in Figure 5. The Within "slit" condition is shown in the left panel and the Between condition is shown in the right panel. In the Between condition, the expected trading relation was initially absent but emerged in the second and third test blocks, $F(2,16) = 4.6$, $p < .05$, for the Cues-by-Blocks interaction; $F(1,8) = 3.3$, $p > .05$, for the Cues main effect. The reason for this interaction is not known. The Within data are surprising in that they, too, reveal the trading relation in form of a consistent 1-cue superiority, $F(1,8) = 8.1$, $p < .05$. The Conditions-by-Cues interaction was not significant, $F(1,8) = 1.7$, indicating similar patterns of results in the two conditions. The overall advantage for 1-cue trials was significant, $F(1,8) = 9.9$, $p < .02$, and so was, of course, the decrease in scores across test blocks, $F(2,16) = 21.7$, $p < .001$. The performance level in the Within condition was remarkably high and similar to that in the Between condition of Experiment 1 (Figure 3, left panel), which had employed the same silence durations.

At first blush, these results look exactly like those expected if the trading relation had a purely psychoacoustic basis. However, the high performance level in the Within condition gives rise to suspicion. Indeed, the author's observations as a pilot subject suggest an alternative interpretation: It seems that the consistent presence of the 0-msec standard on every trial may have acted as an anchor that shifted the phonetic boundary toward rather short values, so that tokens with only 40, 30, and even 20 msec of silence began to sound like "split." If so, the trading relation evident in the Within condition may derive from phonetic perception, rather than from a psychoacoustic interaction. It was for this reason that Part b of the experiment was run. By using standards with silences closer to the boundary and different standards in each test block, it was hoped that anchoring effects might be reduced. The Within "split" condition was added to gather additional information comparable to that obtained in Experiment 1.

The results of Part b are shown in Figure 6. The conditions in the two panels on the left correspond to those in Figure 5. The change in standards had a quite dramatic effect. In the Between condition, performance was better than previously and exhibited a clear trading relation, $F(1,6) = 8.0$, $p < .05$. Performance in the Within ("slit") condition, on the other hand, was much poorer than previously and showed no significant trading relation, $F(1,6) = 1.2$. The poor performance suggests that the subjects could no longer rely on a phonetic criterion. Consequently, the absence of any trading relation may be interpreted as supporting the hypothesis that the trading relation in the Between condition had a phonetic, rather than psychoacoustic origin.

One possible objection to that conclusion, however, which cannot be dismissed at present, is that the secondary cue (the brief release burst at the onset of the strong "blit") was effectively masked by the preceding fricative noise in 0-msec silence stimuli. Since all comparison stimuli in the Within ("slit") condition were of that kind, the secondary cue may simply have had no opportunity to produce any perceptual effects, be they phonetic or psychoacoustic. This objection cannot be raised against the results of the

Within ("split") condition (right-hand panel in Figure 6), however, which strongly resemble those of the Within ("slit") condition: Again, performance was very poor, and there was no difference at all between 1-cue and 2-cue trials. Thus it appears that subjects did not pay any attention to the secondary cue, unlike the Between condition, where that cue made a large difference (cf. also the labeling data in Figure 4). It seems possible that the lack of any secondary-cue effect in the Within ("slit") condition was likewise due to lack of attention, although the possibility of masking remains.

EXPERIMENT 3: "GA"-"KA"

In the final experiment of this series, another attempt was made to assess within-category discrimination at the short end of a temporal continuum. This time, I chose a voice-onset-time (VOT) continuum for stops with a velar place of articulation, whose phonetic boundary tends to lie at relatively long values of VOT (Lisker & Abramson, 1970). Since the secondary cue was to be the onset frequency of the F1 transition (cf. Lisker, Liberman, Erickson, Dechovitz, & Mandler, 1977; Summerfield & Haggard, 1977), I returned to synthetic stimuli.

Method

The stimuli were created on the Haskins Laboratories parallel resonance synthesizer.[1] All stimuli were 250 msec in duration, had a linearly falling fundamental frequency contour and linear 50-msec formant transitions that, in the case of F2 and F3, went from 1764 to 1230 Hz and from 2025 to 2527 Hz, respectively. The primary cue varied was VOT, i.e., the duration of the initial aspiration phase during which F1 was turned off. The secondary cue was the linear F1 transition whose onset frequency, duration, and extent differed between two versions: In short-transition (high F1 onset) stimuli, F1 started at 407 Hz and reached 765 Hz after 50 msec; in long-transition (low F1 onset) stimuli, it started at 279 Hz and reached 765 Hz after 70 msec, given a VOT of 0 msec. At longer VOTs, F1 started at correspondingly higher values. The two F1 trajectories were chosen so as to have the same slope, making the magnitude of the secondary cue difference constant for different values of the primary cue (VOT).

Because Experiment 2 had revealed strong effects of the choice of standard, the present experimental tapes were immediately recorded in two versions. In the Between condition, version A, the standard had 20 msec of aspiration and a short F1 transition, and the comparison stimuli had VOTs of 40, 35, and 30 msec. In version B, the standard had 50 msec of aspiration and a long F1 transition (of which only the last 20 msec remained, of course), and the comparisons had VOTs of 30, 35, and 40 msec. In the Within ("ga") condition, version A, the standard had no aspiration and a short F1 transition, while the comparisons had VOTs of 20, 15, and 10 msec. In version B, the standard had 20 msec of aspiration and a long F1 transition, and the comparisons had VOTs of 0, 5, and 10 msec. Note that the B versions differed from the corresponding conditions in Experiment 2, Part b, in that the standards were held constant through all test blocks, while the comparisons

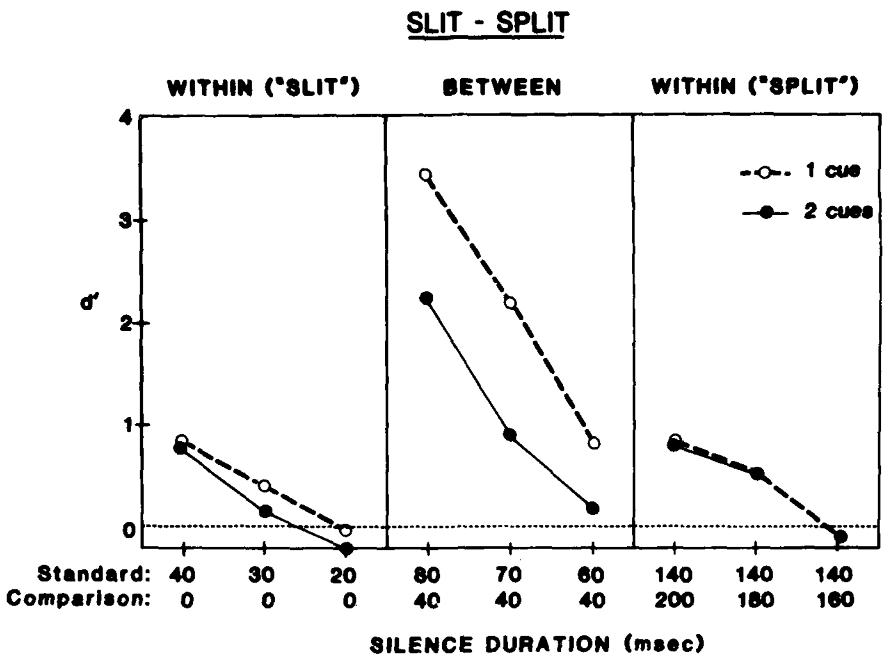


Figure 6. Discrimination results of Experiment 2, Part b.

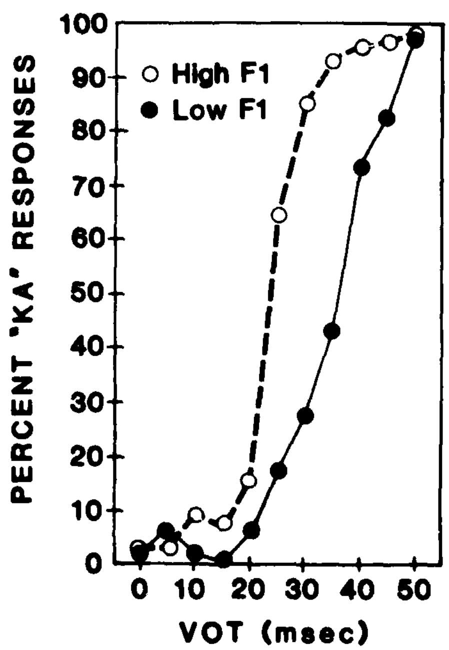


Figure 7. Identification results of Experiment 3.

changed from block to block; this resulted in some differences in the precise VOT comparisons used in versions A and B. A Within ("ka") condition could not be included with these stimuli, for the F1 transition did not extend sufficiently into the "ka" category.

A separate identification test included 10 random sequences of long- and short-transition stimuli with VOTs ranging from 0 to 50 msec in 5-msec steps. Ten paid volunteers participated, four of whom had also taken Part b of Experiment 2. Five subjects took version A, and five took version B. The data of one additional subject were discarded because he apparently wrote "same" for "different" (and vice versa) during part of the experiment and responded randomly elsewhere.

Results and Discussion

Figure 7 shows the identification results. The expected trading relation was clearly present, with category boundaries at approximately 23 and 36 msec of VOT for high and low F1 onsets, respectively.

The results of the discrimination tests are shown in Figure 8. They are plotted separately for versions A (top panels) and B (bottom panels) of the tests, not only because the VOT comparisons were slightly different but also because one of the strongest effects in the overall analysis of variance was the Cues by Versions interaction, $F(1,8) = 26.9$, $p < .001$, which suggested that the relationship between scores for 1-cue and 2-cue trials changed across versions. No other interaction with Versions was significant. The overall analysis also revealed a highly significant Conditions by Cues interaction, $F(1,8) = 33.6$, $p < .001$, which indicates that the pattern of results was different for the Within and Between conditions.

Both these effects are evident in Figure 8. Overall, performance was better on 2-cue trials than on 1-cue trials in version A, while the opposite held in version B. Two-cue trials enjoyed a relative advantage in the Within condition, while 1-cue trials were favored in the Between condition. The last-mentioned finding, of course, is the expected phonetic trading relation; because of the strong Cues by Versions interaction, it was small and nonsignificant in version A but large and significant, $F(1,4) = 12.4$, $p < .05$, in version B. In the Within condition, on the other hand, there was a large 2-cue superiority in version A, $F(1,4) = 52.9$, $p < .01$, but no difference whatsoever in version B. Note also the unexpectedly high level of performance in the Within condition in both versions.

These data present some problems for interpretation, but they are quite clear on the main point: There was no sign of any trading relation in the Within condition. When the trading relation was present in the Between condition, it disappeared in the Within condition (version B); when it was absent in the Between condition, a large advantage for 2-cue trials emerged in the Within condition (version A). This pattern of results suggests that the trading relation between F1 onset and VOT is not psychoacoustic in origin (cf. Summerfield, 1982).

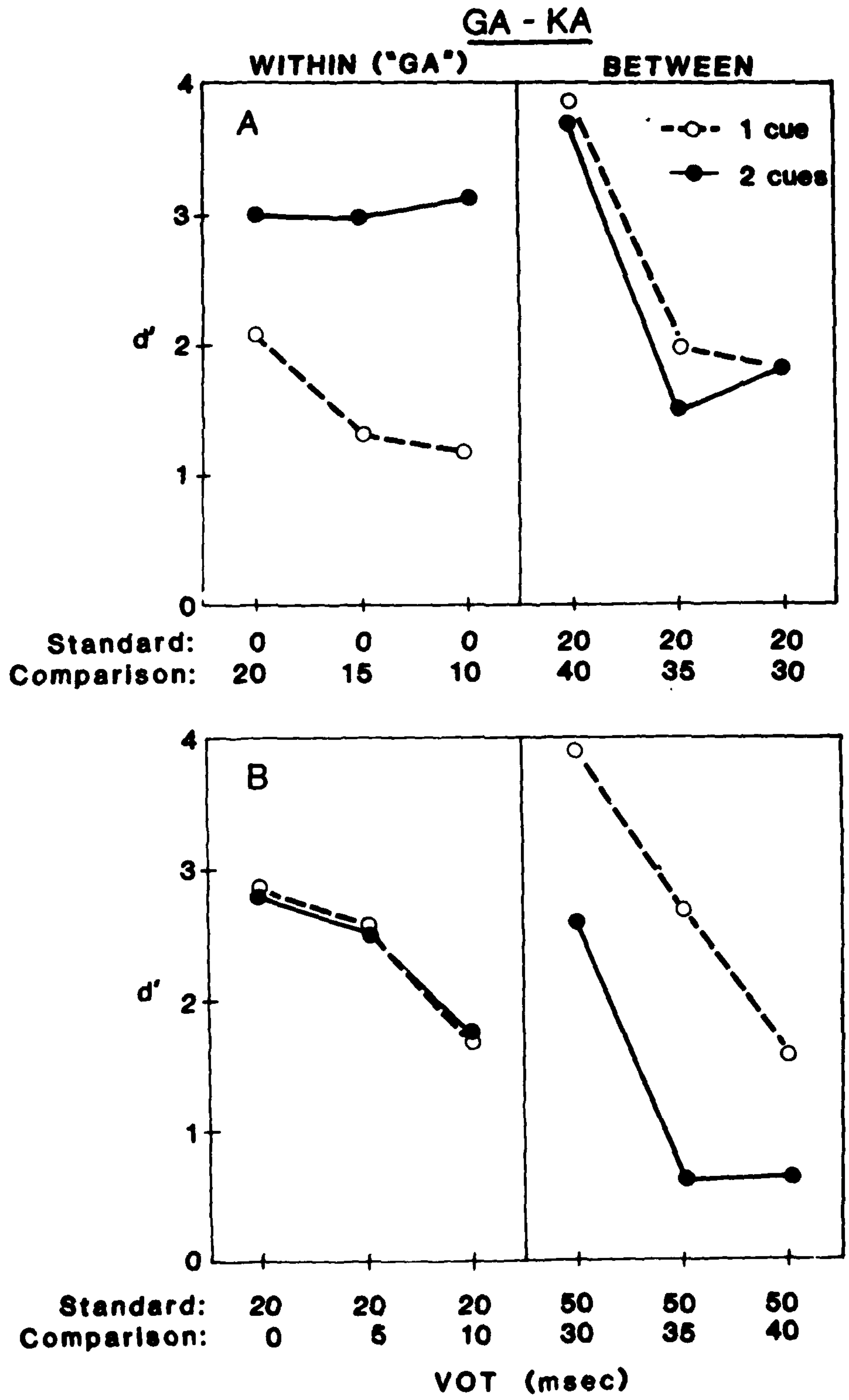


Figure 8. Discrimination results of Experiment 3.

One aspect of the present experiment that has not been considered so far is that, in contrast to the previous studies in this series, the primary and secondary cues were not independent. As VOT increased, the effective onset frequency of F1 rose and the F1 transition got shorter. A quick calculation shows that, in all conditions, the differences in F1 onset frequency between the standard and comparison stimuli were larger on 1-cue trials than on 2-cue trials. In fact, the stimuli on 2-cue trials should have been nearly indistinguishable on the basis of F1 onset or duration alone. This contrasts with the large advantage for 2-cue trials in the Within condition, version A, suggesting that these stimuli were discriminated on a basis other than F1 onset. Note also the absence of a decline in 2-cue discrimination scores over test blocks in that condition, which suggests that the secondary cue that caught the subjects' attention was independent of VOT. The only aspect of the secondary cue that was indeed independent of VOT in the short range was its final portion--the point at which F1 reached asymptote relative to the higher formants. This aspect of the stimuli may have been auditorily salient in the Within condition, even though it is apparently not an important factor in phonetic classification (Summerfield & Haggard, 1977). Why it was so much more salient in version A than in version B, where subjects seemed to attend only to the temporal aspect of VOT, is still a mystery. Considering the small number of subjects, however, it may simply have been a difference in listener strategies that was unrelated to the particular arrangement of stimuli.

GENERAL DISCUSSION

The present three studies extend the four experiments reported by Repp (1981). Although each experiment in this series has its own individual problems, the cumulative evidence does favor the hypothesis that most trading relations between acoustic cues in phonetic perception are phonetically conditioned. That is, they are a direct consequence of distinguishing between members of phonetic categories that are defined by a multiplicity of acoustic attributes. There is no convincing evidence for any significant psychoacoustic interactions between any of the cues varied, with the sole exception of VOT and aspiration amplitude (Repp, 1981: Exp. 3), which also was the only case in which a trading relation was expected to be psychoacoustic in nature.

To summarize the present findings: Experiment 1 investigated the trading relation between silence duration and presence/absence of release burst as cues to the stop manner contrast. While the trading relation was obtained in the Between condition, it was reversed in the Within condition. Because of the unexpected magnitude of the trading relation in identification, subjects may have applied a phonetic strategy in both conditions. The reversal in the trading relation across conditions was shown to be consistent with that hypothesis. The results are also consistent with the hypothesis that the subjects followed an auditory strategy in the Within condition, different from the phonetic strategy used in the Between condition. However, the results are not consistent with the hypothesis that the same auditory strategy was followed in both conditions, for in this case the pattern of results should have been similar in the two conditions. It may be concluded that the trading relation is either phonetic in origin or, if due to a psychoacoustic interaction, specifically limited to the phonetic boundary region.

Experiment 2, varying similar cues, focused on the within-category region at short values of the primary, temporal cue. At first, a similar trading relation was found in the Between and Within conditions. While this result seemed to lend support to the psychoacoustic hypothesis, it was argued that it may have resulted from a phonetic boundary shift due to anchoring in the Within condition, which thereby became another Between condition. Indeed, a change in stimulus arrangement eliminated the trading relation in the Within condition. An added Within condition using long values of the primary cue likewise yielded no trading relation. These results support the hypothesis that the trading relation is of phonetic origin.

Experiment 3 focused on the trading relation between VOT and F1 onset frequency as cues to the voicing contrast, using short values of VOT for the Within condition. Although the results showed some striking effects of stimulus arrangement, overall the trading relation was obtained in the Between condition but was reversed in the Within condition, thus lending further support to the phonetic hypothesis.

In the Introduction, it was pointed out that the phonetic hypothesis, which maintains that trading relations are a byproduct of phonetic categorization, cannot be clearly distinguished from a version of the psychoacoustic hypothesis that postulates that trading relations are due to auditory interactions occurring only at the phonetic boundary. However, this second hypothesis is weakened by at least two considerations. One emerges from the data of Experiments 2 and 3, which suggest that the trading relations studied disappear not only at relatively long values of the temporal dimension (which may suggest the involvement of a temporal threshold or masking) but also at the shortest values of the same dimension. A psychoacoustic explanation of these findings would have to be quite involved, although it is perhaps not impossible. The second, more serious problem for the boundary-specific psychoacoustic hypothesis is, however, that it rests on the assumption that the placement of the phonetic boundary is itself psychoacoustically conditioned--i.e., that it represents an auditory threshold of some sort (Pisoni, 1977; Pastore et al., 1977; Schouten, 1980). However, there is now ample evidence that linguistic category boundaries, while limited in certain ways by auditory acuity, are placed in accordance with the acoustic-phonetic characteristics of a particular language and, moreover, are flexible under a variety of conditions (Repp & Liberman, Note 1). That is, the location of the boundary is itself phonetically conditioned and therefore cannot be part of a purely psychoacoustic hypothesis.

In conclusion, then, the present data lend support to the classic dual-process view of speech perception (in the laboratory), as proposed by Fujisaki and Kawashima (1969, 1970) and Pisoni (1973) and reaffirmed by such recent authors as Samuel (1977), Soli (in press), and Repp (in press). Within the confines of the auditory perceptual system, these two processes represent the bottom-up and top-down components. (Models of word recognition typically lump both together under the heading of bottom-up.) The phonetic component is top-down because it represents the contribution to perception of the past experience of the individual--of the phonetic category prototypes established through speaking and listening. The auditory, bottom-up component, which includes interactions and nonlinearities of various sorts, merely provides the raw material on which the interpretive phonetic component operates. Therefore, to say that a specific trading relation is phonetic in origin is quite analogous to saying that the word "apple" refers to the edible object not because of its acoustic (or even phonetic) properties but because the listener knows the word and its meaning. Once this is acknowledged, phonetic trading relations become merely one of many byproducts of categorical perception in the laboratory whose detailed investigation promises few new insights. Rather, the important questions for theoretical and empirical study become the acquisition of phonetic categories and how to conceptualize their internal representation.

REFERENCE NOTE

1. Repp, B. H., & Liberman, A. M. Phonetic category boundaries are flexible. Paper in preparation.

REFERENCES

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

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

Fujisaki, H., & Kawashima, T. On the modes and mechanisms of speech perception. Annual Report of the Engineering Research Institute (Faculty of Engineering, University of Tokyo), 1969, 28, 67-73.

Fujisaki, H., & Kawashima, T. Some experiments on speech perception and a model for the perceptual mechanism. Annual Report of the Engineering Research Institute (Faculty of Engineering, University of Tokyo), 1970, 29, 207-214.

Lisker, L., & Abramson, A. S. The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences. Prague: Academia, 1970, pp. 563-567.

Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D., & Mandler, R. On pushing the voice-onset-time (VOT) boundary about. Language and Speech, 1977, 20, 209-216.

Miller, J. D., Wier, C. C., Pastore, R., Kelly, W. J., & Dooling, R. J. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. Journal of the Acoustical Society of America, 1976, 60, 410-417.

Pastore, R. E., Ahroon, W. A., Baffuto, K. J., Friedman, C., Puleo, J. S., & Fink, E. A. Common-factor model of categorical perception. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 686-696.

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

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

Repp, B. H. Phonetic and auditory trading relations between acoustic cues in speech perception: Preliminary results. Haskins Laboratories Status Report on Speech Research, 1981, SR-67/68, 165-189.

Repp, B. H. Limits on the power of silence as a stop manner cue. Journal of the Acoustical Society of America, 1982, 72 (Supplement No. 1), S16. (Abstract)

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

Repp, B. H., Healy, A. F., & Crowder, R. G. Categories and context in the perception of isolated steady-state vowels. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 129-145.

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

Schouten, M. E. H. The case against a speech mode of perception. Acta Psychologica, 1980, 44, 71-98.

Soli, S. D. The role of spectral cues in the discrimination of voice onset time. Journal of the Acoustical Society of America, in press.

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

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

FOOTNOTE

[1]To the best of my knowledge, these were the last stimuli created on that distinguished instrument before it went out of commission in May 1982. A serial synthesizer was avoided because of the amplitude changes consequent upon changes in F1 frequency.

LINGUISTIC CODING BY DEAF CHILDREN IN RELATION TO BEGINNING READING SUCCESS

Vicki L. Hanson, Isabelle Y. Liberman,+ and Donald Shankweiler+

Abstract. The coding of printed letters in a task of consonant recall was examined in relation to the level of success of prelingually and profoundly deaf children (median age 8.75 years) in beginning reading. As determined by recall errors, the deaf children who were classified as good readers appeared to use both speech and fingerspelling (manual) codes in short-term retention of printed letters. In contrast, deaf children classified as poor readers did not show influence of either of these linguistically-based codes in recall. Thus, the success of deaf children in beginning reading, like that of hearing children, appears to be related to the ability to establish and make use of linguistically-recoded representations of the language. Neither group showed evidence of dependence on visual cues for recall.

To be able to comprehend text, a reader must hold several words, and their order, in short-term memory long enough for sentence interpretation. The nature of this short-term memory store is a matter of considerable interest. For hearing children, research evidence suggests that success in beginning reading is related to ability to make efficient use of a speech-based code.[1] In tests of short-term memory, hearing second graders who are good readers have been found to be more sensitive to this information than those who are poor readers. For example, in a test of the recall of printed consonant strings, the performance of second grade good readers was found to differ significantly for rhyming and nonrhyming strings (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977). For the poor readers, in

---

+Also University of Connecticut.
Acknowledgment. The research was supported by Grant NS-18010 from the National Institute of Neurological and Communicative Disorders and Stroke and by Grant HD-01994 from the National Institute of Child Health and Human Development. An earlier version of this paper was presented at the Seventh Annual Boston University Conference on Language Development. We wish to thank John Richards, Alvin Liberman, Carol Fowler, and Brian Davidson for valuable comments on earlier drafts of the paper and Margo Carter for drawing the handshapes in Figure 1. We are grateful to the staff and students of the American School for the Deaf in West Hartford, Connecticut for their cooperation with the project. In particular, we wish to express our appreciation to Lynn McGill for her considerable help.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

contrast, performance was similar in the two cases. The difference in error pattern was attributed to the good readers' greater or more efficient use of a speech-based code. This result has been obtained not only with printed letter presentation, but also when the letters' names were spoken (Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). Similar results have also been obtained in tasks of recognition memory for words. Good readers are more likely than poor readers to make errors in recognizing words that rhyme with earlier-occurring words, whether the words are heard (Byrne & Shea, 1979) or read (Mark, Shankweiler, Liberman, & Fowler, 1977). These findings have suggested that for hearing children in the process of acquiring reading skills, the poor readers may be deficient in the use of a speech-based code.

The present research examines short-term memory coding as it relates to the beginning reading success of prelingually, profoundly deaf children. The most comprehensive work that has been done to date on reading in deaf populations is an extensive study by Conrad (1979) of older hearing-impaired students (ages 15-16.5) in England and Wales. In that study, three factors were found to be determinants of reading success: degree of hearing loss, level of intelligence, and use of a speech-based code. Of these factors, the latter is of particular relevance here.

The use of a speech-based code was assessed by Conrad by means of a short-term memory task in which the students were presented short lists of rhyming words (e.g., _do_, _blue_, and _through_) and nonrhyming words (e.g., _bean_, _door_, and _farm_). Students were considered to be using a speech-based code if they made more errors on rhyming lists than on nonrhyming lists. Degree of hearing loss was found to be related to reading achievement (those persons having a loss of 85 dB or greater showing a marked deficiency in reading achievement), but success in reading for a given degree of hearing loss was largely determined by the use of a speech-based code. Individuals who made use of this code tended to be better readers than those who did not. Although the ability to use a speech-based code was correlated with degree of hearing loss and intelligence, use of a speech-based code was also an independent determiner of reading success.

It is of further interest to note that the majority of the profoundly deaf students in Conrad's study had not acquired the use of a speech-based code and, moreover, that those profoundly deaf students who had acquired it were using it less efficiently than their hearing counterparts. This latter finding accords well with results obtained with deaf college students (Hanson, 1982). The question therefore arises as to whether alternative coding strategies might be in use by deaf readers. The most obvious available alternative strategy is a manually-based code. Its use could not be assessed in Conrad's study since the schools from which he drew his subjects were strictly oral in their educational approach.

Research with deaf subjects has indicated that internal representations based on manual language systems can be used in short-term memory. Studies using American Sign Language (ASL) have shown that when sign stimuli are presented to skilled users, short-term recall is mediated by a sign-based code. It has been demonstrated that, for deaf adults, intrusion errors in sign recall tend to be formationally related to sign parameters (Bellugi, Klima, & Siple, 1975). Thus, for example, an error in the recall of the sign

NOON might be the word tree. The ASL sign for TREE is similar to the sign NOON in handshape and place of articulation and differs only in movement. Deaf subjects have also been found to have more difficulty in recalling lists of signs that are formationally similar than lists of unrelated signs (Hanson, 1982; Poizner, Bellugi, & Tweney, 1981; Shand, 1982). Similarly, deaf children tested with a continuous recognition memory procedure tended to recognize formationally similar signs falsely (Frumkin & Anisfeld, 1977).

However, the important question of how a manual short-term memory code might relate to the acquisition of reading in young children has remained largely unexplored. Research with deaf teenage and adult signers has examined short-term memory coding of written letters and words, but these studies have not examined how coding strategy relates to reading success. The results have been somewhat inconsistent in their indications; some studies finding evidence for speech-based coding (Hanson, 1982; Locke & Locke, 1971; Novikova, 1966; Wallace & Corballis, 1973) and others finding evidence of manually-based coding (Conlin & Paivio, 1975; Locke & Locke, 1971; Moulton & Beasley, 1975; Odom, Blanton, & McIntyre, 1970; Shand, 1982). Such variety in outcome is understandable given the differences in subject background characteristics (e.g., degree of hearing loss, educational achievement, and age) and the varied methodologies employed.

Short-term memory coding has been examined in deaf children (Frumkin & Anisfeld, 1977; Liben & Drury, 1977), but once again not in relation to reading success. Deaf children receiving oral education, tested in a task of recognition memory for printed words, have been found to make semantic errors in a task of recognition memory for printed words, as well as making visual/phonetic errors (Frumkin & Anisfeld, 1977). Since visual and phonetic similarity were confounded in the study (as in their stimuli TOY-BOY, MAKE-TAKE), it is impossible to know whether it was phonetic similarity or visual similarity, or both, that led to the errors. Deaf children educated with the Rochester Method, which uses simultaneous speech and fingerspelling, have been observed using simultaneous speech and dactylic rehearsal in a task of short-term memory for printed letters (Liben & Drury, 1977).

The present research examines stort-term memory for written material by young children just beginning to acquire reading skills. Though it derives its motivation from Conrad's (1979) seminal work, it departs from that work in two major respects. First, the children under study are beginning readers, whereas Conrad tested students about to graduate from high school. Secondly, the children have been instructed with simultaneous speech and manual communication, whereas Conrad's subjects had received only oral instruction.

The procedure follows the format of previous studies of short-term memory in which printed strings of letters, varying in their phonetic similarity (rhyming or nonrhyming), are presented for recall by good and poor beginning readers (Liberman et al., 1977; Shankweiler et al., 1979). The task here is expanded by also including stimuli varying in their manual and visual similarity. In selecting items for the manually similar strings of letters, it was, of course, necessary to base similarity on the handshapes of fingerspelling, not on the signs of ASL. That is because the signs of ASL correspond, not to letters, but very roughly to English at the whole-word level (see Klima & Bellugi, 1979). Fingerspelling, as its name implies, is a dactylic system based on a manual alphabet. In the American manual alphabet there is a one-handed configuration for each of the 26 letters of the English alphabet. Words are manually spelled out in fingerspelling by the sequential production of each letter of the word. Fingerspelling thus provides a manual system for representing the orthography of English.

In the present research, the recall of strings of consonants that are phonetically, manually (dactylically), or visually similar was compared to recall of unrelated (control) strings. Differential ability to recall a given experimental set will be presumed to reflect coding strategies in short-term memory. In short-term memory studies, similarity typically produces performance decrements compared with a control condition in which the stimulus items are dissimilar (e.g., Baddeley, 1966; Conrad & Hull, 1964). To anticipate our results, we should note that the procedure of the present experiment differs from the typical short-term memory task in one respect: Each experimental set of letters was limited to only four consonants; moreover, all four consonants of a set were presented on each trial of testing with a set. It might be expected that such a procedure would influence the pattern of results. As will be seen, this was indeed the case. With this repeated presentation of the same sets of consonants, similarity produced improvement in performance relative to the control set, instead of a decrement in performance.

METHOD

Subjects

Background information necessary for subject selection was obtained from the detailed records kept by the school for the deaf where the subjects were enrolled as students. In order to be accepted as subjects, the children had to meet several stringent selection criteria. The criteria required that a child be both prelingually and profoundly deaf (hearing loss of 85 dB or greater in the better ear) and of average or above average intelligence. Children with handicapping conditions other than hearing loss were excluded. The number of children meeting these criteria even at a school for the deaf was limited. A further limiting factor was that only children returning parent permission forms could be included in the study. The experimental subject group finally included 17 children. One was dropped from the study due to unwillingness to complete the task. The remaining 16 subjects were distributed as follows: four children were in a Preparatory class, three in first grade, three in second grade, and six in third grade. The school attended by the subjects uses a Total Communication approach to instruction.

An additional prerequisite for subject selection was that the child know the names of letters of the printed alphabet and know the correspondence between each printed letter and its dactylic representation. The students' teachers were consulted in this regard.

The ratings by the school's Reading Diagnostician were used to differentiate groups of good and poor readers. These ratings were based on the children's measured reading achievement in relation to their ages. The reading achievement results were from the Woodcock Reading Mastery Test for the four youngest children and from the Stanford Achievement Test - Hearing Impaired for all other children. By these criteria, ten of the children were classified as good readers, six as poor readers. Although averaging over results from two different tests is not strictly legal, for purposes of providing a descripton of the reading abilities of these children, such averaging was undertaken. For the good readers, the mean reading achievement was grade 2.2; for the poor readers, grade 1.8. By an analysis of covariance with age as the covariate, this difference in reading ability between the two groups was significant, $F(1,13) = 12.12$, $p < .005$.

Additional background information was obtained regarding each subject's age, speech production skills, and parents' hearing status. The speech intelligibility of each child was based on the ratings of a Speech Pathologist at the school on a scale of 1 to 5 in which 5 represents speech that is completely intelligible and 1 represents speech that is completely unintelligible. The subjects in the good and poor reader groups did not differ significantly in their rated speech intelligibility, $t(14)=.36$, $p > .20$.

A summary of these background characteristics of the subject groups is given in Table 1. For the children in the Preparatory class and in first grade, the IQ score was a combined measure based on the Hiskey-Nebraska Test of Learning Aptitude and the child's chronological age. For the children in the second and third grades, the IQ score was a combined measure based on the performance section of the Wechsler Intelligence Scale for Children (Revised) and the child's chronological age. Since scores for age and IQ were markedly skewed, median scores are presented. Median levels of hearing loss are also presented since mean averages of such scores would be nonsensical.

Four of the subjects had deaf parents; all four were classified as good readers. One subject, classified as a poor reader, had an older deaf sibling.

Table 1

Characteristics of good and poor readers

	Hearing loss (dB)[a]	Age[a]	IQ[a]	Speech Intelligibility[b]
Good readers				
Score	101	8.5	105	2.3
Range	87-110+	6.25-11.0	88-143	1-4
Poor readers				
Score	103.5	9.3	97	2.1
Range	85-107	7.5-11.33	87-111	1-4

(Note: [a]median score; [b]mean score)

Stimuli

The stimuli were individual letters of the alphabet. To examine the possible effects of phonetic, dactylic, and visual similarity, sets of consonants related along each of these dimensions were constructed. In constructing sets that vary in similarity along three dimensions, it is to be expected that the degree of similarity between dimensions may vary. Thus, it may be argued, for example, that the visually similar items are not as similar as the phonetically similar items. Such potential disparity in relative similarity would be difficult to assess reliably and, for now, will not be considered.

Due to the limitations of a 26-letter alphabet and a need to manipulate phonetic, dactylic, and visual similarity independently, it was necessary to modify the procedure of earlier studies somewhat (Conrad, 1972; Liberman et al., 1977; Shankweiler et al., 1979). The major modifications were that sets were limited to only four consonants each and that the same four consonants were presented on each trial using each set.

The phonetically similar set consisted of four rhyming consonants, B C P V, which have been rated as phonetically similar (Wolford & Hollingsworth, 1974) and which are a subset of the stimuli used by others to investigate the use of a phonetic code (e.g., Liberman et al., 1977). The dactylically similar set consisted of the four letters M N S T. The manual handshapes for these letters, which are pictured in Figure 1, have been found to be dactylically similar as rated by adult native signers of ASL (Richards & Hanson, Note 1). The visually similar set consisted of the letters K W X Z, which have been rated as visually similar (Wolford & Hollingsworth, 1974) and are a subset of letters previously used to measure visual coding (Conrad, 1972). In addition, a control set of four letters, G J R L, was constructed. The letters of this set are dissimilar along all three dimensions studied here.

As much as possible, letters of each set were selected to be similar only along the relevant dimension. That is, for example, the letters of the visually similar set were selected to be dactylically and phonetically dissimilar. There were unavoidably some confoundings, however, if sets truly high in phonetic and dactylic similarity were to be used. The alphabet does not permit a complete independence of phonetic, dactylic and visual similarity. As a result, in the phonetically similar set the letters B and P are also visually similar (Wolford & Hollingsworth, 1974), and in the dactylically similar set the letters N and M are also phonetically and visually similar (Wolford & Hollingsworth, 1974).

While these stimuli were chosen on the basis of judged similarity in sorting tasks (Wolford & Hollingsworth, 1974; Richards & Hanson, Note 1), their similarity can be evaluated on the basis of confusability scores from other studies on auditory, dactylic, and visual perception. As shown in Table 2, the measured auditory confusability is highest for the phonetically similar set, the measured dactylic confusability is highest for the dactylically similar set, and the measured visual confusability is highest for the visually similar set. The confounding of phonetic similarity and dactylic similarity on the letters M and N is apparent in these confusability ratings. The

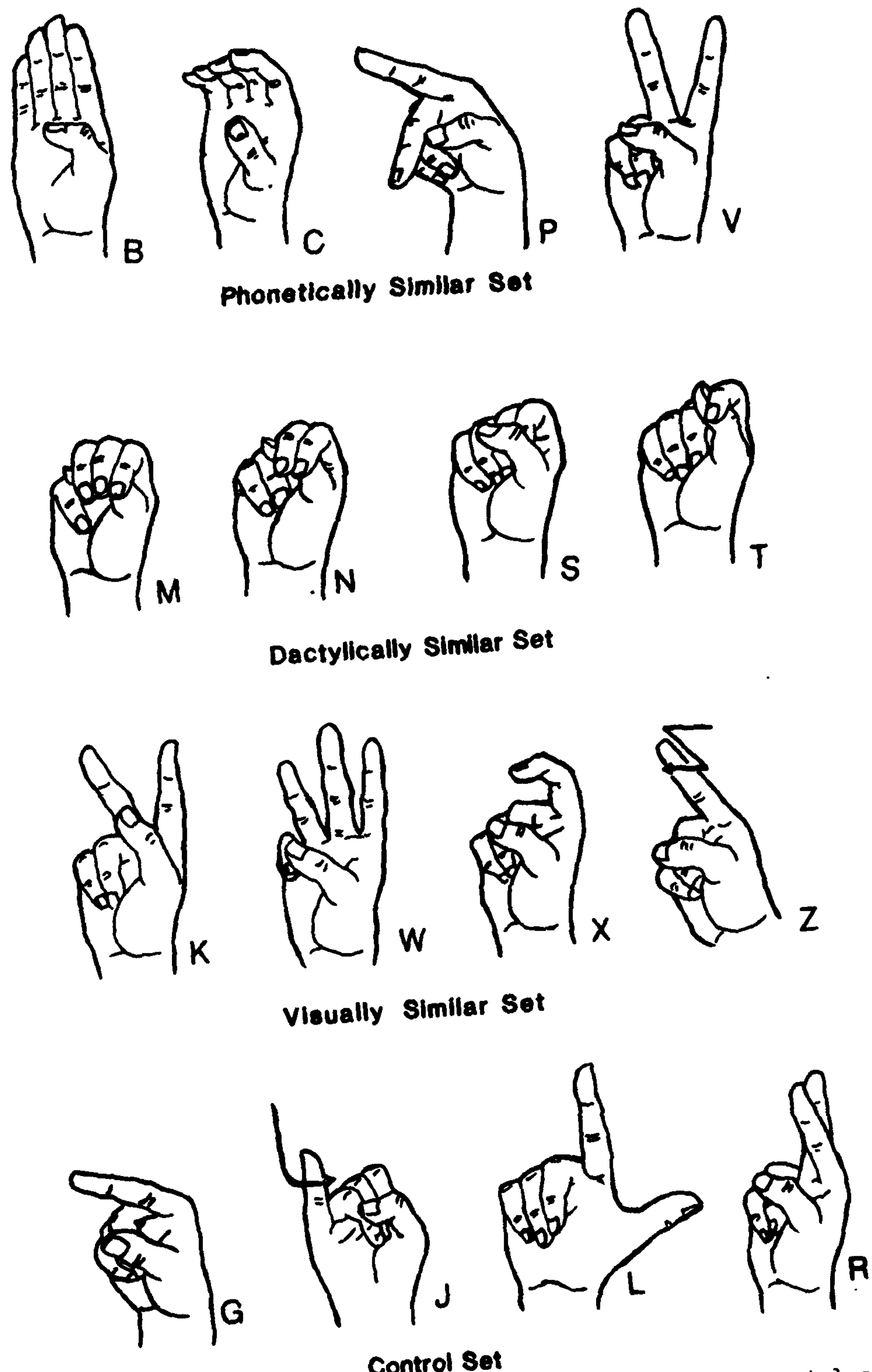


Figure 1. The handshapes of the letters in the four experimental sets.

similarity of M and N account for 86% of the auditory confusability of the dactylically similar set. Thus, the relatively high auditory confusability of the dactylically similar set results from the confusability of these two letters. The auditory confusability of these two letters with the other letters of the dactylically similar set, however, is low.

Table 2

Auditory, dactylic, and visual confusions of the four stimulus sets based on previous studies.

	Auditory[a] Confusions	Dactylic[b] Confus ions	Visual[c] Confusions
Phonetically similar set			
BCPV	1321 (45.2%)	2 (1.4%)	8 (18.6%)
Dactylically similar set			
MNST	989 (33.8%)	121 (86.4%)	8 (18.6%)
[MN	846 (28.9%)]		
Visually similar set			
KWXZ	294 (10.0%)	16 (11.4%)	21 (48.8%)
Control set			
GJLR	321 (11.0%)	1 (.7%)	6 (14.0%)
Total	2925 (100%)	140 (100%)	43 (100%)

[a]From Conrad (1964)
[b]From Weyer (Note 2)
[c]From Fisher, Monty, & Glucksberg (1969), 400 msec presentation

---

The test consisted of 16 trials--four presentations of each of the four sets of stimuli. Each letter of a set appeared once in each of the four possible serial positions. Trials were randomized with the constraint that the same stimulus set was not tested on consecutive trials.

Each letter was typed in uppercase and slides of the individual letters were made.

Procedure

Stimuli were presented at the rate of one consonant every 2 sec. That is, each slide was displayed for 1 sec with a 1 sec blank interval following.

The children, who were tested individually, were instructed that on each trial they would see four letters, one after the other. They were to watch carefully as each cf the four letters was presented and try to remember the letters in order. Following presentation of the items, they were to write them in correct order in their answer booklets. The answer booklets were prepared so that answers to each trial could be written on a separate page. On each page, four lines were drawn to indicate that four letters were to be recalled. Two practice trials were presented, using letters not appearing in the four stimulus sets. *Instructions were simultaneously signed* and spoken by the experimenter.

RESULTS

Responses were scored in two ways: order-strict scoring, in which a response was considered correct only if the correct letter appeared in the correct serial position; and order-free scoring, in which a response was considered correct if a correct letter for that trial was written, regardless of serial position. The mean number of errors for the two reader groups in each condition for both scoring procedures is shown in Table 3. The two scoring procedures produced a similar pattern of results; An analysis of variance performed on the number of errors for the between-subjects factor of group (good or poor readers) by the within-subjects factors of stimulus set (phonetic, dactylic, visual, or control sets) and scoring procedure (order-strict or order-free scoring) produced no significant interactions involving scoring procedure ($p>.25$). There was, however, a main effect of scoring procedure, $F(1,14)=55.40$, $p<.001$, with significantly more errors occurring in the order-strict than in the order-free scoring.

Table 3

Mean number of errors (out of 16 possible) for good and poor readers. Given in parentheses are the standard deviations.

	Phonetically Similar Lists	Dactylically Similar Lists	Visually Similar Lists	Control Lists
Good readers				
Order free	3.5 (3.1)	3.6 (3.1)	5.8 (3.8)	5.8 (3.8)
Order strict	5.7 (4.1)	6.0 (4.4)	8.2 (3.4)	7.5 (5.7)
Poor readers				
Order free	7.5 (4.6)	6.7 (4.2)	6.5 (3.6)	7.3 (3.7)
Order strict	10.0 (7.5)	9.3 (5.2)	9.2 (4.2)	11.0 (5.1)

Good and poor readers were found to be differentially affected by the four stimulus sets as evidenced by a significant interaction of group by stimulus set, $F(3,42)=3.71$, $p<.025$. Post hoc tests were conducted to determine the basis of this interaction. An analysis on the simple effects indicated a significant effect of stimulus set for the good readers,

$F(3,42)=7.71$, $p<.001$, but no significant effect of stimulus set for the poor readers, $F(3,42)=1.20$, $p > .25$. Thus, performance of the poor readers did not significantly vary as a function of stimulus set. For the good readers, in contrast, accuracy for the phonetically and dactylically similar sets was significantly greater than accuracy on the control set (Dunnett's *t*-statistic, $p<.05$, two-tailed). Performance of the good readers on the visually similar set was not significantly different from the control (Dunnett's *t*-statistic, $p > .05$, two-tailed).

An analysis was also undertaken of the types of errors made by good and poor readers. For the responses on the phonetically similar trials, the number of responses that rhymed with the target set was tabulated. These responses were the five letters D, E, G, T, and Z. Using the order-free scoring procedure, 55% of the errors made by the good readers on the phonetically similar set were responses that rhymed with the target set. For the poor readers, only 27.2% of such errors rhymed with the target set. Since a chance response with one of the 22 letters not from the phonetically similar set would produce rhymes for five of the letters (22.7% of the responses), it is apparent that the poor readers were responding randomly when they made an error, while the good readers tended to respond with a letter related to the target set. The dactylically similar set is less suitable than the phonetically similar set for such an analysis because the only two letters that are manually very similar are A and E, both vowels (Richards & Hanson, Note 1; Weyer, Note 2). Since vowels never occurred in the experiment, it might be expected that subjects would have a reluctance to respond with vowels. The pattern of results with the dactylically similar set was, however, consistent with the results of the phonetically similar set: With chance at 9.1%, the errors of the good readers were dactylically related to the target set 22.2% of the time, while the errors of the poor readers were, again, exactly at chance, with a related letter only 9.1% of the time. Thus, the error analysis on the phonetically and dactylically similar sets indicates that only the good readers made errors based on the linguistic similarity of the target sets.

An analysis of the individual responses of good readers is relevant to the question of whether the improved performance of the good readers on the dactylically similar set can be attributed primarily to the phonetic similarity of the letters M and N in that set. This analysis revealed that the improvement was not due solely to better recall of only these two letters. Using the order-free scoring procedure, it was found that the good readers recalled an M on 20% of their responses on dactylically similar test trials, an N on 16% of their responses, an S on 21% of their responses, and a T on 22% of their responses. Thus, it is clearly not the case that the M and N are solely responsible for the improved performance.

Since the good readers vary in age from 6.25 to 11.0 years, it is of interest to determine whether the tendency to use speech-based and manually-based codes changes with age. For hearing children, use of a speech-based code has been shown to increase throughout this age span (Conrad, 1971). For each of the good readers, an index of speech-based and dactylically-based encoding was obtained as the ratio of number of errors with the phonetically or dactylically similar set to the number of errors on the control set. Thus, for example, if a subject made three errors on the phonetically similar sets and four errors on the control sets, the speech encoding index for the subject would be .75. By this measure, the lower the index, the greater the indication of speech encoding. A correlation of -.47 was obtained between age and the speech encoding index, and a correlation of -.56 was obtained between age and the dactylic encoding index. Both of these correlations are in the expected direction in finding that the older the child, the greater the evidence for both speech and dactylic encoding.

Analysis of recall accuracy indicated that use of linguistic coding strategies affected the ability of subjects to recall information about the order in which items were presented. Because a valid comparison of recall accuracy between the two reader groups can only be made on the control sets, these analyses of accuracy were confined to the control sets. It was found that the poor readers were relatively more penalized by order-strict scoring than were the good readers, as demonstrated by a significant interaction of scoring procedure by group in an analysis of the errors, $F(1,14)=5.02$, $p<.05$. To determine the basis of this interaction, additional analyses were undertaken of the accuracy of the two reader groups for the control lists. Since the poor readers were somewhat older than the good readers, an analysis of covariance was performed with age as the covariate. The analysis indicated a significant difference between the groups for order-strict scoring, $F(1,13)=5.08$, $p < .05$, but not for the order-free scoring, $F(1,13)=2.17$, $p > .15$. These results suggest that poor readers have relatively more difficulty than good readers in the recall of order information.

DISCUSSION

The results indicate that the good readers differed from the poor readers in their use of linguistically-based recall strategies. This was shown by the good readers' improved performance on the phonetically and dactylically similar lists as compared with the control lists. In contrast, the performance of poor readers did not vary as a function of stimulus set. Thus, in keeping with results obtained with hearing beginning readers (Byrne & Shea, 1979; Liberman et al., 1977; Mark et al., 1977; Shankweiler et al., 1979), deaf children who are good beginning readers are able to make greater or more efficient use of linguistically-based codes in short-term recall than are deaf children having difficulties in acquiring reading. It should be noted that the better performance of the good readers on the phonetically similar set could not be simply a reflection of differences in speech production capabilities of the good and poor readers. The speech production skills of the two reader groups were not significantly different. This suggests that it is not differences in speech ability, per se, that differentiate good and poor readers, but rather the good readers' more effective use of a short-term memory code based on linguistic features.

The lack of significant influence of linguistic similarity for the poor readers was not due to individual differences among the poor readers obscuring group tendencies. Inspection of the recall errors of the poor readers indicated a consistent pattern--for each of the poor readers, the recall accuracy across the four stimulus sets was comparable. The failure of the accuracy of the poor readers to vary as a function of stimulus set is in marked contrast to the performance of the good readers. The recall accuracy for each of the good readers consistently showed an improvement in <u>both</u> the phonetically and dactylically similar sets as compared with the control.

In the present experiment, phonetic and dactylic similarity were manipulated to investigate potential differences between good and poor readers in linguistic coding. It must be borne in mind that linguistic similarity will facilitate or hinder recall ability depending on task demands. In poetry, for example, as in certain short-term memory tasks (see Watkins, Watkins, & Crowder, 1974), phonetic similarity aids recall. The recall accuracy of the good readers in the present study benefited by the rhyming set, whereas in earlier studies with hearing children the performance of the good readers was penalized by the rhyming set (Liberman et al., 1977; Shankweiler et al., 1979). Since other investigations with deaf subjects have found decrements in serial order recall when sets of words are phonetically similar (Conrad, 1972, 1979; Hanson, 1982; Locke & Locke, 1971; Wallace & Corballis, 1973), it cannot be the case that phonetic similarity affects deaf and hearing subjects differentially. The explanation for the discrepancy between the present results and earlier studies would seem to be differences in procedure. On any *given* trial in a typical short-term memory experiment, the subject is shown only a subset of the set of stimuli. In the present experiment, however, the constraints imposed by the need to manipulate independently the phonetic, dactylic, and visual similarity of the consonant sets limited the available stimuli for each set; on any given trial an entire set of confusable stimuli was presented. If subjects in this situation could determine the similarity principle used in stimulus selection, they could use that principle to aid recall. The finding that good readers, but not poor readers, made errors that were consistent with the target set in the phonetic and dactylic similarity conditions provides strong evidence that the good readers did abstract the linguistic similarity principles used in stimulus list construction and that *they then used this principle* to aid recall. It is just this ability to establish and make use of linguistically-based codes in the recall of letter strings that distinguishes the two groups.

The phonetically similar set consisted of letters whose names were auditorily confusing, but not dactylically or visually confusing. In the construction of the dactylically similar set, however, some confounding was unavoidable. The two letters M and N were also high in auditory confusability. The data nonetheless suggest that this phonetic similarity was not the sole reason for the improvement of the good readers on the dactylically similar set: Though this phonetic similarity applied to only two of the four letters of the dactylically similar set, analyses showed that the improved recall applied to all four letters.

Some comment should be made about the failure to find evidence of the use of visual coding strategies that have so often been considered to be the preferred strategies for deaf individuals (see, for example, Conrad, 1972; Frumkin & Anisfeld, 1977; MacDougall, 1979; Wallace & Corballis, 1973). Caution must always be used in cases of failure to find that the experimental manipulation produces an effect. It is possible that the present experimental situation was inappropriate for detecting a visual strategy, and that such strategies may have been present but were not detected. Although we cannot rule out this possibility altogether, such a possibility does not diminish the major finding of the present study that the good readers differed from the poor readers in their use of linguistically-based codes.

The fact that no evidence was obtained for the poor readers' use of phonetic, dactylic, or visual codes in the present study is consistent with recent findings for hearing children who are poor readers. Although these poor readers are able to recall the letters with better than chance accuracy, when they make an error, their error pattern is random. These findings with poor readers have been interpreted as indicating that poor readers have linguistic codes available to them, but that they make less efficient use of these codes than do good readers (Wolford & Fowler, in press).

In line with such an interpretation, two features of the present study should be noted. First, as indicated earlier, one criterion for subject selection in the present study was that the subjects know the names and handshapes of the letters of the alphabet. Thus, all subjects in the experiment had this linguistic information available to them. Second, the experimenter here observed that nearly all the subjects, whether good readers or poor readers, simultaneously produced the spoken names and the handshapes of the printed letters as each stimulus item was presented. Only the good readers, however, appeared by their performance to have abstracted the system underlying these linguistic performances and to make use of this information in recall. The failure of the deaf poor readers to make effective use of a linguistic representation after deriving the letter names is closely paralleled in research with hearing children. This was demonstrated with hearing beginning readers in a consonant recall task similar to the one used here, in which the children spoke aloud the letter name for each printed letter as it was presented (Wolford & Fowler, in press). In that study, as in the present one, good readers, but not poor readers, displayed errors related to linguistic recall strategies.

The difference between good and poor readers in the use of short-term memory codes was also associated with differences in serial recall ability. The analysis of the control sets demonstrated that the poor readers were relatively more penalized than the good readers by the order-strict scoring procedure. Thus, the poor readers were less able than the good readers to retain information about the order in which items were presented. These results are in accord with research with hearing children in finding that poor readers exhibit specific difficulty in the retention of order information (Katz, Shankweiler, & Liberman, 1981). This difficulty may be understood in terms of the deficient use of a linguistically-based code. It has been hypothesized that a speech-based code is particularly well-suited for carrying information about item order (Baddeley, 1978; Crowder, 1978; Healy, 1975). Indeed, the ability of deaf persons to recall information about order has been found to vary as a function of use of a speech-based code (Conrad, 1979; Hanson, 1982). As the good readers in the present study were found to use both speech-based and manually-based codes, it is not possible here to determine whether it was the speech code alone that was related to ability to recall order information or whether the manual code contributed also. It must remain for future research to determine whether a manually-based code can retain this information as well as a speech-based code.

In summary, the present findings are important in the indications they provide that deaf children need not be limited to reading strategies that involve visual retention; instead they are able to make use of linguistic strategies--derived, it appears, from both spoken and manual language--that could mediate comprehension. Although the language system is accessed via different modalities in the speech-based and manually-based codes used by the good readers, both provide the reader with a means of representing the internal structure of words (see also Hirsh-Pasek, 1981), and, specifically, in terms of the present study, provide a linguistic basis for holding information in short-term memory. These results argue that successful deaf beginning readers differ from their poorly reading deaf counterparts in the use of these linguistic recall strategies. This suggestion is consistent with research on hearing children in indicating that differences in the use of linguistically-based representations in working memory are a relevant factor in learning to read.

REFERENCE NOTES

1. Richards, J. T., & Hanson, V. L. Handshape similarity for the 26 letters of the American manual alphabet. Unpublished data.
2. We er, S. A. Fingerspelling by computer (Technical Report 212). Institute for Mathematical Studies in the Social Sciences, Stanford University, August 17, 1973.

REFERENCES

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

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

Bellugi, U., Klima, E. S., & Siple, P. Remembering in signs. Cognition, 1975, 3, 93-125.

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

Conlin, D., & Paivio, A. The associative learning of the deaf: The effects of word imagery and signability. Memory & Cognition, 1975, 3, 335-340.

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

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

Conrad, R. Short-term memory in the deaf: A test for speech coding. British Journal of Psychology, 1972, 63, 173-180.

Conrad, R. The deaf school child. London: Harper & Row, 1979.

Conrad, R., & Hull, A. J. Information, acoustic confusion, and memory span. British Journal of Psychology, 1964, 55, 429-432.

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

Fisher, D. F., Monty, R. A., & Glucksberg, S. Visual confusion matrices: Fact or artifact? Journal of Psychology, 1969, 71, 111-125.

Frumkin, B., & Anisfeld, M. Semantic and surface codes in the memory of deaf children. Cognitive Psychology, 1977, 9, 475-493.

Hanson, V. L. Short-term recall by deaf signers of American Sign Language: Implications of encoding strategy for order recall. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1982, 8, 572-583.

Healy, A. F. Coding of temporal-spatial patterns in short-term memory. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 481-498.

Hirsh-Pasek, K. A. Phonics without sound: Reading acquisition in the congenitally deaf. Unpublished doctoral dissertation, University of Pennsylvania, 1981.

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

Klima, E., & Bellugi, U. The signs of language. Cambridge, Mass.: Harvard University Press, 1979.

Liben, L. S., & Drury, A. M. Short-term memory in deaf and hearing children in relation to stimulus characteristics. Journal of Experimental Child Psychology, 1977, 24, 60-73.

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

Locke, J. L., & Locke, V. L. Deaf children's phonetic, visual, and dactylic coding in a grapheme recall task. Journal of Experimental Psychology, 1971, 89, 142-146.

MacDougall, J. C. The development of visual processing and short-term memory in deaf and hearing children. American Annals of the Deaf, 1979, 124, 16-22.

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

Moulton, R. D., & Beasley, D. S. Verbal coding strategies used by hearing-impaired individuals. Journal of Speech and Hearing Research, 1975, 18, 559-570.

Novikova, L. A. Electrophysiological investigation of speech. In F. J. McGuigan (Ed.), Thinking: Studies in covert language processes. New York: Appleton-Century-Croft, 1966.

Odom, P. B., Blanton, R. L., & McIntyre, C. K. Coding medium and word recall by deaf and hearing subjects. Journal of Speech and Hearing Research, 1970, 13, 54-58.

Poizner, H., Bellugi, U., & Tweney, R. D. Processing of formational, semantic, and iconic information in American Sign Language. Journal of Experimental Psychology: Human Perception and Performance, 1981, 7, 1146-1159.

Shand, M. A. Sign-based short-term coding of American Sign Language signs and printed English words by congenitally deaf signers. Cognitive Psychology, 1982, 14, 1-12.

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

Wallace, G., & Corballis, M. C. Short-term memory and coding strategies in the deaf. Journal of Experimental Psychology, 1973, 99, 334-348.

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

Wolford, G., & Fowler, C. A. Differential use of partial information by good and poor readers. Developmental Review, in press.

Wolford, G. & Hollingsworth, S. Evidence that short-term memory is not the limiting factor in the tachistoscopic full-report procedure. Memory & Cognition, 1974, 2, 796-800.

FOOTNOTE

[1]The use of the term "speech-based code" here is not meant to imply that the code need be based on auditory or articulatory concomitants of speech, but rather may be an abstract representation of the phonetic or phonological features of the language.

DETERMINANTS OF SPELLING ABILITY IN DEAF AND HEARING ADULTS: ACCESS TO LINGUISTIC STRUCTURE*

Vicki L. Hanson, Donald Shankweiler,+ and F. William Fischer++

Abstract. The extent to which ability to access linguistic regularities of the orthography is dependent on spoken language was investigated in a two-part spelling test administered to both hearing and profoundly deaf college students. The spelling test examined ability to spell words varying in the degree to which their correct orthographic representation could be derived from the linguistic structure of English. Both groups of subjects were found to be sensitive to the underlying regularities of the orthography as indicated by greater accuracy on linguistically-derivable words than on irregular words. Comparison of accuracy on a production task and on a multiple-choice recognition task showed that the performance of both deaf and hearing subjects benefited from the recognition format, but especially so in the spelling of irregular words. Differences in the underlying spelling process for deaf and hearing spellers were revealed in an analysis of their misspellings: Deaf subjects produced fewer phonetically accurate misspellings than did the hearing subjects. Nonetheless, the deaf spellers tended to observe the formational constraints of English phonology and morphology in their misspellings. Together, these results suggest that deaf subjects are able to develop an appreciation for the structural properties of the orthography, but that their spelling may be guided by an accurate representation of the phonetic structure of words to a lesser degree than it is for hearing spellers.

---

*Also Cognition, in press.
+Also University of Connecticut.
++Central Connecticut State University.
Acknowledgment. This research was supported by Grant NS-18010 from the Institute of Neurological and Communicative Disorders and Stroke and by Grant HD-01994 from the National Institute for Child Health and Human Development. The authors wish to thank John Richards, Isabelle Liberman, Ignatius Mattingly, Nancy Fishbein, Dennis Schemenauer, Nancy McGarr, and Bob Katz for their help with the project, either in subject testing, data analysis, or in providing helpful criticisms of earlier versions of the manuscript. This study was made possible through the cooperation of Gallaudet College, the Linguistics Research Laboratory at Gallaudet College, and the National Center on Deafness at California State University, Northridge.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

Those who do research on the psychology of language have not, until recently, displayed much interest in spelling. As long as it is regarded as a low-level, isolated ability that feeds chiefly on rote learning and visual memory, spelling seems remote from a concern with language. Only now is it becoming generally recognized that to understand how people learn to spell is an interesting and challenging problem both linguistically and cognitively (Frith, 1980). There appears to be a growing tendency to progress beyond the notion that the orthography of English is a highly inconsistent system. Rather, it is a mulitileveled system containing regularities that penetrate deeply into the morphophonemic and lexical aspects of language (Chomsky, 1970; Klima, 1972; Venezky, 1970). For the speller who lacks sensitivity to these regularities of the orthography, the spellings of many words must appear arbitrary and opaque.

How the consistencies that the orthography captures actually affect the speller of English is, of course, an empirical question. For present purposes, it will be assumed that there exists a linguistic speller in the same sense that it has been assumed that there exists a linguistic reader (Mattingly, 1972, 1980). The ideally proficient reader-writer is sensitive to various kinds of linguistic information that are contained in the orthographic representation of words in the lexicon. Accordingly, the linguistic reader-writer can unpack this information in the act of reading, and can fully and correctly package it in the act of spelling.

The question raised in the research presented here is to what extent the acquisition of linguistic principles of the orthography is dependent on the spoken language. To examine this question, the pattern of spelling errors for prelingually and profoundly deaf college students is compared to that of hearing college students.

To put this issue in perspective, the research literature that pertains to interpretation of spelling errors both for hearing and deaf persons will first be briefly examined. In general it may be said that hearing spellers appreciate that the orthography maps the phonetic structure of words,[1] but that they sometimes fail to appreciate the other regularities that the orthography captures. Thus, there is much evidence that the predominant form of spelling error for hearing children and adults consists of misspellings consistent with the words's phonetic representation, i.e., their misspellings can be read as phonetically equivalent to the target word (Alper, 1942; Fischer, 1980; Masters, 1927; Sears, 1969). These phonetic misspellings appear to stem from a failure to appreciate fully the phonological and derivational factors that English spelling preserves.

Evidence that some structural principles of the orthgraphy are acquired and used in spelling was found in a study by Fischer (1980). Fischer constructed a spelling test designed to assess spellers' sensitivity to the underlying linguistic structure of words. Hearing college students had little difficulty with words in which the spelling was straightforwardly related to the phonetic structure (e.g., zebra), but had difficulty on words for which the correct spelling could not be fully derived from morphophonemic information (e.g., sergeant). Good spellers, more than poor spellers, were found to be able to make use of linguistic regularities to spell words.

Some investigators have suggested that rote memory and/or visual retentiveness may be a major factor in skilled spelling with spellers relying, at least in part, on stored word images (Baron, Treiman, Wilf, & Kellman, 1980; Barron, 1980; Ehri, 1980; Sloboda, 1980). If success in spelling is highly related to retention of visual patterns, then good spellers would be expected to make more efficient use of such a strategy than poor spellers. It is not the case, however, that good spellers exceed poor spellers in visual retentiveness of every kind of material; Fischer (1980) found no difference between good and poor hearing spellers on a test of memory for nonword abstract patterns, the Recurring Figures Test of Kimura (1963). There is some evidence, however, that spellers can benefit from the presence of visual forms of the word. When the test offers choices among printed alternative spellings of a word, performance has been found in some cases to improve (Simon & Simon, 1973; Tenny, 1980). Whether it does so or not seems to depend on the type of word being tested. Fischer (1980) found that multiple-choice recognition performance is more accurate than spelling to dictation for both good and poor spellers, but that the advantage of the recognition format is limited primarily to words whose spellings are not linguistically derivable (e.g., sergeant).

It is possible that the importance of rote memorization and/or visualization for spelling ability may be greater for deaf spellers than for hearing spellers. The absence of normal experience with the sounds of the spoken language may make acquisition of linguistic regularities difficult. Indeed, early work implicated visual retention as a factor important to spelling success for deaf children (Gates & Chase, 1926), but no comparison between production and multiple-choice recognition with deaf subjects has been carried out to date.

A few studies have examined the ability of deaf subjects to make use of phonetic structure of words during spelling. One such study was carried out by Dodd (1980) on orally-trained deaf children in England. The children (mean age 14.5 years) were required to lipread pseudowords. Analysis of their spoken and written productions indicated that if a consonant was correctly represented in the spoken response, it was generally also correctly represented in the written response. The implication is that these deaf children had acquired the ability to use the alphabet analytically.

Nonetheless, there is evidence that deaf spellers' misspellings are often quite unlike those of hearing persons. In contrast to the misspellings of hearing persons, fewer of the misspellings produced by deaf children and adults can be considered phonetically equivalent to the target word (Dodd, 1980; Hanson, 1982; Hoemann, Andrews, Florian, Hoemann, & Jansema, 1976). The unanimity of the studies is especially striking in that the studies have tested deaf subjects with backgrounds that are quite heterogeneous with regard to many factors--degree of hearing loss, age, and type of schooling, to name a few. The implication from this finding is that the spelling process for deaf persons may be fundamentally different from the spelling process for hearing persons.

Although a study by Cromer (1980) would seem somewhat at odds with this interpretation, since he found that the majority of misspellings by deaf children were "phono-graphical" errors, it must be noted that Cromer's phono-graphical errors are not the same as phonetic misspellings. According to Cromer, a phono-graphical error occurs when the "mis-spelled word resembles in some respect the sound of the target word when pronounced" (p. 412). Errors such as as basking for basket and amanals for animals were, as a result, scored as phono-graphical errors. Clearly, as these examples indicate, this classification system does not distinguish between those responses that are phonetically consistent with the target and those responses that are not. Thus, no direct comparisons between Cromer's study and the other spelling studies with deaf subjects is possible.

For the present study, subjects were chosen who are profoundly deaf from birth. In order to examine deaf and hearing subjects' access to linguistic structure, the tasks of Fischer (1980) were adapted for the present study. These tasks allow for a determination of spelling ability as a function of phonological and orthographic structure. If subjects rely on linguistic structure, then the more orthographically transparent the word spelling, the greater ease subjects should have in spelling the word. Thus, if deaf persons have acquired knowledge of the structure of words and they use this knowledge in spelling, then their spelling accuracy should vary as a function of level of orthographic transparency. As such, words whose spellings are derivable from linguistic principles should be more accurately spelled than irregular words whose spellings are not thus derivable. If deaf persons rely primarily on rote memorization or visual memory in spelling, then, other things being equal, words with linguistically-derivable spellings should be spelled no more accurately than irregular words.

Studies of spelling with hearing subjects most commonly rely on dictated word lists. For deaf subjects, results from this method of presentation would necessarily be ambiguous since errors of spelling would be inextricably confounded with errors of lipreading. To avert this confounding, the spelling test used in the present study provided written cues to elicit the subjects' responses. The performance of the deaf subjects was compared with that of a group of hearing subjects.

METHOD

Subjects

A group of deaf subjects and a group of hearing subjects were tested in a one-hour experiment. Neither group was preselected on the basis of spelling ability.

The deaf subjects were 27 profoundly deaf college students from Gallaudet College and from California State University, Northridge. All were prelingually deaf and had a hearing loss of greater than 85 dB in the better ear. They had no other handicapping conditions. The educational background of the subjects varied as to particular instructional method. All were proficient in the use of sign language (American Sign Language and signed English) and fingerspelling. Fourteen had deaf parents.

The hearing subjects were 37 college students from the University of Connecticut and from Central Connecticut State University.

Procedure

A reading comprehension test and a two-part spelling test (consisting of a Production Task and a Recognition Task) were administered to all subjects. The reading test was always given first, followed by the spelling Production Task and finally by the spelling Recognition Task.

Reading Test. The reading achievement of each subject was tested on the comprehension subtest of the Gates-MacGinitie *Reading Test* (1969, Survey F, Form 2). Survey F of the test is designed for grades 10 through 12. This testing level was chosen as previous work had indicated that deaf college students could be expected to read at the ninth- or tenth-grade level (Reynolds, 1975). For each of the subjects, a standard score on the reading comprehension test was obtained for grade level 10.1. A standard score of 50 on the test represents the mean performance for grade 10.1. Each 10 points on the standard score represents one standard deviation.

Spelling Test. The spelling test required the spelling of 45 English words. Three different classes of words were defined according to criteria framed by Fischer (1980). The classes ranged from Level I, in which the spellings were most transparent and related very straightforwardly to phonetic structure, to Level III, in which the spellings were opaque. In order to ensure that the words were not ones having highly overlearned spellings, all stimulus words were selected to be low in frequency of occurrence in written English. There were 15 words per level.

For Level I words, the correct spelling fairly straightforwardly reflected the phonetic structure: Success with these words requires that the user know the basic conventions of orthographic mapping including, for example, conventions for representing long and short vowels. In addition, the spelling patterns had a high frequency of occurrence in written English. The Level I words were as follows: explode, hardware, harpoon, migrate, plastic, refund, regret, reptile, rodeo, splash, splinter, stampede, tadpole, torpedo, transplant. Mean frequency was 2.27 occurrences per 1,014,232 words of natural language text (Kučera & Francis, 1967).

For Level II words, the correct spelling was not completely reflected in the phonetic structure, but could be obtained by reliance on linguistic principles. In eight of the fifteen Level II words, the phonetic structure reflected the morphophonemic structure, but knowledge of how to form suffixes was required for correct spelling. The words fitting this pattern were the following: beginner, desirable, galleries, heroes, ninety, noticeable, picnickers, thankful. In the other seven of the Level II words, the underlying morphophonemic relation was ambiguously represented in the phonetic structure. For these words, segment(s) were unstressed and thus ambiguous in the phonetic representation of the word and could be disambiguated by reference to a related word that stressed the segment (e.g., grammar-grammatical and digestible-digestion). The following stimuli fit this pattern: condemn, digestible, grammar, imaginary, janitor, permissible, repetition. For the Level II words, mean frequency of occurrence in written English was 8.60 (Kučera & Francis, 1967).

For Level III words, the correct spelling could only be partially derived by use of phonetic and morphophonemic structure. These included some borrowed words that contained spelling patterns infrequent in English. The following words were in the Level III category: ache, cantaloupe, champagne, chauffeur, Fahrenheit, mortgage, moustache (mustache), neighbor, plagiarism, plumber, receipt, sergeant, vacuum, vinegar, yacht. Mean frequency of occurrence was 8.33 (Kučera & Francis, 1967).

In the Production Task, subjects were asked to spell the 45 words using a Cloze procedure, in which a written sentence context was provided for the target word and the first letter of the target word was presented. This procedure had two advantages over spelling from dictation tasks. First, it was advantageous with deaf subjects in that it did not require that stimuli be lipread. Second, for both subject groups it assured that all misspellings were misspellings of words in the subjects' vocabularies. The following is an example of a test sentence:

(1) Temperature is measured in degrees F__________.

Since this experiment was concerned only with spelling processes, not with world knowledge, it was decided that subjects would be provided with additional cues if they were unable to figure out the target word from the sentence context. The following written instructions were given to subjects:

> This experiment is concerned with spelling. For each sentence below, complete the spelling of the word that fits in the blank (the first letter of the omitted word is always given). If you are not sure what word fits in the sentence, ask the experimenter. PLEASE PRINT!!

If subjects had questions about a word to be spelled, the experimenter provided an alternative definition of the word. The word was not spoken for hearing subjects. If a sign existed for the target word, that sign was produced for deaf subjects.

The same 45 words were also used in the Recognition Task. Words were tested in the same order as in the Production Task. On each trial there were three alternative spellings of the target word plus the choice "None of these." The written instructions were as follows:

> Circle the correct spelling for each of the following words. If the correct spelling is not listed, circle "None of these." (These are the same words you just spelled.)

The alternative choices were generally phonetically consistent with the target. Also, since deaf adults sometimes make ordering errors when spelling (Hanson, 1982), an attempt was made to include misspellings that deaf subjects might choose (e.g., roedo for rodeo).

Scoring

A disadvantage of the Cloze procedure is that sometimes the sentence cue fails to elicit the desired word, or it may fail to elicit any word at all.

Since it is inappropriate to score such responses as spelling errors, they were scored as omissions. The following criteria were adopted for classification of a response as an omission:

a) no response.

b) a response that was a correctly spelled word, but was not the target word (e.g., sliver for splinter).

c) a response that did not contain at least 1/2 of the letters of the target word (e.g., phorgery for plagiarism).

d) a morphologically incorrect form of the target in which the target word was not completely represented in the response (e.g., hero for heroes and digestive for digestible). (This was done so as not to confound grammatical abilities with the current test of spelling proficiency.) A morphologically incorrect form in which the target was completely represented in the response was not scored as an omission (e.g., splinters for splinter).

Analysis of the Production Task was based on only those trials that were not scored as omissions. Since the purpose of the Recognition Task was to examine whether subjects would benefit in spelling accuracy from having visually presented alternatives available, analyses in the Recognition Task were based on only those trials that had been analyzed in the Production Task.

RESULTS

Spelling Production Task

Nearly all subjects failed to respond with the correct word on at least one occasion. Because data based on too few responses in each portion of the test are unstable, it was decided to exclude from further analysis the data of those subjects who had as many as 15 responses scored as omissions (i.e., one third of the total number of items). This criterion excluded eleven deaf subjects and no hearing subjects. Those excluded tended to be the poorest readers, but not necessarily the poorest spellers. Indeed, it is the case that the excluded deaf subjects scored significantly worse on the reading comprehension test than did the included deaf subjects, $t(25)=4.41$, $p<.001$, two-tailed, but did not differ significantly in spelling proficiency from those included, $t(25)=1.82$, $p>.05$, two-tailed.

One hearing subject was excluded for failure to complete the Recognition Task. The analysis of spelling proficiency in relation to orthographic transparency was based on the remaining 36 hearing college students, and 16 deaf college students.

Results of the Spelling Production Task for these subjects are shown in Figure 1. An analysis of variance was performed on the percentage correct responses for the two subject groups at the three levels of orthographic transparency. Of major concern to the present study was the finding that there was a significant main effect of level of orthographic transparency,

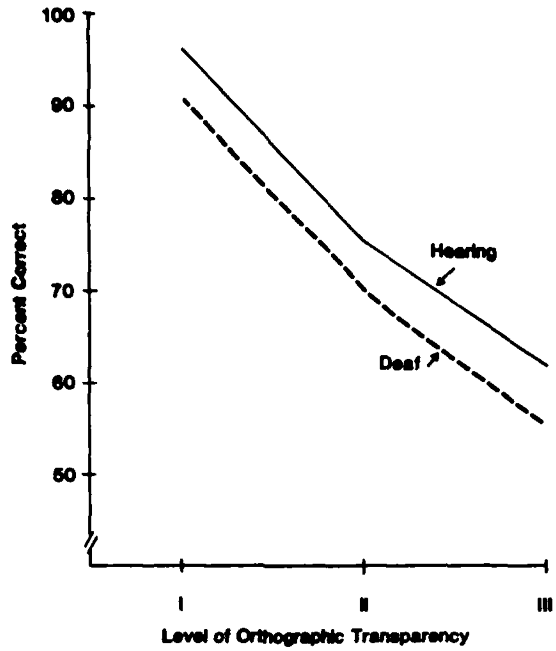


Figure 1. Mean percentage correct responses in the spelling Production Task as a function of level of orthographic transparency.

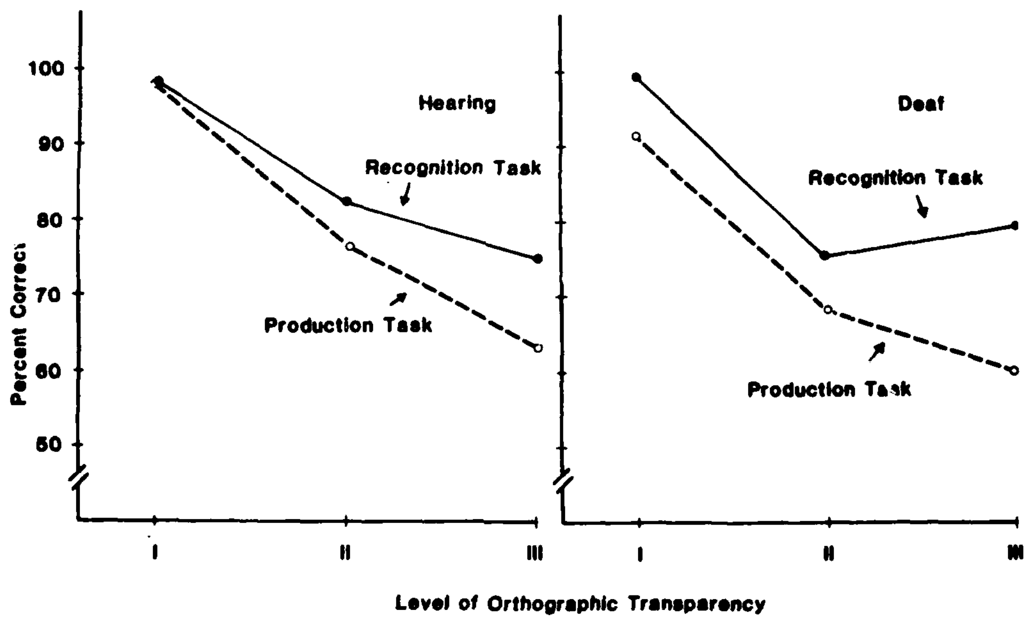


Figure 2. Mean percentage correct responses in the spelling Production and Recognition Tasks as a function of level of orthographic transparency.

$F(2,100)=107.82$, $p<.001$, $MSe=126.36$, that did not interact with subject population, $F<1$. Post hoc analyses demonstrated significant differences between each level of orthographic transparency (Newman-Keuls, $p<.01$). These results indicate that words of different orthographic types differed greatly in difficulty of spelling; in this the present findings are in complete agreement with Fischer (1980). Words of high orthographic transparency are consistently more often spelled correctly than words of low transparency or exception words. What is newly demonstrated is that, by and large, parallel differences in effect of orthographic transparency are shown by deaf and hearing subjects.

Comparison of Production and Recognition Tasks

Results comparing performance on the Production Task and the Recognition Task are shown in Figure 2. An analysis of variance was performed on the percent correct scores with the between-subjects factor of subject population and the within-subjects factors of orthographic transparency and task (Production Task vs. Recognition Task). A significant main effect of task, $F(1,50)=62.63$, $p<.001$, $MSe=90.82$, indicated that spelling performance was more accurate on the Recognition Task than on the Production Task. In addition, subject population interacted with task, $F(1,50)=5.28$, $p<.05$, $MSe=90.82$. This interaction reflected a greater improvement in performance on the Recognition Task for deaf subjects than for the hearing subjects, although a post hoc analysis revealed that there was a significant improvement in the Recognition Task for each group individually [for hearing subjects, $F(1,50)=25.62$, $p<.001$; for deaf subjects, $F(1,50)=37.66$, $p<.001$].

There was also a significant interaction of task by orthographic transparency, $F(2,100)=17.88$, $p<.001$, $MSe=43.15$. Since performance on the Level I words was so accurate, even for the Production Task, this interaction probably reflects to some extent a ceiling effect. The high level of performance on Level I words dramatically illustrates a major point of the present study--that spellers are influenced by orthographic transparency. Orthographically transparent words are not often misspelled by either hearing or deaf spellers. To determine whether there was an interaction of task by orthographic transparency for Level II and III words, neither of which are at ceiling, an additional analysis of variance was performed on these two levels of orthographic transparency alone. Again a significant interaction was obtained, $F(1,50)=14.99$, $p<.001$, $MSe=57.62$. The source of this interaction, as shown in Figure 2, is that there is more improvement with the Recognition Task for Level III words than for Level II words. A significant three-way interaction with population, $F(1,50)=7.17$, $p=.01$, $MSe=57.62$, indicated that deaf subjects improved more on Level III words than did hearing subjects.

To summarize, the comparison of performance on the Production and Recognition Tasks revealed that spelling performance was more accurate on the Recognition Task than on the Production Task, but the advantage of having the printed alternatives available was limited primarily to Level III words. Although both hearing and deaf spellers benefited from the recognition format, deaf spellers appeared to benefit somewhat more.

Error Types

Examination of misspellings can be used to gain insight into the spelling process. With groups of deaf and hearing subjects matched for overall proficiency in spelling, this allows us to ask, given a particular level of competence in spelling, whether it builds on the same underlying cognitive ability for deaf and hearing spellers. This analysis was therefore based on subsets of the two subject populations matched in overall spelling ability on the Production Task. These matched groups consisted of nine subjects each, with the subjects drawn from the deaf and hearing subjects included in the preceding analyses. The spelling proficiency and reading achievement of the resulting subgroups are shown in Table 1. These matched groups did not differ significantly in spelling accuracy on this task, $t(16)=1.10$, $p>.05$, two-tailed, but did differ significantly in reading achievement, $t(16)=4.06$, $p<.001$, two-tailed. These results indicate that the deaf subjects were poorer readers than the hearing subjects of comparable spelling proficiency.

Table 1

Characteristics of the subject groups matched for spelling proficiency. Shown is the mean accuracy on the spelling Production Task and the mean standard scores on the Gates-MacGinitie reading comprehension test.

	Hearing (N=9)	Deaf (N=9)
Spelling	70.5%	69.1%
SD	2.5	3.0
Reading	61.3	49.3
SD	6.0	6.5

Each misspelling was scored in terms of whether or not the misspelled segment(s) of the word constituted a substitution (e.g., janiter for janitor), omission (e.g., chamagne for champagne), or insertion (e.g., torpedeo for torpedo). If multiple errors occurred within a given word, each error was scored separately. For example, two errors were scored when vinegar was spelled as viniger and when digestible was spelled as disgestable. By this analysis, only two misspellings were unclassifiable (the response tad pole for tadpole by a hearing subject and the response puglarism for plagiarism by a deaf subject).

Each segment substitution error was further scored in two respects. First, it was asked whether or not the substitution was a "phonetic" substitution (e.g., vineger for vinegar) or a "nonphonetic" substitution (e.g., redeo for rodeo). Determination as to whether or not a substitution was phonetic was based on Hanna, Hanna, Hodges, and Rudorf's (1966) listing of alternative patterns for the spelling of English phonemes. Using this analysis, spellings were scored in terms of spelling patterns rather than individual letters. Thus, if condemn was spelled as condem, it was scored as a phonetic substitution since mn and m are both legitimate spelling patterns for /m/ in final position. Other examples of phonetic substitutions include grammer for grammar, vacume for vacuum, and champane for champagne. Examples of nonphonetic substitutions include torpado for torpedo and chanpagne for champagne. Secondly, it was asked whether the substitution was a vowel segment substitution (e.g., digestable for digestible) or a consonant segment subsitution (e.g., plummer for plumber and chaufeur for chauffeur).

This analysis indicated that the groups of deaf and hearing subjects matched for spelling proficiency differed considerably in the types of errors they produced. As can be seen from Table 2, segment substitutions predominated for both deaf and hearing spellers, with only a small percentage of the misspellings for either group resulting from segment insertions. However, the deaf subjects made more errors that were not substitutions than did the hearing subjects. For the hearing subjects, only about 9% of the errors were omissions and insertions, while for the deaf subjects 29% of the errors were omissions and insertions. This difference in the percentage of nonsubstitution errors for the two groups was statistically significant, $t(16)=4.45$, $p<.001$, two-tailed. Since substitution errors represent an awareness of the number of phonemic segments of words, this finding suggests that the number of segments in words was not apprehended as accurately by the deaf subjects. Moreover, for those substitution errors that did occur, the deaf subjects had less tendency to produce errors that were phonetically acceptable renderings of the target segments. More than 80% of the errors by hearing subjects were phonetically acceptable substitutions, as compared to fewer than 50% of the errors of deaf subjects. This difference between the two groups was statistically significant, $t(16)=7.90$, $p<.001$, two-tailed.

Table 2

Mean percentage of each error type for the matched subject groups. Standard deviations are given in parentheses.

	Hearing		Deaf	
	Phonetic	Nonphonetic	Phonetic	Nonphonetic
Substitutions	81.6% (9.1)	9.0% (7.6)	46.3% (9.8)	24.7% (13.9)
Omissions		6.4% (5.6)		20.1% (7.8)
Insertions		3.0% (3.6)		8.9% (5.2)
Total	81.6%	18.4%	46.3%	53.7%

Both deaf and hearing subjects were found to make more substitutions on vowel segments than on consonant segments: Hearing subjects made 70.0% of the substitutions on vowels, deaf subjects made 70.6% of their substitutions on vowels. Thus, these hearing and deaf subjects did not differ significantly in their tendency to make vowel substitutions, $t(16)=-.11$, $p>.05$, two-tailed. The greater difficulty on spelling vowel segments here and elsewhere with hearing subjects (Fischer, 1980; Masters, 1927; Seymour & Porpodas, 1980) underscores the greater complexity of vowel representation than consonant representation in English orthography.[2]

Consistent with previous findings (Hanson, 1982), several of the misspellings of the deaf subjects contained an error in ordering of one or more letters of the word, resulting in misspellings that did not preserve the phonetic representation of the target word. Thus, for example, a misspelling of vinegar was vingear, a misspelling of janitor was jaintor, a misspelling of reptile was retiple, and a misspelling of cantaloupe was cantapole. Of the words misspelled by deaf subjects, 13.0% contained such an ordering error. Of the misspellings by hearing subjects, only .9% contained this type of error.

The misspellings were further scored to examine whether or not they were orthographically regular. Only those responses that were pronounceable and had legal letter sequences were considered to be orthographically admissible. Two judges independently scored the responses. Of the 208 misspellings considered in this analysis, the judges agreed on the classification for 94.2%. On those responses for which they originally disagreed, the two judges discussed the misspelling until a classification was agreed upon. Results of this analysis indicated that 91.7% of the misspellings of hearing subjects were considered orthographically regular and that 96.0% of the misspellings of the deaf subjects were considered to be so.

The results of this error analysis thus suggest that deaf spellers are sensitive to structural constraints of the orthography. That they are able to appreciate these constraints is shown by their production of misspellings that are permissible letter sequences in the language, and by the tendency of their substitution errors to be predominantly vowel substitutions.

In spite of their general conformity with the principles of English orthography, the misspellings of deaf subjects were generally not phonetically equivalent with the target words. Inconsistency with the phonetic representation was revealed by the analysis indicating fewer phonetically acceptable substitution errors by deaf than hearing subjects and by the analysis indicating that a few of the misspellings of the deaf subjects represent an inaccurate ordering of the segments of a word. These findings suggest either 1) that deaf spellers have less accurate representations of the phonetic structure of individual words in their lexicons than do hearing spellers, 2) that they do not use the phonetic information in their lexicons when spelling, or 3) that they use this information less accurately than do hearing spellers. Research by Dodd (1980) with deaf children is relevant in distinguishing between these alternatives. Dodd found that the deaf children tended to spell consonant segments accurately that they pronounced accurately. (No analysis of vowel segments was undertaken in that study.) This suggests that the first of the three alternatives presented here may best explain the performance of deaf spellers; that is, the nonphonetic spellings they make may tend to reflect a difficulty in incorporating into their lexicons accurately specified phonetic representations of individual words.

Spelling Proficiency in Relation to Other Language Factors

For the purpose of examining the relationship between spelling and reading, subjects' scores on the reading comprehension test and their percent correct on the spelling Production Task were compared. This analysis was based on the data of all 37 hearing subjects tested and all 27 deaf subjects tested. Table 3 shows the mean percent correct in the spelling task for deaf and hearing subjects together with the mean standard scores on the Gates-MacGinitie Reading Test. Recall that a standard score of 50 on the Gates-MacGinitie test represents a reading level of grade 10.1. Overall, the hearing subjects were more proficient readers, $t(62)=10.22$, $p<.001$, and spellers, $t(62)=3.23$, $p<.01$, than the deaf subjects. For hearing subjects, the reading scores correlated, although only weakly so, with spelling performance, $r=.356$, $t(35)=2.25$, $p<.05$. The direction of correlation suggests that the greater the subject's reading ability, the greater the spelling proficiency. The same trend was true for the deaf subjects, although the resulting correlation was not significant, $r=.275$, $t(25)=1.43$, $p>.05$.[3]

Table 3

Mean accuracy on the Spelling Production task and mean standard scores on the Gates-MacGinitie reading comprehension test.

	Hearing (N=37)	Deaf (N=27)
Spelling	77.5%	69.6%
SD	9.7	9.7
Range	47 - 98	53 - 92
Reading	64.8	45.9
SD	7.4	7.2
Range	42 - 78	33 - 60

A question of interest is how the speech production capabilities of the deaf subjects relate to reading achievement and spelling proficiency. To address this question, speech intelligibility ratings were obtained for the deaf subjects from Gallaudet College. (Scores were not available for the five deaf subjects from the other university.) The ratings were based on a scale of 1 to 5, in which a score of 1 represents speech that is readily understood by the general public and a score of 5 represents speech that cannot be understood by listeners. For the 22 deaf subjects whose data were involved in this analysis, the mean speech intelligibility score was 3.89 (SD=.96, Range=2-5). These speech intelligibility ratings were not significantly correlated with either reading achievement, $r=-.002$, or spelling proficiency, $r=.398$, $t(20)=1.94$, $p>.05$.

DISCUSSION

As in earlier work, deaf spellers in the present experiment were by no means always inferior in spelling accuracy to their hearing counterparts (Cromer, 1980; Gates & Chase, 1926; Templin, 1948). Although the hearing subjects, overall, were somewhat more accurate than the deaf subjects on the spelling Production Task, both groups displayed a wide range of ability levels. The degree of overlap in the distribution of scores for the groups was notable in light of the degree of auditory impairment in the deaf group: All of these subjects were selected for profound deafness extending from infancy. The results provide a convincing demonstration that it is possible for persons with such a background to learn to spell as accurately as many hearing persons at the college level.

To examine the extent to which apprehension of the linguistic regularities of the orthography is dependent on the spoken language, the error patterns of deaf and hearing subjects were compared. In earlier research with hearing adults, Fischer (1980) has shown that a word's difficulty from the standpoint of spelling is chiefly a reflection of the word's formal properties and only secondarily a reflection of its frequency of occurrence. The results here are in complete agreement with Fischer's in that spelling performance was heavily influenced by level of orthographic transparency for both deaf and hearing spellers. Consistent with this evidence that deaf spellers are able to appreciate the structural constraints of the orthography, we found that the misspellings of deaf subjects tend to be orthographically regular in the sense that only legal strings are produced (see also Hanson, 1982). In sum, these data indicate that it is possible for prelingually, profoundly deaf individuals to develop a sensitivity to the phonological and morphological constraints of written English.

Deaf and hearing spellers further exhibited a similar pattern of results on the Recognition Task in that the greatest benefit occurred on irregular words. These were the words in which the correct spelling could not be completely derived by linguistic principles (the Level III words). Thus, consistent with Fischer's findings (1980), these results suggest that visually presented alternative spellings are of primary benefit in allowing the speller to access rote and/or visual information that is otherwise difficult to retrieve.

Thus far, ways in which deaf and hearing subjects resemble each other have been discussed. Now, how they differ must be considered. First, they differ in that deaf subjects appear to benefit more than hearing subjects from having the visual alternatives presented. It appears, therefore, that deaf spellers to a greater extent than hearing spellers, have stored visual knowledge about a word's spelling that they are not able to retrieve in productive spelling, but which they can access when visual alternatives are available.

The groups differ in a major way in the kinds of errors they produce. Our findings strongly confirm earlier indications that deaf subjects, unlike hearing subjects, produce many strings that are not phonetically equivalent to the target word, i.e., nonphonetic misspellings (Dodd, 1980; Hanson, 1982; Hoemann et al., 1976). In the present research, nonphonetic errors occurred nearly three times more frequently with deaf subjects than hearing subjects, even when the comparison was restricted to groups of deaf and hearing subjects matched on overall level of spelling performance.

It is important to note that the misspellings made by the deaf subjects in this study differ markedly from error patterns that are often labeled "visual" or "orthographic"; that is, misspellings in which the letter strings only grossly approximate the target word and that indicate a failure to appreciate the syllabic and segmental structure of words (see, for example, Boder, 1973; Bub & Kertesz, 1982; Seymour & Porpodas, 1980; Wapner & Gardner, 1979). Such misspellings retain some of the characteristics of how the target word looks, as in the example of misspelling broom as beoom (Wapner & Gardner, 1979). The presence of such an error suggests that the speller does not appreciate how the orthography maps onto the spoken language. In contrast, deaf spellers have been found to be able to perform a phonemic analysis of words (Dodd, 1980), and their misspellings here and elsewhere have been shown to be consistent with the structural constraints of English morphology in preserving the rules governing syllable structure within words (Hanson, 1982). Moreover, if the deaf subjects here had not been sensitive to variations that exist in orthographic transparency, they would have performed with comparable accuracy on Level I, II, and III words. It would seem, then, that the nonphonetic misspellings of the deaf subjects arise not because these spellers are unable to appreciate the mapping betwen the written and spoken language, but rather may arise from difficulty in the establishment of an accurate phonetic representation of specific words.

The suggestion here that deaf spellers may have difficulty in the establishment of an accurate phonetic representation of words is in contrast to their ability, so apparent in the findings of this study, to appreciate phonological constraints of the language. Several factors may contribute to such awareness for deaf spellers, of which the most likely candidates are speech-related factors, reading, and fingerspelling.

Turning first to speech-related factors, speech production skills were examined here. The speech intelligibility ratings of the present subjects indicated that, as a whole, they had speech that was judged by skilled listeners to be nearly unintelligible. Although the skills of the individual subjects varied, the present study found that speech production skills were not significantly correlated with spelling proficiency. Since subjects with poorly intelligible speech were often good spellers, this suggests that acquisition of linguistic sensitivity may not necessarily require an ability to produce speech that listeners can readily understand, but only a means of analyzing word structure that the individual can use for acquiring the linguistic principles relating to that structure. Such a means of analysis might also be provided by lipreading (Dodd & Hermelin, 1977) and/or by whatever residual hearing each profoundly deaf person might possess.

Alternatively, just as hearing persons, through experience in reading, may induce phonological and morphological structure from the orthographic representation of written words (Liberman, Liberman, Mattingly, & Shankweiler, 1980), so might deaf readers similarly induce these structural facts. The relationship between the level of performance in reading and spelling is a matter of some interest. The comparison between reading comprehension and spelling proficiency indicated only a tenuous relationship in either population. The low correlations obtained were not artifactual, however. Both deaf and hearing subjects displayed a considerable range of talent on both reading and spelling measures, sufficient to permit a valid assessment of correlation. Moreover, for the hearing subjects the results obtained here are consistent with correlations obtained between reading comprehension and spelling reported for standardized tests (Dunn & Markwardt, 1970). Higher correlations between reading and spelling tend to be obtained when the reading measure is word recognition, particularly for persons in the process of acquiring reading, such as children in the primary grades and adults enrolled in literacy classes (Dunn & Markwardt, 1970; Jastak & Jastak, 1965; Perin, 1982). The low correlations reflect the possibilty that reading comprehension and spelling rely, in part, on different cognitive/linguistic abilities. For example, the reader can manage with a rather tacit knowledge of structural features of the orthography because context at various levels is provided in the text. The speller, on the other hand, must make explicit use of these features.

For the deaf subjects, in particular, there was a dissociation between reading achievement and spelling proficiency. Not only was there no significant correlation obtained between the two tasks, but, as shown in Table 1, the deaf subjects tended to be much poorer readers than the hearing subjects of comparable spelling skill. Thus, while deaf persons appear to be at a disadvantage in acquiring reading when compared with hearing persons, it is of interest that no comparable disadvantage seems to occur for spelling.

For deaf persons with experience in manual communication, reliance on fingerspelling might also provide a means of acquiring an appreciation of the structure of the orthography. Fingerspelling is a manual communication system in which words are spelled out by the sequential production of letters of a manual alphabet. Much as readers might induce phonological rules from reading, deaf persons might also induce these rules from fingerspelling.

Fingerspelling may also serve deaf spellers as a productive system. The deaf subjects were observed to fingerspell extensively during the experiment as a way of trying out spellings on their hands before writing their answers. The role of fingerspelling in writing words cannot be inferred with certainty here, but two possibilities may be suggested. First, fingerspelling may provide visual feedback that could be used much like the alternative spellings of the Recognition Task. The fact that subjects sometimes fingerspelled under the table (thus blocking their view of their hands) suggests, however, that the feedback may not always, or even mostly, be visual. It suggests that kinesthetic feedback may be used instead. This feedback could serve both as a check of a particular word's spelling against a stored representation of the word, and also to monitor legal letter sequences.

In summary, deaf spellers in the present research were found to display an ability to appreciate the structure of English orthography. This finding is inconsistent with the hypothesis that deaf spellers are limited to rote memorization or visual retention as spelling strategies. Obviously, it cannot be assumed that all deaf spellers (or hearing spellers) are sensitive to the linguistic structure reflected in the orthography. It is relevant here that the present subjects were all college students; it might be expected that persons with little education would rely on different strategies. The present results are important, however, in indicating the extent to which acquisition of linguistic structure is possible given limited acquaintance with the spoken language.

REFERENCES

Alper, T. G. A diagnostic spelling scale for the college level: Its construction and use. Journal of Educational Psychology, 1942, 33, 273-290.

Baron, J., Treiman, R., Wilf, J. F., & Kellman, P. Spelling and reading by rules. In U. Frith (Ed.), Cognitive processes in spelling. London: Academic Press, 1980.

Barron, R. W. Visual and phonological strategies in reading and spelling. In U. Frith (Ed.), Cognitive processes in spelling. London: Academic Press, 1980.

Boder, E. Developmental dyslexia: A diagnostic approach based on three atypical reading-spelling patterns. Developmental Medicine & Child Neurology, 1973, 15, 663-687.

Bub, D., & Kertesz, A. Deep agraphia. Brain and Language, 1982, 17, 146-165.

Chomsky, N. Phonology and reading. In H. Levin & J. P. Williams (Eds.), Basic studies on reading. New York: Basic Books, 1970.

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

Dodd, B. The spelling abilities of profoundly pre-lingually deaf children. In U. Frith (Ed.), Cognitive processes in spelling. London: Academic Press, 1980.

Dodd, B., & Hermelin, B. Phonological coding by the prelinguistically deaf. Perception & Psychophysics, 1977, 21, 413-417.

Dunn, L. M., & Markwardt, F. C. Peabody Individual Achievement Test. Circle Pines, Minn.: American Guidance Service, Inc., 1970.

Ehri, L. C. The development of orthographic images. In U. Frith (Ed.), Cognitive processes in spelling. London: Academic Press, 1980.

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

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

Frith, U. (Ed.), Cognitive processes in spelling. London: Academic Press, 1980.

Gates, A. I., & Chase, E. H. Methods and theories of learning to spell tested by studies of deaf children. Journal of Educational Psychology, 1926, 17, 289-300.

Gates-MacGinitie reading tests. New York: Columbia Teachers' College Press, 1969.

Hanna, P. R., Hanna, J. S., Hodges, R. E., & Rudorf, E. H., Jr. Phoneme-grapheme correspondences as cues to spelling improvement. Washington, D.C.: U.S. Government Printing Office, 1966.

Hanson, V. L. The use of orthographic structure by deaf adults: Recognition of fingerspelled words. Applied Psycholinguistics, 1982, 3, 343-356.

Hoemann, H. W., Andrews, C. E., Florian, V. A., Hoemann, S. A., & Jensema, C. J. The spelling proficiency of deaf children. American Annals of the Deaf, 1976, 121, 489-493.

Jastak, J. F., & Jastak, S. R. Wide Range Achievement Test. Wilmington, Del.: Guidance Associates, 1965.

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

Klima, E. S. How alphabets might reflect language. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye. Cambridge, Mass.: MIT Press, 1972.

Kučera, H., & Francis, W. N. Computational analysis of present-day American English. Providence, R.I.: Brown University Press, 1967.

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

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

Liberman, I. Y., Shankweiler, D., Orlando, C., Harris, K. S., & Bell-Berti, F. Letter confusions and reversals of sequence in the beginning reader: Implications for Orton's theory of developmental dyslexia. Cortex, 1971, 7, 127-142.

Masters, H. V. A study of spelling errors. Iowa City, Iowa: University of Iowa Studies in Education, IV(4), 1st Series #138, September 1, 1927.

Mattingly, I. G. Reading, the linguistic process and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye. Cambridge, Mass.: MIT Press, 1972.

Mattingly, I. G. Reading, linguistic awareness and language acquisition. Haskins Laboratories Status Report on Speech Research, 1980, SR-61, 135-150.

Meadow, K. P. Early manual communication in relation to the deaf child's intellectual, social, and communicative functioning. American Annals of the Deaf, 1968, 113, 29-41.

Perin, D. Spelling strategies in good and poor readers. Applied Psycholinguistics, 1982, 3, 1-14.

Reynolds, H. N. Development of reading ability in relation to deafness. Proceedings of the VIIIth World Congress of the World Federation of the Deaf, Washington, D.C., 1975.

Sears, D. A. Engineers spell acoustically. College Composition and Communication, 1969, 20, 349-351.

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

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

Simon, D. P., & Simon, H. A. Alternative uses of phonemic information in spelling. Review of Educational Research, 1973, 43, 115-137.

Sloboda, J. A. Visual imagery and individual differences in spelling. In U. Frith (Ed.), Cognitive processes in spelling. London: Academic Press, 1980.

Templin, M. C. A comparison of the spelling achievement of normal and defective hearing subjects. Journal of Educational Psychology, 1948, 39, 337-346.

Tenney, Y. J. Visual factors in spelling. In U. Frith (Ed.), Cognitive processes in spelling. London: Academic Press, 1980.

Venezky, R. The structure of English orthography. The Hague: Mouton, 1970.

Vernon, M., & Koh, S. D. Effects of oral preschool compared to early manual communication on education and communication in deaf children. American Annals of the Deaf, 1971, 116, 569-574.

Wapner, W., & Gardner, H. A study of spelling in aphasia. Brain and Language, 1979, 7, 363-374.

FOOTNOTES

[1]The levels of structure described here as "phonetic" denote a level considerably more abstract than sound. Unfortunately, linguistic disciplines offer no terms that have won general acceptance to capture differences in level of abstractness. It must be noted at the outset, however, that alphabets do not map sound as such, and could not, if they are to function as intended; i.e., no writing system in general usage captures details of the speech sound pattern associated with dialect and idiolect, or those associated with coarticulation and environment (see Klima, 1972, and Liberman, in press, for discussions of these points).

[2]The greater number of substitutions on vowel segments than consonant segments in spelling is consistent with research on misreading; this research on misreading has shown that (hearing) readers are much more likely to have difficulty in correctly reading vowel segments than in correctly reading consonant segments (Fowler, Liberman, & Shankweiler, 1977; Liberman, Shankweiler, Orlando, Harris, & Bell-Berti, 1971; Shankweiler & Liberman, 1972).

[3]Although the present study was not designed to assess differences between deaf subjects with deaf parents and deaf subjects with hearing parents, this question is of some interest as it is generally found that deaf children of deaf parents outperform deaf children of hearing parents on reading tests (Meadow, 1968; Vernon & Koh, 1971). No significant difference was obtained here as a function of parents' hearing status for either reading, $t(25)=.78$, $p>.05$, two-tailed, or spelling, $t(25)=.48$, $p>.05$, two-tailed, probably due to the fact that the present sample was restricted to college students--those persons who, by definition, are already the more academically successful deaf persons.

A DYNAMICAL BASIS FOR ACTION SYSTEMS*

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

1. INTRODUCTION

Students of the neural basis of cognition might well take as their dictum the first phrase in the gospel according to St. John: "In the beginning was the word." In this chapter we beg to differ and side instead with Goethe's Faust who, not satisfied with the accuracy of the biblical statement, proposed a rather different solution: "Im anfang war die tat"--"In the beginning was the act."[1] Certainly, if there is a lesson to be learned from the field of neuroembryology, it is that motility precedes reactivity; there is a chronological primacy of the motor over the sensory.[2] Although one of our main premises is that any distinction between "sensory" and "motor" is an artificial one (cf. Kelso, 1979), this brief sojourn into developmental embryology affords what we take to be a main contrast between the topic of concern in this chapter--the control and coordination of movement--and the subject matter of the rest of this book.

Our goals in this chapter are twofold. First, we want to describe some of the main developments in the field of movement control (as we see them) that have occurred in the last six to seven years. The developments hinge around a central problem that has continued to plague the physiology and psychology of movement almost since its inception, viz., the identification of significant units of coordination and control. In the last Neurosciences Research Program Bulletin that dealt specifically with motor control, Szentagothai and Arbib (1974) suggested that:

> "While the term synergy has not been explicitly defined here, it is evident that the traditional Sherringtonian usage is too restrictive to capture the concepts...One now awaits a redefinition of synergies to revitalize motor systems research along the behavioral lines of investigation successfully used in the visual system." (p. 165)

---

*Chapter to appear in M. S. Gazzaniga (Ed.), Handbook of cognitive neuroscience. New York: Plenum, in press.

+Also University of Connecticut.
++Also New York University Medical Center.

Acknowledgment. This work was supported by NINCDS grant NS-13617, BRS grant RR-05596, and Contract No. N0014-83-C-0083 from the U.S. Office of Naval Research to Haskins Laboratories, NINCDS Grant NS-17778 to Cornell University Medical College, and by an NIH Postdoctoral Fellowship to the second author. We thank Carol Fowler, David Ostry, and Elliot Saltzman for helpful comments.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

Much earlier of course, the Soviet school under Bernstein's dominant influence (cf. Bernstein, 1967) had advocated the synergy as a significant unit, and the idea was taken up seriously in this country by Greene (1972), Boylls (1975), Fowler (1977), Turvey (1977), Kelso (1979), and Saltzman (1979), among others. In fact, Boylls (1975) provides an elegant definition of synergy (or, "linkage" in his terms), which contrasts sharply with the traditional Sherringtonian concept: A "linkage" is a group of muscles whose activities covary as a result of shared afferent and/or efferent signals, deployed as a unit in a motor task.

A number of laboratories, including our own, have been working out the details of functional synergies (or, synonomously, muscle linkages or coordinative structures). In the first part of this chapter we shall explain briefly why the synergy concept is necessary, how synergies can be identified in many different activities, what their chief characteristics are, and how they are modulated by various sources of contextual information. All along we will try to show that there is a subtle and mutually dependent relationship between the small scale, neural, informational aspects of the system, and the large scale, power producing machinery—the muscle dynamics. The first part of this chapter is largely review, with a few novel nuances, but some of the organizational features that emerge are worthy of note in that they compare in an interesting way to recent theorizing about neuronal assemblies and brain functions (cf. Edelman & Mountcastle, 1978). At the end of the chapter, we shall make these comparisons explicit because they suggest a common ground for understanding the coherent behavior of muscle and neuronal ensembles.

Although we can supply a solid justification for the use of the synergy concept, and although we can provide hints--from the motor control literature--for how synergies can be regulated to accomplish particular acts, a principled basis is still required for understanding how the many free variables in the motor system can be harnessed in the first place. How do stable spatiotemporal organizations arise from a neuromuscular basis of many degrees of freedom? And what guarantees their persistence and stability? What principles underlie the cooperative behavior among muscles that is evident during coordinated activity?

In the second part of the chapter we take up these and related questions seriously. In contrast to "machine theories," which consider the many degrees of freedom to be regulated as a "curse" (cf. Bellman, 1961), and nonlinearities as a source of complication (cf. Stein, 1982), we advocate a set of "natural" principles gleaned from systems that require many degrees of freedom and in which nonlinearities are requisite conditions for the emergence of ordered phenomena (cf. Kelso, 1981; Kelso, Holt, Kugler, & Turvey, 1980; Kugler, Kelso, & Turvey, 1980, 1982; Turvey, 1980; see also Carello, Turvey, Kugler, & Shaw, in press). This "natural" perspective (Kugler et al., 1982) takes its impetus from (and is parasitic upon) contemporary physics,[3] and views the problems of coordination and control as continuous with, and a special case of, the more general problem of cooperative phenomena (cf. Haken, 1977). In this view, autonomy, self-organization, and evolution of function are stressed as system attributes. Our guess is that these attributes will prove difficult—in the long run—for the student of action to ignore, and, to the extent that they pertain to a theory of brain function, the cognitive neuroscientist as well.

2. A FUNDAMENTAL PROBLEM: THE SELECTION OF UNITS

2.1 The General Problem of Units

It is the time-honored thesis of classical physics that macroscopic states can be explained through microscopic analysis. The basic structure of nature is thought to be understood, first and foremost, through recourse to elementary units.[4] With the addition of a set of derived concepts (the laws of nature), natural phenomena can be explained. Biology has largely followed this paradigm by partitioning living systems into atomistic entities and laws of combination. Witness, for example, the dramatic successes in genetics, molecular biology, and neurophysiology: in some circles, units such as genes, molecules, and neurons, when synthesized appropriately, are thought to provide the basis of biological order.

One problem with this view, pointed out by Goodwin (1970), is that the analytical reductionist program with its accompanying resynthesis works only when there is a simple and direct relationship between the units of a system and its higher level behavior.[5] In biological systems, however, the units themselves are complex and thus there are many ways for higher order phenomena to arise. The scientist is then faced with the mammoth task of exploring all possible interactions among units and discovering those that could produce the observed higher order behavior. Even if this dubious strategy were possible, the problem of explaining the "macro" from the "micro" is not simply one of specifying <u>interactions</u> among elemental units. This is because at each level of complexity novel properties appear whose behavior cannot be predicted from knowledge of component processes. Paraphrasing Anderson (1972), there is a shift from quantitative to qualitative; not only do we have more of something as complexity increases, but the '<u>more</u>' <u>is different</u>. This is a physical fact (but eminently applicable to biology and psychology) arising from the theory of <u>broken symmetry</u>: As the number of microscopic degrees of freedom increases, matter undergoes sharp, discontinuous phase transitions that violate microscopic symmetries (and even macroscopic equations of motion), and leave in their wake only certain characteristic behaviors. As we shall see, symmetry breaking is a natural property of systems whose constraints are subject to change. We shall make much of this later on, because it is a central theme that may allow us to envision how coordination might arise in systems with many degrees of freedom. That is, how we can take a multivariable system and control it as if it had just one or a few degrees of freedom.

2.2 Units in Action Versus Units of Action

A great hindrance to the development of a theory of motor control and coordination has been the confusion between units <u>in</u> and units <u>of</u>. The unit is analyzed as if it were a piece in a puzzle or an ingredient in a cake, rather than in terms of its relational properties. For example, a pendulum consists of a number of components that can be thought of as the units <u>in</u> a pendulum system, but it is the <u>relations among components</u> that define the function of the pendulum system (cf. Ghiselin, 1981, for an informed discussion of units). With a few notable exceptions, students of action have classified units in terms of their anatomy rather than their function. Yet if there is a truism about action, it is that significant units are differentiated according to their function rather than according to the neuromuscular machinery that constitutes them.

Witness, for example, Gallistel's (1980) "new synthesis of the organization of action", in which the reflex arc is chosen as a major building block or unit of behavior because it contains "...all the elements necessary to explain the occurrence of muscular contraction or relaxation or glandular secretion." According to Gallistel, "...the necessary elements are those Sherrington recognized: an effector, a conductor, and an initiator" (1980, p. 399). Would that this connectionist metaphor provided the necessary criteria for units of action! Gallistel's Cartesian attitude of decomposing *the system into its parts* (configured in a fixed arrangement) and his offering some glue (in the form of neural potentiation and depotentiation) to stick them together again must, if our discussion of units is relevant, be off the mark. Admittedly, Sherrington was the main figure in reflex physiology, but even he recognized that the reflex was a "probable fiction" or at best a "purely abstract conception" (Sherrington, 1906). Aside from the recognition that a pure reflex is seldom, if ever, observed as a unique part of an act, few of us would want to build a theory of movement's control with fictions as the substrate (cf. Kelso & Reed, 1981).

Decomposing the system into arbitrarily defined analytical units evokes serious consequences for measurement. In all likelihood, the physical decomposition obscures the system's dynamics so that the unit's observable properties are no longer relevant. A good example is the three-body problem in physics (cf. Rosen, 1978), such as the earth-sun-moon system. Decomposing the system into analytically tractable single and two-body subsystems brings us no closer to an analytic solution for the original three-body problem. To solve the three-body problem, new sets of analytic units must be discovered that are defined by new observables, such that the partitioning respects the original dynamics. These may look nothing like the units that we have chosen for so-called "simplicity," or that we refer to as basic "building blocks." The functional units of behavior that we shall discuss are not anything like simple reflexes, and, only in certain very restrictive cases do they correspond to other proposed units of analysis such as "...single muscles or groupings of muscles acting normally around a joint" (Stein, 1982). Moreover, the criteria underlying their selection are not at all like those employed by Gallistel--or Sherrington, for that matter. As Reed (in press) points out, the units of action are not triggered responses that can be chained together by central or peripheral processes, but postures (which he calls "persistences in an animal-environment relation") and movements (transformations of one posture into another). In fact, one of the claims we shall try to substantiate is that a unit of action at any level of analysis must be so designed that persistence of function is guaranteed.

3. UNITS OF ACTION IN MULTIVARIABLE SYSTEMS

3.1 The Concept of Coordinative Structure

As we have already intimated, the problem of identifying units of action has long been a thorny issue, and continues to be debated in both the neural and behavioral literature. The elegant remarks of Greene (1971), made over a decade ago, still seem to apply in many circles:

"The masses of undigested details, the lack of agreement and the inconclusiveness that mark the long history of investigations of motor mechanisms arise from our limited ability to recognize the significant informational units of movement." (Greene, 1971)

There are signs, however, that some consensus is being reached concerning the units of action. This may reflect a growing appreciation of the fundamental problem of control and coordination identified by Bernstein (1967); namely, that of regulating a system with many degrees of freedom. Bernstein's key insight was that the large number of potential degrees of freedom of the skeletomuscular system precludes the possibility that each is controlled individually at every point in time. He then proposed a scheme whereby many degrees of freedom could be regulated through the direct, executive control of very few. In this view, individual variables of the motor system are organized into larger functional groupings called "linkages" or "synergies" (Boylls, 1975; Gurfinkel, Kots, Pal'tsev, & Fel'dman, 1971), "collectives" (Gel'fand, Gurfinkel, Tsetlin, & Shik, 1971), or "coordinative structures" (Easton, 1972a; Fowler, 1977; Kelso, Southard, & Goodman, 1979; Turvey, 1977). During a movement, the internal degrees of freedom of these functional groupings are not controlled directly but are constrained to relate among themselves in a relatively fixed and autonomous manner. The functional group can be controlled as if it had many fewer degrees of freedom than comprise its parts, thus reducing the number of control decisions required.

One example of a functional constraint on movement, a coordinative structure, is exhibited by people performing the task of precision aiming. When a skilled marksperson aims at a target, the wrist and shoulder joints do not change independently but are constrained to change in a related manner. Specifically, any horizontal oscillation in the wrist is matched by an equal and opposite oscillation in the shoulder, thus reducing the variation around the target area (Arutyunyun, Gurfinkel, & Mirsky, 1969). In an unskilled *marksperson*, movement at the wrist joint is unrelated to movement at the shoulder, allowing the arm to wander.

As the foregoing example reveals, coordinative structures are units of action, emphasizing the functional aspects of movement. Constraints are thought to arise temporarily and expressly for particular behavioral purposes (Boylls, 1975; Fitch & Turvey, 1978). The same degrees of freedom may be constrained in different ways to achieve different purposes, and different degrees of freedom may be constrained to achieve the same goal. Thus, coordinative structures are significant units not by virtue of their shared degrees of freedom, but by their capability of achieving a common goal. In this regard, the way we use the term "coordinative structure" differs from that of Easton (1972a), who views them as reflex based. Indeed, there is evidence that even reflexes exhibit functional specificity, adjusting to the phase of movement the animal is in when the reflex is elicited. For example, Forssberg, Grillner, and Rossignol (1975, 1977) examined reflex behavior in the spinal cat. A tap to the paw during the stance phase of stepping was associated with increased activity in the extensor muscles; a tap applied during the transfer phase enhanced activity in the flexor muscles. Such behavior is significant in that it performs an adaptive function for the animal, lifting the paw over an obstacle (see also, Fukson, Berkenblit, & Fel'dman, 1980) Thus, movements are seldom simply reactive; they are adaptive, functionally specific, and context sensitive (for many motor examples in the ethological literature, see Bellman, 1979; Reed, in press).

Note also that the coordinative structure perspective differs from open-loop models of control, which give privileged status to efference, as well as from closed-loop models, in which afference is dominant. The state of the marksperson's wrist joint, for example, is not only viewed as providing information about its own position (afference), but also as specifying the appropriate positions of the linked elements (efference). Thus, afference and efference both provide information relevant to the linkage, and neither one has priority over the other (Kelso, Holt, Kugler, & Turvey, 1980; Kugler, Kelso, & Turvey, 1980).

3.2 Coordinative Structures as Dynamic Linkages Defined Over Units of Action

Although constraining skeletomuscular variables results in an increase in control, it does so at the expense of range of motion. The number of possible trajectories of the limb is reduced, but the individual trajectory is not uniquely determined by constraints. When free variables are linked to perform a function, a balance exists between the linkage's flexibility, or freedom to undergo change, and limitations on its flexibility (Pattee, 1973; see also Fowler, 1977; Fowler, Rubin, Remez, & Turvey, 1980). Systems that do not perform functions are either too tightly constrained (e.g., rigid objects) or hardly constrained at all (e.g., an aggregate of grains of sand). Systems that perform functions are *selectively limited* in their actions, not uniquely determined.

In our earlier discussion of units (Section 2.1) we pointed out that complex systems exhibit discontinuities in structure and behavior (broken symmetry); that is, new modes of organization and behavior appear that are not easily predictable from the preceding modes. These new spatiotemporal structures are sometimes referred to as *emergent properties*. In the domain of movement, there is a tendency to account for the appearance of new phenomena--such as a novel movement pattern to accomplish some goal--by reference to the generativity embodied in a generalized motor program (e.g., Schmidt, 1975), motor engram (e.g., Heilman, 1979), or schema (cf. Head, 1926; Pew, 1974; Schmidt, 1975).

Rather than adopt this latter strategy, it may be better to recognize that all that's really happened is that our mode of description has failed at the point at which the novelty appears, requiring us to adopt a new mode of description that may be quite unrelated to the old one[6] (cf. Rosen, 1978). The main difficulty with an analysis of emergent properties lies, as Rosen (1978) cogently remarks, "...*in the tacit assumption that it is appropriate to describe a natural system by a single set of states*" (p. 91, italics his). This strategy necessarily restricts the observables that are possible and eliminates the possibility for new ones. However, when dynamical interactions occur, either among the states of a system, or when the system interacts with its environment, new observables are possible that were meaningless or invisible in the absence of coupling. As a consequence, an entirely new set of state descriptions of the system is possible because the observables have changed.

Let us bring these abstractions down to earth and back to the domain of movement. A coordinative structure, as we have defined it, is a functional linkage among previously unrelated entities--it is a prototypical example of an emergent phenomenon. By the arguments given above, a coordinative structure offers an alternative description of a system because it is defined on observables that bear little or no relationship to those of its components. By being a dynamic coupling among component variables, its state space offers a much richer set of trajectories than is possible in a system having the identical set of components but described by a single set of states.

3.3 Coordinative Structures as Nonlinear Vibratory Systems

Dynamical linkages (equations of constraint) selectively reduce the number of independently controlled degrees of freedom, thereby allowing a rich set of trajectories. But what kind of system is produced when elements of the motor apparatus are linked dynamically? Recent work on motor systems has identified functional units of action with nonlinear mass-spring systems. An attractive feature of such systems (among some others) is that they are intrinsically self-equilibrating: When the spring is stretched or compressed and then released, it will always equilibrate at the same resting length. Thus, the final equilibrium position is not affected by the amount that the mass is displaced--a property called equifinality (cf. von Bertalanffy, 1973).

In its more detailed (but we would add, unevenly interpreted) version, a given joint angle may be specified according to a set of muscle equilibrium lengths (cf. Fel'dman, 1966a, 1966b). Once these are specified, the joint will achieve and maintain a desired final angle at which the torques generated by the muscle sum to zero. Such a system exhibits equifinality in that desired positions may be reached from various initial angles and in spite of unforeseen perturbations encountered during the motion trajectory. Thus, if the length of a muscle at a joint is currently longer than the equilibrium length, active tension develops in the muscle; if the current length is shorter than the equilibrium length, the muscle relaxes. We can see how this concept is akin to a coordinative structure. Control of many variables (e.g., degree of activation in various muscles at a joint) is simplified by establishing a constraint: Given a set of muscle equilibrium lengths, the torque generated by tension in each muscle is dependent on its current length.

Recent support for this account comes primarily from work on limb and head movements. For example, Kelso (1977) and Kelso, Holt, and Flatt (1980) have shown that normal and functionally deafferented humans are more accurate in reproducing the final position of a limb from varying initial positions than in reproducing movement amplitude. In addition, Bizzi and his colleagues (Bizzi, Dev, Morasso, & Polit, 1976; Polit & Bizzi, 1978) have shown that normal and rhizotomized monkeys can reproduce learned target positions of the head or arm even when the movement trajectory is perturbed by application of a load. Similar results have been found in humans (Kelso & Holt, 1980), and predictable effects of changing effective mass of a limb have also been observed (e.g., Fel'dman, 1966b; Schmidt & McGown, 1980). The findings are not easily accounted for by traditional motor control models. For example, closed-loop models could account for the accurate reproduction of final position in spite of changes in initial position of the limb, or perturbations of the limb trajectory, but they could not explain why equifinality holds when the limb is deafferented. In theory, open-loop programming models could handle the deafferentation findings but--at least in conventional form--are unable to explain satisfactorily adjustments to unanticipated perturbations.

A fundamental point, from our perspective, is that considering final limb position as the equilibrium state of a constrained collective of muscles allows for the independence of final position from initial position without requiring processes of measurement and comparison. Although we could describe a dynamical system like a mass-spring in terms of externally imposed reference levels, and though we could mathematize it into canonical feedback form, little would be gained by doing so (cf. Yates, 1980, for additional remarks). A muscle collective _qua_ spring system is intrinsically self-equilibrating: Conserved values such as the equilibrium point are a consequence of the systems' parameterization and consequently there is no need to introduce a "representation" anywhere. Such systems belong to a generic class of dynamical systems called _point attractors_, that is, those characterized by an equilibrium position to which all trajectories tend.

3.4 The Importance of Dynamical Analogy

We should make our position clear on the identification of functional units of action with nonlinear vibratory systems such as mass-springs. It is obvious that a muscle has spring-like properties (the length-tension properties of an isolated muscle, for example, are well-known, e.g., Rack & Westbury, 1969), and hence it is tempting to treat each individual muscle participating in an activity as a separate mass-spring system. The resulting system would likely require large look-up tables for the purpose of specifying parameters such as stiffness and equilibrium length for each muscle (cf. Sakitt, 1980). Moreover, such a strategy emphasizes the model's material embodiment--the structural characteristics of muscle--which, though quantifiable and relatively easy to measure, tell us nothing about the nature of the organization among muscles when people perform tasks. In the spirit of Rashevsky's (1938) _relational biology_, and its enlightening extensions by Rosen (1978), we view the importance of the mass-spring analogy not in terms of the system's material structure but as indicative of a particular _functional organization_. The key insight for us is recognizing the dynamical analogy between a mass-spring system and a constrained collective of muscles and joints in terms of their functionally similar behavior (Kelso, Holt, Kugler, & Turvey, 1980; Kugler et al., 1980; Saltzman & Kelso, in press). In this respect, as Fel'dman (1966b) remarked:

> "The motor apparatus...is similar to many physical systems, for example, a spring with a load; although its movement as a whole is determined by the initial conditions, the equilibrium position does not depend on them and is determined only by the parameters of the spring and the size of the load" (p. 771).

Thus, if one ignores the question of _what_ oscillates (the material structure) and instead asks what the functional organization is, it becomes clear that many physical and biological systems (including muscles and mass springs) admit common dynamical descriptions even though they consist of utterly diverse structures. Their dynamical equivalence--to belabor the point--lies not in their physicochemical likeness but in their sharing an abstract organization. Note that this dynamical description of the cooperative behavior among muscles has little to do with the individual behavior of a muscle or its sarcomeres and fibrils. The power of the approach, however, is that it allows one to see how a wide variety of different systemic behaviors can obey the same dynamical laws. In fact, dynamical analogy may be a basic strategy open to any natural science whose "ultimate aim" in Planck's words "[is] the correlating of various physical observations into a unified system." (Planck, 1926; cited in Saunders, 1980).

Nonlinear systems of masses and springs have been traditional characterizations of many different phenomena ranging from the vibrational modes of atoms to the behavior of vocal tracts and hearts. The deep relationship among the behavior of all such structures is that they are realized by the same abstract functional organization. In a later section we shall explore this regularity in more detail, for it can be argued that the principles governing the cooperation of many subsystems are identical regardless of the structure of the subsystems themselves (cf. Haken, 1977).

4. MODULATION OF COORDINATIVE STRUCTURES

4.1 Some Remarks on Functional Nonunivocality

A second fundamental insight of Bernstein's (1967) was the realization that actors are mechanical systems, subject to gravitational and inertial forces as well as to reactive forces created by movements of links in the biokinematic chain. A consequence of this fact is that the relationship between motor impulses and their outcome in movement must be indeterminate (nonunivocal). This problem may be considered as the mirror image of a problem that perceptual theorists have long recognized--that is, the lack of a simple one-to-one relationship between a physical stimulus and a psychological percept. In speech perception, for example, many different acoustic patterns may, in different contexts, be perceived as the same phoneme and the same acoustic pattern may be perceived as different phonemes (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Rakerd, Verbrugge, & Shankweiler, 1980; among many others). In motor control, different contextual conditions may require very different patterns of innervation in order to bring about the same kinematic movement, whereas the same pattern of innervation may produce very different movement outcomes. The different "contextual conditions" of a movement depend not only on environmental changes, but also on the dynamic state of component segments. This problem is magnified in biokinematic chains (such as humans): The body segments have mass and, once impelled, gather momentum and develop kinetic energy, which may in turn provide forces acting on other segments in the chain.

Consider this anatomical/mechanical source of indeterminacy in a bit more detail. The fact that a link in a biokinematic chain is accelerating does not necessarily imply that the movement is under direct muscular control. Acceleration of a link may also be a function of reactive forces contingent on movements of adjacent links. Further, the force that one link exerts on another is not only dependent on muscle forces exerted on the first link, but also on the manner in which the first link is moving relative to the second. For example, during locomotion the limb transition from hip flexion through hip and knee extension is largely due to passive forces. The inertial torque generated by flexing the hip is sufficient to continue the forward movement of the leg from the hip and to extend the knee and ankle (Arshavskii, Kots, Orlovskii, Rodionov, & Shik, 1965; Grillner, 1975). Such is the case even when the hip musculature is slightly active, a condition that, in the absence of other forces, would bring the leg backwards (Bernstein, 1967).

Another (very different) source of indeterminacy between central commands and movement consequences is of physiological origin. Most fibers of the pyramidal motor system of primates, once thought to synapse directly on the motoneurons, actually synapse on spinal or brainstem interneurons (cf. Dubner, Sessle, & Storey, 1978; Evarts, Bizzi, Burke, Delong, & Thach, 1971). The "state" of the interneurons is dependent on the combined influence of supraspinal descending pathways, spinal interactions, and afferent nerve impulses. Thus, the interneuronal system may provide an excitatory or inhibitory bias of the motoneurons. If the bias is such that the membrane potential of the motoneuron is close to threshold, a very small additional depolarization results in its firing. As Granit (1977) remarks, "...the inte nuncial apparatus does what the gamma motor fibers do for the muscle spindle by contracting their intrafusal fibers; it determines the motoneuron's bias from moment to moment as required by the task at hand" (p. 162). Thus, the same descending activity might encounter very different "states" in the spinal interneurons, with considerable variation in the motor effect. Central influences, then, are thought to serve an organizing function by biasing lower-level systems toward producing a class of actions, but the lower-level systems can adjust autonomously to varying contextual conditions. We consider in more detail below some forms that modulation or tuning of coordinative structures might take.

4.2 "Tuning" Coordinative Structures

Constraints--analogous to the grammar of a language--do not uniquely determine a movement's trajectory, but rather allow a rich set of controlled trajectories. How then can actions be modulated according to changing environmental circumstances, yet still maintain their fundamental form? A clue may be gleaned from Gel'fand and Tsetlin's (1971) argument that well-organized functions allow a mutable partitioning of variables into those that preserve qualitative aspects of a movement's structure (termed "essential") and those that produce quantitative, scalar changes (termed "nonessential"). Bernstein (1967) argued along similar lines, noting that for living things, qualitative characteristics of space configurations and of the form of movement predominate over quantitative ones. For example, a birch leaf differs from a maple leaf by qualitative properties of the first order, whereas all maple leaves belong to the same class in spite of the large amount of biometric variation among members of the class.

Boylls (1975) has formalized a set of constraints on the electromyographic (EMG) activity of linked muscles that could preserve relational aspects of an action over scalar change. First, the timing of activity in components of a functional unit will be relatively independent of the amplitude of activity. Second, the ratios of EMG activity among muscles will remain roughly fixed relative to the time frame and the absolute levels of individual activity. Thus, according to Boylls, most actions can be partitioned into three relatively independent descriptions: 1) a temporal description that refers to the relative timing of activity in components of the linkage; 2) a structural description that defines the ratio of activity among linked variables and changes slowly with respect to real time; and 3) a metrical specification that operates as a scalar multiplier of activity in the linkage. As we shall see, it is the relationships among muscles that persist (hence "essential") over metrical variation.

The foregoing characterization of constraints immediately suggests three important questions. First, can we see constancies in the timing relations among components of diverse activities across metrical changes? Second, do these constraints hold only at the level of muscle activity, or do they also describe the kinematics of movement? Third, what are the sources of metrical modulation? With regard to the first question, because the timing of an act is hypothesized to be independent of the force requirements, one should be able to uncover timing constancies, by altering the metrics (e.g., to change the speed or force of production). Those variables that are unaltered across scalar change may prove crucial if a given motor pattern is to be characterized as an instance of a certain class of actions.

This strategy has proved successful in uncovering coordinative structure styles of organization in many different types of activities. The most well-known and abundant data come from studies of locomotion. For example, when a cat's speed of locomotion increases, the duration of the "step cycle" decreases (cf. Grillner, 1975; Shik & Orlovskii, 1976) and an increase in activity is evident in the extensor muscles during the end of the support phase of the individual limb (when the limb is in contact with the ground). Notably, the increase in muscle activity (and the resulting increase in propulsive force) does not alter the relative timing of activity among functionally linked extensor muscles, although the duration of their activity may change markedly (Engberg & Lundberg, 1969; MacMillan, 1975; Madeiros, 1978; see also Schmidt, 1980, and Shapiro & Schmidt, 1982, for further reviews).

Constancy of timing relationships in muscle activity has been reported for other obviously cyclical activities, such as mastication and respiration (see Grillner, 1977, for review). More recently, however, the stability of the timing prescription over metrical change has been shown to characterize muscle activity associated with less obviously cyclical or stereotyped activities, such as postural control (Nashner, 1977) and voluntary arm movements (Lestienne, 1979). Limited electromyographic evidence exists as well that this style of organization is characteristic of speech production. Tuller, Kelso, and Harris (1982a) found that the relative timing of activity in various articulatory muscles is preserved across the large changes in duration and amplitude of activity that accompany suprasegmental variations in syllable stress or speaking rate.

With regard to the question of generalizability to kinematics, there is a growing empirical base in which kinematic descriptions of motor actions are qualitatively similar to the electromyographic descriptioons we have been discussing. For example in handwriting, a highly developed motor skill, the relative timing of major features within a word does not change with variations in writing speed (Viviani & Terzuolo, 1980). In speech production, the relative timing of articulatory movements in a given utterance is stable across different speaking rates and stress patterns (Tuller, Kelso, & Harris, 1982b). A similar situation occurs in bimanual movements--relative timing between the limbs is preserved even when they are performing different spatial tasks with different force requirements (Kelso, Southard, & Goodman, 1979a, 1979b). This organizational style may also apply to the kinematics of coordinated systems with very different physical structures. For example, when subjects are asked to produce a string of monosyllables while tapping a finger, they have no trouble with the task. But when subjects are asked to perform the tasks at different rates, they do so by small integer sub- or superharmonics. A true dissociation in the timing of speech and manual gestures does not appear to be possible when both tasks are involved (see Kelso, Tuller, & Harris, 1983, for details).

It seems obvious that our first two questions can be answered in the affirmative: Timing relations among electromyographic and kinematic events appear stable over metrical change. But what are the sources of metrical change? Can coordinative structures be "tuned" by sources other than direct, central nervous system command? Put another way, what can we get "for free" or with minimal computational cost before we burden the nervous system with sole responsibility for control? For example, turning the head seems to bias the system for extension of limbs on the side to which the head is turned, and for flexion of limbs on the opposite side. Similarly, Easton's (1972b) experiments show that when cats look up, stretching their eye muscles, there is spinal biasing that facilitates extension of the forelimbs. When the cat looks down, there is a bias toward forelimb flexion. Such tuning relationships may be exploited by athletes (Fukuda, 1961) or under conditions of fatigue (Hellebrandt, Houtz, Partridge, & Walters, 1956). The exploitation of systemic relations may also help account for certain details of ipsilateral eye-hand coordination in split-brain monkeys. Gazzaniga (1966, 1969) reported that split-brain monkeys had to orient the eyes, head, and neck toward the target food in order to reach accurately, although the reach itself did not appear to be under moment-to-moment visual control. Although this interpretation is ours and not Gazzaniga's, it may be that the monkeys were exploiting systemic biasing relations to facilitate arm extension.

Another source of physiological tuning that is currently receiving much attention is the biasing of spinal organization that occurs before and during voluntary movements (cf. Gottlieb, Agarwal, & Stark, 1970; Kots, 1977). Such experiments examine changes in excitability of motoneuronal pools by eliciting a monosynaptic Hoffman reflex and recording its amplitude over time. Gottlieb et al. required subjects to track a visual target by controlling the amount of force on a foot plate. Approximately 60 msec prior to any evidence of voluntary EMG activity in the agonist muscle for the upcoming movement there is a progressive increase in the agonist muscle's reflex excitability. In other words, the increase in reflex excitability acts to facilitate the upcoming movement. Simultaneous with increased excitability in the agonist muscle, the level of excitability in the antagonist muscle is depressed (Kots & Zhukov, 1971). Thus, prior to any actual movement, boundary conditions arise that predispose the nervous system to produce one of a restricted class of movements (see also Fowler, 1977; Kelso, 1979; Lee, 1980; and Saltzman, 1979, for a more expansive review of preparatory tuning).

The relationships among muscle systems are not the only sources of tuning for movement. The different perceptual systems can be extremely rich sources of modulation. Dietz and Noth (1978), for example, provide convincing evidence for optical information as a source of control in motor actions. In their experiment, subjects were asked to fall forward, hands first, onto a platform that could be tilted so that different falling distances were required. Electromyographic activity was monitored in the triceps brachii, which were used to extend the arms for bracing against the fall. When subjects were able to see the platform, the onset of EMG activity began a constant amount of time before impact (and thus a variable amount of time aft:r starting the fall), regardless of how far away the platform was. When the subjects were blindfolded, the muscle response began at the beginning of the fall (see also Lee, 1976, 1978; Lee & Lishman, 1974).

Orientation-specific optical change can also bias an actor towards performing a class of movements, although no movement actually occurs. For example, when a large disk of colored dots is placed in a cat's line of sight and rotated to the left (optically indicating a tilt of the cat to the right) the extensor reflexes on the cat's right side and the flexor reflexes on the left side are enhanced (Thoden, Dichgans, & Savadis, 1977). Had the cat actually been tilted in the direction specified by the optical flow, the reflex changes would facilitate the cat's regaining an upright position.

The perceptual tuning of the action system is not tied to a particular sense modality. For example, one vision substitution device for the blind transmits a pattern of intensity differences from a camera to a bank of mechanical vibrators on the "viewer's" back. In this situation, rapid expansion of the tactile array specifies a large, rapidly approaching surface that the viewer moves to avoid (White, Saunders, Scadden, Bach-Y-Rita, & Collins, 1970; for details concerning how global expansion of the optical array might specify movements of the observer, or of large objects in the environment, see Gibson, 1950, 1966). Other sources of tuning of the action system may be vestibular (e.g., Melville Jones & Watt, 1971a, 1971b) or auditory (Davis & Beaton, 1968; Pal'tsev & El'ner, 1967; Rossignol, 1975; Rossignol & Melville Jones, 1976).

In summary, we have seen how constraints defining coordinative structures preserve relationships among components but still enable flexibility by allowing variables to take on different values. The chief characteristic of coordinated activity, we have argued, is that it exhibits relational invariance over metrical change. Metrical specification, as we have noted, amounts to a tuning of the coordinative structure. As emphasized by Greene (1972) and later by others (e.g., Fitch, Tuller, & Turvey, 1982), tuning an otherwise invariant structure is an efficient way of producing flexibility with a minimal amount of reorganization.

5. UNITS OF ACTION AS RATIONALIZED BY NONLINEAR SYSTEMS ANALYSIS

We have noted that a chief feature of units of action rests in a mutable (functionally-specific) partitioning of component variables into those that preserve the structural or "topological" (in the Bernstein sense) organization of movement and those capable of effecting scalar transformations on the structure. Here we address briefly--because it is laid out in more detail elsewhere (cf. Kelso, 1981; Kelso et al., 1980; Kugler et al., 1980, 1982)--the theoretical framework that may best rationalize units of action.

Moreover, the framework that we shall elaborate allows us to identify other criterial properties of action units that are crucial from a biological perspective, though seldom if ever recognized. Fundamentally, a functional unit at any level can be defined as a cluster of elements of various kinds that is just sufficiently organized to produce a persistent function (cf. Iberall, 1978). Unlike currently popular theories that view control as effected through a preestablished arrangement among component parts (a cybernetic machine) or due to a set of prescribed orders (an algorithmic machine), this definition recognizes that first and foremost biological systems belong to a class of physical systems that are open to fluxes of energy and matter with their surround. In contrast, cybernetic and algorithmic machines are closed to exchanges of energy and matter with their environment and hence are likely to apply to a very limited set of circumstances. The order and regularity observed in living organisms are brought about, in Bertalanffy's (1973) words, "by a dynamic interplay of processes," based upon the fact that living things obey the laws of open, irreversible thermodynamics. Unlike machines, open systems can actively evolve toward a state of higher organization.

The recognition that the flow of energy through the system plays an active organizing role and that stability can only be maintained at the price of energy dissipation (e.g., Haken, 1977; Iberall, 1977; 1978; Iberall & Soodak, 1978; Katchalsky, Rowland, & Blumenthal, 1974; Morowitz, 1978, 1979; Prigogine & Nicolis, 1971; Yates, 1980), provides a key to understanding the temporal stability that we have highlighted as a main feature of units of action. Energy dissipated, of course, must be replaced if persistent function is to be possible; it is this requirement that allows us to see that the stability is not a static one in the equilibrium sense, but a dynamic stability consisting of stable periodicities and cycles. Morowitz's (1978, 1979) theorems offer a needed insight: Work is accomplished any time there is a flow of energy from a source of high potential energy to a lower potential sink; this source-sink flow will lead to at least one cycle in the system (for numerous biological examples, see Yates, 1980, and for a detailing of neural periodicities, see Iberall & Cardon, 1964). A clarification of the type of cycle that characterizes biological systems affords a unique opportunity to identify fundamental properties of action units. Specifically, we shall see that action units are persistent, temporally stable, and autonomous entities (cf. Iberall, 1975; Yates, 1980; Yates & Iberall, 1973; Kugler et al., 1980, for applications to movement).

Consider the ideal, linear harmonic oscillator as a class of device that exhibits repetitive motion. Once started, such a system can continue indefinitely without dissipative losses. But for that reason, it is not a realistic physical entity, because all real systems dissipate energy. We can introduce a dissipative term (such as damping due to friction) into the following equation of motion:

$$(1) \qquad m\ddot{x} + b\dot{x} + kx = 0$$

where x=displacement, m=mass, k=stiffness, b=damping. However, the motion that results will run down, because no means are provided to overcome the energy losses. To obtain persistence of motion in a dissipative system, that is to compensate for energy losses due to friction, a nonlinear coupling term must be introduced. The latter constitutes an "escapement" forcing function that permits a pulse of energy, ε, to be drawn from a continuously available source of potential, and injected into the system at appropriate phase, θ:

(2) $m\ddot{x} + b\dot{x} + kx = \varepsilon(\theta)$

It is important to emphasize that the "escapement" forcing function (like the escapement in a grandfather clock) is not strictly time-dependent; hence it is autonomous in the conventional mathematical sense; it is an intrinsic timing mechanism in the sense that $\varepsilon(\theta)$ is drawn from a potential energy source that is part of the system itself. There is no ghost driving the machine from the outside or providing instructions to the oscillatory component (cf. Minorsky, 1962; Yates, 1980).

Equation (2) can be rewritten to reveal that the escapement pulse exactly offsets the energy loss averaged over each cycle, so that periodic motion is assured:

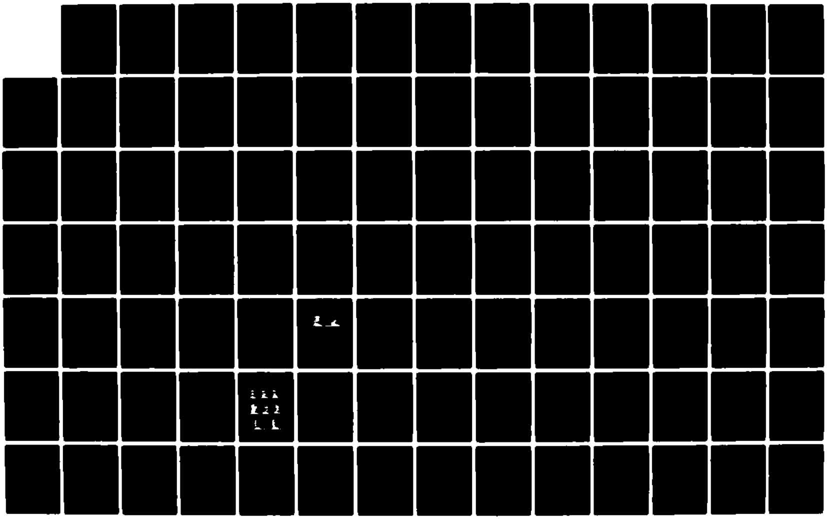
(3) $m\ddot{x} + kx = \overline{\varepsilon(\theta) - b\dot{x}} = 0$

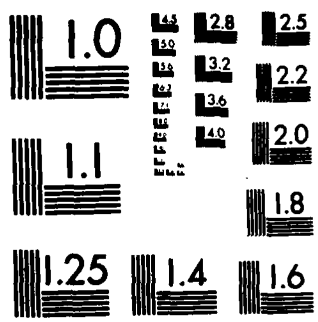
where the bar expresses an average. Systems described by nonlinear equations such as (2) and (3) are called <u>limit cycles</u> because they will settle into steady, near isochronous motion of fixed amplitude independent of sporadic disturbances and initial conditions (see also Section 3.3). Thus, if an oscillatory component is displaced with a push of large amplitude, its loss of energy will be greater than the escapement pulse can provide to offset it. The system will lose amplitude until energy balance (orbital stability) is achieved. Similarly, a small change in initial displacement is associated with smaller frictional losses than the energy pulse injected. Amplitude therefore will grow until the system reaches a balanced state, characterized by limit cycle behavior, that is, a closed cycle of events on the phase plane (cf. Jordan & Smith, 1977; Minorsky, 1962). The limit cycle, then, constitutes a <u>periodic attractor</u>, in current terminology (see Gurel & Rössler, 1979, for many examples) to which all deviated states tend. Limit cycles have been used to model many different neural phenomena, from EEG (Basar, Demir, Gönder, & Ungan, 1979; Freeman, 1975; Kaiser, 1977) to excitatory and inhibitory interactions in neurons (cf. Wilson & Cowan, 1972). More fundamentally, however, the persistent, self-sustaining, autonomous, and orbitally stable trajectories of nonlinear, limit cycle systems are manifestations of thermodynamic engines. Such engines sustain cyclic motion by absorbing over the course of each cycle an amount of free energy that just balances the energy dissipated per cycle. Without this energy balance, the system would simply decay toward a static equilibrium state (Iberall, 1977, 1978a, 1978b; Yates, 1980; Yates & Iberall, 1973).

As far as the control and coordination of movement are concerned, the implication of this discussion is that a unit of action at any scale of analysis must fulfill thermodynamic criteria (cf. Kugler et al., 1980). Moreover, the chief distinguishing features of a coordinative structure, namely, the dissociation of power and timing and the fixed proportioning of activity among elements (see Section 4), are neither arbitrary nor exotic. To the contrary, the phase-dependent energy input pattern guarantees that the timing and duration of energy inputs will be independent of the magnitude

AD-A129 713 STATUS REPORT ON SPEECH RESEARCH A REPORT ON THE STATUS AND PROGRESS OF S..(U) HASKINS LABS INC NEW HAVEN CT M STUDDERT-KENNEDY ET AL. 1983 SR-73(1983) 3/4

UNCLASSIFIED N00014-83-C-0083 F/G 6/16 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

within a fixed time frame (a period of oscillation). Also, the magnitude of the input or 'squirt' will be a fixed proportion of the power supply. The stability regime realized by a nonlinear system such as a coordinative structure is asymptotic and orbital; the limit cycle "quantizes" action (formally, the product of energy and time, cf. Iberall, 1978a, 1978b) and the system's conserved values or equilibrium operating conditions are specified in the loose coupling among limit cycle processes (see for example Goldbeter, 1980; Kawahara, 1980; Smith, 1980).

Extending the foregoing identification of coordinative structures with limit cycles may allow us to intuit how the dynamic organization of the action system for a particular activity may constrain where and when perceptual information can be most effectively "picked-up" (Gibson, 1950, 1966, 1979). We have seen that the design of the system, with its source of potential energy, nonlinear escapement, and oscillatory component, determines when in the cycle the energy source will be tapped. The mathematical description of this is an autonomous one in which time itself is not formally represented; no "extrinsic" timing mechanism is required (see Fowler, 1980, for a comparison of models of "extrinsic" and "intrinsic" timing). Such a description fits the work we have already mentioned on so-called "reflex reversal" (Forssberg et al., 1977) in which the same input can have very different behavioral effects when it occurs in different phases of the step cycle. Similarly, in Orlovskii's (1972) work on cat locomotion, neural stimulation of Deiter's nucleus in the mesencephalon of a stationary cat results in limb extension. Continuous stimulation of the same nucleii in a walking cat enhances extension only during the extensor phase of the step cycle. Neural stimulation (perceptual information?) is gated according to the nature of the systemic organization, and limited to that phase of the cycle where its effect is adaptive.

The identification of functional units of action, coordinative structures, with limit cycle mechanisms offers a number of attractive features for a programmatic approach to problems of coordination and control. Chief among those undergoing empirical exploration (see Kelso, Holt, Kugler, & Turvey, 1980; Kelso, Holt, Rubin, & Kugler, 1981; Kelso, Tuller, & Harris, 1983) are stability (in the face of unforeseen perturbations), persistence (as a rhythmical pattern), mutual entrainment (between like and different anatomical structures), and capability to exhibit new modal forms (see Section 6 below). Our perspective interfaces nicely with earlier (e.g., von Holst, 1937/1973) and newly emerging oscillator theoretic views of neural control (cf. Delcomyn, 1980; Gallistel, 1980; Grillner, 1977; Stein, 1977) although it differs in important and nontrivial ways. The attributes we have articulated here arise—not necessarily because of special biological mechanisms (like central programs)—but because living systems belong to a particular class of open, physical system.

Currently dominant model constructs for movement control stress the reflex arc and the servomechanism as basic building blocks. The reflex arc is composed of effector, conductor, and initiator elements (Gallistel, 1980). Modern servocontrol theory keeps the effector (output) and the initiator (input as referent level) and adds additional processes such as feedback, comparison, and error correction. But in the present view, machine concepts having to do with adaptive controllers, feedback, and programs are not likely to be useful to our accounts of the order and regularity displayed by biological systems (cf. Kelso, 1981; Kugler et al., 1982, for more detailed arguments). Living things, as Yates (in press) cogently remarks, "...are not hard-wired, hard-programmed, hard-geared, or hard-molded. They persist, as ill-defined systems, marginally stable in a nonlinear sense (while being linearly unstable)." As dynamical systems with active, interacting components and large numbers of degrees of freedom, they are capable of spontaneous organization and evolution of function.

Up to now we have been concerned with those principles that guarantee structurally stable modes of coordination in the face of quantitative variation in control parameters. Now we address the other side of the coin, namely, how do new forms of spatiotemporal organization come about? How do old "kinetic forms" give way to new ones?[7] We first consider some examples in nature that may allow us to intuit an answer (cf. Haken, 1977; Katchalsky et al., 1974; Kugler et al., 1982, for more details); we then consider some specific examples that are continuous with our earlier discussion of oscillatory systems, and that are based on our own and other's movement research. A fundamental feature of all these examples is that qualitatively new modes of organization emerge when certain parameters are scaled past critical bounds. Importantly, these new modal behaviors may reduce the requirement for a priori programs in the sense of a prescription for a phenomenon existing before the phenomenon appears.

6. DYNAMICS OF NATURAL SYSTEMS

We are concerned here--as we have been all along--with systems of many degrees of freedom that somehow cooperate with each other to produce regular and orderly behavior (at a macroscopic level). Cooperative phenomena are well known in physical systems and have provided a basis for many technical applications. Common to all of these (e.g., the laser, tunnel diodes, ferromagnetism) is a transition from a disordered state to a more highly ordered one. Unlike say, semiconductors, which achieve ordered states when temperature is lowered toward equilibrium, systems such as the laser undergo phase transitions only when they are driven far from equilibrium--they are dissipative or synergetic structures by virtue of degrading a good deal of free energy (cf. Haken, 1977; Katchalsky et al., 1974; Prigogine, 1980; see Kelso, Holt, Kugler, & Turvey, 1980, and Kugler et al., 1980, 1982, for empirical and theoretical treatment of a dissipative structure perspective on action). Although it is a minor point, elsewhere (after Katchalsky et al., 1974) we have preferred the term "dynamic pattern" to "dissipative structure" because it removes any ambiguity between classical notions of the term structure and Prigogine and colleagues' dissipative structure (Kelso et al., 1983). Both terms, however, are synonymous and refer to a functional or dynamic organization.

6.1 Physical Examples of Emergent Modes

Several examples will allow us to demarcate the main features of dynamic patterns and the conditions under which they arise. Some of these attributes have been considered already in Section 5. These examples will necessarily be sketchy from a mathematical point of view but they allow us to convey a flavor of the approach.

Consider the simple example of turning on a faucet. At low levels of water pressure (flow through the nozzle), the flow of water is nonturbulent, or laminar. Although laminar flow seems well ordered, in fact the movement of water molecules follows a random statistical law. As the tap is opened more, and water pressure is increased, the flow may no longer be laminar in appearance. In fact, at a critical point of pressure water takes on a turbulent or "muscular" appearance (in accord with the theme of this chapter) in which molecules now display coherence in the form of powerful streams. If the tap is opened still more, other abrupt changes--vortices and the like--are possible. The theme that emerges here is that the continuum of atomisms (laminar flow) becomes unstable and, at a point at which inertial forces greatly predominate over viscous ones (characterized by a dimensionless ratio called a Reynolds number), gives rise to a new stability (observed as turbulence).

The convection instability of Bénard allows us to secure these ideas more firmly. When a fluid layer (such as spermacetti oil) is placed in a large pan, heated uniformly from below, and kept at a fixed temperature from above, initially--if the temperature gradient is small--the fluid will remain quiescent. In this case, heat spreads through the fluid by heat conduction, a process in which molecules undergo thermal vibrations and transfer a part of their thermal energy in collisions without, on the average, changing their positions. As the temperature gradient is increased, a state of thermal nonequilibrium is reached and convection occurs. At the beginning, small convection streams (macroscopic motions) are suppressed, but as the temperature gradient is increased to a critical value, fluctuations are amplified and macroscopic motions occur. These take the form of rolls or hexagons, depending on boundary conditions (cf. Koschmeider, 1977). The new ordered states are themselves open to increased structuralization, because at higher values of the temperature gradient further patterns, such as oscillatory 'spokes' are possible. Fluctuations play a vital role, because without them higher order states cannot evolve. Moreover, the nature of the fluctuations themselves significantly affects the new order that is established (e.g., polygons, hexagons; cf. Koschmeider, 1977, for many more details of a much more complicated story than that relayed here). One interesting aside to the Benard effect that is relevant to our earlier discussions of equifinality in the motor system and to dynamic patterns in general is that given patterns need not relate to a unique mechanism; conversely, different mechanisms may generate a common pattern (cf. Katchalsky et al., 1974). Thus, biological systems are not unique in displaying convergence (many-to-one mappings) and divergence (one-to-many mappings) (see Section 7).

6.2 Summary 1

There are several lessons to be learned from the foregoing examples in physical systems before we consider matters of biology. First is the notion mentioned earlier, that systems at many scales of magnitude exhibit transitions from one state to another that are discontinuous even though the factors controlling the process change continuously. Second, and relatedly, transitions from one mode to another are discontinuous, not because there are no possible intervening states but because none of them is stable. Thus, the transition from one state to another is likely to be brief compared to the time spent in stable states. Third, and in the Poincare-Thom tradition, for new modes to appear, all that need change mathematically is the qualitative shape of the potential curve that occurs only when an equilibrium condition is created or destroyed. A consequent implication is that there may be a relatively large number of ways for a system to exhibit continuous change, but only a relatively small number of ways for it to change discontinuously. We associate the discontinuities with nonlinear properties that are revealed when the system is scaled (putatively a continuous process) to some critical value.

6.3 Biological Examples of Emergent Modes

Let us see how the foregoing style of inquiry is relevant to matters of greater interest to the motor physiologist and cognitive psychologist. Consider first the forms of gait that an animal might display and the causal basis for transitions among gaits. Relatively little is known about locomotory patterns or the transitions among them. It is tempting, however, to assume that a given gait is governed by a central program (or in noncomputer jargon, a central pattern generator) that prescribes the kinematic details for cyclical flexion and extension of limbs. Switching among gaits could be accounted for by assigning a "gait selection process" to the animal (Gallistel, 1980). There are good reasons to be skeptical of such a view, which ranks in the "just so" category. A primary one stems from a remarkable experiment by von Holst (1937/1973) in which he amputated the legs of a centipede (Lithobius), leaving only three pairs of legs intact (see also von Buddenbrock, 1921, for a similar but less drastic manipulation). Regardless of how large an anatomical gap was left between remaining legs (up to five segments), the centipede (which normally walks with adjacent legs about one-seventh out of phase) assumed the gait of a six-legged insect. Furthermore, the asymmetric gaits of the quadruped were displayed when all but two pairs of legs were amputated. Von Holst (1937) used these experiments to argue against "any fixed reflex locomotor relationship between the legs"--but the message surely applies equally to central pattern generators. It is facetious to suggest that the animal stored all possible representations of locomotory patterns in anticipation of some innovative experimenter (or a small boy) performing an amputation! It seems more likely--and a route for the scientist to explore--that the design of the animal places considerable constraints on which locomotory states are dynamically stable in the equilibrium sense and which are not.

What then of gait transitions? In the case of the quadruped it is well established that there are only a few modes of locomotion. At low speeds, the common mode is one of asymmetry between limbs of the same girdle characterized by a half period (180 degrees) difference in phase. At higher speeds, the limbs of the front and rear girdle shift--in a fairly abrupt way--to an in-phase, symmetrical mode. How might the gait transition be interpreted? A first clue comes from observations that horses (Hoyt & Taylor, 1981) and migrating African gnus (Pennycuick, 1975) use a restricted range of speeds within each gait that corresponds to minimum energy expenditure. In fact, for the horse, the minimum oxygen cost per unit distance is almost the same for walking, trotting, and galloping (cf. Hoyt & Taylor, 1981). As speed is increased, however, the locomotory mode (say walking) becomes unstable; it becomes extremely costly to maintain that mode at a given rate. The walking mode becomes unstable, as it were, and "breaks" into a trotting mode. Similarly, it is energetically expensive to maintain a trotting mode at slow locomotory speeds, a fact that appears to dictate a switch into the walking mode. The discontinuous nature of these transitions suggests--like some of the physical examples earlier--that when a critical value is reached, the system bifurcates, revealing a qualitative change in its topological structure. More generally, the different gaits may be interpreted as those few stable modes that can arise as a consequence of scaling up on muscle power (see also Kugler et al., 1980, for more on topological approaches). The stable range of speed for each modal gait corresponds to regions of minimum energy dissipation. It should be emphasized that there is a good deal of overlap between the locomotory modes (see Hoyt & Taylor, 1981, Figure 2) and that the account given here is not that locomotory modes are hard-wired and deterministic. Horses can trot at speeds at which they normally gallop, but it is metabolically expensive to do so.

The account of gait shifts in terms of nonequilibrium dynamics would be enhanced if qualitatively similar types of phenomena were observed in other types of activities--activities perhaps of a less stereotypic kind.[8] In our final examples we discuss voluntary manual activities and speech. Consider an experiment (reported briefly in Kelso, 1981) in which a subject is asked to cycle the hands at the wrist using asymmetrical muscle groups. Thus, direction of movement is the same for each hand; flexion (extension) of one is accompanied by extension (flexion) of the other. The only instruction to the subject is to increase rate of cycling--provided either verbally (at approximately 15 sec. intervals) or by a pulsing metronome. An example of the data is given in Figure 1, which plots the displacement-time profile of the hands singly (top half) and against each other (bottom half). It can be seen that the hands shift from an out-of-phase pattern (asymmetrical muscles) to an in-phase pattern between points H and T. The shift is evident in the Lissajous figure below, where it can be seen that within a cycle the hands 'kick' into a different mode. The same data are shown in Figure 2, except that it is easier to see what is going on as one steps through the data file shown on the upper left of the figure. It can be seen that the phase relations between the hands are very stable in Figures 2A and 2B. Were the two motions perfectly sinusoidal with phase $= \pi$, a straight line would be observed. In Figure 2C, the phase difference between the two hands has undergone a modest increase and also become more variable, as evident in the widening of the Lissajous trajectories. However, it is also clear that a fairly abrupt change of phase occurs; descriptively, the left hand "slips in" an extra half-cycle while the right hand waits, and then both perform synchronously (symmetrical muscle groups). Figure 3 represents the same data on the phase plane in which position is plotted against velocity for each hand. It can be seen in the center portion of the figure that the two hands start out in different quadrants of the phase plane but end up in the same quadrant (with approximately the same position-velocity coordinates for each hand; see figure caption for full description).

Although this example warrants more detailed analysis than that given here, it is nevertheless quite clear that a similar qualitative picture emerges for voluntary hand movements as for the gait transitions discussed earlier. That is, a qualitatively new modal pattern emerges as a function of continuously scaling on a single parameter (in this case rate). The change in phase occurs relatively quickly compared to the time spent in the modes themselves--often within a single cycle. Importantly, these data suggest

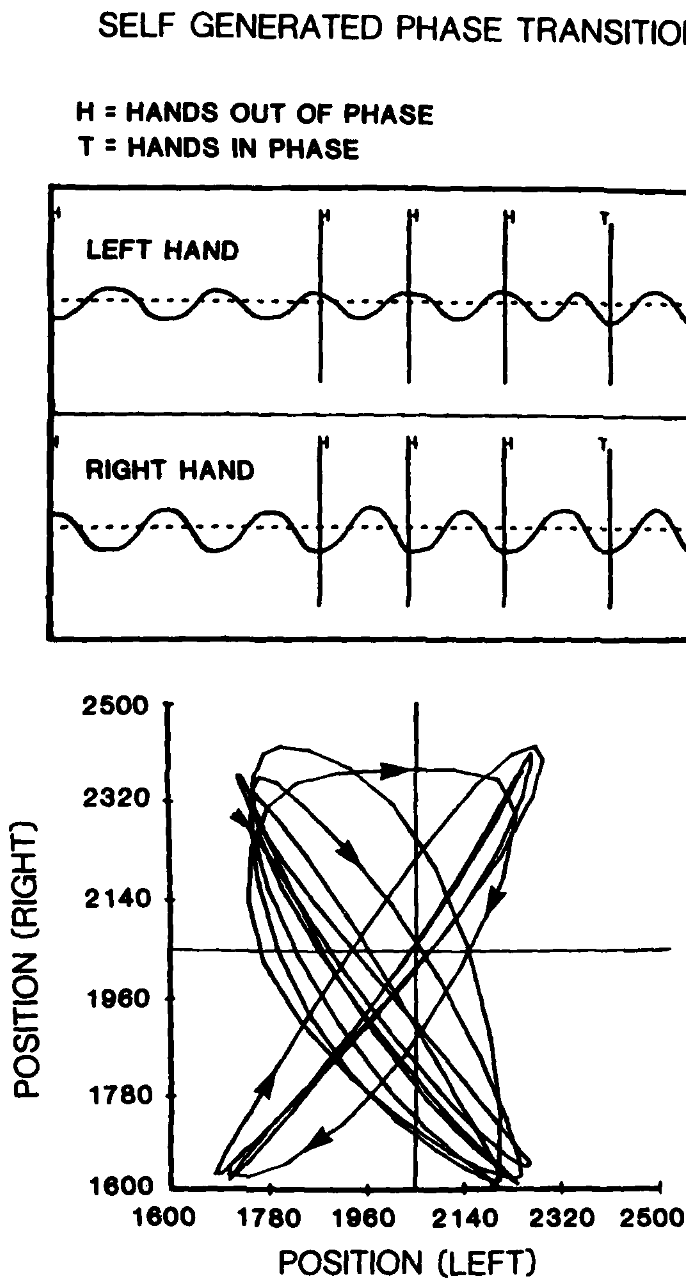


Figure 1. Displacement-time profiles of left and right hands (top) and position of each plotted against each other (bottom) as a Lissajous figure. "Hands out of phase" means that flexion of one hand is accompanied by extension of the other and vice-versa. That is, direction of movement is the same for each hand (ignore plotting convention). "Hands in phase" means that both hands flex and extend at about the same time. The figure shows a shift from out of phase to in phase as rate increases (that is, as one examines the data file from left to right).

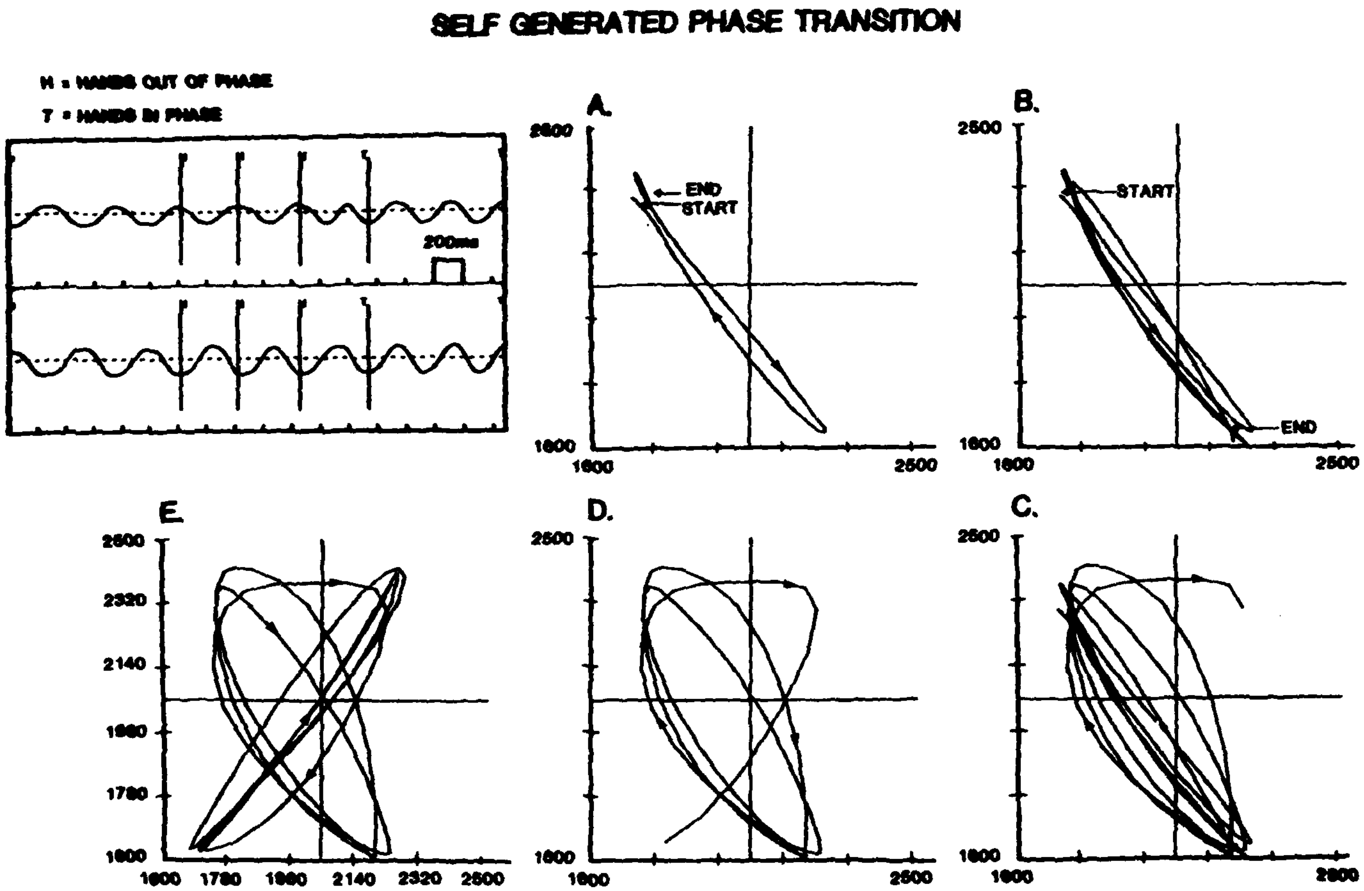


Figure 2. *Same data* as displayed in Figure 1, but Lissajous *figure of* left vs. right hand is plotted as one steps through the data file (A-E). For description see text.

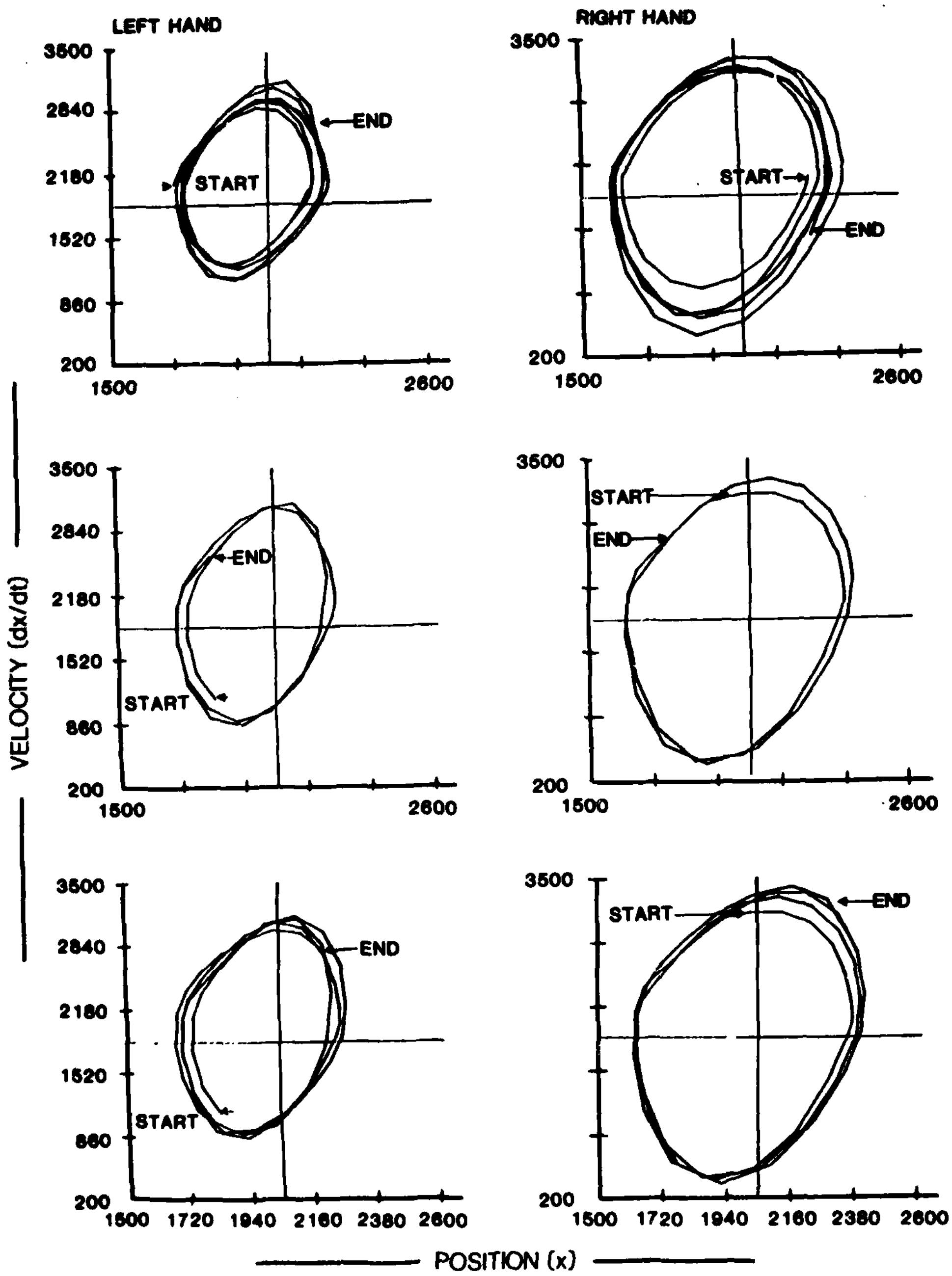


Figure 3. Same data as displayed in Figures 1 and 2, but plotted as phase plane trajectories for left and right hands. Position and velocity are expressed as arbitrary units. Top third shows the phase plane trajectories of the two hands prior to a phase transition. The hands start and end in different quadrants of the plane. Middle third shows the transition itself with the trajectories sampled over the same time window. It is clear that the left hand produces an extra half cycle so that the hands end up in phase. Bottom third of the figure overlaps with middle third and proceeds to end of file.

rather strongly that the new mode is revealed by scaling on a system sensitive parameter. It appears also that only two modes are stable; other phase relations--at least in unpracticed subjects--appear highly unstable.

We turn now to a final example, one that offers a potentially rich but little explored domain for the style of inquiry being advanced here. We refer to speech production and perception and in doing so draw principally from the observations and discussions by Catford (1977) and Stevens (1972, 1977).

Speech, of course, is a complex process arising from the interactions among articulators at several levels--respiratory, laryngeal, and supralaryngeal. A good deal of effort has been directed toward the identification of distinctive acoustic attributes as they may underlie the phonetic categories described by linguists (e.g., Chomsky & Halle, 1968; Halle & Stevens, 1971; Stevens, 1972; Stevens & Blumstein, 1978). For us, however, the acoustic attributes are of interest only to the extent that they shed light on the articulatory dynamics that produced them. It is important to recognize immediately, however, that the postures and movements of the articulators structure the sound but do not themselves generate sounds. To return to a recurring theme, articulatory configurations create the necessary aerodynamic conditions, as a consequence of which sound generation is possible. In this regard, our earlier discussion of turbulence as a highly ordered space-time phenomenon is appropriate: The presence or absence of turbulence in the vocal tract plays a significant role in the production of speech sounds such as fricatives. Below a certain critical velocity, airflow through an articulatory channel such as an open glottis will be laminar and noiseless (so-called 'nil' phonation, cf. Catford, 1977), as in the phonation of [f, s, ʃ]. Above a critical value, turbulent, noisy flow occurs, as in the phonation of stressed initial voiceless sounds [p^h, t^h, k^h].

The Reynolds number, it will be recalled, depends on the diameter of the channel (more generally, the various forms of constriction in the vocal tract), the velocity of flow, and the viscosity of air: It is the ratio of inertial to viscous forces. Beyond a certain value of the ratio, two types of turbulence arise; one, a more general type of channel turbulence (discussed above) and the other a vortex-producing wake turbulence. Wake turbulence occurs when a high velocity jet of air is produced against the edges of the upper and lower teeth, for example in production of /s/ or /ʃ/ as in 'sip' or 'ship,' respectively. Wake turbulence also plays a role in various laryngeal modes such as voiceless falsetto (or so-called 'glottal whistle'), which appears to be due in part to periodic vortex formation that develops past the thinned edges of the vocal folds (cf. Catford, 1977).

The nonlinear distinctive effects of turbulence are only one aspect of what may be a larger design principle, one in which gradual, linear changes in certain variables can lead to discontinuous, distinctive outcomes. Continuous adjustments of the vocal folds (e.g., in terms of their positioning in relation to each other, effective mass, and stiffness) also give rise to distinct modes that occur as discontinuous jumps. Like the gaits of the quadruped, there seem to be relatively few stable modes. Whisper, for example, occurs at a much smaller critical flow velocity than the production of voiceless fricatives as a consequence of much smaller glottal constriction. The voicing mode occurs when the vocal folds, in a suitably tensile state, form a narrow glottal chink, while the pressure drop across the glottis creates a Bernoulli effect. As a result, the vocal folds are set into vibration--they snap together and are forced open again by subglottal pressure, only to close once more because of their elastic properties and the Bernoulli effect (at least according to myoelastic-aerodynamic theory, see Titze, 1980, for a good review). If the vocal folds are further constricted, so-called creaky voice is evident (though not well understood), and then, when the folds are constricted to a point at which subglottal pressure can no longer drive them apart, the conditions for the production of glottal stops are created. Thus, we see in these examples of laryngeal function that from an apparent continuum of vocal fold maneuvers, a variety of modes arise. These dramatically different modes (and the story is actually much longer than we can tell here) are indicative of 'preferred stabilities' (see Section 5 on structural stability, and earlier gait and hand movement examples), and the transitions among the modes can be characterized as unstable.

To bring this discussion into the realm of the speaker/hearer, if we know anything about speech it is that "...the diverse, continuous and tangled sounds are...perceived as a scant handful of discrete and variously ordered segments" (Liberman, 1982). What befuddles the scientist is that there is no apparently direct relationship--in a linear sense--between the parameters responsible for structuring the sound (the articulatory system) and the acoustic output arising from the source. In certain cases, large changes in articulatory parameters have minimal acoustic consequences, as in Kakita and Fujimura's demonstrations that for production of the vowel /i/ a wide variety of contractile values on the tongue muscles will yield relatively invariant formant structure (Fujimura & Kakita, 1979; Kakita & Fujimura, 1977; see Kelso & Tuller, 1982, for fuller discussion). In other cases, small changes in relevant observables, such as voice onset time (Lisker & Abramson, 1964), can result in one phonemic class being replaced by another. The former constitute structurally stable articulatory parameterizations; the latter refer to unstable regions (in the topologist Thom's terms, they belong to the catastrophe set; Thom, 1975).

The existence of these complex relations (apparently at every level of the speech system and probably the ear as well) may only be a problem for the scientist who seeks out one-to-one correspondences between particular acoustic "cues" and that which is perceived. It seems to us--if the parallels we have drawn among the various examples here are appropriate--that the issue is not really one of specifying acoustic attributes that map onto a linguistic featural description (e.g., Halle & Stevens, 1971; Stevens & Blumstein, 1978). As some phoneticians and motor control researchers have remarked, this is a particularly Procrustean strategy in that it forces the data into some preestablished linguistic categorization scheme. Rather, it seems to us that the perspective offered here dictates the fairly unexplored strategy of determining which articulatory parameterizations are structurally stable and which are not (and why). More generally, it is to understand those dynamical transformations among articulators that reveal, and ultimately 'freeze out,' as it were, the modes and phonetic segments of a language.

6.4 Summary 2 (with Due Homage to Haken, 1975, 1977).

In this section we have tried to provide a flavor for what we believe to be deep analogies among many different subsystems when they cooperate to produce coherent functions. Characteristic of all the examples is that new "modes" or spatiotemporal regularities emerge when the system is scaled on certain parameters to which it is sensitive. [As an aside, if this view is viable, we suspect a good deal of work will have to be devoted to identifying what these parameters are--an enterprise that is closely affiliated to the ecological approach to perception and action advocated by Gibson (1966, 1979) and his school (Shaw & Turvey, 1981; Turvey & Shaw, 1979; Turvey, Shaw, & Mace, 1978).] In the various cases we have described, the initial modal pattern becomes unstable, and it is this instability that is a prerequisite for the emergence of new modes. "Mode" is a concept for the collective behavior of many degrees of freedom; it is characterized by a macroscopic description that is not known at a more microscopic level (see also Section 3.2). Thus, an oscillating string made up of 10^{22} atoms is described by "macro" quantities like wavelength and amplitude, which are entirely different from the description at an atomistic level (Haken, 1977). Similarly, the relevant observables for coordinative structures (and we would argue, the control and coordination of movement) are relational in time and space, they have little to do with descriptions of the firing properties of motor units.

Unlike machines that are designed by people to exhibit special structures and functions, the functions and structures discussed here develop, as it were, spontaneously--they are self-organizing. Importantly, during the scaling up process there is no a priori specification or representation of the new structure (Kelso, 1981; Kugler et al., 1980). In fact, a new mode often emerges when a random event occurs in an unstable region, when a fluctuation becomes amplified. Such is the case, one suspects, in the gait of a horse (and perhaps the singer at a particular point in the voice range--close to the passagio, Teaney, Note 2). Near the unstable region--where it is energetically costly to maintain a given mode--a small change in, say, walking speed, will have dramatic effects: a new mode will arise. Literally, a phase transition occurs.

When we see new forms of organization occur, we are addressing systems possessing many degrees of freedom that are intrinsically nonlinear and dissipative; systems that operate in "preferred" regions of their state space; systems that are structurally stable on the one hand, and capable of a fair degree of flexibility on the other--in short, systems in which variance plays on invariance. The bottom line for systems that display so-called critical behavior is that the same fundamental principles pertain regardless of the dimensionality of the system or its material structure, and that these principles are the ones that a theory of action might embrace to account for the emergence of new forms of space-time patterns displayed by the cooperative behavior of muscles and joints. The alternative--when push comes to shove--is a hermeneutic device that prescribes new orderings. If nothing else, the approach offered here promises to try to reduce Hermes' role to a minimum.

7. CONCLUSIONS: INTEGRATING PRINCIPLES OF HIGHER BRAIN FUNCTION AND PRINCIPLES OF MOTOR SYSTEM FUNCTION

Our discussion of units of action as displaying limit cycle behavior with all their attractive features, and our focus on spontaneously organizing systems with their inherent nonlinear, multimodal properties, offer potentially exciting possibilities for a deeper understanding of movement coordination and control. They represent a new and perhaps speculative development in the theory of action systems. They lead to new research directions (what are the modes of the action system and their stabilities; how limited are they; what conditions give rise to stability and instability; can transitional behavior be classified, etc., etc.). In seating action systems in physical biology, there is the promise of adequate theory. What constitutes a "new direction" or an "interesting research problem" is obviously a matter of choice. All we have done here is to make our biases apparent.

In our concluding remarks we want to end on a "tamer" note by bringing some of the ideas expressed here (mostly in Part 1) into the more standard nomenclature and conventions of neuroscience. Our vehicle is a comparison of some of the principles we have elaborated in this chapter (which, as we have intimated, have a long standing heritage) with some recently developed views of higher brain function (Edelman & Mountcastle, 1978). Although we cannot go into any great detail at this point, we will try to show by way of summary (see Table 1) that many of the kernel ideas in Edelman's "group theory" of higher brain function (Edelman, 1978) have been in the motor system's literature for some time. Our view all along has been that nature operates with ancient themes, and in Edelman's compendium, combined with certain notions expressed here, we see some consensus emerging on what these themes might be. We are encouraged to elaborate these themes in part because of an awareness that several noted neuroscientists have become disenchanted with the reductionist paradigm (e.g., Bullock, 1980; Schmitt, 1978; Selverston, 1980). In the past it has been commonplace for the neuroscientist to talk of neural circuits controlling behavior, but even in the simplest networks (and we use the term "simple" guardedly here; see below) it has proved difficult to relate specific patterns of neural activity to behavioral action. Surely there is a message here: If the strategy is deemed questionable for small circuits in terms of the number of ganglia involved--and there is informed consensus that this is the case (see commentary on Selverston, 1980)--then what hope is there for understanding a brain complex of 15 billion elements?[9] Even if we knew all the parts and their properties, we would still not know how the system operated. As Schmitt (1978) remarks:

> Many theories of higher brain function have been proposed...These theories usually rely heavily upon processes subserved by spike action potential waves travelling in hard-wired circuits...Such circuits usually consist of neurons that are large enough to permit easy impalement by microelectrodes and that possess long axons forming tracts connecting processing centers in general regions of the brain that have been characterized as sensory, motor, associational, frontal, temporal, parietal, and occipital.
>
> Theories based on partial systems are subject to the component-systems dilemma that bedevils all attempts at biological generaliza-

Table 1

Some predictions of motor control metatheory compared with some predictions of Edelman's group-degenerate[11] theory of higher brain function (Page numbers in the left column refer to Edelman, 1978.)

(1) "Groups of cells, not single cells are the main units of selection in higher brain function." (p. 92)	Ensembles of muscles and joints—called coordinative structures or functional synergies—not single muscles or joints are the significant units of control and coordination of action (Section 3)
(2) "Such cell groups will be found to be multiply represented, degenerate and isofunctionally overlapping. Many-one interactions...will be found, with extensive divergence as a sign of degeneracy." At the same time, multiple inputs...will be found to converge on the same cell group leading to abstract cell-group codes."[11] (p. 93)	Motor equivalence/equifinality is a property of action systems (Section 3). The same output can be achieved using different muscle ensembles, and different outputs can be accomplished using the same muscle ensembles. One to many (divergence, degeneracy) and many to one (convergence, abstraction) are common features of multi-degree of freedom systems (see (4) below).
(3) "No pontificial neuron, or single-neuron "decision unit" will ever be found at the highest levels of a system of any large degree of plasticity." (p. 93)	Action systems work most efficiently under assumptions of executive ignorance and addressless, distributed control—a minimally intelligent executive intervening minimally. (Sections 3 and 4).
(4) "Selection will be found to play a large, but not inclusive role in forming a first repertoire during embryogenesis...no sizeable, precommitted molecular repertoire will be found to explain cell-cell interaction in the developing nervous system." (p. 93)	Certain so-called fundamental patterns of movement may constitute a first repertoire for action systems. But fixed actions at a joint, preassembled reflexes or central pattern generators (programs) are not the principal bases of action systems. The latter are differentiated by their <u>functional</u> significance, not by their anatomical specificity. (Sections 1 and 2).

(5) "Correlations will be found that suggest phased reentrant signaling on degenerate neuronal groups with periods of 50-200 msec." (p. 93)	The behavior of muscle-joint ensembles or coordinative structures expresses a design that is fundamentally cyclical in nature as a consequence of which persistence of function, stability, autonomy, entrainment, and emergence of function (e.g., modal changes) are possible. (Sections 5 and 6)

Postscript

According to Edelman (1978) "...the selective theory of higher brain function requires no special thermodynamic assumptions and is free of mentalistic notions" (p. 94). We welcome this, but stress that the units of action must be motivated on the grounds of (irreversible) thermodynamics (see prediction 5). Indeed, any unit of brain function (like any unit of action) must not only be defined in terms of its neural structures but also the metabolic machinery that supplies energy and removes by-products. Many of the attractive attributes of action systems elaborated here follow from a dynamic, homeokinetic scheme in which the many degrees of freedom are regulated by means of coupled ensembles of limit cycle, thermodynamic engines (Iberall, 1978a, 1978b). It is this basic characterization, with appropriate extensions, that may allow us, in Edelman's terms, to "...avoid an infinite regression of hierarchical states...to provide for planning and motor output without a programmer...[to] mitigate the need for programming" (p. 94). That has been--and continues to be--the goal of so-called action theory (e.g., Fowler et al., 1980; Kelso, Holt, Kugler, & Turvey, 1980; Kugler et al., 1980; Reed, in press). Although there are obvious differences between group theory and action theory, this shared aim is not one of them.

> tion. Such theories fail to articulate and effectively deal with the essence of the problem, which is the distributive aspect that emerges from the complex interaction of functional units...in the brain. (p. 1)

Although it is clear that much still remains to be known about the parts--and we may have to wait for technology for much of this--it is equally clear that the behavior of large and complex aggregates cannot be understood in terms of extrapolations from so-called simple circuits. As we remarked earlier in this paper, constructionism breaks down in the face of scale and complexity (see Section 2.1). At each level of complexity, novel properties appear whose behavior cannot be predicted from knowledge of component processes alone. This is why the form of reductionism that we have taken here--advocated in contemporary physics and an emerging physical biology--is a reductionism to a minimum, but universal set of principles, rather than to elemental properties. This is why we see an interesting link between Edelman's theory[10] and those ideas that have over the years emerged in the area of motor systems. In this chapter, we have tried to reveal the rich heritage involved in the movement domain--stemming from the Bernstein tradition--as well as the important syntheses by people like Greene, Boylls, Turvey, and others. Only in the search for common principles can we see a true integration of very disparate disciplines--a true science of natural systems.

Throughout this paper we have remarked on the qualitative likeness--in terms of dynamical behavior--exhibited by complex, dissipative systems in spite of dramatic variations in material composition and the scale at which they are observed. Given this state of affairs, the overlap between some of the main postulates of Edelman's theory (but not all of them) and those expressed here is hardly surprising--at least to us. Thus, the principles relate to the behavior of complex systems and cooperative phenomena rather than to any particular structural embodiment. It is understanding coherent behavior that takes precedence here--not whether that coherent behavior is of ensembles of neurons, or muscles, or anything else.

REFERENCES

Adey, W. R. Tissue interactions with nonionizing electromagnetic fields. Physiological Reviews, 1981, 61, 435-514.

Anderson, P. W. More is different. Science, 1972, 177, 393-396.

Arshavskii, Y. I., Kots, Ya. I., Orlovskii, G. N., Rodionov, I. M., & Shik, M. L. Investigation of the biomechanics of running by the dog. Biophysics, 1965, 10, 737-746.

Arutyunyun, G. H., Gurfinkel, V. S., & Mirsky, M. L. Investigation of aiming at a target. Biophysics, 1969, 13, 536-538.

Basar, E., Demir, N., Gönder, A., & Ungan, P. Combined dynamics of EEG and evoked potentials. I. Studies of simultaneously recorded EEG-EPograms in the auditory pathway, reticular formation, and hippocampus of the cat brain during the waking stage. Biological Cybernetics, 1979, 34, 1-19.

Bellman, K. The conflict behavior of the lizard, Sceloporus occidentalis, and its implication for the organization of motor behavior. Unpublished doctoral dissertation, University of California, San Diego, La Jolla, Calif., 1979.

Bellman, R. E. Adaptive control processes: A guided tour. Princeton, N.J.: Princeton University Press, 1961.

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

Bertalanffy, L. von. General system theory. Harmondsworth, England: Penguin University Press, 1973.

Bizzi, E., Dev, P., Morasso, P., & Polit, A. Effects of load disturbance during centrally initiated movements. Journal of Neurophysiology, 1978, 41, 542-556.

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

Buckley, P., & Peat, F. D. A question of physics: Conversations in physics and biology. Toronto: University of Toronto, 1979.

Buddenbrock, E. von Der rhythmus der schreitbewegungen der stabheuschrecke Dyxippus. Biologisches Zentralblatt, 1921, 41, 41-48.

Bullock, T. H. Reassessment of neural connectivity and its specification. In H. M. Pinsker & W. D. Willis (Eds.), Information processing in the nervous system. New York: Raven press, 1980.

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

Catford, J. C. Fundamental problems in phonetics. Bloomington, Ind.: Indiana University Press, 1977.

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

Davis, C. M., & Beaton, R. D. Facilitation and adaptation of the human quadriceps stretch reflex produced by auditory stimulation. Journal of Comparative and Physiological Psychology, 1968, 66, 483-487.

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

Dennett, D. C. Intentional systems. The Journal of Philosophy, 1971, 118, 87-106.

Dietz, V., & Noth, J. Preinnervation and stretch responses of triceps bracchii in man falling with and without visual control. Brain Research, 1978, 142, 576-579.

Dubner, R., Sessle, B. J., & Storey, A. T. The neural basis of oral and facial function. New York: Plenum Press, 1978.

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

Easton, T. A. Patterned inhibition from single-eye muscle stretch in the cat. Experimental Neurology, 1972, 34, 497-510. (b)

Edelman, G. M. Group selection and phasic reentrant signalling: A theory of higher brain function. In G. M. Edelman & V. B. Mountcastle (Eds.), The mindful brain. Cambridge, Mass.: MIT Press, 1978.

Edelman, G. M., & Mountcastle, V. B. The mindful brain. Cambridge, Mass.: MIT Press, 1978.

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

Evarts, E. V., Bizzi, E., Burke, R. E., DeLong, M., & Thach, W. T. (Eds.) Central control of movement. Neurosciences Research Program Bulletin, 1971, 9(1).

Fel'dman, A. G. Functional tuning of the nervous system with control of movement or maintenance of a steady posture. II. Controllable parameters of the muscle. Biophysics, 1966, 11, 565-578. (a)

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

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

Fitch, H. L., Tuller, B., & Turvey, M. T. The Bernstein perspective: III. Tuning of coordinative structures with special reference to perception. In J. A. S. Kelso (Ed.), Human motor behavior: An introduction. Champaign, Ill.: Human Kinetics, 1982.

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

Forssberg, H., Grillner, S., & Rossignol, S. Phasic gain control of reflexes from the dorsum of the paw during spinal locomotion. Brain Research, 1977, 132, 121-139.

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

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

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

Freeman, W. J. Mass action in the nervous system. New York: Academic Press, 1975.

Fujimura, O., & Kakita, Y. Remarks on quantitative description of the lingual articulation. In B. Lindblom & S. Öhman (Eds.), Frontiers of speech communication research. London: Academic Press, 1979.

Fukson, O. I., Berkinblit, M. B., & Fel'dman, A. G. The spinal frog takes into account the scheme of its body during the wiping reflex. Science, 1980, 209, 1261-1263.

Fukuda, T. Studies on human dynamic postures from the viewpoint of postural reflexes. Acta Oto-Laryngologica, 1961, Suppl. 161.

Gallistel, C. R. The organization of action: A new synthesis. New York: Erlbaum, 1980.

Gazzaniga, M. S. Visuomotor integration in split brain monkeys with other cerebral lesions. Experimental Neurology, 1966, 16, 289-298.

Gazzaniga, M. S. Eye position and visual motor coordination. Neuropsychologia, 1969, 7, 379-382.

Gel'fand, I. M., Gurfinkel, V. S., Tsetlin, M. L., & Shik, M. L. Some problems in the analysis of movements. In I. M. Gel'fand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge: MIT Press, 1971.

Gel'fand, I. M., & Tsetlin, M. L. Mathematical modeling of mechanisms of the central nervous system. In I. M. Gel'fand, V. S. Gurfinkel, S. V. Fomin,

& M. L. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge, Mass.: MIT Press, 1971.

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

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

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

Gibson, J. J. An ecological approach to visual perception. Boston: Houghton-Mifflin, 1979.

Goldbeter, A. Models for oscillations and excitability in biochemical systems. In L. A. Segal (Ed.), Mathematical models in molecular and cellular biology. Cambridge: Cambridge University Press, 1980.

Goodwin, B. Biological stability. In C. H. Waddington (Ed.), Towards a theoretical biology. Chicago: Aldine, 1970.

Gottlieb, C., Agarwal, G., & Stark, L. Interactions between voluntary and postural mechanisms of the human motor system. Journal of Neurophysiology, 1970, 33, 365-381.

Gould, W. J. Effect of respiratory and postural mechanisms upon action of the vocal cords. Folia Phoniatrica, 1971, 23, 211-224.

Granit, R. The purposive brain. Cambridge, Mass.: MIT Press, 1977.

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

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

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

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

Gurel, O., & Rössler, O. E. Bifurcation theory and applications in scientific disciplines. Annals of the New York Academy of Science, 1979, 316.

Gurfinkel, V. S., Kots, Y. A., Paltsev, E. I., & Fel'dman, A. G. The compensation of respiratory disturbances of the erect posture of man as an example of the organization of interarticular interaction. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge, Mass.: MIT Press, 1971.

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

Haken, H. Synergetics: An introduction. Heidelberg: Springer Verlag, 1977.

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

Hamburger, V. The developmental history of the motor neuron. Neurosciences Research Program Bulletin, 1977, 15, 1-37.

Head, H. Aphasia and kindred disorders of speech. Cambridge: Cambridge University Press, 1926.

Heilman, K. M. Apraxia. In K. M. Heilman & E. Valenstein (Eds.), Clinical neuropsychology. New York: Oxford University Press, 1979.

Hellebrandt, F. A., Houtz, S. J., Partridge, M. J., & Walters, C. E. Tonic neck reflexes in exercises of stress in man. American Journal of Physical Medicine, 1956, 35, 144-159.

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

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

Iberall, A. S. On nature, man and society: A basis for scientific modeling. Annals of Biomedical Engineering, 1975, 3, 344-385.

Iberall, A. S. A field and circuit thermodynamics for integrative physiology: I. Introduction to general notions. American Journal of Physiology/Regulatory, Integrative, & Comparative Physiology, 1977 2, R171-R180.

Iberall, A. S. A field and circuit thermodynamics for integrative physiology: II. Power and communicational spectroscopy in biology. American Journal of Physiology/Regulatory, Integrative, & Comparative Physiology, 1978, 3, R3-R19. (a)

Iberall, A. S. A field and circuit thermodynamics for integrative physiology: III. Keeping the books - a general experimental method. American Journal of Physiology/Regulatory, Integrative, & Comparative Physiology, 1978, 3, R85-R97. (b)

Iberall, A. S., & Cardon, S. Z. Control in biological systems--A physical review. Annals of the New York Academy of Sciences, 1966, 117, 445-518.

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

Jordan, D. W., & Smith, P. Nonlinear ordinary differential equations. Oxford: Clarendon Press, 1977.

Kaiser, F. Limit cycle model for brain waves. Biological Cybernetics, 1977, 27, 155-163.

Kakita, Y., & Fujimura, O. A computational model of the tongue: A revised version. Journal of the Acoustical Society of America, 1977, 62, S15. (Abstract)

Katchalsky, A. K., Rowland, V., & Blumenthal, R. Dynamic patterns of brain cell assemblies. Neurosciences Research Program Bulletin, 1974, 12 (1).

Kawahara, T. Coupled Van der Pol oscillators--A model of excitatory and inhibitory neural interactions. Biological Cybernetics, 1980, 39, 37-43.

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

Kelso, J. A. S. Motor-sensory feedback formulations: Are we asking the right questions? The Behavioral and Brain Sciences, 1979, 2, 72-73.

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

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

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

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

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. Patterns of human interlimb coordination emerge from the properties of non-linear oscillators: Theory and data. Journal of Motor Behavior, 1981, 13, 226-261.

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

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

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

Kelso, J. A. S., & Tuller, B. Exploring the information support for speech. Haskins Laboratories Status Report on Speech Research, 1982, SR-69, 43-54.

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

Koschmeider, E. L. Instabilities in fluid dynamics. In H. Haken (Ed.), Synergetics: A workshop. New York: Springer-Verlag, 1977.

Kots, Ya. M. The organization of voluntary movement. New York: Plenum, 1977.

Kots, Ya. M., & Zhukov, V. I. Supraspinal control of the segmental centers of muscle antagonists in man--III "Tuning" of the spinal apparatus of reciprocal inhibition in the period of organization of voluntary movement. Biophysics, 1971, 16, 1129-1136.

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

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

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

Lee, D. N. On the functions of vision. In H. Pick & E. Saltzman (Eds.), Modes of perceiving. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Lee, D. N., & Lishman, J. R. Visual proprioceptive control of stance. Journal of Human Movement Studies, 1974, 1, 87-95.

Lee, W. Anticipatory control of postural and task muscles during rapid arm flexion. Journal of Motor Behavior, 1980, 12, 185-196.

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

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

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

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

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

Madeiros, J. Investigation of neuronal mechanisms underlying human locomotion: An EMG analysis. Unpublished doctoral dissertation, University of Iowa, 1978.

Melville-Jones, G., & Watt, D. G. D. Observations on the control of stepping and hopping movements in man. Journal of Physiology, 1971, 219, 709-727. (a)

Melville-Jones, G., & Watt, D. G. D. Muscular control of landing from unexpected falls in man. Journal of Physiology, 1971, 219, 729-737. (b)

Minorsky, N. Nonlinear oscillations. Princeton, N.J.: Van Nostrand, 1962.

Morowitz, H. J. Foundations of bioenergetics. New York: Academic Press, 1978.

Morowitz, H. J. Energy flow in biology. Woodbridge, Conn.: Oxbow Press, 1979.

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

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

Orlovskii, G. N. The effect of different descending systems on flexor and extensor activity during locomotion. Brain Research, 1972, 40, 359-371.

Pal'tsev, Y. I., & El'ner, A. M. Change in the functional state of the segmental apparatus of the spinal cord under the influence of sound stimuli and its role in voluntary movement. Biophysics, 1967, 12, 1219-1226.

Pattee, H. H. Physical problems of the origin of natural controls. In A. Locker (Ed.), Biogenesis, evolution, homeostasis. Heidelberg: Springer-Verlag, 1973, 41-49.

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

Pew, R. W. Human perceptual motor-performance. In B. H. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition. New York: Lawrence Erlbaum Associates, 1974.

Pinsker, H. M., & Willis, W. D., Jr. (Eds.) Information processing in the nervous system. New York: Raven Press, 1980.

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

Prigogine, I. From being to becoming. San Francisco: W. H. Freeman, 1980.

Prigogine, I., & Nicolis, G. Biological order, structure and instabilities. Quarterly Reviews of Biophysics, 1971, 4, 107-148.

Rack, P. M. H., & Westbury, D. R. The effects of length and stimulus rate on tension in the isolated cat soleus muscle. Journal of Physiology (London), 1969, 204, 443-460.

Rakerd, B., Verbrugge, R. V., & Shankweiler, D. P. Speaking rate, syllable stress, and vowel identity. Haskins Laboratories Status Report on Speech Research, 1980, SR-62, 149-159.

Rashevsky, N. Mathematical biophysics. New York: Dover Press, 1938.

Reed, E. Outline for a theory of action systems. Journal of Motor Behavior, in press.

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

Rossignol, S. Startle responses recorded in the leg of man. Electroencephalography and Clinical Neurophysiology, 1975, 39, 389-397.

Rossignol, S., & Melville-Jones, G. Audio-spinal influence in man studied by the H-reflex and its possible role in rhythmic movements synchronized to sound. Electroencephalography and Clinical Neurophysiology, 1976, 41, 83-92.

Sakitt, B. A spring model and equivalent neural network for arm posture control. Biological Cybernetics, 1980, 37, 227-234.

Saltzman, E. Levels of sensorimotor representation. Journal of Mathematical Psychology, 1979, 20, 92-163.

Saltzman, E. L. Interactions between volitional and automatic sensorimotor processes: Lateral stimulus-response compatibility. Unpublished doctoral dissertation, University of Minnesota, 1979.

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

Saunders, P. T. An introduction to catastrophe theory. Cambridge: Cambridge University Press, 1980.

Schmidt, R. A. A schema theory of discrete motor skill learning. Psychological Review, 1975, 82, 225-260.

Schmidt, R. A. On the theoretical status of time in motor program representations. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland, 1980.

Schmidt, R. A., & McGown, C. Terminal accuracy of unexpectedly loaded rapid movements: Evidence for a mass-spring mechanism in programming. Journal of Motor Behavior, 1980, 12, 149-161.

Schmitt, F. O. Introduction. In G. M. Edelman & V. B. Mountcastle (Eds.), The mindful brain. Cambridge, Mass.: MIT Press, 1978.

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

Shapiro, D. C., & Schmidt, R. A. The schema theory: Recent evidence and developmental implications. In J. A. S Kelso & J. Clark (Eds.), The development of movement control and coordination. New York: Wiley, 1982.

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

Shaw, R., & Turvey, M. T. Coalitions as models for ecosystems: A realist perspective on perceptual organization. In M. Kubovy & J. Pomerantz (Eds.), Perceptual organization. Hillsdale, N.J.: Erlbaum, 1981.

Sherrington, C. S. The integrative action of the nervous system. London: Constable, 1906.

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

Smith, W. R. Hypothalamic regulation of pituitary secretion of luteinizing hormone--II. Feedback control of gonadotropin secretion. Bulletin of Mathematical Biology, 1980, 42, 57-78.

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

Stein, P. S. G. Application of the mathematics of coupled oscillator systems to the analysis of the neural control of locomotion. Federation Proceedings, 1977, 36, 2056-2059.

Stein, R. B. What muscle variables does the central nervous system control? The Behavioral and Brain Sciences, 1982, 5, 585-541.

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

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

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

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

Thoden, U., Dichgans, J., & Savadis, T. Direction specific optokinetic modulation of monosynaptic hindlimb reflexes in cats. Experimental Brain Research, 1977, 30, 155-160.

Thom, R. Structural stability and morphogenesis (D. H. Fowler trans.). Reading, Mass.: W. A. Benjamin, 1975.

Thompson, D. A. W. On growth and form (2nd Ed.). Cambridge: University Press, 1917/1942.

Titze, I. R. Comments on the myoelastic-aerodynamic theory of phonation. Journal of Speech and Hearing Research, 1980, 23, 495-510.

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

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

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

Turvey, M. T. Clues from the organization of motor systems. In U. Bellugi & M. Studdert-Kennedy (Eds.), Signed and spoken language: Biological constraints on linguistic form. Weinheim: Verlag Chemie, 1980.

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

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

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

White, B. W., Saunders, F. A., Scadden, L., Bach-Y-Rita, P., & Collins, C. C. Seeing with the skin. Perception & Psychophysics, 1970, 7, 23-27.

Wilson, H. R., & Cowan, J. D. Excitatory and inhibitory interactions in localized populations of model neurons. Biophysical Journal, 1972, 12, 1-24.

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

Yates, F. E. Systems analysis of hormone action: Principles and strategies. In R. F. Goldberger (Ed.), Biological regulation and development, Vol. 3: Hormone action. New York: Plenum, in press.

Yates, F. E., & Iberall, A. S. Temporal and hierarchical organization in biosystems. In J. Urquart & F. E. Yates (Eds.), Temporal aspects of therapeutics. New York: Plenum, 1973, 17-34.

FOOTNOTES

[1]With due deference to the celebrated embryologist V. Hamburger (1977).

[2]It is well established that the basal plate, the motor part of the spinal cord, proliferates and differentiates long before the altar plate, or dorsal part, that receives sensory input. This observation has led some to speculate on the primacy of motor function, in a way that might provoke the cognitive neuroscientists: "The elemental force that embryos and fetuses can express freely in their spontaneous motility, sheltered as they are in the egg and uterus, has perhaps remained throughout evolution the biological mainspring of creative activity in animals and man and autonomy of action is also the mainspring of freedom" (Hamburger, 1977, p. 32)

[3]Emerging primarily from Iberall and colleagues' Homeokinetic Theory (e.g., Iberall, 1977; 1978; Soodak & Iberall, 1978; Yates, 1980) but drawing also on Prigogine and colleagues' Dissipative Structure Theory (e.g., Prigogine, 1980; Nicolis & Prigogine, 1977), Haken's Synergetics (Haken, 1977; 1978), Morowitz's Bioenergetics (Morowitz, 1978; 1979), and Rosen's Dynamical Systems' Theory (Rosen, 1970; 1978). A synthesis of these theories appears in Kugler, Kelso, and Turvey (1982).

[4]Physical science still pursues this strategy with some vigor in certain circles, although not without its skeptics. Thus, some have remarked that "elemental units"--as the least divisible parts--are not necessarily "fundamental units," and that indivisibility is no criterion for fundamentality (cf. Buckley & Peat, 1979).

[5]A good example is that of a gas, whose molecular kinetic energy can be averaged to provide a macrostate observable such as temperature.

[6]In the case of perception, for example, we find it hard to understand how extensive, physical variables (like decibels) give rise to intensive, psychological effects (like roaring jets and rock bands). As Shaw and Cutting (1980) point out, this is a "structure-creating" transfer function that maps continuous variation of linear variables onto discontinuous categorical changes that, by definition, are nonlinear. At least two solutions can be offered to this problem: One is to assume that the perceptual apparatus is creative in nature and gives meaning to meaningless sensations (much like a schema for movement rearranges the spatiotemporal orderings of muscles in a creative, generative way); another is to adjust the basis of measurement so that it is common to the perceiver (producer) and the perceived (that which is produced).

[7]The sentiment here follows that of the great Canadian ice skating champion, Toller Cranston, who in a television interview (NBC, January 31,

1982) remarked that he has always considered his work to be artistic "fundamentally as kinetic form." Of course the science of form continues to be a hotly pursued area of study (e.g., Gould, 1971; Rosen, 1978; Thompson, 1917/1942).

[8]We balk, of course, at the fairly common description (at least among some psychologists) of locomotion as stereotypical and low-level. Many of the examples we have given in this paper attest to the generativity and context sensitivity of actions, and locomotion is a prime example. We are still at the tip of the iceberg as far as understanding these attributes is concerned--in locomotion or any other "less stereotyped" activity.

[9]Sometimes number is sufficient to indicate degree of complexity and we take the modularity idea of brain design to be--in part--an effort to come to grips with the problem of dealing with individual neuronal elements. But, to put it mildly, number is only a small aspect of complexity. Lest we think otherwise, consider the following list of factors, all of which are part of the domain of neuroscience:

1) Aside from elementary particle physics, neuroscience deals with the molecular and ionic events in cells, aspects of which are the mechanisms of molecular excitability and ion selectivity. The latter involves understanding--among other things--the mechanisms of ionic pumps, release and binding of neurotransmitters, growth of neurons, the structure of membranes, and the conductance properties of membrane channels.

2) Neuroscience attempts to analyze membrane circuitry and the geometry of cell membranes (little is known about the detailed anatomy of the cell being recorded in physiological studies or the distribution and type of conductance channels in cell membranes; cf. Pinsker & Willis, 1980).

3) The response properties of cells have been the staple diet of neuroscience. These vary on many different dimensions including threshold, latency, firing rate, tonic vs. phasic, brisk vs. sluggish, receptive field, refractory period, filter properties, transfer functions, etc.

The list we have provided here refers only to events at the cellular level, but it is enough to illustrate our point; namely, that number of elements is only one--and perhaps not the major--dimension of complexity.

[10]We have made no attempt to provide all the details of Edelman's theory. We represent here only "the main predictions" (Edelman 1978, pp. 92-93) because of their striking parallels, evolved independently, with principles synthesized from the movement literature, and complex, multivariable systems in general. We should also stress that the list of movement principles presented in the table is far from complete (however see Sections 2 through 5), and that we view cooperative phenomena--of neurons, muscles, or whatever--in a much larger context (see Section 6).

[11]Roughly, degeneracy refers to the capability of different structures or elements to perform similar functions.

ON THE SPACE-TIME STRUCTURE OF HUMAN INTERLIMB COORDINATION*

J. A. Scott Kelso,+ Carol A. Putnam,++ and David Goodman+++

Abstract. In three experiments, using behavioral measures of movement outcome as well as movement trajectory information and resultant kinematic profiles, we show that there is a strong tendency for the limbs to be coordinated as a unitary structure even under conditions where the movements are of disparate difficulty. Environmental constraints (an obstacle placed in the path of one limb, but not in the other) are shown to modulate the space-time behavior of both limbs (Experiment 2). Our results obtain for symmetrical (Experiment 1) as well as asymmetrical movements that involve non-homologous muscle groups (Experiment 3). These findings suggest that in multijoint limb movements, the many degrees of freedom are organized to function temporarily as a single coherent unit that is uniquely specific to the task demands placed on it. For movements in general, and two-handed movements in particular, such units are revealed in a partitioning of the relevant force demands for each component (a force scaling characteristic) and a preservation of the internal "topology" of the action, as indexed by the relative timing among components. These features, as well as systematic deviations from perfect synchrony between the limbs, can be rationalized by a model that assumes the limbs behave qualitatively like nonlinear oscillators.

INTRODUCTION

Many of the actions that humans perform require the cooperation of the upper limbs, but generally speaking, little attention has been devoted to

---

*To appear in a special issue on motor control, Quarterly Journal of Experimental Psychology.

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

++Dalhousie University, Halifax, Nova Scotia.

+++Simon Fraser University, Burnaby, British Columbia.

Acknowledgment. This research was supported by NINCDS Grant NS-13617, BRS Grant RR-05596 and Contract No. N00014-83-C-0083 from the U.S. Office of Naval Research to Haskins Laboratories. We thank Betty Tuller for very helpful comments on an earlier version, Stephen Monsell and two anonymous reviewers for their thought-provoking reviews, and Elliot Saltzman for many useful discussions on the mysteries of movement and the nature of oscillators. We also thank the many individuals (too numerous to name) who have talked with us about these data at various colloquia in which they have been presented.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

seeking principles that might underlie human interlimb coordination. Although some interesting studies of bimanual tapping performance have appeared recently (e.g., Peters, 1981; Yamanishi, Kawato, & Suzuki, 1980), by far the greatest research effort has been directed toward understanding the mechanisms associated with single limb movements, most involving only one degree of freedom (e.g., Bizzi, Dev, Morasso, & Polit, 1978; Cooke, 1980; Fel'dman, 1966, 1980; Kelso, 1977; Kelso & Holt, 1980).

Of course, there is a long history of work on the coordination among the appendages of vertebrates and invertebrates, the results of which have been especially impressive (for review, see Delcomyn, 1980). As an instance, Wilson's research on insect locomotion revealed, in principle, how the many surface kinematic details of gait could be synthesized out of a tonically activated network of coupled oscillators (Wilson, 1966; see also Grillner, 1975; Stein, 1977). Even here however, the nature of coupling processes among limbs remains somewhat obscure, a situation that may be remedied when nonlinear oscillator theory is more fully developed and exploited (cf. Pavlidis, 1973; Winfree, 1980). Indeed, some preliminary steps have already been taken to apply this framework to an understanding of human rhythmical movement (Kelso, Holt, Rubin, & Kugler, 1981; Yamanishi et al., 1980).

Although the work on animal neuromotor systems is obviously important to gain a fuller understanding of biological coordination in complex systems possessing many degrees of freedom, it seems useful to proceed with investigations on the human front as well, in the hope that general principles may emerge. With this in mind, in 1979 we introduced a paradigm that we felt might have broad potential for exploring the processes underlying the control of both limbs when they work together to accomplish a task (Kelso, Southard, & Goodman, 1979a, 1979b). The question that we asked was a very simple one: How will subjects respond if required to produce movements of the upper limbs toward targets of widely disparate difficulty as quickly and accurately as possible? A formulation developed for reciprocal tapping tasks by Fitts (1954) relating movement duration, movement amplitude, and target precision demands allowed us to examine the issue experimentally. The equation relating these variables is:

$$MT = a + b \log_2 (2A/W)$$

where A is the amplitude of movement
W corresponds to target width
a and b are constants, and
MT is movement time

For limbs operating singly, the obvious prediction from the above relationship is that movement time depends on the ratio of movement amplitude to movement precision. But now consider a situation in which one limb, say the left, moves a short distance to a large target (termed easy) while the other moves a longer distance to a small target (termed hard). For the single limb case, movement time in the easy condition, according to Fitts' Law, will obviously be much shorter than in the hard condition. However, when the two conditions are combined, Kelso et al. (1979a, 1979b) did not find that the limb producing a short movement to an easy target arrived earlier than its more difficult counterpart as one might expect. Instead, there was a strong tendency for both movements to be initiated and terminated synchronously. Indeed, an examination of the movement times indicated that the hand moving to the difficult target moved more rapidly in the combined, easy-hard condition than its single limb control, while the easy hand obviously slowed down--as if the limbs were adopting a common temporal metric.

It is important to point out that the limb moving to the easy target did not appear to "hover" over the target or "wait" for its difficult counterpart, but rather moved at a quite different speed. High-speed cinematography (200 frames/sec) and consequent examination of horizontal displacement, velocity, and acceleration patterns over time revealed that the limbs under easy-difficult target conditions reached peak velocity and peak acceleration at practically the same time during movements. Thus, although different spatial demands for the two limbs affected the magnitude of forces produced by each limb, the absolute timing and the segmental durations of movement components, that is, the timing relations between the two limbs remained quite constant.

The idea that motor coordination involves a reduction of the degrees of freedom of the sensorimotor system, not into prefabricated sets of reflexes, but into functional groupings of muscles constrained to act as a single unit (termed functional synergies [e.g., Gelfand, Gurfinkel, Tsetlin, & Shik, 1971; Saltzman, 1979] or coordinative structures [e.g., Fowler, 1977; Turvey, Shaw, & Mace, 1978]) stems originally from Bernstein (1967) and has undergone theoretical extension by Greene (1972), Boylls (1975), Turvey (1977) and others. To paraphrase Boylls (1975), functional synergies are collectives of muscles, all of which share a common pool of afferent and/or efferent information that are deployed as a unit in a motor task. In spite of powerful logical arguments that they are the significant units of action, it is only recently that rigorous analysis of muscle-joint collectives has taken place (cf. Kelso, 1981, for recent review of their existence in activities ranging from posture and locomotion to speech and handwriting).

The Kelso et al. (1979a, 1979b) experiments reveal what appears to be the chief signature of a functional synergy, namely that when a group of muscles cooperate as a single, coherent structure to accomplish a task, the internal timing relations among muscles and kinematic components are preserved invariantly over changes in the magnitude of activity in individual components. However, it is fair to say that the kinematic evidence on which this claim is based is rather sparse. In the early experiments (Kelso et al., 1979a, 1979b) we were restricted by limitations imposed by high speed cinematography and tedious frame-by-frame analysis. In fact, only the kinematics on the horizontal plane were examined over a series of six trials on a single subject. One of the goals of the present experiments was to supplement this very preliminary evidence with a much more detailed analysis of the movement trajectories of two limbs and their kinematic behavior on both horizontal and vertical planes. The first experiment reported here is a 'behavioral replication' of our earlier work, but used a pulsed light emitting diode (LED) technique to capture the space-time trajectories of the limbs. A second experiment explored more directly the influence of environmental constraints on the dynamical behavior of the hypothesized functional unit. If indeed the action system solves the two-handed task by controlling the limbs as a single structure, then the introduction of an obstacle that one limb must "jump over" to reach the target, may have (at least initially) concomitant modulatory effects on the other unconstrained limb.1 The obstacle in this case can be interpreted as placing a contextual constraint on the degrees of freedom of the unit rather than the individual limb.

All our experiments up to now have examined symmetrical movements of the upper limbs primarily involving extension of the forearm-wrist-hand linkages away from the body midline (Kelso et al., 1979a, Experiment 1), flexion toward the body midline (Experiment 2) or forward reaching movements in the sagittal plane (Experiment 3). The symmetry constraint is a powerful one in human movement, manifested, for example, in the so-called "mirror movements" exhibited by small children and certain brain-damaged populations (cf. Woods & Teuber, 1978). It is also omnipresent in the two-handed signs of American Sign Language. According to Klima and Bellugi (1979), "The symmetry constraint specifies that in a two-handed sign, if both hands move and are active, they must perform roughly the same motor acts" (p. 64). It would seem an important extension of the work on symmetrical limb movements to examine as well the coordination of asymmetrical movements that involve non-homologous muscle groups. In Experiment 3, we show that they too exhibit a space-time structure similar to that observed for symmetrical movements.

EXPERIMENT 1

Method

Subjects. The subjects were seven right-handed unpaid volunteers ranging in age between 18 and 25 years.

Apparatus. We have described the apparatus in detail in previous papers (Kelso et al., 1979a). It consists of a Plexiglas base mounted on a standard table with two home keys and two movable target keys. The home keys are centered in the base, 4.5 cm apart. In Experiment 1, two combinations of target size by target distance were used. The easy target was 7.2 cm wide and was positioned 6 cm from its corresponding home key. The hard target was 3.6 cm wide and was positioned 24 cm from its corresponding home key. A single target was used in one-handed conditions and two targets were used in the two-handed conditions. Thus, four different two-handed conditions were possible: a) two-handed easy, b) two-handed hard, c) two-handed mixed, hard target on right, easy on left, and d) two-handed mixed, hard target on left, easy on right. A red LED served as the warning light and the sound from a Minisonalert provided the stimulus to move. The onsets of warning light and stimulus tone were controlled by a Digital Equipment Corporation PDP 8/A computer that also collected initiation times, movement times, and total response times. The targets were painted white and were perfectly visible even though the experiment took place in a dimly lit room in order to facilitate the collection of photographic data on movement trajectories.

LEDs were firmly attached to the dorsal side of the index fingertip of each hand. The LEDs were set to pulse synchronously at a calibrated frequency of 200 Hz. In addition, two LEDs were attached to the target apparatus a known distance apart and within the field of view of the camera in order to provide a linear scale and horizontal reference line. A 35 mm Yashika camera, fitted with a Vivitar 50 mm lens (F stop 2.8) was positioned 2.0 m from the target apparatus so that its optical axis was perpendicular to the plane containing the midpoints of the starting buttons and the targets. The camera was loaded with Kodachrome color slide film (tungsten ASA rating 160). To film each trial, the camera shutter (set on bulb stop) was opened just prior to the start of each movement and was closed immediately after the targets were contacted. As a result, all LED flashes for the duration of any one trial were exposed on a single frame.

Task. The subject's task was identical to the one used in our previous studies of interlimb coordination (Kelso et al., 1979a, 1979b). Instructions to subjects were to move their index fingers from the home keys to the target keys as fast and as accurately as possible after receiving a stimulus to move. There were no instructions to move simultaneously in two-handed conditions. The movements themselves primarily involved extension of the forearm-wrist-hand linkage in the lateral plane. For one-handed conditions, the subject depressed the left home key with the left index finger, or the right home key with the right index finger, and, on receiving the stimulus to move proceeded to the designated target, touching it only with the index finger. For two-handed conditions, the subject depressed both home keys with the index fingers and proceeded to hit the respective targets following the onset of the auditory stimulus.

Procedure. As in our previous two-handed studies, eight experimental conditions were used that varied depending on whether a single limb or both limbs were involved or whether the movement was easy or hard. All subjects performed 20 trials preceded by 5 practice trials in each of the eight conditions. The last four trials of each condition were photographed using the procedures outlined above. Each stimulus was preceded by a 1 - 3 sec variable foreperiod; there was an intertrial interval of 5 sec. A 3 min rest period was given between each condition.

A within-subject design was used with all seven subjects performing in all experimental conditions, whose order was randomized. From the 20 trials in each condition, mean initiation time, movement time, and total response time were computed for each hand. Individual trials initiated prior to or within 30 msec of the stimulus to move were considered anticipations and excluded from the analysis. Similarly, trials with an initiation time greater than 800 msec, or trials in which a target was missed, were also excluded. There were four one-handed and four two-handed conditions, making a total of 12 separate means for each subject and each dependent variable.

For the kinematic analysis, each film frame was projected perpendicularly on an opaque screen of a Graf/Pen sonic digitizer. The X and Y coordinates were recorded from the image of the LEDs, each representing the location of a fingertip at the end of successive 5 msec intervals. Each XY coordinate was scaled to the actual displacement and stored on tape. The digitized data were smoothed by fitting cubic spline functions to the horizontal and vertical displacement-time data for each hand. An International Mathematical and Statistical Libraries subroutine called ICSSCU was used to perform data smoothing. Finally, the smoothed displacement-time data functions were mathematically differentiated every 5 msec to arrive at horizontal and vertical velocity-time and acceleration-time functions.

TOTAL RESPONSE TIME	MOVEMENT TIME	INITIATION TIME	LEFT TARGET	HOME KEYS	RIGHT TARGET	INITIATION TIME	MOVEMENT TIME	TOTAL RESPONSE TIME
					1	223	93	316
317	91	226	2					
					3	245	159	404
396	160	236	4					
321	96	225	6		5	219	101	320
421	169	252	8		7	249	176	425
406	158	248	10		9	247	138	385
374	131	243	12		11	236	158	394

Figure 1. Mean initiation time, movement time, and total response time (in msec) for single and two-handed movements directed away from the body midline. (For actual dimensions of targets and their distances from the home keys, refer to text.)

Results and Discussion

Two separate aspects of the data are addressed below. The first involves an analysis of the behavioral data and speaks to the issue of whether or not subjects initiate and terminate movements simultaneously, especially under conditions in which the task demands are quite different. The second aspect concerns the kinematic analysis, which allows us to examine the space-time trajectories of the movements themselves.

Analysis of the Behavioral Data

The mean initiation times, movement times, and total response time are shown for each condition in Figure 1. Pre-planned contrasts using Dunn's procedure (Kirk, 1968, p. 79) were used to assess the contrasts of interest. This procedure consists of splitting up the alpha level among a set of planned comparisons and does not require a prior significant overall F-ratio. The mean square error was computed for all dependent variables and then, depending on the number of means (in this case 12), the number of desired comparisons (in this case 6) and the degrees of freedom for experimental error (in this case 77), a d-value was calculated that must be exceeded by a given difference between means to be significant.

a) Initiation time analysis. For initiation time, MSe was 318.8, d = 26 msec, $p < .05$. No significant overall hand differences (left versus right, mean differences < 5 msec, $p > .05$) were found. In two-handed conditions of equal difficulty, the hands initiated the movements at approximately the same time, as revealed by the non-significance of all comparisons (all ps > .05). The average time difference in initiating the movements of the separate hands in the two-hand easy trials (5 versus 6) was 6 msec, while in the two-hand difficult trials (7 versus 8) it was only 3 msec. In the conditions in which each hand was performing tasks of varying difficulty, the easy hand was initiated 3 msec earlier on the average than the difficult one (9 and 12 versus 10 and 11), a finding that replicates our earlier work (Kelso et al., 1979a, 1979b).

It is conceivable, however, that these small differences between the hands are in part artifactual because they reflect algebraic differences that may have cancelled each other out when the mean was calculated over 20 trials. In a further analysis of the initiation time data, absolute time differences between each hand were tabulated and placed into time bins. A survey of Table 1 indicates that the hands were initiated within 20 msec of each other on over 93% of the valid individual trials, even in conditions of mixed difficulty.

Further evidence for the cooperation of the limbs is provided by the correlations between the two hands computed for each individual subject and presented in Table 2. These correlations were extremely high with only one out of a possible 28 below r = .97. The similarity in initiation behavior of the two limbs that we have found has also been obtained by others. Peters (1981), for example, has shown in a high speed cinematographic analysis of bimanual tapping that the hands are initiated near simultaneously, a result that he interprets as evidence in favor of a common activation source for the two hands.

Table 1

Number of individual trials (and percent of total trials) in which the absolute time differences between hands was less than the tabled value (in msec).

CONDITION	ABSOLUTE DIFFERENCE BETWEEN HANDS <10	<20	<30	<40	<50	>50	PERCENT INVALID TRIALS
INITIATION TIME							
Easy-Easy	101(85)	117(98)	119(100)			0(0)	6
Hard-Hard	97(79)	115(94)	121(98)	122(99)	123(100)	0(0)	8
Easy-Hard	96(77)	122(98)	123(99)	123(99)	124(100)	0(0)	11
Hard-Easy	89(71)	111(88)	123(98)	125(99)	126(100)	0(0)	10
MOVEMENT TIME							
Easy-Easy	77(63)	110(89)	120(98)	122(99)	123(100)	0	6
Hard-Hard	58(49)	87(73)	103(87)	112(94)	116(98)	3(3)	8
Easy-Hard	34(28)	63(51)	88(72)	109(89)	119(97)	5(4)	11
Hard-Easy	33(27)	59(48)	86(69)	108(87)	116(94)	9(7)	10
TOTAL RESPONSE TIME							
Easy-Easy	99(81)	118(96)	123(100)	123(100)	123(100)	0	6
Hard-Hard	64(54)	94(79)	119(92)	114(96)	116(98)	3(3)	8
Easy-Hard	38(37)	77(62)	89(72)	110(89)	117(94)	7(6)	11
Hard-Easy	42(33)	80(64)	106(84)	115(91)	118(94)	8(6)	10

Table 2

Correlations of left versus right hand for each subject over the valid trials in each of the four two-handed conditions.

	EASY-EASY			HARD-HARD			EASY-HARD			HARD-EASY		
SUBJECT	IT[a]	MT[b]	TRT[c]	IT	MT	TRT	IT	MT	TRT	IT	MT	TRT
S_1	.99	.84	.98	.99	.78	.98	.97	.67	.95	.97	.92	.98
S_2	.99	.50	.98	.99	.78	.95	.97	.83	.95	.98	.45	.87
S_3	.99	.98	.98	.98	.94	.98	.99	.82	.97	.99	.59	.98
S_4	.97	.56	.74	.99	.67	.97	.99	.72	.98	.99	-.11	.67
S_5	.99	.77	.99	.99	.96	.98	.92	.28	.89	.99	.76	.82
S_6	.99	.88	.99	.99	.65	.93	.99	.49	.97	.99	.75	.97
S_7	.99	.75	.99	.99	.75	.93	.98	.51	.96	.97	.76	.94

a = Initiation time in msec.
b = Movement time in msec.
c = Total response time in msec.

b) Movement time analysis. The pre-planned contrasts of the movement time data produced results consistent with our previous findings (Kelso et al., 1979a, 1979b). MSe was 227.2, d = 22 msec, $p < .05$; d = 26.6 msec, $p < .01$ for single mean comparisons. One-handed easy movement times (1 and 2) were much faster than their difficult counterparts (3 and 4) as Fitts' formulation (Fitts, 1954) predicts (mean difference = 67.5 msec, $p < .01$). This effect was also evident when examining two-handed movements of the same difficulty (5 and 6 vs. 7 and 8, mean difference = 74 msec, $p < .01$). As expected, the movement times of each hand when performing two-handed tasks of similar difficulty were not significantly different (mean difference for the easy-easy task = 5 msec, $p > .05$, and for the hard-hard task = 7 msec, $p > .05$). Moreover, the mean difference of 23 msec between the two hands when performing tasks of differing difficulty was also nonsignificant ($p > .05$), although there is a clear tendency for the easy hand to reach its target first.

Some insight into the interpretation of the null effect under mixed conditions is obtained by noting that the movement time of the hand performing the easy task of the mixed difficulty task (9 and 12) is considerably elevated over the easy-easy counterpart (5 and 6) (mean difference = 36 msec, $p < .01$). In contrast, when examining the hand performing the hard task in the same conditions, the movement times, while not significantly different (mean difference = 14.5 msec, $p > .05$), are reduced compared to their hard-hard counterpart movements. As in our previous experiments, these data suggest that it is not only the easy hand that slows to the level of its more difficult counterpart, but rather, both hands adjust, admittedly to varying degrees, as if the motor system were adopting a common time scaling for two-handed movements.

As with the initiation times, the absolute difference between movement times for each hand in the paired movements was tabulated (see Table 1). The proportion of trials in which movements were made within 10 msec of each other was somewhat lower for the condition of mixed difficulty (27%) than for the conditions of equal difficulty (62% for easy-easy; 49% for the hard-hard). However, even in the conditions of mixed difficulty, approximately 70% of the movements were made within 30 msec of each other. The movement time correlations for each hand in the two-handed condition are presented in Table 2. Although not as high as the correlations for initiation times, 20 of the 28 individual correlations were significant ($p < .05$), with no significant differences across the four conditions.

c) Total response time. The outcome of the total response time analysis was very similar to that of the movement time data. All significant effects in the movement time analysis were also significant in the total response time analysis. For the combined condition, the mean time difference between easy and difficult targets was 20 msec, which mirrors our earlier data (Kelso et al., 1979a) and is not significant at the .05 level (MSe = 628.0, d = 36msec, $p < .05$). Coordinating the movements of both hands in the combined condition eliminated 80% of the difference in total response time found between the easy-easy and hard-hard conditions.

With respect to the tabulation of the absolute time differences of each hand (see Table 1), since the initiation times for each hand were so similar, the total response time effects were almost identical to those of the movement times. As expected, the individual subject correlations for response times were high (see Table 2), and all were significant at the .05 level.

Kinematic Analysis

The last four trials of each subject in each condition were filmed as described previously. We have chosen to illustrate the results of 2 subjects, although we used mean data (over all 7 subjects) for the analysis of kinematic features. The trajectories for subjects MB and PH are shown in Figures 2 and 3, respectively. These trajectories, with minor exceptions, were typical of all subjects. Although we have made no attempt to quantify the shape of the trajectories themselves, it is clear that the patterns for each limb are extremely reproducible from trial to trial. Moreover, the trajectories between limbs are very similar under conditions in which the target difficulty is identical for each limb. Even in the combined easy-hard condition, although the paths of the two trajectories are obviously different, their form looks remarkably alike as if one were an expanded (or contracted) version of the other. A further notable feature of all the trajectories is that they are smooth and continuous (as judged by the relative spacing between dots) and exhibit no evidence of any "feedback" corrections, an observation that fits the rapid movement times in this experiment.

Knowing the time course of the trajectories, the horizontal and vertical components of the displacement, velocity, and acceleration over time were derived as described in the Methods Section. These are depicted in Figures 4 and 5, again for the same two subjects (see figure legends for plotting convention). In both conditions in which the left and right hands perform the same task, it is apparent that the kinematics are quite similar. Of greater interest, however, are the conditions of mixed difficulty. Note in Figures 4 and 5 that there is remarkable similarity in each pair of displacement curves, as if one curve is scaled to the other. There are a number of other kinematic parameters that remain relatively invariant between the limbs. One is the time of peak velocity in the horizontal direction, i.e., the time at which the movement changes from positive to negative acceleration (the same temporal locus as the zero crossing of the acceleration-time curve), which is almost coincidental for both hands in each separate condition. Thus the limbs start their braking action at approximately the same time (see also Lestienne, 1979).

A second kinematic descriptor is the point of maximum vertical displacement that corresponds to the transition between the ascent and descent of the movement and the time of zero vertical velocity. Note in Figures 4 and 5 that once again this point in time is also virtually coincident for both hands. Two further kinematic descriptors of interest are the times of peak vertical velocity in the positive (upward) and negative (downward) directions. Once again, we see a relatively tight correspondence in timing across both limbs.

The mean times-to-peak of the four kinematic variables discussed above are presented in Table 3. Note that in the single hand conditions the times to peak of these parameters are quite disparate from each other. As expected, the difference is also apparent in two-handed movements of equal difficulty. However, when the hands move to different targets, the time differences between the two hands are reduced considerably. For instance, the time to

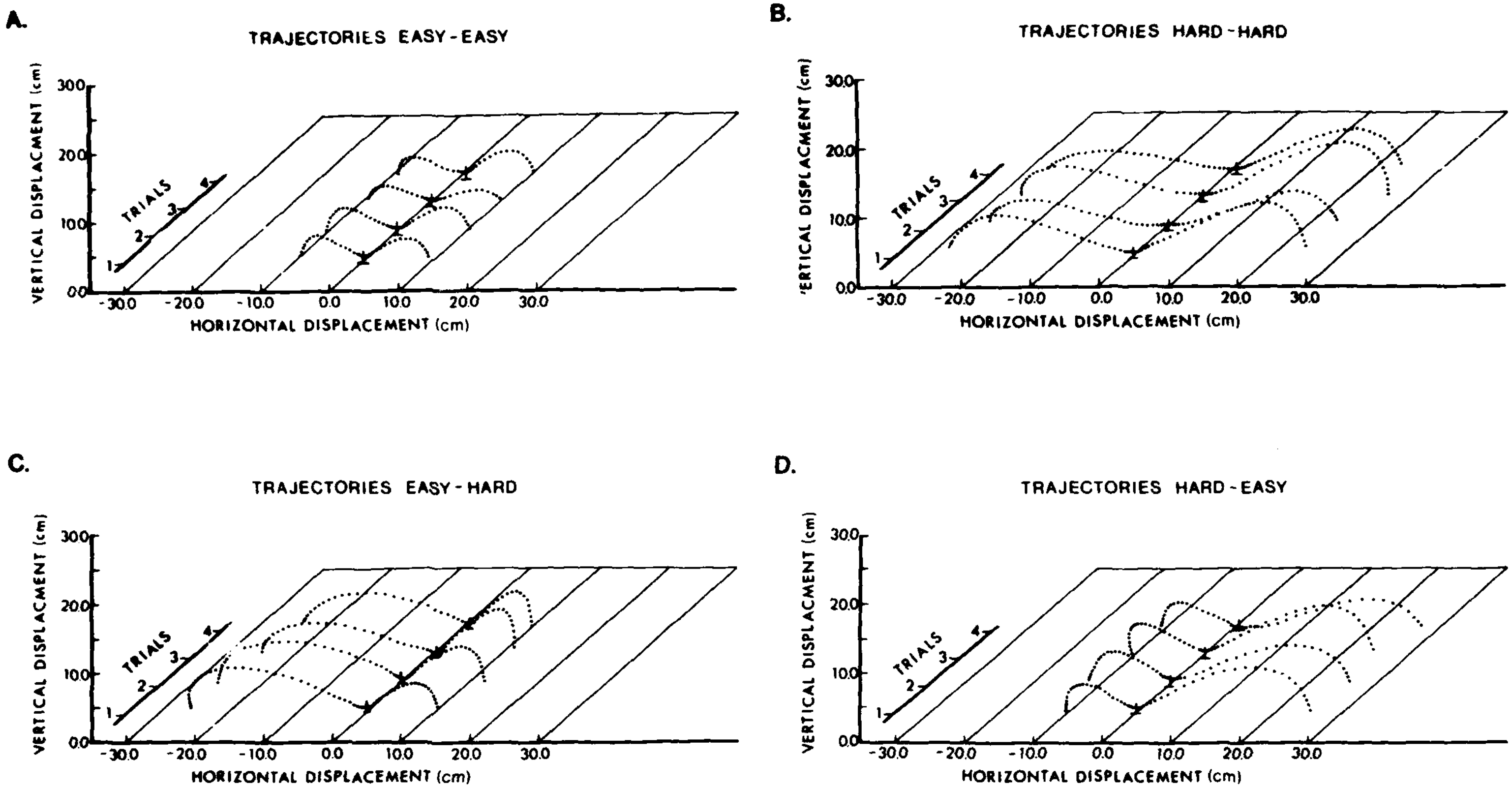


Figure 2. Movement trajectories for subject M.B. plotted on the horizontal and vertical displacement plane for the four two-handed conditions. Dots refer to light-emitting-diode pulses sampled at 200 Hz.

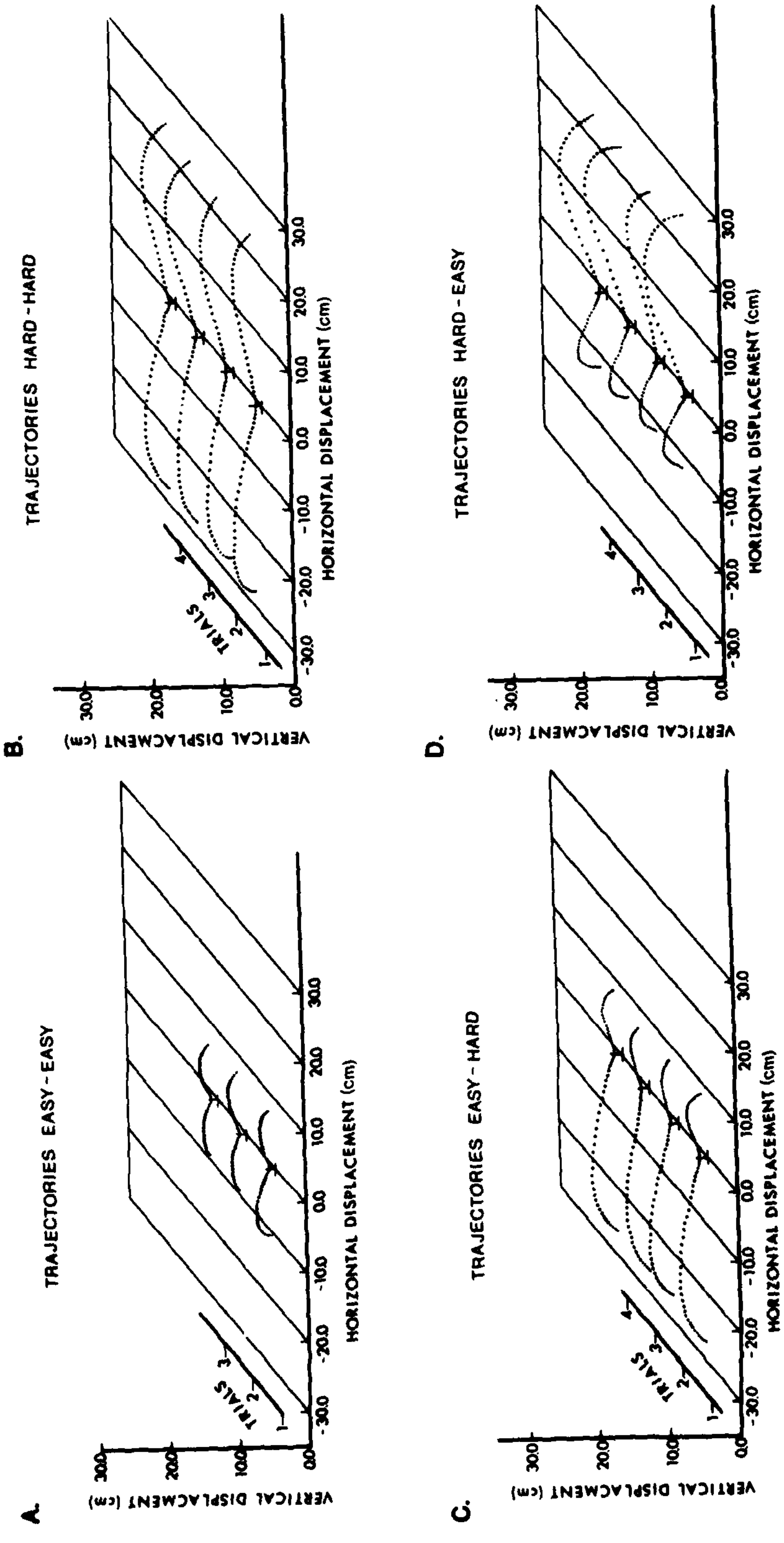


Figure 3. Movement trajectories for subject P.H. plotted on the horizontal and vertical displacement plane for the four two-handed conditions. Dots refer to light-emitting-diode pulses sampled at 200 Hz. One trial was lost in the *easy-easy* condition due to poor film processing.

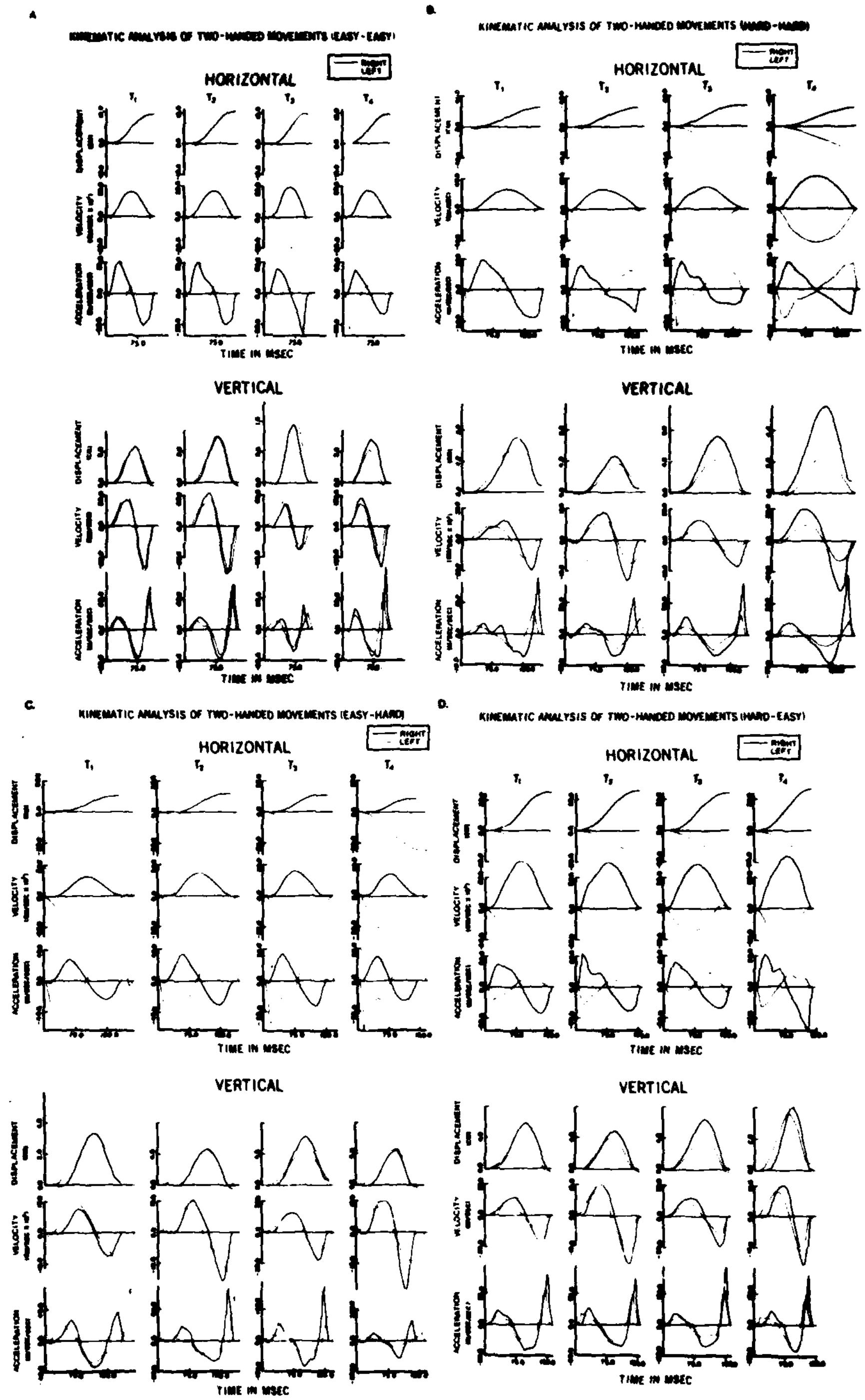


Figure 4. The patterns of horizontal and vertical displacement, velocity and acceleration over time for the two-handed movement trajectories of subject M.B. (for derivation procedures refer to text). The last four trials in each condition were filmed and are displayed here.

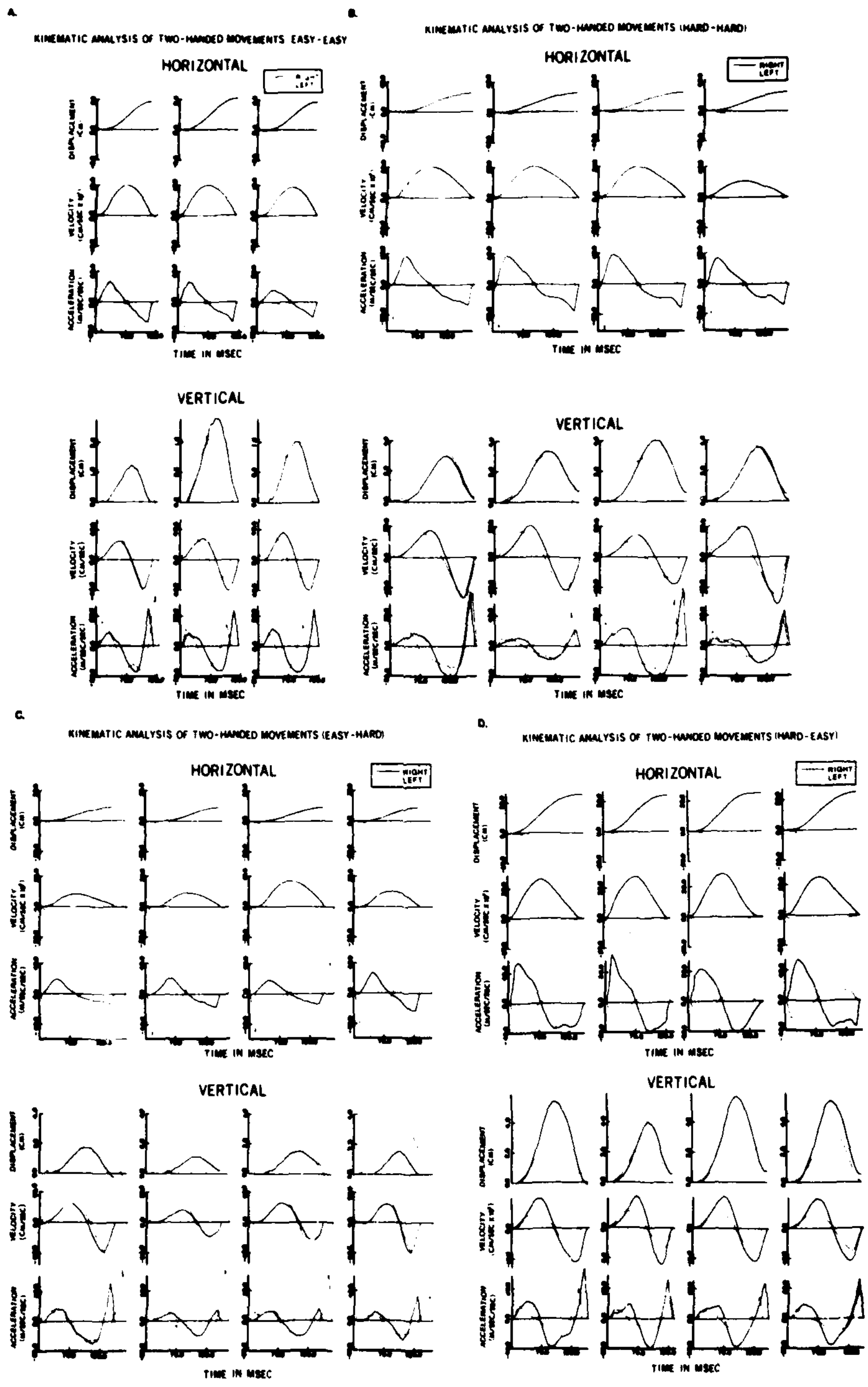


Figure 5. The patterns of horizontal and vertical displacement, velocity and acceleration over time for the two-handed movement trajectories of subject P.H. (for derivation procedures refer to text). The last four trials in each condition were filmed and are displayed here.

peak vertical velocity difference is reduced from 21.0 msec in the single hand condition to 8 msec in the two-handed condition. Like the behavioral data, the two limbs exhibit a kind of "mutual synchronization" under mixed difficulty conditions, with the easy hand slowing down to a much greater degree than the hard hand speeding up.

Table 3

Mean times to peak (in msec) of kinematic descriptors[a]

MOVEMENT CONDITIONS	KINEMATIC DESCRIPTOR: VERTICAL DISPLACEMENT	HORIZONTAL VELOCITY	VERTICAL VELOCITY 1	VERTICAL VELOCITY 2
Single-Easy	47	33	21	72
Single-Hard	90	64	42	137
Two-Hand Same				
Easy	43	39	19	68
Easy	53	40	24	74
Hard	90	62	43	139
Hard	91	62	43	139
Two-Hand Mixed				
Easy	68	42	30	106
Hard	82	57	38	129

[a]Refer to text for details.

EXPERIMENT 2

One obvious test of the claim that the limbs, under certain conditions, are coordinated and controlled as a single unitary structure is to manipulate a part of the structure to determine if the behavior of the unit or only the part is modulated. We have examined this idea in other work on rhythmical hand movements (Kelso et al., 1981) by perturbing one limb mechanically (a torque that changed the direction of motion) and then observing if the phase relations of the limbs were affected by the perturbation. Quite remarkably, both limbs returned to synchrony almost immediately. The tack in the present experiment was a little different. Rather than introducing a perturbation, we placed an obstacle in the path of one limb while requiring both limbs to move to their respective targets. Although obstacle height was somewhat arbitrarily chosen (about the height of a beer bottle), and was the same for all subjects, we predicted nevertheless that the obstacle would exert a mutual influence on both limbs, that is, the unit as a whole.

Methods

Subjects. Seven subjects, all of whom had participated in the previous experiment, served as subjects in Experiment 2.

Apparatus. The apparatus used in the first experiment was also employed in this experiment with the following two modifications. First, only one target size by target distance was utilized (3.6 cm target, 24 cm from the home keys). Second, a barrier (18 cm high by 7.5 cm wide) was placed mid-way between the home key and the target key. (We will refer to this as the 'hurdle' condition.) As in the first experiment, LEDs were attached to the fingers in order to provide trajectory information.

Task. Instructions to the subject were to move from the home key to the target key as quickly and as accurately as possible, without touching the barrier, following the onset of a stimulus to move. Again, nothing was said to the subject regarding simultaneity in the dual-limb case. There were two conditions: a) a single-hand condition over the barrier, and b) a two-hand condition, with the barrier erected only on one side.

Procedure. All subjects performed both of the conditions in a random order. Four of the subjects had the hurdle on the left side, while the other three had the hurdle on the right side. Twenty trials, which were not preceded by any practice trials, were performed in each of the conditions. The first two trials, two of the middle trials (trials 8 and 9), and the final two trials were filmed in the two-handed condition. For each trial there was a ready light followed by a 1 to 3 sec variable foreperiod, and the stimulus to move. Each trial was separated by a 5 sec inter-trial interval.

Results and Discussion

As in Experiment 1, first we present the behavioral findings followed by the kinematic data. Mean initiation time, movement time, and total response time are shown for the four conditions in Figure 6. In two-handed movements, the limb moving over the hurdle was initiated slightly before the contralateral limb (mean difference = 9.5 msec). This early departure, however, was offset by a longer movement time for the limb traversing the hurdle (mean difference = 54 msec, $p < .01$), which was reflected in a significant total response time difference of 45 msec, $p < .01$.

Thus, while we find that the imposition of a hurdle in the movement trajectory of the limbs disrupts the simultaneity effects we had witnessed in Experiment 1 and in our previous studies (Kelso et al., 1979a, 1979b), it is also apparent that there is a compensatory effect on the non-hurdle limb. This observation comes about by comparing times in the hurdle condition to those in the non-hurdle conditions of Experiment 1. For instance, the movement times and total response times of the non-hurdle hand in the hurdle conditions were elevated 38.5 and 57 msec, respectively, over the counterpart conditions of Experiment 1 (7 and 8 in Figure 1).

Further observation of each subject's data (see Table 4) reveals a large disparity between the timing relationships of the limbs across the different subjects. The mean difference in total response times for the hurdle versus

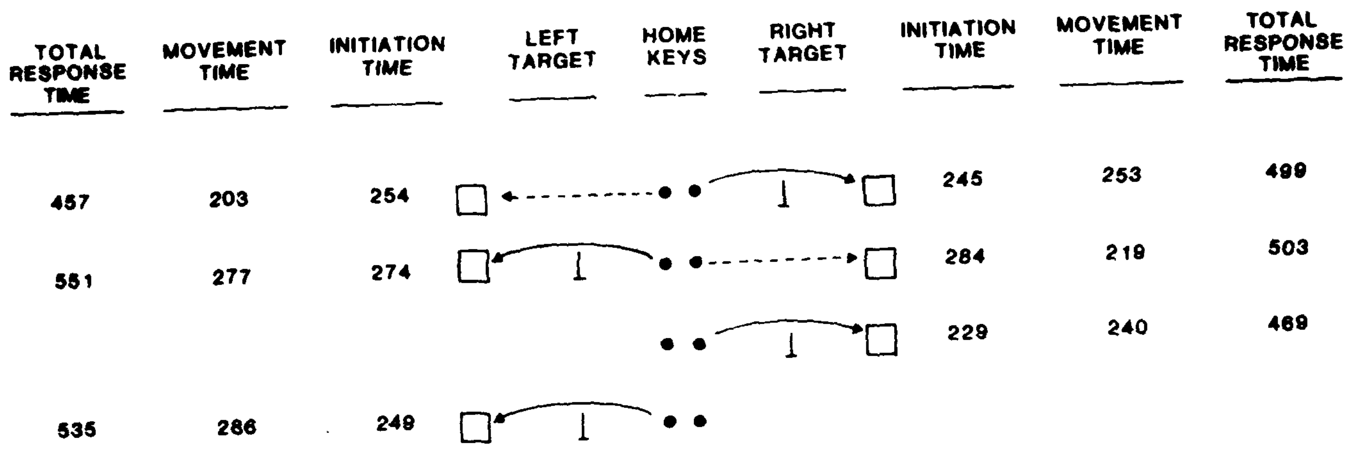


Figure 6. Mean initiation time, movement time, and total response time (in msec) for experiment in which one limb must traverse an obstacle (solid line) while the other (dashed line) is left free to vary (refer to text for distances, target dimensions, and obstacle height).

the non-hurdle limb ranged from a low of 10 msec (subject MB) to a high of 99 msec (subject PH). This suggests that at least some subjects (e.g., TH, GH, and especially PH) may have adopted a rather different strategy from the one adopted by subjects in our earlier studies (Experiment 1 and Kelso et al., 1979a, 1979b). As indicated in Table 4, initiation times for PH show a sizable temporal disparity between the hands, with the hurdle hand being initiated some 19 msec before its non-hurdle counterpart. Rather than initiating the movements simultaneously, subject PH appears to perform the two movements in a 1-2 manner rather than as a unified pair.2 This may be one of the reasons for the differences observed among subjects. In addition, the movement times of subjects TH, GH, and PH are sufficiently different between the hurdle and non-hurdle limbs to suggest that the parameters for the two limbs may be specified separately. The movements required by the task may have been perceived as sufficiently different from each other that the powerful symmetry constraint between the limbs no longer holds, hence the two hands may not participate in the same coordinative structure.

Table 4

Individual mean data in msec for hurdle and non-hurdle trials.

	Initiation Time		Movement Time		Total Response Time	
Hurdle on Right	Non-hurdle	Hurdle	Non-hurdle	Hurdle	Non-hurdle	Hurdle
SP	280	274	204	233	484	507
RH	287	277	258	292	545	570
TH	215	203	188	245	402	447
GH	233	229	165	244	398	473
Mean	254	246	204	254	457	499
Hurdle on Left						
MB	249	239	240	261	489	499
SB	332	331	255	290	587	621
PH	272	253	163	280	434	533
Mean	284	274	219	277	503	551

On the other hand, other subjects do appear to coordinate the limbs as a single unit. The movement time and total response time differences between the limbs are much smaller for subjects SP, RH, MB, and SB (means = 32 msec and 23 msec, respectively) than for PH, GH, and TH (means = 81 msec and 73 msec, respectively). Although the trajectories of both limbs are modified by the hurdle, the effects are much stronger for the former grouping of subjects than the latter. To illustrate, the limb trajectories and consequent kinematics are presented for subjects PH and MB in Figures 7 and 8. There are dramatic differences between the two displays. For PH, shown in Figure 7, the non-hurdle limb reaches a maximum vertical displacement of less than one-half of the limb traversing the hurdle. Even so, and especially on the first trial, the vertical displacement for the non-hurdle limb is amplified more

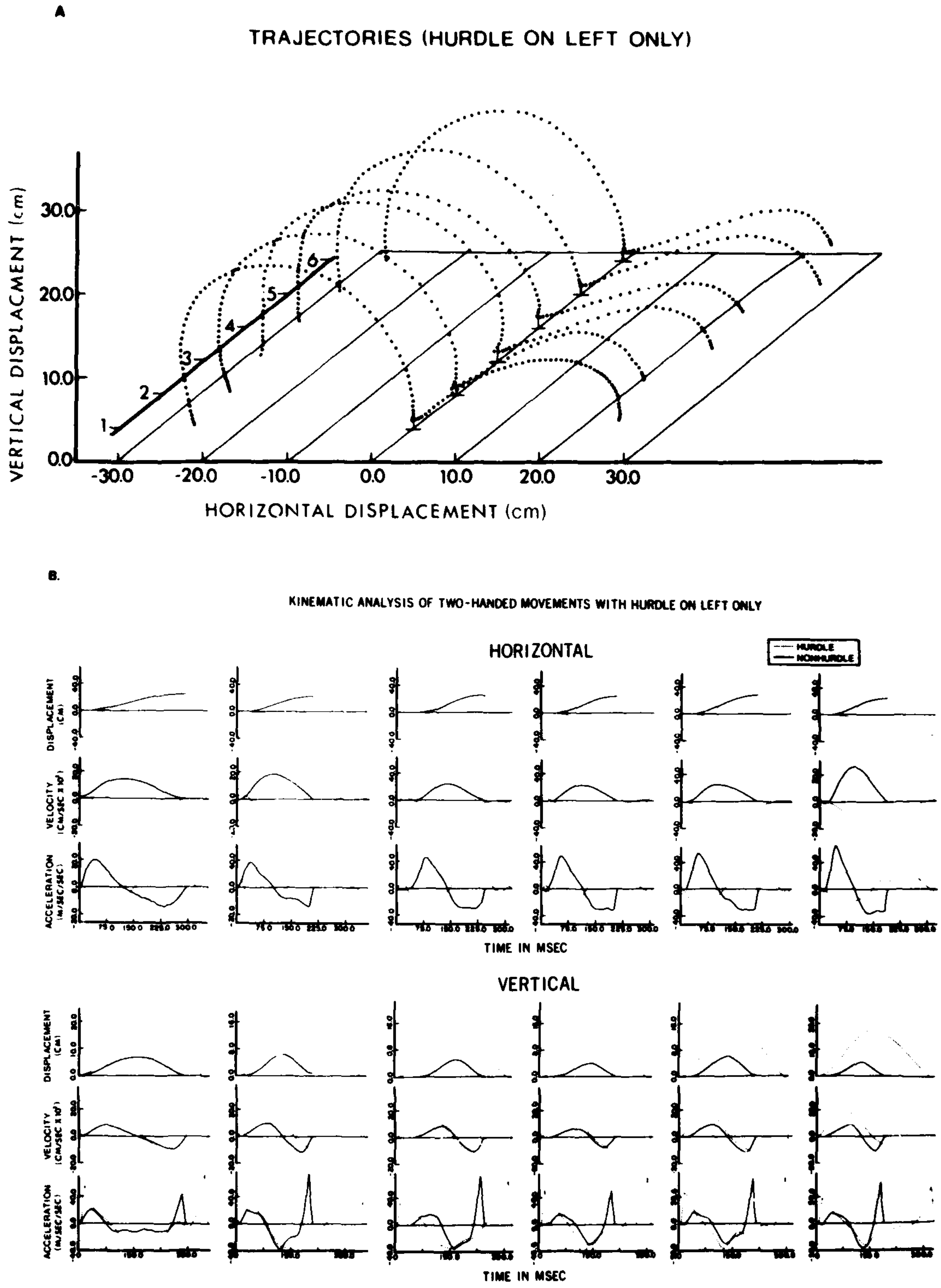


Figure 7. (A) Movement trajectories and (B) consequent kinematic profiles of subject P.H. Trials 1 through 6 on z-axis correspond to filmed trials 1 and 2, 8 and 9, and 19 and 20.

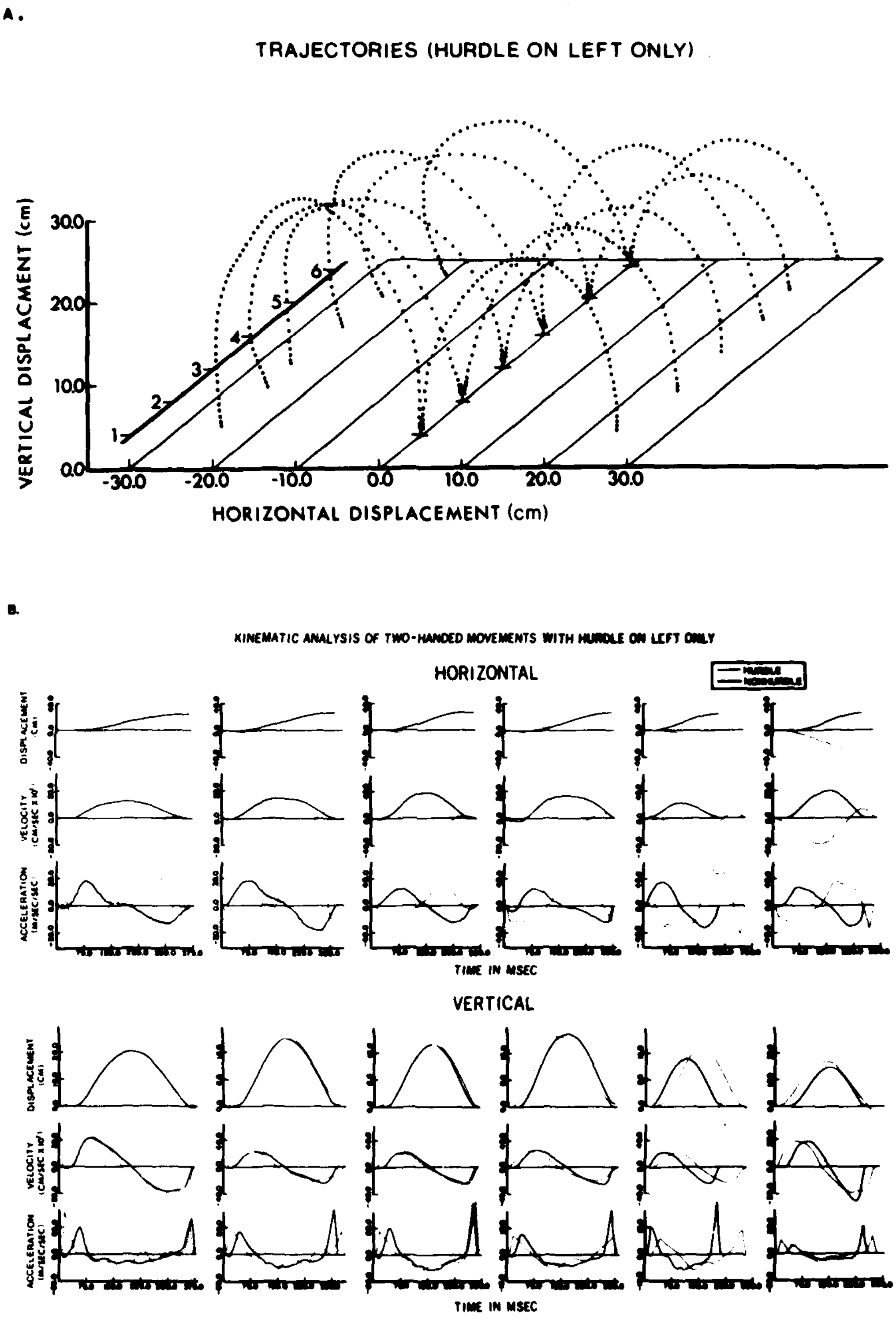


Figure 8. (A) Movement trajectories and (B) consequent kinematic profiles of subject M.B. Trials 1 through 6 on z-axis correspond to filmed trials 1 and 2, 8 and 9, and 19 and 20.

than usual (compare Figure 5 for the same subject performing under hard-hard conditions). In contrast, for subject MB, shown in Figure 8, the trajectories of both limbs are very much alike across trials, and the kinematic similarities between both limbs are strikingly apparent.

EXPERIMENT 3

Because all the published experiments using this paradigm have examined symmetrical movements of limbs, and because the symmetry constraint seems to be such a powerful one in movement (see Introduction), we felt that it would be useful also to examine asymmetrical movements that involve non-homologous muscles. On the face of it, there are not too many reasons to predict different results for such movements. Skilled pianists, for example, appear to be able to move their hands in the same or different directions with equal facility. It is still possible, however, that non-homologous muscle groups may be less effectively controlled as a functional unit in our task, or indeed that they are controlled in a more independent way. We explore this issue in the final experiment of this series.

Methods

Subjects. Subjects were ten right-handed volunteers between the ages of 20 and 32 years, none of whom had participated in any of the previous two-handed experiments.

Task. The two-handed apparatus described previously was modified somewhat for this experiment, which involved asymmetrical movements of the limbs. The base of the apparatus was split into two identical halves, such that each housed a home key and a target key that was positioned either near or far from the home key. The two bases were then placed side by side and oriented so that the home keys were located opposite the left shoulder of the subject, and the target keys extended laterally to the right. Thus, movements of both hands were always to the right, and involved primarily flexion of the left arm and extension of the right. As in our previous studies, two distance by target sizes were used, resulting in both an easy task (7.2 cm target, centered 6 cm from the home key) and a hard task (3.6 cm target, centered 24 cm from the home key). Filming was not conducted for this experiment. Other than these modifications, the apparatus remained identical to that of Experiment 1.

All combinations involving single and two hands and easy and hard targets were performed by each subject. Instructions to subjects were identical to those described previously. In each of the eight resulting conditions, there were 25 trials; the first five were considered to be practice trials and excluded from statistical analysis. One half of the subjects performed the task such that the right hand was always associated with the home key-target key arrangement closest to the body, while the left hand was assigned to the home key-target key farthest from the body. This assignment was reversed for the remaining subjects.

Results and Discussion

Mean initiation times, movement times, and total response times are shown for each condition in Figure 9. Our main concern was whether the findings of simultaneity of initiation and termination of movement found in our previous work extended to asymmetrical movements in which non-homologous muscle groups were used. The basic findings were indeed replicated. No significant differences in initiation times were found between hands, the largest mean difference = 8 msec, $p > .05$ (MSe = 395.2, d = 23 msec, $p < .05$).

As expected, movements to the hard target took longer than movements to the easy target, both in the single hand conditions (mean difference = 64 msec, $p < .01$), and the two hand conditions in which the movements were identical (mean difference = 66 msec, $p < .01$, MSe = 912.2, $d_{.01} = 48$ msec). This rather large difference in movement times between the easy and hard conditions was reduced considerably when the two movements were executed under conditions of mixed difficulty (mean difference = 15 msec, $p > .05$). These results then mirror the major aspects of our earlier work on symmetrical movements, and provide little reason to assume that the organization for asymmetrical movements is qualitatively different.

GENERAL DISCUSSION

Our intent in these experiments was to elaborate the processes underlying the control and coordination of both limbs when they cooperate together in a task that places very different spatial demands on each limb. A key feature of the approach was to combine behavioral measures of movement outcome (e.g., initiation time, movement time) with information about space-time trajectories, followed by a kinematic analysis of the movement trajectories themselves. Although there is a long history of work on the analysis of human motion (e.g., Marey, 1894), only quite recently have engineers and neuroscientists come to recognize its importance for understanding the logical operations through which the nervous system participates in the organization of skilled movements (e.g., Abend, Bizzi, & Morasso, in press; Soechting & Lacquaniti, 1981).

A central and ongoing aspect of our work, following the lead of Bernstein (1967), is to examine movements in which many degrees of freedom are involved, in an attempt to identify the "significant functional units" of coordination (cf. Greene, 1971). After Gelfand and Tsetlin (1971, see also Bernstein, 1967, Chapter 6), we envisage the variables that define these functional units or coordinative structures as falling into two classes: essential variables that determine the form of the function (also referred to as the structural prescription of movement, cf. Boylls, 1975; Kelso et al., 1979a, 1979b; Turvey et al., 1978) and non-essential variables that specify marked changes in the values of the function, but leave its topological properties essentially unchanged (the metrical prescription).

A main way to discover the signature of coordinative structures is to alter the metrics of the motor activity (e.g., speed it up, do it more forcefully, alter its spatial requirements) and observe which variables are modified and which variables or relations among variables remain unchanged. Note that changing the metrical properties of an action could obscure its

CONDITION	HOME KEYS	EASY TARGET	HARD TARGET	INITIATION TIME	MOVEMENT TIME	TOTAL RESPONSE TIME
1	●	1		240	120	360
2	●	2		224	103	327
3	●		3	228	183	411
4	●		4	233	169	402
5	●	5		249	141	390
	●	6		253	139	392
6	●		7	253	207	460
	●		8	248	205	453
7	●	9		249	176	425
	●		10	241	183	424
8	●		11	271	229	500
	●	12		273	204	477

Figure 9. Mean initiation time, movement time, and total response time (in msec) for single and two-handed lateral movements to the right. Two-handed conditions require asymmetrical movements involving non-homologous muscle groups.

basic form by altering properties of individual components that might otherwise remain stable. Alternatively, these changes may index the major ways that invariance can be observed: Some variables must change but others must remain the same if the internal structure of the action is to be preserved. This strategy has proved successful in uncovering coordinative structure styles of organization in many different types of activities (Boylls, 1975; Fowler, 1977; Kelso, 1981; Kelso & Tuller, in press; Kugler, Kelso, & Turvey, 1980). The most well-known examples come from studies of locomotion. For example, when a cat's speed of locomotion increases, the duration of the "step cycle" decreases (cf. Grillner, 1975; Shik & Orlovskii, 1976). Changes in the speed of locomotion are known to be accomplished by distributing more force into the support or stance phase of the cycle. That is, there is an increase in the activity of extensor muscles in an individual limb when it is in contact with the ground. Significantly, an increase in propulsive force during the stance phase does not disrupt the relative timing among linked extensor muscles, even though their absolute magnitudes and durations change considerably (Engberg & Lundberg, 1969; see also Madeiros, 1978, and Shapiro, Zernicke, Gregor, & Diestal, 1981, for human evidence).

Constancy of timing relationships across scalar changes in rate has been reported for other activities of a cyclical kind, such as mastication and respiration (see Grillner, 1977, for review). However, the stability of temporal relationships over metrical change has also been shown to characterize less obviously cyclical activities including postural control (Nashner, 1977), voluntary arm movements (Lestienne, 1979) and handwriting (Viviani & Terzuolo, 1980). Similarly, Freund and Budingen (1978) demonstrate that the rise time of voluntary contraction in rapid, discrete movements is constant no matter how strong the contraction is or how far the limb has to move. According to Freund and Budingen (1978), "...the independence of the time of contraction of skeletal muscles from the final force level or angle of movement is regarded as a necessary condition for the synchrony of synergistic action" (p. 2).

From the overall results of the experiments reported here there is good reason to believe that the motor system solves the problem posed in the present task by constraining the limbs to function as a single, synergistic unit within which component elements vary in a related manner. The behavioral data in Experiments 1 and 3 indicate that the large and highly significant differences in movement time found between easy and hard conditions are reduced considerably when the hands are combined. The small but consistent tendency for the easy limb to strike its target first was further reduced when total response time was the dependent measure.

Although their experimental conditions were rather different from ours (10 and 30 cm movements with a weighted stylus to a 1 mm target), Marteniuk and MacKenzie's (1980) results are similar to the present findings as well as our earlier studies. Their data also reveal a significant slowing of the easy hand and a speeding up of the difficult one under mixed conditions compared to two-hand controls. Although they make much of the statistical fact that the easy hand reaches its target earlier, the average difference between the two limbs was only 20 msec, which is in sharp contrast to the difference between the two-hand control conditions (mean difference = 68 msec, see Marteniuk & MacKenzie, Table 2).3 In addition, Marteniuk and MacKenzie (1980) report a "dramatic overshoot" in terms of spatial error for the easy hand under mixed conditions compared to its control, further suggesting a strong coupling between the limbs both spatially and temporally.

The picture of interlimb coordination becomes clearer in the present work when the space time trajectories and consequent kinematic characteristics are examined. A number of features of the kinematic data emerge that are worthy of note and implicate certain underlying processes. In Experiment 1 it is obvious that the net forces produced in the horizontal direction are different in magnitude for each limb under conditions of varying spatial demand, as revealed by peak accelerations. Moreover, there is considerable inter-trial variability in these values. Even though the metrics change, however, times to peak velocity and acceleration are quite stable; the temporal structure remains remarkably invariant (cf. Figures 4 and 5). When an obstacle is placed in the way of one limb (Experiment 2), there is still a strong tendency for the limbs to preserve their relative timing, although it is clear that this is not absolutely mandatory for some subjects. It seems apparent, nevertheless, that the scaling requirements on one limb influence the other; what we cannot provide at present is a principled reason for why the effects are greater for some subjects than others. One idea, which we are exploring, is that there may be a critical scaling value on obstacle height to which subjects are perceptually sensitive, that influences whether the limbs are treated as a symmetrical unit or not. The analogy here comes from recent work on locomotion, in which it can be shown that at certain critical values of velocity (related to minimum energy criteria) horses shift from one locomotory pattern to another, e.g., walking to trotting (Hoyt & Taylor, 1981). In our experiments, there may be a critical value of obstacle height in relation to the limb dimensions of the performer that specifies which coordinative structures are to be marshalled.

Although we have not paid much attention to the initiation time data (since it was not the main concern here), it is interesting that there is a general elevation in initiation time in the obstacle experiment, particularly when two limbs are involved. Recent work in this area (see Keele, 1981, for review) suggests that the time to prepare a movement (as reflected in initiation time) is a function of the upcoming movement's complexity (cf. Henry & Rogers, 1960; Sternberg, Monsell, Knoll, & Wright, 1978). Moreover, Keele (1981, p. 1410-11) suggests that preparatory time increases when two elements are timed differently. To the extent that this occurs in the present Experiment 2, there is support for Keele's (1981) view; certainly the effects on initiation time are much smaller when the limbs share common timing (cf. Kelso et al., 1979a, 1979b).

The strong tendency for the temporal structure of two-handed movements to be preserved in the face of scalar variation in kinematic values provides strong support for the Bernstein view that it is not individual muscles that are controlled, but rather muscle linkages that govern the interaction between limbs in a relatively autonomous way. As we have emphasized elsewhere, these are neither fixed motor programs nor prefabricated reflexes; they are modulable and functional units of action directed toward accomplishing particular goals.

In a remarkable, but not widely known treatise on cerebellar function, Boylls (1975) argues that the structural aspects of movement--as indexed by qualitative ratios and relative timing among linked muscles and kinematic events--are specified in terms of the relative amounts of activity distributed among descending tracts from the anterior cerebellar lobe. Absolute activities in these tracts specify values on metrical parameters. Obviously we cannot measure neural activity in our paradigm, but we do have some data that are consistent with Boylls' theory. In a study identical to Experiment 1, Tuller and Kelso (Note 2) examined interlimb coordination in split-brain patients. Although the movements were slower overall than in normal subjects, the relative timing between the limbs in the easy-difficult conditions was again near synchronous (mean movement time difference = 13 msec). These data suggest that the details of timing may not be prescribed at higher cortical levels, but rather arise from the functioning of autonomous structures, perhaps at the level of cerebellum and below. Interestingly, Orlovskii's (1972) research has shown that cerebellar stimulation during cat locomotion affects only the magnitude of muscle contraction, leaving the timing among muscles unchanged relative to the step cycle (cf. Shik & Orlovskii, 1976, for review).

The discovery of coordinative structures (or muscle linkages) and their rigorous analysis continues to be the goal of much of the Russian work on motor control (e.g., Gelfand et al., 1971) and seems crucial if we are to understand how the many degrees of freedom of the motor system are regulated. Investigations have begun of the space-time characteristics of single limb movements to targets (e.g., Abend et al., in press; Soechting & Lacquaniti, 1981) and the present work is an extension to the localization behavior of both limbs. It seems reasonable to propose that in our task the equilibrium positions of both limbs can be defined independently as a function of the spatial demands of the task (Kelso et al., 1979a, 1979b; Marteniuk & MacKenzie, 1980). Recent work on single-limb movements suggests that final position can be specified in terms of a balance (or equilibrium point) between the length-tension ratios of agonist and antagonist muscles (e.g., Bizzi et al., 1978; Cooke, 1980; Fel'dman, 1966, 1980; Kelso, 1977; Kelso & Holt, 1980; Lestienne, Polit, & Bizzi, 1981). In localizing limbs, the muscle-joint ensemble behaves dynamically like a nonlinear oscillatory system with specifiable parameters of equilibrium length and stiffness (cf. Bizzi et al., 1978; Fel'dman, 1966; Kelso, 1977; Kelso, Holt, Kugler, & Turvey, 1980). The fact that, in our task, the magnitude of force produced by each limb is different adds support to the notion that stiffness and equilibrium length are potentially modulable parameters of two-handed movements.

We strongly suspect, however, that the relatively invariant timing relations between the limbs arise from parameter specification of the muscle-joint linkage system rather than special timing mechanisms. In identifying the behavior of muscle collectives with autonomous nonlinear oscillators, observables such as time and trajectory are not explicitly represented. Instead, they are a consequence of the system's dynamic parameterization (e.g., equilibrium lengths, stiffnesses).

In our final remarks let us consider how the oscillator-theoretical framework might accommodate the present data on the cooperative behavior of two limbs producing movements of different amplitude. Two main claims would seem to require evaluation. The first strong claim (one that we have not actually made) says that the behavior of the two limbs is perfectly synchronized. The second claim (one based on empirical fact) says that there are small, but systematic departures from synchrony that are often not statistically significant. That is, there is a tendency in our data for the limb moving to the near target to arrive slightly earlier than the limb moving to the more distant target. These small departures from perfect synchrony may be amplified when high accuracy demands are placed on subjects (e.g., Marteniuk & MacKenzie, 1980) or if the movements are of widely different amplitudes. However, both claims of perfect synchrony between the limbs and of near-synchrony between the limbs may be accounted for in a principled way by the same type of model.

Consider the perfect synchrony claim first. Let us assume that each limb can be treated as a single-dimensional system and that the stiffness parameterization is the same for each limb. The equilibrium points, however, must be differentially specified to conform with task requirements. In this case, if both limbs behaved as linear systems, they would necessarily produce identical movement times. In linear mass-spring systems, for example, amplitude and frequency are independent. Thus, assuming constant stiffness over the range of motion, small and large movements must have the same period; the movements will be perfectly isochronous.

Deviations from isochrony can be explained if one makes the additional assumption of stiffness nonlinearity, that is, that the average stiffness is not absolutely constant throughout the motion. In "soft" nonlinear springs, for example (e.g., Jordan & Smith, 1977), stiffness actually decreases with increasing distance from the equilibrium point. Extrapolating to the present case, movements of large amplitude will be slightly slower than those of short amplitude, because they have smaller average stiffnesses over the range of motion. Moreover, the greater the amplitude difference between the two limbs the greater should be the deviations from isochrony. Thus, if the limbs are viewed as behaving like linear oscillatory systems, perfect isochrony is predicted. Consistent deviations from isochrony, however, can be accommodated by the assumption that the limbs in this case behave as "soft" nonlinear oscillators in which stiffness is defined differentially for short and long movements.

In conclusion, the present data reveal a dissociation between force scaling and timing that is indexical of muscle-joint ensembles when they are temporarily constrained to function as a single unit. Such units appear to share the same abstract functional organization as autonomous nonlinear oscillatory systems.

REFERENCE NOTES

1. Marteniuk, R. G. Personal communication, October 1980.
2. Tuller, B., & Kelso, J. A. S. Interlimb coordination in split-brain patients. Manuscript in preparation.

REFERENCES

Abend, W., Bizzi, E., & Morasso, P. Human arm trajectory formation. Brain, in press.

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

Bizzi, E., Dev, P., Morasso, P., & Polit, A. Effects of load disturbance during centrally initiated movements. Journal of Neurophysiology, 1978, 41, 542-556.

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

Cooke, J. D. The organization of simple, skilled movements. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North Holland, 1980.

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

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

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

Fel'dman, A. G. Superposition of motor programs--II. Rapid forearm flexion in man. Neuroscience, 1980, 5, 91-95.

Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 1954, 47, 381-391.

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

Freund, H. J., & Budingen, H. J. The relationship between speed and amplitude of the fastest voluntary contractions of human arm muscles. Experimental Brain Research, 1978, 31, 1-12.

Gelfand, I. M., & Tsetlin, M. L. Mathematical modeling of mechanisms of the central nervous system. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge, Mass.: MIT Press, 1971.

Gelfand, I. M., Gurfinkel, V. S., Tsetlin, M. L., & Shik, M. L. Some problems in the analysis of movements. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge, Mass.: MIT Press, 1971.

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

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

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

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

Henry, F. M., & Rogers, D. E. Increased response complexity for complicated movements and a memory drum theory of neuromotor reaction. Research Quarterly, 1960, 31, 448-458.

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

Jordan, D. W., & Smith, P. Nonlinear ordinary differential equations. Oxford: Clarendon Press, 1977.

Keele, S. W. Behavioral analysis of movement. In V. B. Brooks (Ed.), Handbook of physiology: Section 1: The nervous system. Volume II. Motor Control, part 2. Baltimore, Md.: American Physiological Society, 1981.

Kelso, J. A. S. Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology, 1977, 3, 529-543.

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

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

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

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

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

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

Kelso, J. A. S., & Tuller, B. A dynamical basis for action systems. In M. S. Gazzaniga (Ed.), Handbook of cognitive neuroscience. New York: Plenum, in press. (Also Haskins Laboratories Status Report on Speech Research, 1983, SR-73, this volume.)

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, Calif.: Wadsworth Publishing Co., 1968.

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

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

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

Lestienne, F., Polit, A., & Bizzi, E. Functional organization of the motor process underlying the transition from movement to posture. Brain Research, 1981, 230, 121-131.

Madeiros, J. Investigation of neuronal mechanisms underlying human locomotion: An EMG analysis. Unpublished doctoral dissertation, University of Iowa, 1978.

Marey, E. Le mouvement. Paris: G. Masson, 1894.

Marteniuk, R. G., & MacKenzie, C. L. A preliminary theory of two-handed movement control. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North Holland, 1980.

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

Newell, K. M. The speed-accuracy paradox in movement control: Error of time and space. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland, 1980.

Orlovskii, G. N. The effect of different descending systems on flexion and extension activity during locomotion. Brain Research, 1972, 60, 359-371.

Pavlidis, T. Biological oscillators: Their mathematical analysis. New York: Academic Press, 1973.

Peters, M. Handedness: Coordination of within- and between-hand alternating movements. American Journal of Psychology, 1981, 94, 633-643.

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

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

Schmidt, R. A., Zelaznik, H. N., Hawkins, B., Frank, J. S., & Quinn, J. T. Jr. Motor-output variability: A theory for the accuracy of rapid motor acts. Psychological Review, 1979, 86, 415-451.

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

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

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

Stein, P. S. G. Application of the mathematics of coupled oscillator systems to the analysis of the neural control of locomotion. Federation Proceedings, 1977, 36, 2056-2059.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), Information processing in motor control and learning. New York: Academic Press, 1978.

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

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

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

Wilson, D. M. Insect walking. Annual Review of Entomology, 1966, 11.

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

Woods, B. T., & Teuber, H.-L. Mirror movements after childhood hemiparesis. Neurology, 1978, 28, 1152-1158.

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

FOOTNOTES

[1]We do not claim that the types of constraints observed in our two-handed movement task cannot be broken down with practice, or by instructional strategies, or by loading the limbs differentially, or by removing visual information, etc. We do claim that, faced with the task of controlling many muscles in the two-handed task, the perceptual-motor system tends to solve this particular problem naturally, by coordinating the limbs as a single unit. These experiments are directed toward an understanding and classification of natural constraints on multidegree of freedom systems. They do not speak to the many apparently arbitrary activities that subjects can perform in laboratory situations.

[2]It is worth noting that subject PH had considerable ballet experience; as a consequence, she may have been more capable of controlling the limbs independently in this task.

[3]As a relevant aside, none of our subjects (and we have tested over 70) in the original Kelso et al. (1979a, 1979b) studies and in the present Experiments 1 and 3 perceived that the movements were non-simultaneous under combined conditions as revealed through post-experiment interviews. The same has been the case in Marteniuk's work (Note 1), suggesting further that the small differences between the limbs, though occasionally statistically different, are not meaningfully different.

SOME ACOUSTIC AND PHYSIOLOGICAL OBSERVATIONS ON DIPHTHONGS*

René Collier,+ Fredericka Bell-Berti,++ and Lawrence J. Raphael+++

Abstract. This paper presents an analysis of some articulatory properties of (Dutch) diphthongs, attempting to correlate articulatory inferences based on perceptual and acoustic data with more direct physiological measurements (recordings of EMG activity). Evidence is presented that supports a distinction between "genuine" and "pseudo" diphthongs: the two classes appear to differ (1) in openness and advancement at their onsets and offsets, (2) in the harmony of tongue position between the beginning and ending configurations, and (3) possibly also in the number of articulatory gestures involved.

INTRODUCTION

It has long been customary to transcribe diphthongs using two phonetic symbols that, used separately, represent simple vowel and semivowel segments. To judge from these impressionistic transcriptions, any two diphthongs may differ minimally in either their onset or offset qualities. For example, in Dutch the diphthong /ɛi/ is said to end with a high front vowel, whereas the diphthong /aj/ is said to end with an acoustically similar semivowel. In such instances one might ask whether these transcriptions--that reflect perceptual differences between two sounds--also reflect measurable differences in acoustic structure and articulatory strategy. Furthermore, we might ask whether the symbols used in the impressionistic transcription of the diphthongs have the same acoustic and articulatory values as do the simple vowel and semivowel segments that they represent. Finally, does conventional transcription practice reflect the perceptual impression that these sounds are composed of two separate segments and, if so, are they produced as a sequence of two articulatory gestures? These questions may best be addressed in a language containing the simple vowels and semivowels used in transcribing its diphthongs.

We have chosen to study Dutch because it is a language containing a sufficient number of diphthongs to allow one to answer the questions we have

---

*Also Language and Speech, in press.
+University of Antwerp and Institute for Perception Research (IPO).
++Also St. John's University.
+++Also Herbert H. Lehman College, The City University of New York.
Acknowledgment. This work was supported by NINCDS grants NS-13617 and NS-05332 and BRS grant RR-05596 to Haskins Laboratories and by the National Science Foundation of Belgium (NFWO).

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

raised above. In fact, it is claimed that Dutch has two types of diphthongs: "genuine" (/εi, ʌy, ɑu/) and "pseudo" (/aj, oj, uj, iw, ew/) diphthongs.1 There has been little consensus among Dutch phoneticians and phonologists as to what characterizes each class of diphthong. Matters have been further complicated by the existence in Dutch of "long" or "tense" vowels that tend to be diphthongized as well, [ei, øy, ou], possibly in still a different way (Koopmans-van Beinum, 1969; 't Hart, 1969).

A good survey of how phoneticians and phonologists have interpreted the nature of Dutch diphthongs is given in Zonneveld and Trommelen (1980). It appears that from the end of the nineteenth century until about 1940, most phoneticians did not make a principled distinction between diphthongs and (long) vowels, and--*a fortiori*--did not differentiate between genuine and pseudo diphthongs. Yet they realized that diphthongs consist of two (or more) elements and can be classified according to the relative openness of their first component and (or) the frontness vs. backness of their second. There was also some discussion as to whether the components correspond to vowels that can occur in isolation. The structural phonologists of the thirties raised the question of whether diphthongs should be given a monophonemic or biphonemic representation. They tended to agree that the genuine diphthongs are single phonemes whereas the pseudo ones consist of two phonemes each. This point of view was still endorsed by Van den Berg (1959), whereas Cohen, Ebeling, Eringa, Fokkema, and van Holk (1959) considered all diphthongs to be biphonemic. Generative phonologists, too, have generally preferred a biphonemic underlying representation for the Dutch diphthongs, but they have shown a wide divergence of opinion as to the nature of the two segments involved.

In recent years, better instrumental and experimental techniques have produced a more reliable phonetic specification of the genuine Dutch diphthongs. A perceptual analysis has resulted in the following characterization:

> [εi] is the Dutch vowel [ε], followed by *movement in the direction of* [i]; [ʌy] is the English vowel [ʌ] (as in "cup") -- and not the Dutch [oe] -- followed by *movement in the direction of* [y]; [ɑu] is the Dutch vowel [ɑ] -- not [ɔ] -- followed by *movement in the direction of* [u]. The endpoints [i, y, u] are reached only in careful, isolated pronunciation, with no final consonant. Usually the endpoints are [ɩ], [ø] and [o]. ('t Hart, 1969, p. 172. Our translation, his italics)

Thus, we find a new emphasis on the dynamic character of the genuine diphthongs and a shift away from the traditionally assumed importance of onset and offset qualities. Spectrographic analysis has revealed that the genuine diphthongs are mainly characterized by a relatively unchanging F_2 and an avalanche-like decrease of F_1 (Mol, 1969).

Cohen (1971, p. 288) summarizes the results of these acoustic and perceptual studies as follows:

> There are a number of arguments...for accepting the diphthongs of the Dutch *εi, ʌy, ɑu* type as vocoids, recognizable as such and distinguishable from the other vocoids of the long and short classes, on account of their peculiar, dynamic character.

Pols (1977, p. 103) summarizes his own recent findings by noting that:

> "A diphthong can be described as quite a long steady-state onset part followed by a fast specific transition to an offset area where no steady-state part is necessary. The diphthong [ɑu] starts at [ɑ] and terminates at [ɔ, o]; [ɛi] starts at [ɛ] and goes to [ɩ, e]; and [ʌy] starts at [a] and goes to [œ, ø]. So, none of the three Dutch diphthongs reaches the vowel position indicated in its phonetic transcription.

Pols also notes that the acoustic variability of these diphthongs is very large. This variability correlates well with the fairly large perceptual tolerance observed by Slis and van Katwijk (Note 1), who studied the acceptability of two-formant synthetic diphthongs having a great variety of beginnings and endpoints in the F_1-F_2 plane.

As for the pseudo diphthongs, there has been little or no controversy over their essential characteristics. They have been and still are considered to be sequences of a "tense" vowel and a semivowel. They start with a vowel whose quality is the same as that of the separately occurring vowels [a, e, o, i, y] and move into the glides [j] and [w]. Phonetically they are the sum of their components.

Comparing the characteristics of the genuine and the pseudo diphthongs, we find that they differ in a number of respects, including: (1) the degree of "openness" at onset; (2) the degree of change in tongue advancement between onset and offset; and (3) the degree of harmony between lip position at onset and offset. For example, each of the genuine diphthongs starts with a relatively open vocal tract and ends with a relatively closed one. A pseudo diphthong, on the other hand, may start with an open, half open, or closed vocal tract, before ending with a semivowel. Furthermore, each of the genuine diphthongs ends with a vocal tract shape in which tongue advancement and lip position are approximately the same as they were at the beginning of the diphthong. Each pseudo diphthong, however, ends with a vocal tract shape in which tongue advancement and, usually, lip position are different than they were at the start of the diphthong.[2] In addition, the genuine diphthongs are characterized by relatively continuous and gradual changes in formant structure, whereas the pseudo diphthongs are produced with more abrupt changes in formant structure (Figure 1).

Since there were no physiological data on the production of Dutch diphthongs, the available acoustic and perceptual information led us to hypothesize that there must also be significant differences between the two classes of diphthongs in the articulatory domain. Therefore, the primary aim of our study was to explore how changes in vocal tract configuration are brought about in each of these diphthongs, in order to determine whether physiological descriptions would support their traditional separation into two classes on the basis of acoustic, perceptual, and articulatory phonetic descriptions. To this end, we have, necessarily, described their production in some detail, to provide a base for making the relevant comparisons.

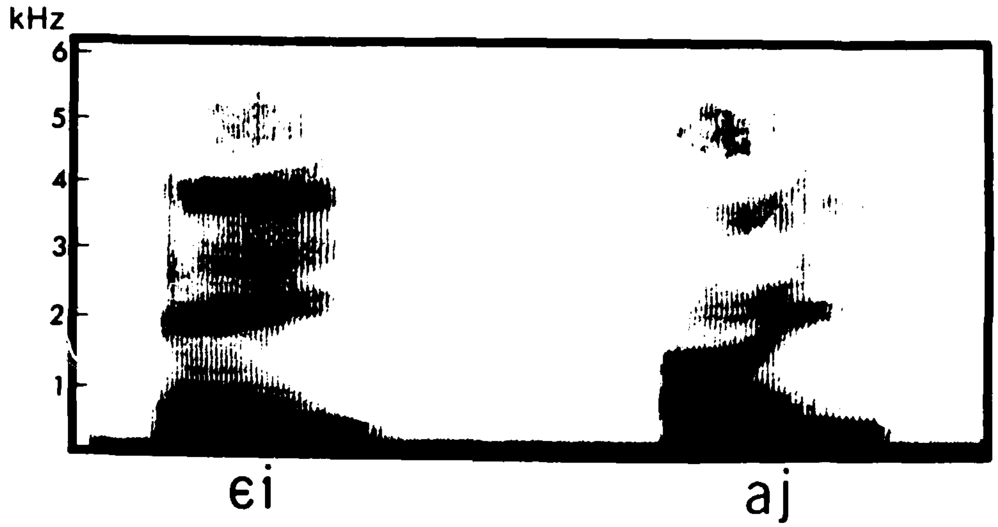


Figure 1. Single token examples of the genuine diphthong [ɛi] and the pseudo diphthong [aj], spoken in isolation.

PROCEDURES

We simultaneously recorded both acoustic and electromyographic (EMG) signals from one speaker of Dutch.3 The EMG potentials were recorded from four muscles known to affect the position of the tongue and the mandible: the genioglossus, styloglossus, mylohyoid, and anterior belly of the digastric. Previously reported physiological data have led us to three groups of assumptions.

Assumptions concerning the functions of the muscles studied. The genioglossus is the only muscle known to contribute significantly to tongue advancement (Alfonso & Baer, 1982; Kakita, 1976; Smith, 1971). It has also been implicated in tongue bunching/raising gestures, although its activity in this regard accompanies activity of other intrinsic and extrinsic tongue muscles (Miyawaki, Hirose, Ushijima, & Sawashima, 1975; Raphael & Bell-Berti, 1975; Raphael, Bell-Berti, Collier, & Baer 1979). The styloglossus is primarily responsible for retraction of the tongue body (Raphael & Bell-Berti, 1975; Smith, 1971). Both genioglossus and styloglossus act with the mylohyoid to elevate the tongue, with the mylohyoid providing the greatest portion of the vertical thrust (Raphael et al., 1979). The mylohyoid may also act to stabilize the hyoid bone, in conjunction with the activity of the anterior belly of the digastric, which assists in lowering the mandible (Raphael et al., 1979).[4]

Assumptions concerning the relationship between the acoustic signal and vocal tract shape for vocoids. It is possible to calculate formant frequencies from a given vocal tract shape and, given a set of formant frequencies, to infer characteristics of the vocal tract shape that produced it (Chiba & Kajiyama, 1941; Delattre, 1951; Fant, 1970; Stevens & House, 1955, 1961). The methods of calculating formant frequencies have been sufficiently refined over the years to generate a near-unique solution for any tract shape. Although the inference of tract characteristics from formant frequencies is less certain, it is widely accepted that the frequency of F_1 is primarily dependent upon the degree of vowel openness, and the frequency of F_2 is primarily dependent upon the length of the front cavity (Fant, 1970; Kuhn, 1975; Stevens & House, 1955, 1961). Thus, for instance, a more open vowel will have a higher F_1 than a more closed one, a fronted vowel will tend to have a higher F_2 than a retracted one, and a rounded vowel will tend to have a lower F_2 than an unrounded one.

Assumptions concerning temporal relationships between EMG potentials and movement. EMG potentials precede their mechanical effect (cf. Harris, 1981). The "contraction times" for the muscles included in this study are on the order of 70-100 msec; that is, movements associated with EMG potentials begin about 70-100 msec after the electrical activity begins.

Pairs of bipolar hooked-wire electrodes were inserted into the genioglossus (anterior fibers), mylohyoid, styloglossus, and anterior belly of the digastric muscles, using standard procedures that are described elsewhere (Hirose, 1971; Raphael & Bell-Berti, 1975). The nonsense test utterances were of the form [də'pDpɑpə], where D=/aj, oj, uj, ew, iw, ɛi, ʌy, ɑu/, and [ə'pVp], where V=/i, u, e, ɛ, a, ɑ, y, oe, o, ɔ/. The subject read from randomized lists of the utterances until he had produced 16 tokens of each. The recordings of all tokens of each of the eight utterance types were aligned with reference to the onset of vocal fold vibration in the diphthong. The EMG potentials were rectified, integrated, and computer sampled, and ensemble averages of the EMG potentials were then calculated for each channel for each utterance type. The EMG data processing system is described in greater detail in Kewley-Port (1973).

In addition to the EMG analysis, we performed acoustic analyses with a digital waveform and spectral-analysis system. Ensemble averages of both the amplitude envelope of the audio waveforms and of digital spectrograms were also calculated.

RESULTS

We shall describe the EMG and acoustic data in relation to the traditonal articulatory phonetic, perceptual, and acoustic descriptions, provided above, concerning the differences between the genuine and pseudo diphthongs. As we have explained above, there is no one-to-one correlation between articulator position and muscle potentials, nor may a unique vocal tract shape be derived from a set of formant values. Hence, we will not attempt to specify absolute articulator position (i.e., vocal tract shape) on the basis of our acoustic or physiological data. Rather, we will compare the data on the diphthongs among themselves and with the data on simple vowels, to infer relative differences in the articulatory parameters. We shall consider first the hypothesis that the two groups differ in openness and advancement at their onsets and offsets. In addition, the onsets and offsets of the genuine diphthongs differ in openness and advancement from the simple vowels and semivowels described as their starting and ending positions, whereas the pseudo diphthongs do not. The second hypothesis is that the groups differ in the harmony of tongue position between the beginning and ending configurations.5 Finally, we shall examine the hypothesis concerned with whether or not the two groups of diphthongs are specified as different numbers of discrete gestures; that is, that the genuine diphthongs are specified as single gestures whereas the pseudo diphthongs are specified as two discrete, concatenated gestures.

A. Hypothesis 1: Openness and Advancement

1. Openness

Traditionally, the genuine diphthongs of Dutch were described as proceeding from relatively open to relatively close articulatory positions, whereas the pseudo diphthongs proceed from various degrees of open to close articulatory positions. Thus, the articulations of the genuine diphthongs were said to begin with relatively open positions (similar, to those of /ɛ, ɔ, oe/) and to end with the close positions of [i,u,y], respectively. In contrast, the articulations of the pseudo diphthongs were said to begin with the appropriate degrees of openness for the vowels /a,o,u/ and /e,i/, and to end with the close positions of the semivowels /j/ and /w/, respectively.

a. Genuine diphthongs. As stated in the introduction, perceptual analyses of the genuine diphthongs have revealed that these diphthongs—especially [ɑu] and [ʌy]—tend to be more open at their beginnings than are the simple vowels used in former transcriptions. This point is fairly well supported by our acoustic and physiological data. The acoustic data in Figure 2a and Table 1 indicate that [ɑu] and [ʌy] have higher F_1 values at their onsets than the simple vowels [ɔ] and [oe]. Hence they are likely to be more open at their beginnings. In fact, [ɑu] has the same F_1 onset value as [ɑ].6 On the other hand [ɛi] has about the same onset F_1 value as [ɛ]. As far as the EMG data for [ɛi] are concerned (Figure 3), there is more anterior belly of the digastric activity at its onset than for [ɛ], but this tongue lowering action may be compensated for by stronger genioglossus activity. At the onset of [ʌy] there is far more anterior belly of the digastric activity than for [oe]. The tongue lowering effect of this action is only partly counterbalanced by the high peak in mylohyoid activity, because this comes late, mainly associated with the later portion of the diphthong. Therefore [ʌy] is likely to have a more open onset position than [oe]. The onset of [ɑu] is very similar to that of [ɑ]: the peaks of mylohyoid and anterior belly of the digastric activity are roughly the same.

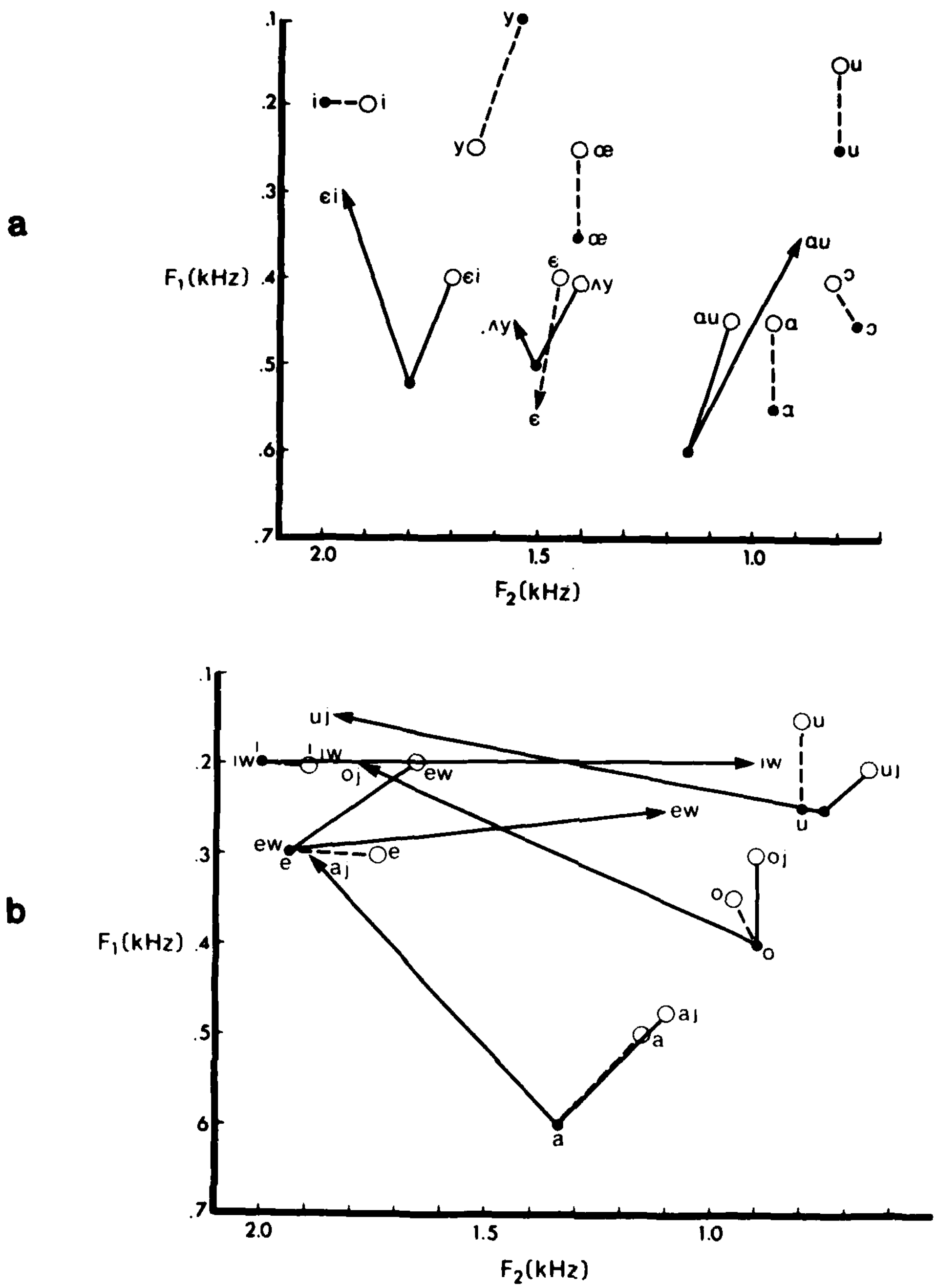


Figure 2. Formant trajectories, in F_1-F_2 plane, of genuine diphthongs (a) and pseudo diphthongs (b), and simple vowels traditionally said to begin and end them. Open circles indicate onset values, filled circles indicate midpoint values, arrowheads indicate offset values (shown only for diphthongs). Solid lines connect diphthong values, dashed lines connect simple vowel onset and midpoint values.

Table 1

Averaged Formant Values (One Speaker, Sixteen Repetitions) for Three Genuine and Five Pseudo Diphthongs, Recorded During the Experiment. Measurements Are Based on Sections at Onset, Midpoint, and Offset, and Are Compared With Formant Values of Simple Vowels Recorded During the Same Experimental Session.

A. Genuine Diphthongs

	Onset			Midpoint			Offset	
	[ɛi]	[ɛ]		[ɛi]	[ɛ]		[ɛi]	[i]
F_1	400	400		525	550		300	200
F_2	1700	1450		1800	1500		1950	2000
	[ʌy]	[oe]		[ʌy]	[oe]		[ʌy]	[y]
F_1	400	250		500	350		450	250
F_2	1400	1400		1500	1400		1550	1650
	[ɑu]	[ɑ]	[ɔ]	[ɑu]	[ɑ]	[ɔ]	[ɑu]	[u]
F_1	450	450	400	600	550	450	350	250
F_2	1050	950	800	1150	950	750	900	800

B. Pseudo Diphthongs

	Onset		Midpoint		Offset	
	[aj]	[a]	[aj]	[a]	[aj]	[i]
F_1	475	500	600	600	300	200
F_2	1100	1150	1350	1350	1900	2000
	[oj]	[o]	[oj]	[o]	[oj]	[i]
F_1	300	350	400	400	200	200
F_2	900	950	900	900	1800	2000
	[uj]	[u]	[uj]	[u]	[uj]	[i]
F_1	200	150	250	250	150	200
F_2	650	800	750	800	1850	2000
	[iw]	[i]	[iw]	[i]	[iw]	[u]
F_1	200	200	200	200	200	250
F_2	2000	1900	2000	2000	900	800
	[ew]	[e]	[ew]	[e]	[ew]	[u]
F_1	200	300	300	300	250	250
F_2	1650	1750	1950	1950	1100	800

At their ends, the first formant frequencies of the genuine diphthongs reflect degrees of openness greater than those of the simple vowels [i, y, u]. Interpreting the EMG data, we must assume that the strong, early, jaw-and-tongue lowering activity of the anterior belly of the digastric is not entirely compensated for by the strong, later, tongue-raising activity of the genioglossus, styloglossus, and mylohyoid. In other words, [ɛi, ʌy, ɑu] do not terminate with the target vowels suggested in their transcriptions. This finding is in agreement with the perceptual analysis by 't Hart (1969).

b. Pseudo diphthongs. The pseudo diphthongs appear to achieve relatively stable first formant frequency values (Figures 5b and 5c), which reflect openness positions equivalent to those of the simple vowels said to begin them (Figure 2b). The EMG data (Figure 4), while somewhat less straightforward, do not contradict the inferences drawn from the acoustic measurements. At the onset of [aj], the EMG values are very similar to those for [a], except that genioglossus activity begins later for the diphthong. [ew,iw,oj] appear to begin with the same balance of tongue raising and lowering activity as [e], [i], and [o], respectively. For instance, at the onset of [iw], there is less tongue fronting and raising activity in the genioglossus than for [i], but much stronger mylohyoid contraction. Similarly, the antagonistic forces of styloglossus and anterior belly of the digastric are reversed at the beginning of [oj] as compared to [o]. At the beginning of [iw] the earlier and stronger mylohyoid activity probably compensates for the reduced genioglossus activity in comparison with [i]. Only in the case of [uj] is there no apparent compensation for the reduced activity of the mylohyoid when compared with [u], but possibly the early onset of genioglossus contraction (associated with [j]) contributes to early tongue raising for this diphthong.

2. Advancement

a. Genuine diphthongs. Acoustically (in terms of F_2 values), the genuine diphthongs [ɛi] and [ɑu] appear to begin with a more fronted tongue position than do the simple vowels said to begin them ([ɛ] and [ɑ]) (Figure 2a). The second formant frequency of [ɑu] indicates that it ends with slightly more fronted tongue position than does the simple vowel [u]. On the other hand, F_2 measurements imply that [ɛi] and [ʌy] end with slightly more retracted tongue positions than do [i] and [y]. That is, all the genuine diphthongs appear to be centralized at their endpoints, when considered in relation to the simple vowels [i,y,u].

The EMG activity (Figure 3) of the genioglossus and styloglossus support the acoustically-based observation that the tongue is more fronted at the beginning of [ɛi] and [ɑu] than [ɛ] and [ɑ]: genioglossus activity is stronger for the early part of [ɛi] than it is for [ɛ], and styloglossus activity (which retracts the tongue) is weaker for [ɑu], especially in its earlier portion, than for [ɑ]. The EMG data also support the acoustically-based inferences about tongue position at the ends of these diphthongs: genioglossus activity is much weaker for [ɛi] than [i] and for [ʌy] than [y], implying less extreme fronting for the diphthongs. In parallel with this difference is the slightly weaker styloglossus activity for [ɑu] than for [u], implying slightly less tongue retraction (i.e., more fronting) for this genuine diphthong.

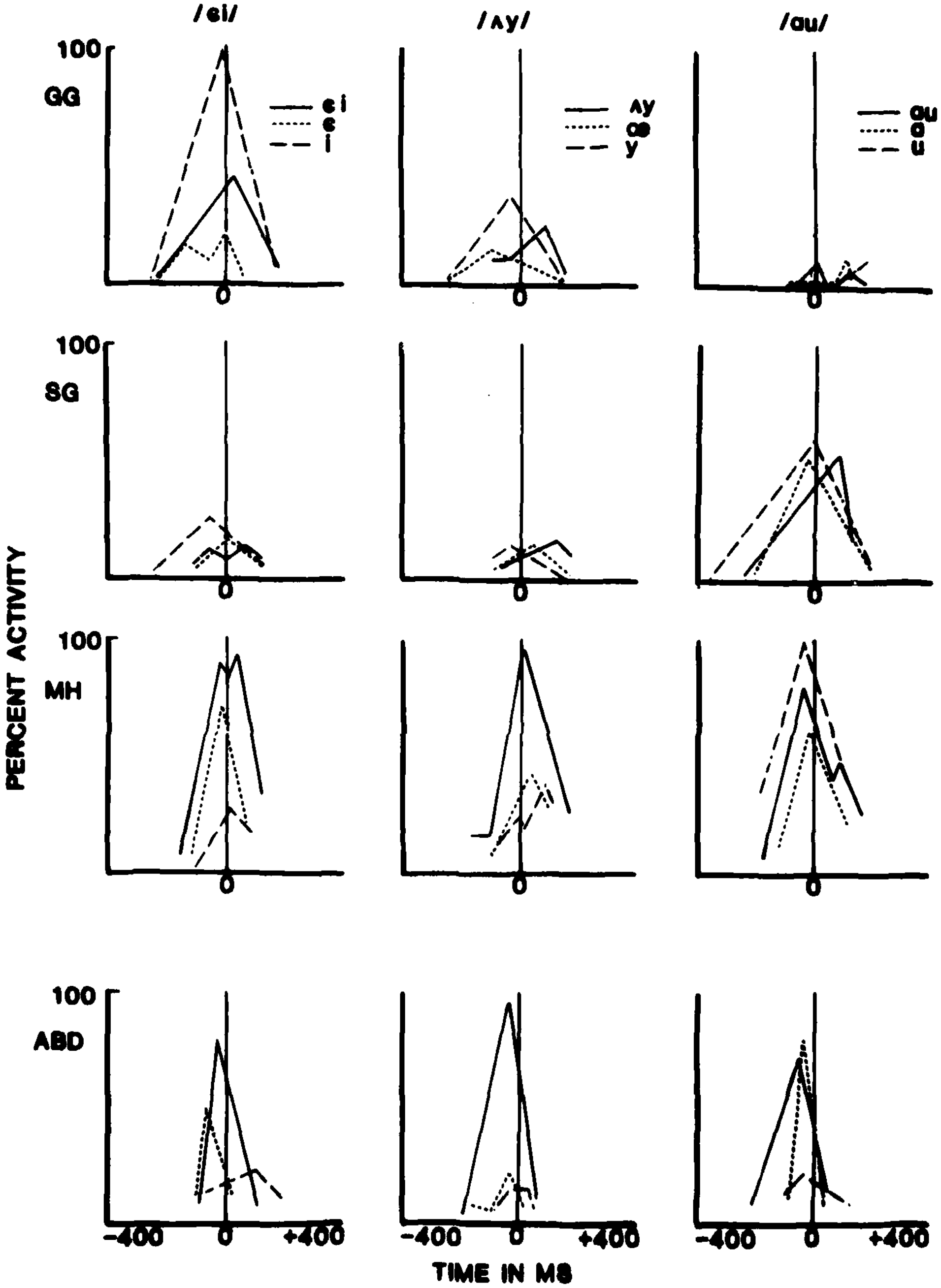


Figure 3. EMG data for genuine diphthongs and simple vowels used in describing them. Each graph is a schematized representation of the time course of EMG activity in a given muscle, expressed as a percentage of the overall range of that muscle's activity across utterance types. Zero on the abscissa represents the acoustic onset of the diphthongs and the simple vowels said to begin and end them.

b. Pseudo diphthongs. The acoustic analyses, in particular the F_2 values, indicate that the first portion of each pseudo diphthong reaches the formant frequencies of the simple vowel said to begin it, but that the second portion of each falls short of its expected semivowel endpoint: the "fronting" diphthongs [aj,oj,uj] fail to reach the second formant frequency values of [i] (Figure 2b), and the "retracting" diphthongs [ew,iw] fail to reach those of [u] (Figure 2b).

Electromyographically, relative activity of the genioglossus and styloglossus for the early part of the fronting diphthongs [aj,oj,uj] is essentially the same as found for the simple vowels [a,o,u] (Figure 4a). The relative activity levels of these muscles for [ew,iw], on the other hand, might lead one to expect slightly less fronting than is inferred for [e] and [i], respectively (Figure 4b). The greater activity of the mylohyoid (which raises the tongue) at the beginnings of [ew] and [iw], than of [e] and [i], suggests that the genioglossus is devoted primarily to tongue advancement, although contributing secondarily to tongue raising.

All five of these diphthongs end 'short' of the F_2 values for [i] or [u], (Figure 2b), and this, too, is reflected in the relative EMG activity level of the genioglossus and styloglossus muscles (Figures 4a and 4b).

Kakita, Hirose, Ushijima, and Sawashima (1976) have observed that there is less genioglossus activity for [j] than for [i], and their X-ray data indicate that the tongue root is indeed less advanced for the semivowel. In our own data this more centralized tongue position for [j] may explain why there is less genioglossus activity for the offset of [aj] and [oj]. Of the fronting diphthongs only [uj] has genioglossus activity as strong as that for [i]; this activity is comparatively brief, however, and follows shortly after strong retracting action by the styloglossus. Among the retracting diphthongs, styloglossus activity is not nearly so strong as that found for [u]. That [iw] and [ew] probably end with a relatively retracted tongue position despite the low level of styloglossus activity at their offset may be due to the fact that the tongue has been strongly raised in their first part (for [i] and [e]), so that it requires less styloglossus action to pull the tongue back for their second part. In short, the EMG data suggest that all the pseudo diphthongs end more centrally than the vowels [i] and [u].

Finally, let us consider the observations made above, in so far as they relate to the basic distinction between diphthong types, viz., that the realization of fixed targets is essential for the pseudo, but not for the genuine diphthongs. We take this to mean that there should be acoustic and EMG differences between the patterns of the genuine diphthongs and those of the simple vowels that, at least in older phonetic transcriptions, are said to compose them. Further, such differences should not be found between the purported simple vowel components of the pseudo diphthongs and the pseudo diphthongs themselves.

Looking for these differences in the acoustic data for the various vowels, we find some support for this distinction between diphthong groups. As we have already seen, the averaged F_1 and F_2 values for the genuine diphthongs differ from those of their simple initial "components," whereas there is a very close correspondence between the F_1 and F_2 values of the

Figure 4a

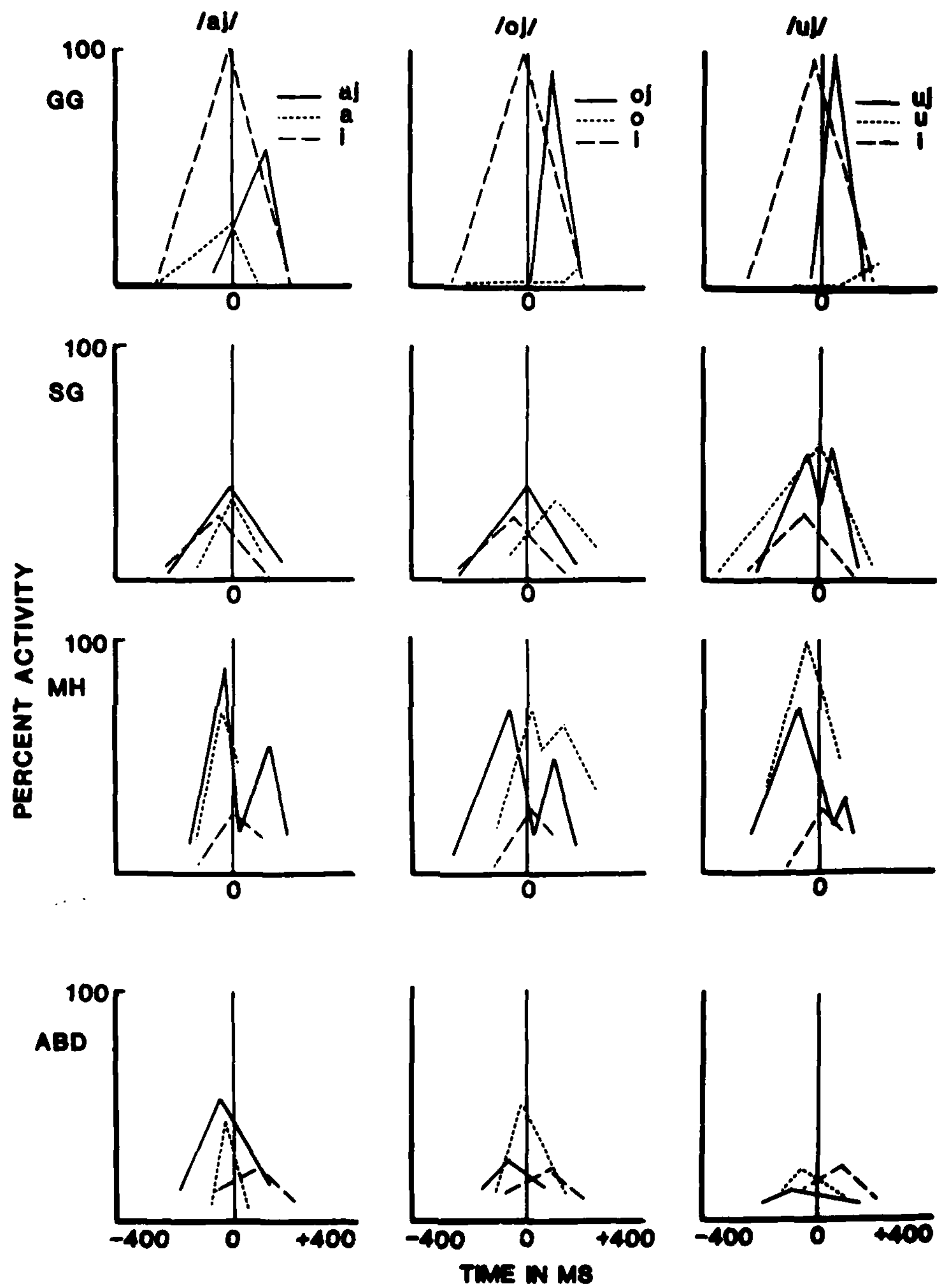
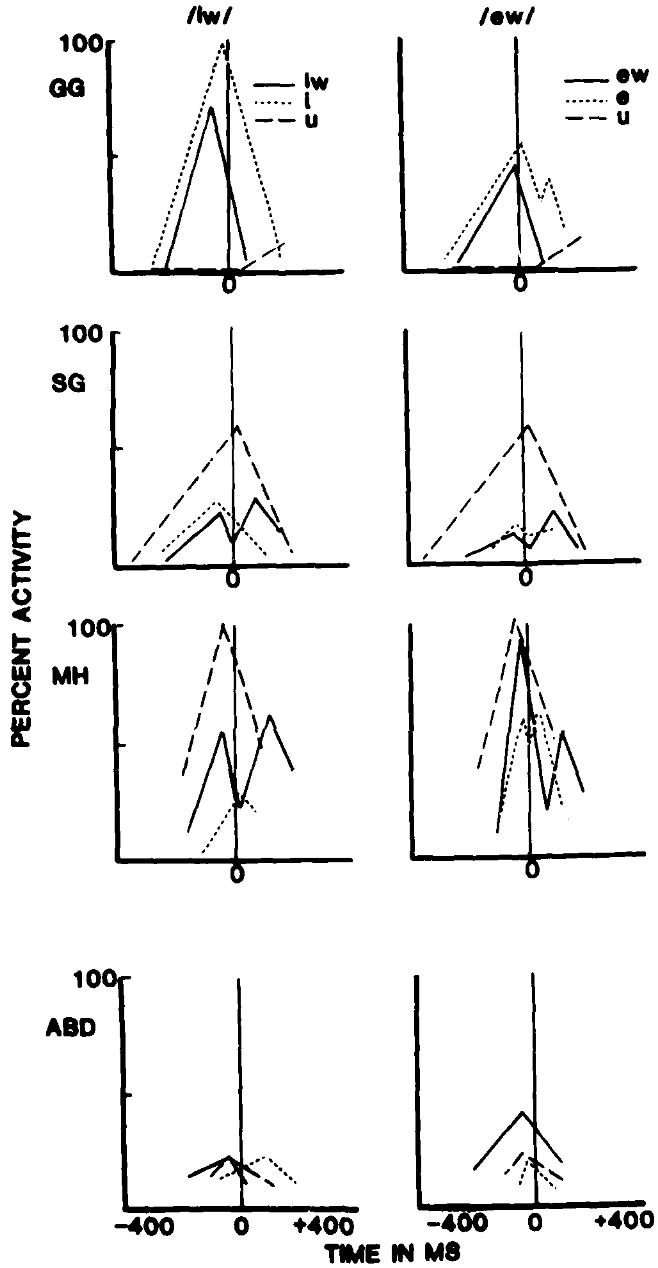


Figure 4. EMG data for the pseudo diphthongs and simple vowels: [aj, oj, uj] in (a), [ew, iw] in (b). Each graph is a schematized representation of the time course of EMG activity in a given muscle, expressed as a percentage of the overall range of that muscle's activity across utterance types. Zero on the abscissa represents the acoustic onsets of the diphthongs, the simple vowels said to begin them, and the vowels /i/ and /u/ that approximate the glides that end them.



/iw/
/ew/
100
GG
iw
i
u
ew
e
u
0
100
SG
100
MH
100
ABD
PERCENT ACTIVITY
-400
0
+400
TIME IN MS

Figure 4b

pseudo diphthongs and their simple initial components (just before the abrupt change in second formant frequency).

Comparing the offsets of the pseudo diphthongs with their simple components yields data sets that are not strictly comparable, because these diphthongs end in semivowels, of which there are no other examples in our data. In Table 1, however, we have included the frequencies of the first two formants at the midpoints of the vowels [i] and [u] on the assumption that the semivowels [j] and [w], respectively, might well approximate these simple vowels acoustically. We find no exact matches and, in several instances, considerable discrepancies in formant values, particularly for the second formants. That is, the second formant frequencies of the pseudo diphthongs fall short of those of [i] and [u], suggesting that the diphthongs are more centralized than are these simple vowels. On the other hand, with the exception of [aj], the first formant values for four of the five pseudo diphthongs are equal to or smaller than those for [i] and [u], suggesting a degree of opening at least as small as that of the most closed vowels.

The acoustic data for the genuine diphthongs, on the other hand, suggest that they end with a more open and central articulation than [i,y,u], supporting the claim that the genuine diphthongs do not match the qualities of the simple vowels that conventional transcriptions suggest as their initial and terminal components. The pseudo diphthongs, in contrast, do match the qualities of the simple vowels that are said to initiate them, although the greatest acoustic similarities occur near the midpoints of the diphthongs and the simple vowels, and not at their onsets. Their offsets approximate the semivowels [j] and [w] rather closely in terms of openness, but tend to be more centralized.

With few exceptions, the EMG data support the inferences drawn from the acoustic data about the differences in starting and ending positions between the genuine and pseudo diphthongs. It is worth noting that the strong correlation between the acoustic and physiological data holds not only for rather gross differences between the two groups of diphthongs. Details of these data support the differentiation of the members of each diphthong class as well. For instance, the F_1 values at the end of the fronting pseudo diphthongs indicate an increasing degree of openness from [uj] to [oj] to [aj]. This gradation is reflected in decreasing levels of genioglossus activity associated with the semivowel. Also the formant values for the offset of [iw] suggest that this diphthong ends with a somewhat higher and more retracted tongue position than [ew]. This correlates with the more pronounced second peak of styloglossus and mylohyoid activity for the former.

These detailed correspondences between the acoustic and the physiological parameters lend support to our assumptions concerning the functions of the muscles studied.

B. Hypothesis 2: Harmony

The claim that there is harmony of tongue advancement for the genuine, but not necessarily for the pseudo, diphthongs is also substantiated by both acoustic and EMG data. The second formants of [ɛi], [ʌy], and [ɑu] display minimal changes in frequency, indicating an absence of extreme changes in tongue advancement (Figures 2a and 5a). In contrast, the second formants for [aj], [oj], [uj], [iw], and [ew] show dramatic frequency shifts, implying the presence of considerable horizontal tongue movement (Figure 2b).

The activity of the muscles responsible for tongue fronting (genioglossus) and backing (styloglossus) also indicates that there is less horizontal tongue movement for the genuine than for the pseudo diphthongs. The genioglossus is moderately active throughout [ɛi], while the styloglossus exerts almost no backward pull; for [ʌy] both muscles are relatively inactive, suggesting a predominance of vertical movement (which is positively indicated by mylohyoid and anterior belly of the digastric activity); and for [ɑu] the styloglossus is moderately active throughout, while the genioglossus is relatively inactive. In contrast, among the pseudo diphthongs we see patterns of activity in which the genioglossus and styloglossus muscles are alternately active. Thus, for [aj], [oj], and [uj], we find early peaks of styloglossus activity and late peaks of genioglossus activity, indicating fronting of the tongue from a backed position; for [iw] and [ew], we find the reverse sequence of genioglossus and styloglossus activity, indicating that the tongue is being retracted from a fronted position.

In summary, we find that our data support claims that distinctions between genuine and pseudo Dutch diphthongs include differences in harmony between the first and second elements with regard to tongue advancement.

C. Hypothesis 3: Single or Concatenated Gestures

Let us turn next to the description that maintains that a genuine diphthong is best characterized by a single articulatory gesture whereas a pseudo diphthong is best characterized as a sequence of two articulatory gestures. The EMG data suggest that there is a difference in the number of gestures for each of the two types of diphthongs. The data cited above, concerning the alternation of genioglossus and styloglossus activity for the pseudo diphthongs, are also relevant here. They depict articulations controlled by two muscles, acting successively first to retract and then to front the tongue ([aj], [oj], [uj]) or to front and then to retract the tongue ([iw], [ew]). The reciprocal timing in activity of these muscles reflects a sequence of opposing motor commands. Further, each pseudo diphthong is produced with two discrete peaks of mylohyoid activity (only in the case of [uj] is the second peak somewhat less pronounced). Each of these peaks is closely aligned in time with a peak of activity in either the genioglossus or the styloglossus muscle, suggesting that the mylohyoid muscle discretely supports the successive fronting and retracting tongue gestures.

In contrast, we would conclude from the EMG data that the genuine diphthongs are characterized as single gestures dominated by the activity of the genioglossus in the case of [ɛi] or by the styloglossus in the case of [ɑu], supported by mylohyoid activity. This supporting activity is less evidently "double peaked" than with the pseudo diphthongs. In the case of [ʌy], where both muscles, as we have noted earlier, are relatively inactive and vertical movement predominates, the mylohyoid muscle displays a single peak of activity, suggesting, once again, a single articulatory gesture.

Further research, using articulatory synthesis techniques, is neeeded to strengthen this hypothesis. Meanwhile, some support for it can be derived from the acoustic data.

The acoustic analysis reveals abrupt changes in second formant frequency of the pseudo diphthongs. For instance, over the first half of its duration the F_2 of [aj] shows a gradual rise in frequency of 250 Hz; over the second half of its duration the increase is 550 Hz, suggesting a rapid movement of the articulators. The analogous frequency changes for [oj] are 100 Hz and 800 Hz; for [uj], 100 Hz and 1200 Hz; for [iw], no change over its first half, and then a decrease of 1100 Hz; and for [ew], an increase of 300 Hz over its first half, and then a drop of 850 Hz. The genuine diphthongs show no such rapid shift in formant frequency in either half of their duration (Figure 5a). Acoustically, then, we do find support for the notion that the pseudo diphthongs are sequences of articulatory gestures.[7]

DISCUSSION

The articulatory data tend to support the acoustic and perceptual separation of the diphthongs into two groups. The genuine ones are characterized by a gradual increase in the activity of those muscles that either cause or support the smooth movement of the tongue in an upward and forward or backward direction. The pseudo diphthongs are characterized by a rather sharp increase in the activity of those muscles that either cause or support the abrupt movement of the tongue from a vowel into a semivowel in which the tongue moves horizontally across the vowel space. In other words, genuine diphthongs behave more like "unitary" segments, while pseudo diphthongs behave like sequences of two segments.

The observed articulatory differences cannot be explained by the difference in the distances an articulator must move between the beginning and the end of the diphthongal gesture in the two groups of diphthongs. Rather, we find that in [εi, ʌy, ɑu], tongue movement is primarily vertical, while in the pseudo diphthongs, tongue movement is primarily horizontal.[8] In terms of "articulatory distance," therefore, the two classes are not necessarily very different. However, the "closing" gesture of the genuine diphthongs is achieved through synergistic action of the mylohyoid and genioglossus or styloglossus, whereas fronting or backing gestures of the pseudo diphthongs are achieved through the sequential antagonistic actions of the genioglossus and styloglossus. This synergism versus antagonism is reflected in the differences in temporal pattern of formant frequency change between the two classes of diphthongs. In the genuine diphthongs, formant frequency change is nearly continuous throughout the entire course of the diphthong; in the pseudo diphthongs, a nearly stable initial portion of substantial duration is followed by a period of rapid formant frequency change (especially in F_2).

The contrastive muscle activity patterns associated with genuine and pseudo diphthongs lends support to Cohen's (1971) proposal to treat [εi, ʌy, ɑu] as "unitary segments, requiring a feature specification of their own, rather than allow for this problem to be circumvented in a treatment which results in a phonetically arbitrary segmentation by assigning one part as dominated by a vocalic and a second one by a sonorant (i.e. non-vocalic, non-

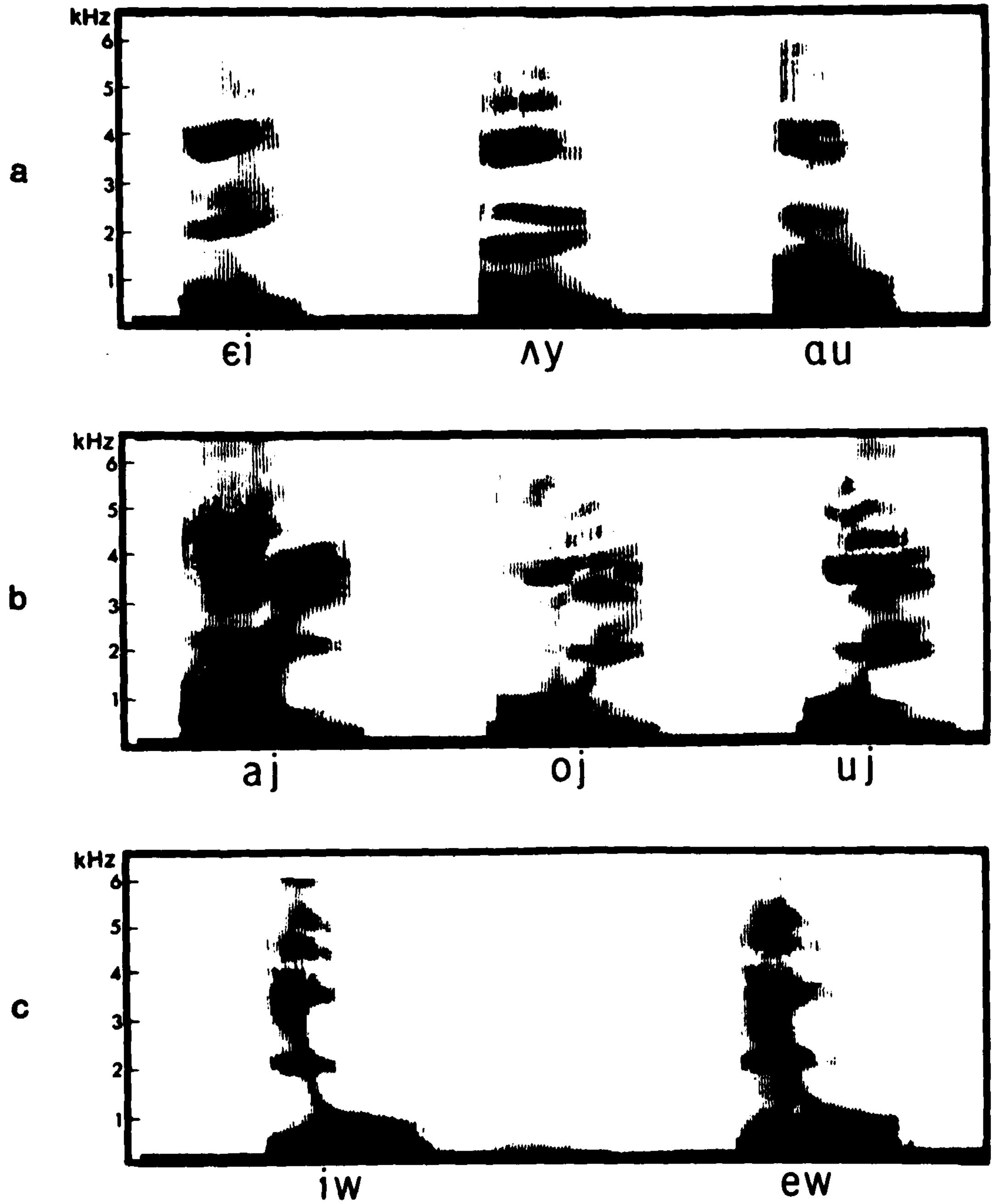


Figure 5. Single tokens of each of the diphthongs studied, pronounced in isolation. Note difference in the rate of formant frequency change in genuine diphthongs (a) and pseudo diphthongs (b and c).

consonantal) feature" (p. 288). A biphonemic interpretation only seems plausible for the pseudo diphthongs.

REFERENCE NOTE

1. Slis, I. H., & van Katwijk, A. F. W. Onderzoek naar de Nederlandse tweeklanken. Institute for Perception Research, Eindhoven, 1963, Report No. 31.

REFERENCES

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

Chiba, T., & Kajiyama, M. The vowel: Its nature and structure. Tokyo: Tokyo-Kaiseikan, 1941.

Cohen, A. Diphthongs, mainly Dutch. In L. Hammerich, R. Jakobson, & E. Zwirner (Eds.), Form and substance. Copenhagen: Akademisk Forlag, 1971, pp. 278-289.

Cohen, A., Ebeling, C. L., Eringa, P., Fokkema, K., & van Holk, A. G. F. Fonologie van het Nederlands en het Fries. The Hague: Martinus Nijhoff, 1959.

Collier, R., & 't Hart, J. The perceptual relevance of formant trajectories in diphthongs. In M. van den Broecke & V. van Heuven (Eds.), Studies for A. Cohen. Dordrecht: Foris Publications, in press.

Delattre, P. The physiological interpretation of sound spectrograms. Publications of the Modern Language Association, 1951, 66, 864-875.

Fant, C. G. M. Acoustic theory of speech production. (2nd Ed.) The Hague: Mouton, 1970.

Harris, K. S. Electromyography as a technique for laryngeal investigation. ASHA Reports, 1981, 11, 70-87.

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

Kakita, K. Activity of the genioglossus muscle during speech production: An electromyographic study. Unpublished doctoral dissertation, University of Tokyo, 1976.

Kakita, K., Hirose, H., Ushijima, U., & Sawashima, M. A preliminary report on the electromyographic study of the production of the Japanese semivowel /j/. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo), 1976, 10, 37-46.

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

Koopmans-van Beinum, F. J. Nog meer practische zekerheden. Nieuwe Taalgids, 1969, 62, 245-250.

Kuhn, G. M. On the front cavity resonance and its possible role in speech perception. Journal of the Acoustical Society of America, 1975, 58, 428-433.

Miyawaki, K., Hirose, H., Ushijima, T., & Sawashima, M. A preliminary report on the electromyographic study of the activity of lingual muscles. Annual Report of the Research Institute of Logopedics and Phoniatrics (University of Tokyo), 1975, 9, 91-106.

Mol, H. Fonetische steunpunten. Nieuwe Taalgids, 1969, 62, 161-167.

Pols, L. C. W. Spectral analysis and identification of Dutch vowels in monosyllabic words. Unpublished doctoral dissertation, University of Amsterdam, 1977.

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

Raphael, L. J., Bell-Berti, F., Collier, R., & Baer, T. Tongue position in rounded and unrounded front vowel pairs. Language and Speech, 1979, 22, 37-48.

Smith, T. St. J. A phonetic study of the function of the extrinsic tongue muscles. Unpublished doctoral dissertation, University of California, Los Angeles, 1971.

Stevens, K. N., & House, A. S. Development of a quantitative description of vowel articulation. Journal of the Acoustical Society of America, 1955, 27, 484-493.

Stevens, K. N., & House, A. S. An acoustical theory of vowel production and some of its limitations. Journal of Speech and Hearing Research, 1961, 4, 303-320.

't Hart, J. Fonetische steunpunten. Nieuwe Taalgids, 1969, 62, 168-174.

Van den Berg, B. Foniek van het Nederlands (1st Ed.). The Hague: Van Goor, 1959.

Zonneveld, W., & Trommelen, M. Egg, onion, ouch! On the representation of Dutch diphthongs. In W. Zonneveld, F. van Coetsem, & O. W. Robinson (Eds.), Studies in Dutch phonology (Dutch Studies, Vol. 4). The Hague: Martinus Nijhoff, 1980.

FOOTNOTES

[1]Possible occurrences of these diphthongs in Dutch words:

/ɛi/	kei	(pebble)	/aj/	maai	(mow)
/ʌy/	lui	(lazy)	/oj/	mooi	(beautiful)
/ɑu/	rauw	(raw)	/uj/	snoei	(trim)
			/ew/	leeuw	(lion)
			/iw/	nieuw	(new)

Another pseudo diphthong, /yw/ as in duw (push), was not included in our utterance set.

[2]Although /aj/ is said to begin with the low front vowel [a], the data we offer below imply a substantial back-to-front movement during this diphthong.

[3]Our subject, the senior author, speaks the Belgian variant of Standard Dutch.

[4]We recognize, of course, that more than four muscles are involved in positioning and shaping the tongue, and that the articulatory description provided here is, of necessity, a simplified one.

[5]We will not address the question of whether the genuine and pseudo diphthongs differ in maintaining harmony of lip position between starting and ending configurations.

[6]We are unable to compare the relative openness, or frontness, of the beginning of [ʌy] with [ʌ] because of the absence of this latter vowel in Dutch.

[7]This difference in the rate of change of the formant frequencies is perceptually less relevant than the correct timing of the onset of that change (Collier & 't Hart, in press).

[8]We should note that even in the case of [aj] our acoustic and EMG data indicate that [a] is articulated more similarly to back vowels, such as [ɑ], than to front vowels, such as [ɛ]. Indeed, the change in F_2 for [aj] is more than twice as great as the largest change in F_2 for the genuine diphthongs. Thus, a more accurate transcription of our subject's version would be [ɑːj].

RELATIONSHIP BETWEEN PITCH CONTROL AND VOWEL ARTICULATION*

Kiyoshi Honda

INTRODUCTION

It is widely recognized that phonatory functions of the larynx are primarily regulated by the intrinsic laryngeal muscles. The extrinsic muscles of the tongue and the larynx, however, play an essential role in ensuring a wide range of laryngeal function by directly and indirectly influencing the position of the hyoid-larynx complex and the intra-laryngeal configuration. These extrinsic muscles function, in addition, as speech muscles to produce articulatory gestures. Hence, articulation and phonation inevitably interact with each other.

The present study is focussed upon hyoid bone movement associated with pitch control and articulatory gestures. There is little information on the mechanism controlling hyoid bone movement in the literature. This may be due partly to the complexity of its supportive structures, and partly to the lack of interest engendered by its ambiguous function. There are more than ten pairs of muscles attached directly and indirectly to the hyoid bone. These muscles have links with articulatory organs such as the mandible and the tongue. In addition, the ligaments and membranes connecting the hyoid bone, the thyroid cartilage, and the surrounding tissues and organs to each other act like a network of springs. The hyoid bone, as a supportive structure of the larynx, is influenced by these forces, and its position is affected by both pitch control and articulatory gestures.

Pitch raising mechanisms have been attributed traditionally almost exclusively to cricothyroid activity, which creates an angular change between the cricoid and the thyroid cartilage. EMG studies of the extrinsic laryngeal muscles have been concerned with their effects on the tilt of the thyroid cartilage or the lowering of the entire larynx, even though the mechanism of larynx elevation is not clear. Recently, a few physiological studies have reported an association between geniohyoid activity and fundamental frequency (F_0). Erickson, Liberman, and Niimi (1977) note that geniohyoid activity during sentence reading with several different intonations is positively correlated with fundamental frequency and/or cricothyroid activity. Sapir, Campbell, and Larsen (1981) report, in an animal experiment using rhesus macaques, that electrical stimulation of the geniohyoid muscle causes a substantial increase of the voice fundamental frequency. In addition, some

*A version of this paper was presented at the Vocal Fold Physiology Conference, Madison, Wisconsin, May 31 - June 4, 1981.
Acknowledgment. This work was supported by NIH grant NS-13870.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

radiographic studies have noted a positive correlation between fundamental frequency and forward translation of the hyoid bone (Colton & Shearer, 1971; Sapir, 1978). These observations suggest that in high pitch the geniohyoid pulls the hyoid bone forward and thus helps to tilt the thyroid cartilage forward.

Figure 1 shows a schematic representation of the relevant anatomy. The role of the hyoid bone in the pitch control mechanism can be explained as follows. The effect of any forward shift of the hyoid bone is passed on to the thyroid cartilage and the intra-laryngeal tissue through the muscles and connective tissues: the thyrohyoid muscle, the lateral and median thyrohyoid ligaments, the hyoepiglottic ligament and the thyrohyoid membrane. The hyoid bone also functions to support the tongue base, and it moves with articulation. The posterior fibers of the genioglossus, whose action is to draw the tongue root forward, have some connections with the hyoid bone, and the effect of its contraction also moves the hyoid bone forward. The median fibrous septum, the hyoglossus muscle, and their related structures may also be involved in pulling the hyoid bone forward. Furthermore, the inferior fibers of the genioglossus, in addition to the posterior fibers, are inserted directly into the body of the hyoid bone (Miyawaki, 1974). Because of these connections, contractions of the geniohyoid and the genioglossus may tilt the thyroid cartilage forward and help increase the longitudinal tension of the vocal folds by drawing the hyoid bone forward.

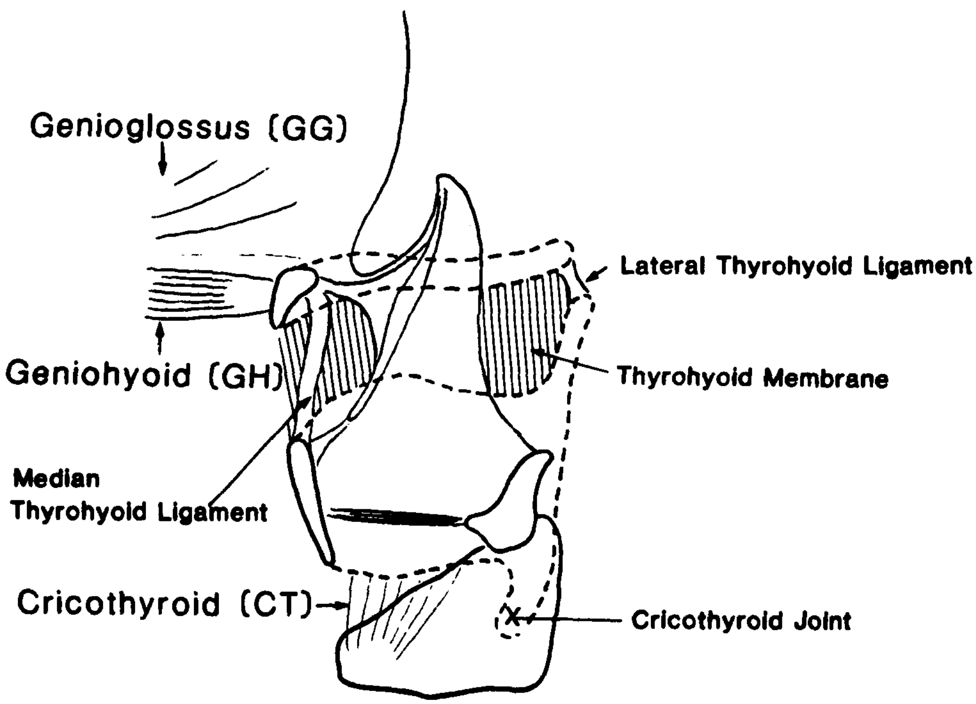


Figure 1. Schematic view of laryngeal framework

METHOD

Electromyographic (EMG) signals from some external laryngeal muscles and movement data of the hyoid bone were collected from a Japanese subject. The utterances used in this experiment were Japanese nonsense two-mora words that consisted of a combination of a high vowel /i/ and a low vowel /a/, with and without intervocalic /m/ (e.g., /ai/ and /ami/). These words were spoken in isolation with three different pitch accent patterns: flat (constant F_0), rising (low-to-high step), and falling (high-to-low step). This experiment was performed in two sessions. In the first session, EMG recording alone was performed for ten repetitions of the utterances so that ensemble averages could be calculated. In the second part, EMG recording and measurement of the hyoid bone movement were performed simultaneously, and analyzed separately for each token. Audio signals were used to extract pitch contours by computer using an auto-correlation method.

The EMG signals from the genioglossus, the geniohyoid, and the cricothyroid were used as data. Since, in the first part of the experiment, the data varied in timing, four tokens that have the most similar utterance timing were selected for ensemble averaging for each utterance type. Audio envelopes and the EMG signals from other muscles, the orbicularis oris, the anterior digastric, and the sternohyoid, were used as timing indicators for selecting these tokens. EMG recording was performed by insertions of paired hooked-wire electrodes, which were prepared by a modification of Miyata, Honda, and Kiritani's (1980) method: the insulation of the wires was thermally removed by an electrically heated nichrome string to obtain a relatively wide electrode area. Paired wires were glued together to stabilize inter-electrode distance. The length of exposed area was approximatory 1mm at the cut end of each wire, and the inter-electrode distance was about 1mm measured from edge to edge of insulation.

The movement of the hyoid bone was measured by an optical tracking system similar to Sel Spot (Lindholm & Oeberg, 1974). Figure 2 shows a schematic diagram of the measuring method. An infra-red LED was attached to the notched end of a plastic tube. The subject held the other end of the tube so that the notch remained fixed to the lower edge of the body of the hyoid bone. The LED is driven by current pulses from the main unit. A two-dimensional diode photo detector outputs currents corresponding to the position of the focussed light spot. The analog operational circuit of the main unit returns DC signals corresponding to the X and Y coodinates of the position of the LED.

RESULTS

EMG of the Geniohyoid and the Cricothyroid

The average EMG of the geniohyoid and the cricothyroid muscles in falling and rising accent patterns is shown in Figure 3. While the cricothyroid muscle shows consistent EMG activity with each pattern and shows no effect of articulation, the geniohyoid muscle has two components: continuous activity in high pitch and relatively low, transient activity in jaw opening. The activity of the geniohyoid associated with jaw opening tends to rise synergistically with the anterior digastric and the sternohyoid when pitch remains flat or rises with jaw opening (e.g., /ia/ and /i'a/), and rise after a

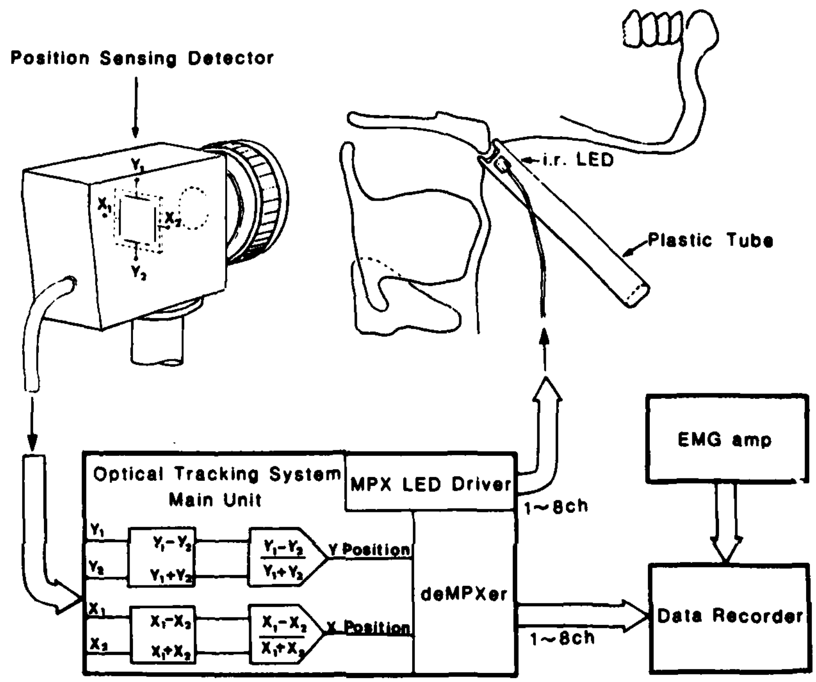


Figure 2. Method for measuring hyoid bone movement. The infra-red LED is driven by current pulses from the main unit, which can drive up to eight LEDs simultaneously by time multiplexing. The light beam from the LED is focused on the position sensing detector, which consists of a photo diode plate with registive surfaces and pairs of edge electrodes. The focused spot causes a depletion of the diode and induces pairs of currents on each surface toward opposite edges depending on the distance from each electrode to the spot. The analog operational circuit of the main unit converts each pair of currents into DC voltages corresponding to the X and Y coordinates of the position of the LED.

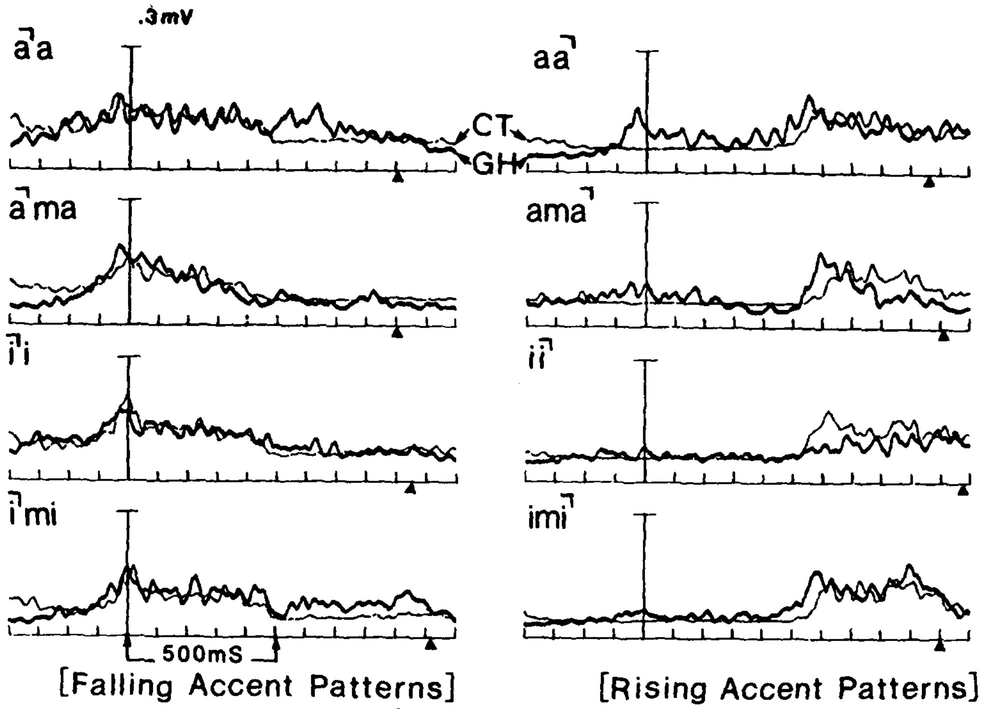


Figure 3. Comparison of average EMG activity of the geniohyoid (GH) and the cricothyroid (CT) in falling and rising accent patterns. Vertical lines indicate voice onset and triangles (▲) represent voice offset.

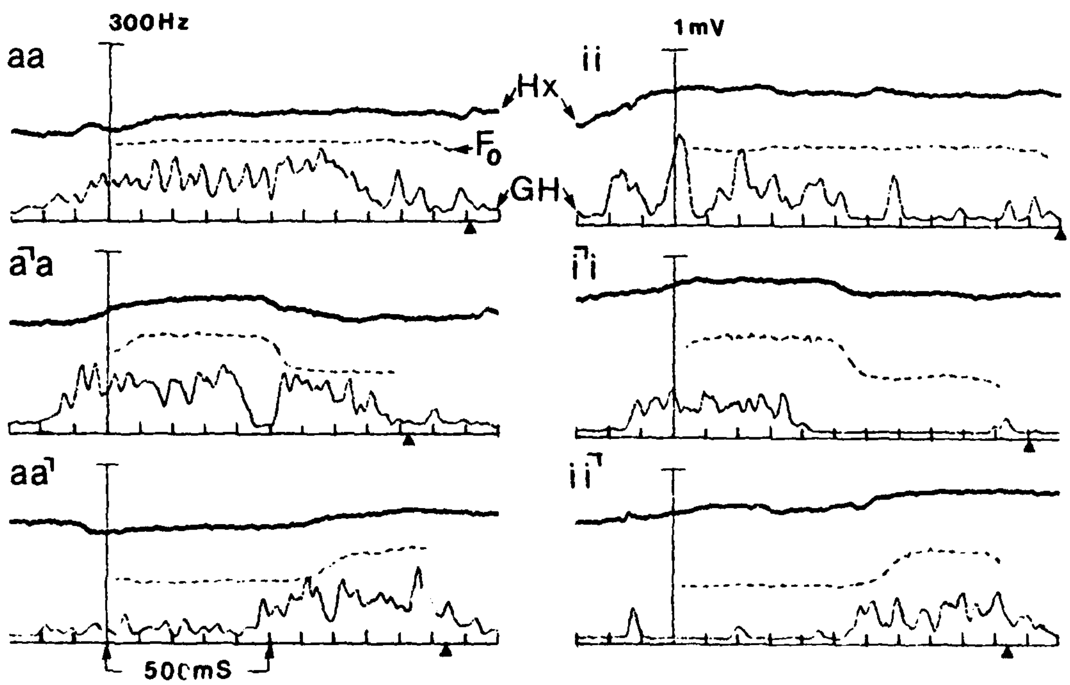


Figure 4. Hyoid bone movement (Hx), fundamental frequency (Fo) and EMG of the geniohyoid (GH) in different accent patterns. Positive slopes of the thick line represent forward movement of the hyoid bone in arbitrary units.

suppression associated with the peaks of the anterior digastric and the sternohyoid when the pitch falls with jaw opening (e.g., /'ia/). The overall pattern of geniohyoid activity resembles that of the cricothyroid, and does not appear to have a consistent correlation with vowel quality in the steady-state portion of the vowels. Both muscles show peak activity associated with voice onset in falling accent patterns, but, in rising accent patterns, the geniohyoid tends to start earlier than the cricothyroid.

These data suggest that the action of the geniohyoid is to draw the hyoid bone forward when the mandible is fixed, and help to depress the mandible when the hyoid bone is fixed. This muscle shows consistent activity with the cricothyroid during pitch change. However, in jaw opening, it appears that the geniohyoid acts cooperatively with other muscles to stabilize hyoid bone position. From the temporal relations between two muscles, it seems that the geniohyoid starts with the cricothyroid in voice initiation, and anticipates cricothyroid activity in pitch raising.

Movement of the Hyoid Bone

(a) In different accent patterns with the same vowels. Figure 4 shows single token data of horizontal movement of the hyoid bone, fundamental frequency, and EMG of the geniohyoid muscle in different accent patterns with the same vowels. While the position of the hyoid bone is stable during utterances with flat accent patterns, its movement follows the curves of the fundamental frequency in utterances with falling and rising accent patterns, moving forward in high pitch and backward in low pitch. Horizontal movement of the hyoid bone tends to precede the changes in fundamental frequency slightly. During falling and rising accent patterns, the EMG activity of the geniohyoid is consistent with pitch accent patterns. If the accent pattern is flat, its activity depends on jaw activity, probably compensating the effect of jaw opening on hyoid bone position.

(b) In vowel change with different accent patterns. Horizontal position of the hyoid bone changes with vowel quality. Figure 5 shows data for the utterances /ai/ and /ia/ with "flat" accent patterns.[1] The high-front vowel /i/ is accompanied by forward position of the hyoid bone and the low-back vowel /a/ is accompanied by back position. Figure 5 shows that the position of the hyoid bone is not affected by geniohyoid activity. In vowel articulation, hyoid bone movement is affected by the activity change of the tongue muscles, most significantly the posterior fibers of the genioglossus. The function of the genioglossus posterior is to raise the tongue dorsum for high vowels by drawing the tongue root forward. Thus, high vowels are associated with forward position and low vowels with back position of the hyoid bone due to anatomical connections with the tongue root. The low-back vowel /a/, however, probably involves other muscles to retract the tongue body, in addition to the lack of genioglossus activity. These may also affect hyoid bone position.

When there is both pitch and articulatory change, the position of the hyoid bone is affected by both. Figure 6 shows the data for the utterances /ai/ and /ia/ with two different accent patterns, falling and rising. In the utterances /'ai/ and /i'a/, the movement of the hyoid bone is nearly flat. The horizontal position of the hyoid bone is almost the same in high-pitched

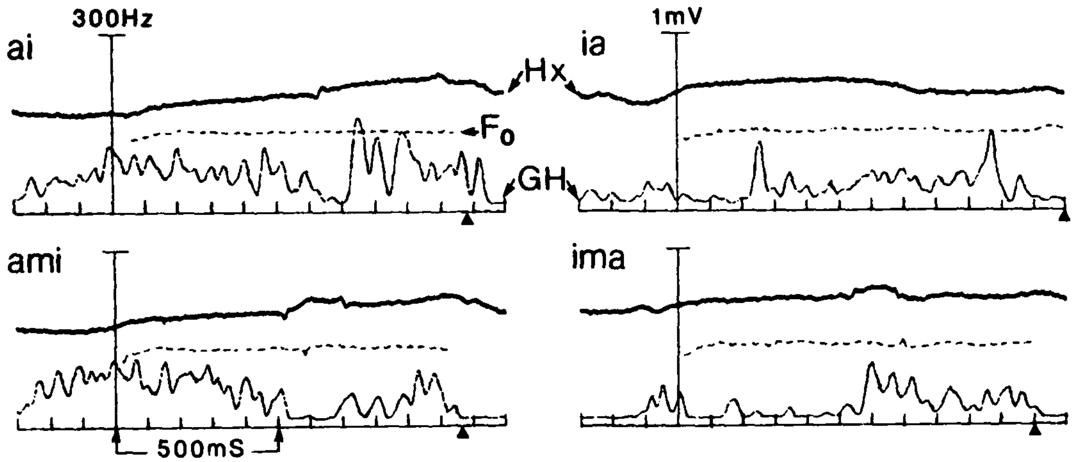


Figure 5. Hyoid bone movement (Hx), fundamental frequency (Fo) and EMG of the geniohyoid (GH) in vowel changes.

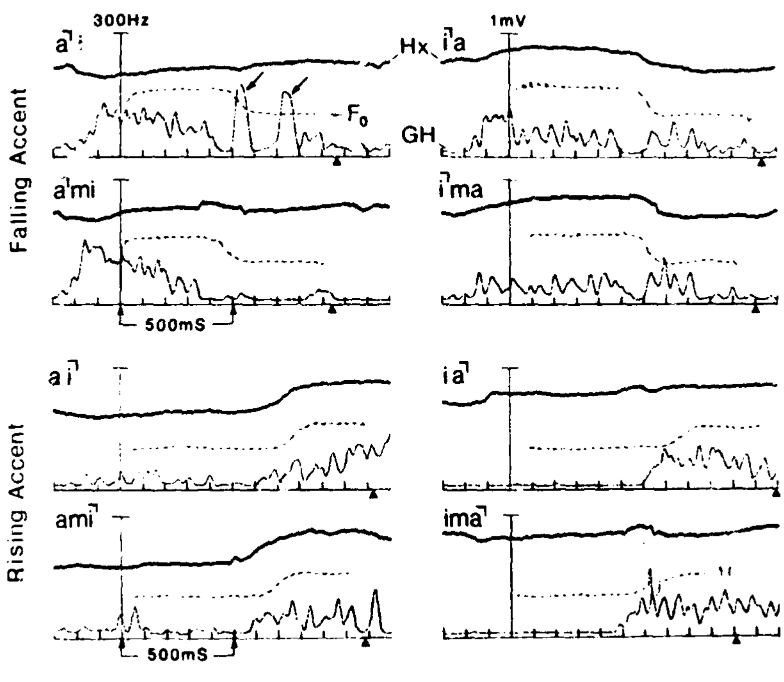


Figure 6. Hyoid bone movement (Hx), fundamental frequency (Fo) and EMG of the geniohyoid (GH) in both vowel and pitch change. The arrows (↙) indicate artifact potentials.

/a/ and low-pitched /i/, and the effects of pitch control and vowel articulation are counterbalanced. On the other hand, the utterances /a'i/ and /'ia/ show the maximum displaceuent of the hyoid bone. The effects of pitch control and vowel articulation reinforce each other in these utterances. The large displacement of the hyoid bone may be related to the fact that activity of the genioglossus increases in high pitch. Figure 7 shows average EMG of the genioglossus and the geniohyoid in the utterances /ai/ and /ia/ with different accent patterns. Genioglossus activity for the vowel /i/ is increased in high pitch compared with that in flat accent patterns. In this experiment, increased activity of this muscle in high pitch was observed only in the vowel /i/. However, Sawashima, Hirose, Honda, and Sugito (1980) note that the genioglossus muscle shows remarkable activity for high pitch in the vowel /a/. These differences seem to depend on the position of the electrode in the muscle, although differences in speaker may also be important. From these results, it is inferred that a high vowel in a stressed syllable has the maximum longitudinal tension of the vocal folds if other factors are the same.

(c) Vertical movement of the hyoid bone. In this experiment, vertical movements of the hyoid bone were also measured. In pitch change, there is a tendency for the hyoid bone to rise with fundamental frequency. In rising accent patterns (e.g., /i'i/ and /a'a/), the hyoid bone rises with fundamental frequency. However, in falling accent patterns, it does not consistently fall with fundamental frequency. With respect to articulation, its position is higher in the vowel /a/ than in the vowel /i/, in agreement with other studies (Menon & Shearer, 1971; Perkell, 1969). The extent of the vertical movement was found to be larger in vowel change than in pitch change. The hyoid bone, as a whole, moves forward and slightly upward (ventro-cranially) in pitch change and moves forward and downward (ventro-caudally) in vowel transition of the utterance /ai/.

DISCUSSION

The geniohyoid and the genioglossus muscles, in animals, function clearly as laryngeal elevators because of their vertical (cranio-caudal) insertion and because of direct connection between the hyoid bone and the thyroid cartilage, and they play an important role in swallowing (Hirano, 1975; Shin, Hirano, Maeyama, Nozoe, & Ohkubo, 1981). In humans, these muscles run rather horizontally and their action turns to pull the hyoid bone forward. Furthermore, larynx position is lowered and the pharyngeal cavity is elongated in humans. These changes in anatomical configuration increase the freedom of tongue movement, which is also ensured by the detachment of the hyoid bone from the thyroid cartilage. Thus, the separation of the tongue and the larynx provides the ability for a wider range of independent control over phonation and articulation. Still, there are interconnections between the tongue and the larynx, and articulatory movement of the tongue can influence phonatory function, and vice versa.

In this study, we are concerned with forward movement of the hyoid bone, its muscular control, and its effect on laryngeal functions, in particular voice pitch change. The results obtained in this experiment may be summarized as follows:

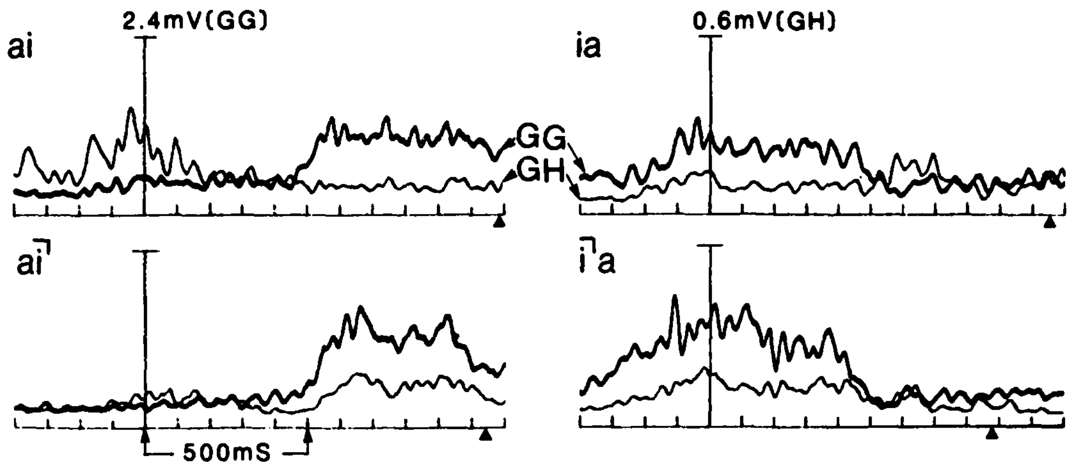


Figure 7. Average EMG of the genioglossus (GG) and the geniohyoid (GH). The genioglossus shows increased activity during high-pitched vowel /i/.

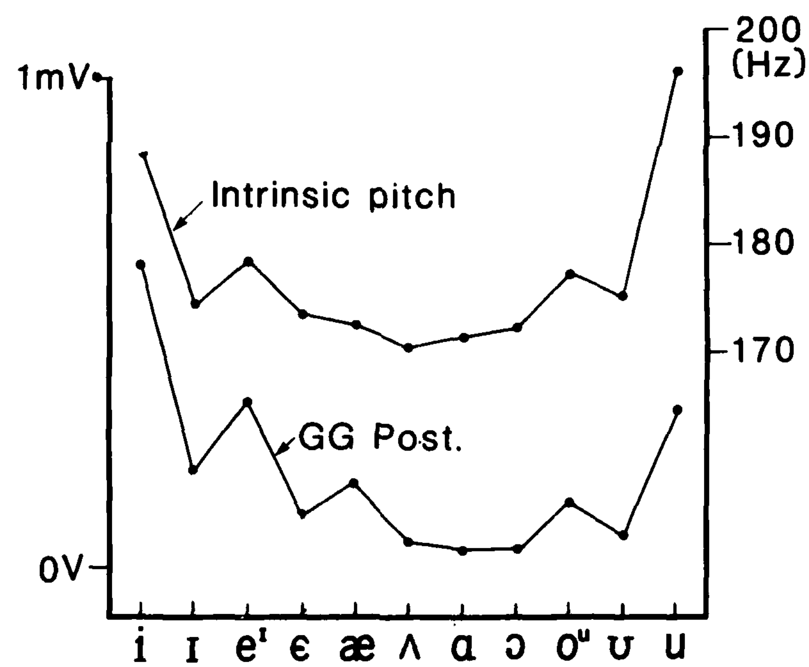


Figure 8. Intrinsic pitch (above) and EMG of the posterior genioglossus (below) in English vowels. In this figure, the data of intrinsic pitch are taken from Lehiste and Peterson (1961), and average fundamental frequencies of the vowels with preceding consonants /p/,/t/ and /k/ are shown. The EMG data were collected from a native speaker of American English. The peak values of the integrated and averaged signals during /əpVp/ utterances are plotted.

1. The geniohyoid muscle shows increased activity in pitch raising and produces forward translation of the hyoid bone.

2. Horizontal position of the hyoid bone is influenced by tongue root position, which is determined by the activity of the posterior fibers of the genioglossus. A high vowel has a forward position of the hyoid bone.

3. The effects of pitch control and vowel quality are superimposed to determine the overall pattern of the hyoid bone movement in utterances containing both pitch change and articulatory movement.

Considering their effects on the "external frame," it is likely that the geniohyoid and the genioglossus pull the hyoid bone and rotate the thyroid cartilage forward. Both muscles seem to participate in pitch raising by increasing the longitudinal tension of the vocal folds. This assumption suggests that the longitudinal tension of the vocal folds may be increased by forward shift of the tongue root to produce high vowels. This is related to the mechanism of the intrinsic pitch of the vowel.

It is generally acknowledged that there is a consistent relation between vowel quality and average fundamental frequency associated with it (Lehiste, 1970; Peterson & Barney, 1952). High (close) vowels such as /i/ and /u/ have higher fundamental frequency than low (open) vowels such as /a/ and /ae/. This phenomenon, the "intrinsic pitch of the vowel," tends to correlate with tongue height. If we assume active participation of the hyoid bone in the pitch raising mechanism, the intrinsic pitch is determined by the activity of the posterior fibers of the genioglossus. The relationship between the intrinsic pitch and the activity of the posterior fibers of the genioglossus is shown in Figure 8. The data for the intrinsic pitch in English were taken from Lehiste and Peterson (1961); the EMG data were obtained in a recent experiment at Haskins Laboratories. This figure shows that posterior genioglossus EMG activity and intrinsic pitch are grossly correlated. However, this relationship is less obvious for the vowel /ae/, which implies that other unknown mechanisms also exist.

In the present study, the effects of the extrinsic muscles of the tongue and the larynx are discussed in relation to movements of the external frame. However, these muscles also influence the intra-laryngeal configuration. The articulatory movements of the tongue may affect other intra-laryngeal events, such as the tension of the aryepiglottic folds via the "functional chain" described by Zenker (Zenker & Zenker, 1960; cited in Sonninen, 1968), or the vertical tension of the vocal folds (Ohala, 1977). Figure 9 summarizes the possible factors that can affect the tension of the vocal folds. The first factor is the force on the external frame as hypothesized in this study: forward movement of the hyoid bone rotates the thyroid cartilage forward. The second and the third factors are derived from the position of the epiglottis, which is determined by the positions of the tongue root and the hyoid bone. The tension of the aryepiglottic folds may apply a force to pull the apex of the arytenoid cartilage up and forward, although it is not clear whether its effect is to lengthen the vocal folds, enhance medial compression, or stabilize the position of the arytenoid cartilage. The vertical tension theory is based on the X-ray finding that the ventricular size is wider in high vowels than in low vowels. It is likely that movement of the epiglottis

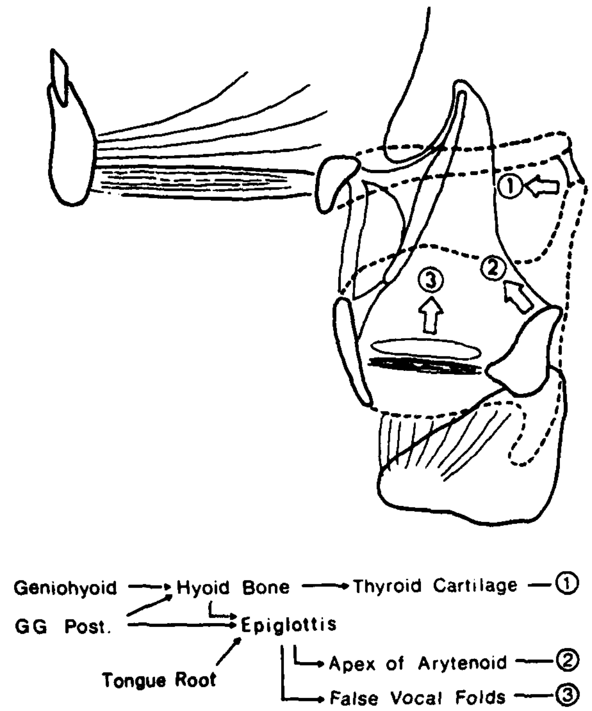


Figure 9. The possible effects of the tongue movement on the larynx. 1. Anterior pull of the thyroid cartilage. 2. Tension of the aryepiglottic folds. 3. "Vertical tension" of the vocal folds.

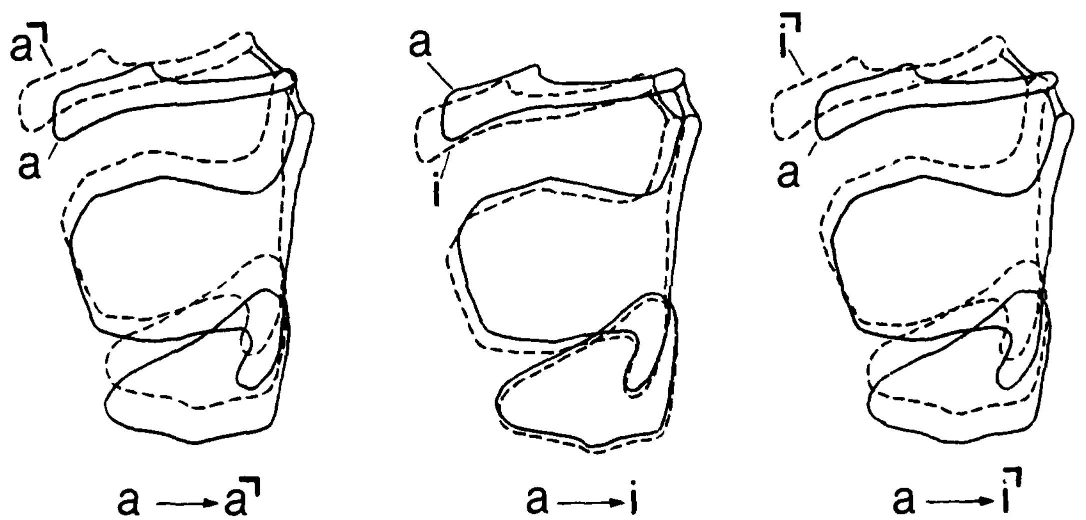


Figure 10. Postulated movements of the laryngeal framework. These figures illustrate the speculated laryngeal frame movements: the cricoid cartilage moves vertically, and the thyroid cartilage rotates around the cricothyroid joint. Activities of the cricothyroid and the thyrohyoid muscles are not considered.

increases the vertical tension of the intra-laryngeal tissue, but there is little physiological evidence on this point.

Vertical movements of the hyoid-larynx complex are associated with pitch change, and the larynx tends to rise with fundamental frequency. The effect of the vertical movement of the entire larynx on the external frame is not yet clear. However, the cricopharyngeus muscle, a sphinter of the esophageal orifice, may explain the relationship between vertical movement of the larynx and pitch change (Sonninen, 1956, 1968). When the cricopharyngeus is contracted, it produces a torque around the cricothyroid joint that rotates the posterior cricoid plate upward to reduce vocal fold tension, as long as the functional center of the cricothyroid joint does not change substantially. As larynx position deviates further from the neutral position towards the lower extreme of its total movement range, the effect of the cricopharyngeus becomes significant. The sternohyoid muscle, which is sometimes considered as a pitch lowering muscle, may realize this function by pulling the entire larynx downward.[2] However, during natural speech, its activity does not always show a close relation with the fundamental frequency, but shows consistency only at the lower extreme of pitch range (Sawashima, Kakita, & Hiki, 1973). The cricopharyngeus cannot easily explain the relationship between larynx elevation and pitch raising unless a considerable sliding of the cricothyroid joint is taken into account. (It may be reasonable to speculate that larynx elevation results from thyrohyoid activity to approximate the thyroid cartilage to the hyoid bone, so that hyoid bone movement may be transmitted more efficiently to the laryngeal framework.)

Vertical movements of the larynx are also associated with vowel articulation. There is a tendency for larynx position to be lower for high vowels than for low vowels, although this is a controversial point.[3] Larynx elevation in low vowels is suggested to be due to hyoglossus muscle activity. Contrarily, larynx depression might be caused indirectly by the transformation of tongue tissue. Contraction of the posterior fibers of the genioglossus raises the tongue dorsum and at the same time pushes the hyoid bone and the tongue base downward, since the insertion point of the posterior fibers of the genioglossus is just above the hyoid bone. The volume of the tongue mass being constant, decreases in the horizontal dimension of the tongue result in increase in its vertical dimension, both raising the dorsum of the tongue and lowering its base. This transformation of the tongue seems to be primarily relevant to vertical movement of the larynx in vowel articulation.

Figure 10 represents a summary of these various factors by showing postulated typical movement of the laryngeal framework associated with different pitches and vowels. The direction of the movement of each component is schematically represented: The thyroid cartilage is assumed to be suspended from the thyoid bone and the effects of the contractions of the cricothyroid and the thyrohyoid are not considered. The relative movement of the hyoid bone and the thyroid cartilage is supposed to be most restricted at the lateral thyrohyoid ligament. The information of hyoid bone tilt was obtained from x-ray films of a different subject. In summary, this figure suggests that pitch control and vowel articulation have an interactive effect on the laryngeal framework.

REFERENCES

Amenomori, Y. Some studies on the suspending mechanism of the larynx. Studia Phonologica, 1961, 1, 95-110.

Atkinson, J. E. Correlation analysis of the physiological factors controling fundamental voice frequency. Journal of Acoustic Society of America, 1978, 63, 211-222.

Colton, R. H., & Shearer, W. Hyoid position as a function of fundamental frequency in the modal and falsetto register. Technical Report (Department of Otolayngology Laboratories and Clinics, Experimental Phonetics Laboratory, State University of New York, Upstate Medical Center, Syracuse, New York), 1971, 9 (November).

Erickson, D., Liberman, M., & Niimi, S. The geniohyoid and the role of the strap muscles. Haskins Laboratories Status Report on Speech Research, 1977, SR-49, 103-110.

Ewan, W. G., & Krones, R. Measuring larynx movement using the thyroumbrometer. Jounal of Phonetics, 1974, 2, 327-355

Hirano, M. Phonosurgery: Basic and clinical investigations. Otologia (Fukuoka), Suppl. 1, 1975, 21, 263-295.

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

Lehiste, I., & Peterson, G. E. Some basic considerations in the analysis of intonation. Journal of the Acoustic Society of America, 1961, 33, 419-423.

Lindholm, L. -E., & Oeberg, K. E. T. An opto-electronic instrument for remote on-line movement monitoring. Biotelemetry, 1974, 1, 94-95.

Menon, K. M. N., & Shearer, W. M. Hyoid bone position during repeated syllables. Journal of Speech and Hearing Research, 1971, 14, 858-864.

Miyata, H., Honda, K., & Kiritani, S. Electrical chracteristics of wire electrodes for EMG measurement in speech research. Transaction of the Committee on Speech Research, Acoustic Society of Japan, 1980, S-79-81.

Miyawaki, K. A study on the musculature of the human tongue. Annual Bulletin of Research Institute of Logopedics and Phoniatrics, University of Tokyo, 1974, 8, 23-50.

Ohala, J. Speculation on pitch regulation. Phonetica, 1977, 34, 310-312.

Perkell, J. S. Physiology of speech production: Results ond implications of a quantitative cineradiographic study. Cambridge, Mass.: M.I.T. Press, 1969.

Peterson, G. E., & Barney, H. L. Control methods used in a study of vowels. Journal of the Acoustic Society of America, 1952, 24, 175-184.

Sapir, S. The motion of the hyoid bone and thyroid cartilage during changes in the voice fundamental frequency: A cinefluorographic pilot study. Masters thesis, University of Washington, 1978.

Sapir, S., Campbell, C., & Larson, C. Effect of geniohyoid, crocothyroid and sternothyroid muscle stimulation voice fundamental frequency of electrically elicited phonation in rhesus macaque. Laryngoscope, 1981, 91, 457-468.

Sawashima, M., Hirose, H., Honda, K., & Sugito, M. Relative timing of articulatory and phonatory controls in Japanese word accent: A study of nonsence two-mora words. Annual Bulletin of Research Institute of Logopedics and Phoniatrics, 1980, 14, 139-147.

Sawashima, M., Kakita, Y., & Hiki, S. Activity of extrinsic laryngeal muscles in relation to Japanese word accent. Annual Bulletin of Research Institute of Logopedics and Phoniatrics, University of Tokyo, 1973, 7, 19-25.

Shin, T., Hirano, M., Maeyama, T., Nozoe, I., & Ohkubo, H. The function of the extrinsic laryngeal muscles. Vocal Fold Physiology, Tokyo, University of Tokyo Press, 1981, 171-180.

Sonninen, A. The role of the external laryngeal muscles in length-adjustment of the vocal cords in singing. Acta Otolaryngologica, Supplement 130, 1956.

Sonninen, A. The external frame function in the control of pitch in the human voice. Annals of the New York Academy of Science, 1968, Vol. 155, Art. 1, 68-90.

Zenker, W., & Zenker, A. Uber die Regelung der Stimmlippenspannung durch von aussen eingreifende Mechanismen. Folia Phoniatrica, 1960, 12, 1-36.

FOOTNOTES

[1]In Japanese (Tokyo dialect), the "flat" accent pattern is phonetically realized as a low-to-high pattern in pitch, whereas when the first mora is accented (as in /'ima/), the pitch pattern is high-to-low. However, in this experiment, the flat accent pattern was produced as a physically monotonous pattern in fundamental frequency, neglecting such phonetic reality.

[2]The data in the literature are not in good agreement on sternohyoid activity in pitch change. Atkinson (1978) reports that the sternohyoid shows a high, consistent, negative correlation with fundamental frequency. However, such a good correlation has not been obtained in natural speech by many other investigators. Sawashima et al. (1973) note that these discrepancies may result from differences in the test words and individual differences in speech gesture. In the present experiment, the sternohyoid showed a transient activity in transition of pitch lowering and sporadic low-level discharges in the following steady-state period of low pitch. This EMG pattern indicates that sternohyoid activity is not monotonically related to pitch lowering. The transient activity of this muscle seems to be coupled with the offset of pitch raising muscles to guarantee the degree or the rate of pitch lowering.

[3]According to Perkell's data (1969), larynx height is inversely correlated with vowel height, and higher vowels have lower position than low vowels. However, Ewan and Krones' data (1974) show that larynx height is not consistently correlated with vowel height, and larynx height for the vowels /i/ and /a/ is sometimes reversed. In addition, Amenomori (1961) notes that hyoid bone position is influenced by pitch, vowel, and intensity. His data on Japanese vowels during sustained phonation indicate that the hyoid position is usually higher in the vowels /i/ and /e/ than /a/, /o/, and /u/; and sometimes lowest in the vowel /a/. Larynx height associated with vowel articulation is affected by several factors: head position, degree of jaw opening, neutral position of the larynx (the degree of laryngeal descent associated with age), mode of phonation, and so on. This implies that the articulatory system has redundancy. For example, tongue height for the vowel /i/ may be accomplished by a predominant contraction of either the genioglossus muscle or the mylohyoid muscle. Acoustic characteristics of the vowel /i/ can be enforced by widening the pharyngeal cavity using genioglossus activity or elevating the tongue base by mylohyoid activity.

LARYNGEAL VIBRATIONS: A COMPARISON BETWEEN HIGH-SPEED FILMING AND GLOTTOGRAPHIC TECHNIQUES*

Thomas Baer, Anders Löfqvist,+ and Nancy S. McGarr++

Abstract. This study was designed to compare information on laryngeal vibrations obtained by high-speed filming, photoglottography (PGG), and electroglottography (EGG). Simultaneous glottographic signals and high-speed films were obtained from two subjects producing steady phonation. Measurements of glottal width were made at three points along the glottis in the anterior-posterior dimension and aligned with the other records. Results indicate that PGG and film measurements give essentially the same information for peak glottal opening and glottal closure. The EGG signal appears to indicate vocal-fold contact reliably. Together, PGG and EGG may provide much of the information obtained from high-speed filming as well as potentially detect horizontal phase differences during opening and closing.

INTRODUCTION

High-speed films are most commonly used to monitor details of the glottal cycle. However, this technique is not only difficult and expensive, but it cannot be performed under natural conditions because a laryngeal mirror must be used. It is therefore desirable to use glottographic monitoring techniques such as photoglottography and electroglottography in place of the more difficult and more invasive technique of high-speed filming.

Photoglottography (PGG), or transillumination, is a semi-invasive technique for monitoring laryngeal behavior. Briefly, transillumination involves directing a light source toward the glottis from above or below and measuring

---

*To appear in the Journal of the Acoustical Society of America.
+Also Department of Phonetics, Lund University, Sweden.
++Also Center for Research in Speech and Hearing Sciences, Graduate School and University Center, The City University of New York.
Acknowledgment. We are indebted to our colleagues at the National Technical Institute for the Deaf for generously sharing their time and laboratories: Robert L. Whitehead, Chairperson of Communication Research Department, and Dale E. Metz whose special expertise as cameraman made the study possible. We are also grateful to Jo Estill and Rosemary Szczesiul for assisting in film measurements. 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-73 (1983)]

glottal width by monitoring the intensity of the light source on the other side (Sonesson, 1960). This technique has proven extremely useful for studying the coordination of glottal movements with those of the supralaryngeal articulators (Löfqvist & Yoshioka, 1981; McGarr & Löfqvist, 1982). For studies of phonation, PGG may supply measures of opening and closing time during the glottal cycle that may be clinically or pedagogically useful. Transillumination may also be useful for monitoring glottal activity preparatory to phonation or at its initiation. In comparison with filming--especially high-speed filming--transillumination can be performed more easily and under more natural conditions, including natural speech. Perhaps more importantly, the transillumination signal is more easily analyzed in parallel with other instrumental measures of vocal fold activity. In combination with these other measures, such as electroglottography and EMG, we believe transillumination can be valuable for examining the relationship between vibratory performance and acoustic output on one hand, and between glottographic signals and those such as EMG that can be obtained more invasively on the other hand.

Although photoglottography has been in practical use for several years, there is some question about its reliability and validity. Notably, many authors seem to agree that it can reliably indicate timing of peak glottal opening and closure, although there may be some uncertainty about the moment of glottal opening (Hutters, 1976; Kitzing & Sonesson, 1974). In studies comparing glottal area variations measured by transillumination and from high-speed films, Harden (1975) found good correspondence during most of the glottal cycle. However, in a similar study, Coleman and Wendahl (1968) challenged the reliability of the technique. The different results obtained in these two studies may be due to different apparatus and techniques employed in the two investigations. For example, differences in the size of the sensor and its placement may be significant. A comparison between glottal width measures obtained by transillumination and from simultaneous fiberoptic filming during voiceless obstruent production showed that temporal information supplied by the two methods was virtually identical (Löfqvist & Yoshioka, 1980; Yoshioka, Löfqvist, & Hirose, 1981). To compare smaller, faster movements during phonation, however, a high-speed filming system is required in place of the fiberoptic endoscope.

While photoglottography, or transillumination, carries information about the pattern of glottal opening, electroglottography (EGG) is thought to convey information about the patterns of vocal fold contact. Briefly, the technique involves the transmission of an electrical field between electrodes placed bilaterally on the neck of the subject so that the electrical impedance is expected to vary as a function of the degree of vocal fold contact. That is, impedance should decrease as the area of vocal fold contact increases, other factors remaining the same. While it is clear that the pattern of electroglottographic signals is related to the patterns of laryngeal vibrations, there has been some disagreement whether the EGG signal accurately represents vocal fold contact area. Most studies indicate good agreement between apparent vocal fold contact and deflections of the EGG signal, with either normal (Baer, Titze, & Yoshioka, in press; Childers, Smith, & Moore, in press; Fant, Ondráčková, Lindqvist, & Sonesson, 1966; Fourcin, 1974; Kitzing, 1977) or excised (Lecluse, Brocaar, & Verschurre, 1975) larynges. On the other hand, Smith (1981) argues that the EGG registers acoustic and mechanical effects and that the conventional interpretation of the EGG signal is untenable. This evidence is, however, unconvincing and not very well documented. We thus believe that the conventional interpretation is still valid until disproven in a more convincing way.

In general, the EGG and PGG signals provide information about complementary parts of the glottal cycle--PGG about the open period and EGG about the closed period. As noted by Rothenberg (1981), however, the glottis rarely either opens or closes abruptly over its entire length. Rather, for part of the cycle, the folds are likely to be in contact or separated over only part of their length. Thus, EGG and PGG signals are likely to overlap. Baer et al. (in press) argued that by obtaining both glottographic signals in parallel, and observing the overlap, the usefulness of each is increased because horizontal phase differences can be detected. A comparison between high-speed film and these measures is still needed to validate this assertion, however.

It therefore seemed appropriate to perform a validation study using our own equipment and techniques for transillumination and electroglottography in collaboration with the high-speed filming system provided by colleagues at the National Technical Institute for the Deaf. Specifically, the validity of glottographic techniques, namely photoglottography (PGG) and electroglottography (EGG), are examined to assess comparable information available in high speed films.

METHOD

The subjects were one female and one male with no evidence of laryngeal pathology. Because of the requirements for effective glottal illumination, each of the subjects was asked to produce steady phonation of the vowel /i/.

During these productions, high speed laryngeal films at 4000 frames/sec were taken using procedures described by Metz, Whitehead, and Peterson (1980). Briefly, this system provides a xenon arc light source coupled with an optical system to project a high intensity light beam on the vocal folds. Reduction of infra-red and ultra-violet radiation in the light source is accomplished by filtering. The cold light is then projected paraxial to the camera lens to intersect on a laryngeal mirror positioned in the oropharynx of the subject. During the positioning and filming, the subject was able to view the vocal folds by means of extrinsic mirrors mounted on the equipment housing. Similarly, the view of the vocal folds could be monitored throughout the filming by means of a reflex viewfinder installed on the camera lens.

High quality acoustic recordings were obtained at the time of the filming. The microphone was positioned on the shaft supporting the laryngeal mirror so that the subject maintained a lip-to-microphone distance of about 7 cm. The acoustic propagation delay between the glottis and the microphone was thus expected to be about 0.7 msec. Noise from the camera and optical-filming system was virtually eliminated since the subject was isolated in a sound treated room separate from the equipment.

Glottographic signals--transillumination and electroglottography--were obtained simultaneously with the high-speed films. Light from the filming system passing through the glottis was sensed by a phototransistor placed on the surface of the neck just below the cricoid cartilage and coupled to the skin by a light-tight enclosure. Electroglottographic signals were obtained from one subject (BW) using the FJ Electroglottograph, and from the other subject (KH) using the Fourcin Laryngograph. According to Lecluse et al. (1975), there is no substantial difference between the signals recorded with those two instruments. The electrodes were placed on the neck at the level of the thyroid prominence. All glottographic signals were recorded on FM channels of an instrumentation tape recorder with a bandwidth of 2.5 kHz. Audio and timing codes were recorded on parallel direct channels. The timing codes were also recorded photographically on the film and were subsequently used for synchronization.

Using a computer-assisted measuring system, frame-by-frame measurements were made from the films during those portions where the film speed was constant at about 4000 frames/sec. Measures of glottal width (WID) were made at the widest point along the anterior-posterior dimension of the glottis for each frame for purposes of comparison with the other glottographic records. Three additional measures of glottal opening were made along the anterior-posterior dimension as follows. The first (ANT) was made as close to the anterior commissure as possible. Since the view of the anterior commissure was sometimes blocked, the exact location of the point used for measurements differed slightly between films. The second measurement (MID) was made in the middle of the membraneous glottis, and the third (POS), close to the vocal processes.

Audio and glottographic signals as well as timing codes were sampled and digitized at 10K samples/sec. Records from each of these were aligned with the film measurements.

RESULTS

Figure 1 shows data for about 3 cycles of steady phonation at 145 Hz for the male speaker (KH). Records are, from top to bottom, the film measurements, photoglottography (PGG), electroglottography (EGG), and the audio signal, respectively. First, measures of glottal width from the films and transillumination (PGG) are shown to be practically identical. Both signals produce the same measures of onset (line A), peak glottal opening (line B), and glottal closure (line C). The EGG signal is plotted with increasing transconductance upwards. As expected, the EGG signal is complementary to the other records. Deflections in the EGG signal correspond roughly with glottal closure indicated by the other two methods. Due to technical problems, the EGG signal, for this subject, is somewhat noisy. Simultaneous audio has been sampled with pre-emphasis and has been shifted by 0.7 msec to compensate for the delay due to acoustic propagation from the glottis to the microphone. It can be noted that acoustic excitation appears to correspond with the end of the open period.

Looking in more detail, deflection in the EGG signal occurs slightly before the glottis is completely closed, as evidenced in the film records and the PGG signals. Peak deflection, corresponding to maximum area of contact, appears to occur about the moment of glottal closure. In examination of the films, the period of overlap in the three records corresponds to the interval when the region of contact between the folds moves from the anterior-posterior ends towards the center for this speaker. As indicated by line D, the descent

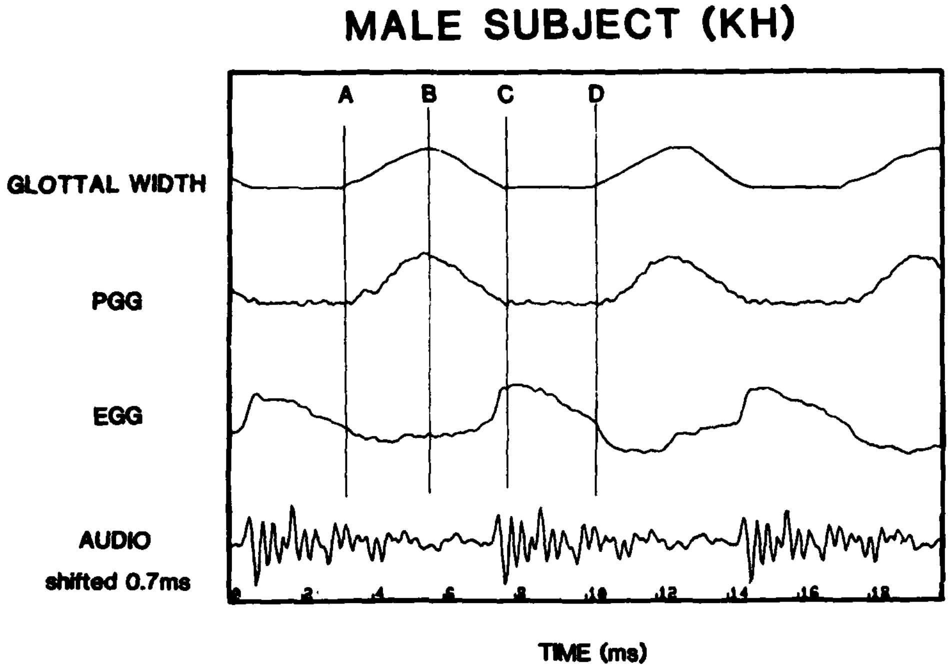


Figure 1. Results for subject KH. The curves represent, from top to bottom, glottal width measured from film, photoglottogram, electroglottogram, and audio signal.

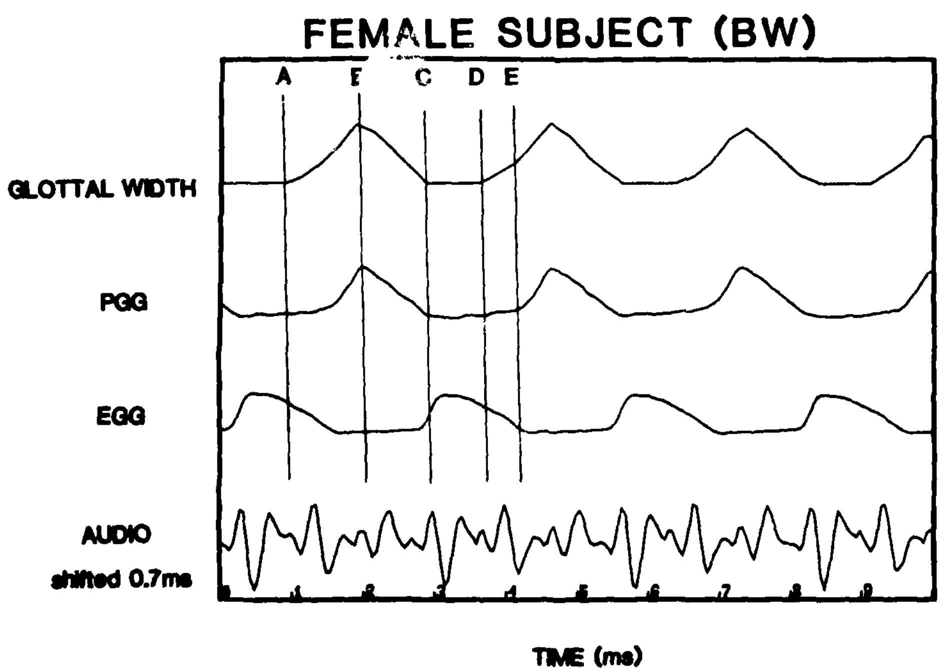


Figure 2. Results for subject BW. Curves as in Figure 1.

of the EGG becomes rapid at the point of glottal opening, producing a knee in the curve. Examination of the film shows that glottal opening propagates from the center to the anterior-posterior ends during the interval between the knee in the EGG curve and its return to baseline (cf. also Figure 3 below).

Figure 2 shows about 4 cycles of steady phonation at 250 Hz. for the female speaker (BW). The moments of opening (lines A and D), peak opening (line B), and glottal closure (line C) as indicated by the film measurements are marked. As with the other speaker, the moments of peak glottal opening (line B) and glottal closure (line C) indicated by the film records and PGG are similar. However, the correspondence between the film records and PGG for this speaker is more subtle. That is, the relative slope of glottal opening in the interval D-E, is greater when measured by glottal width of the films than when indicated by PGG. In the PGG signal, the onset is so gradual that it is difficult to identify a single point as the moment of opening. Further, the EGG signal does not show a knee as in Figure 1. Thus, correspondence between the film and PGG, as well as PGG and EGG at opening, indicates that the glottal opening was gradual and showed large horizontal phase differences in these records (cf. also Figure 4 below). This gradual opening could explain the absence of the "knee" in the EGG. Again there is acoustic excitation at the end of the open period.

Figure 3 shows the glottograms and the three measurements from the film (ANT, MID, and POS, respectively), as well as the measures of glottal width (WID) for speaker KH. From the film measures, two observations are apparent. First, the glottis does not open simultaneously along its entire length. Opening occurs slightly earlier in the medial region, and then propagates to the anterior and posterior ends. Glottal closure, on the other hand, occurs almost simultaneously along the entire length of the glottis. Second, the relative duration of the closed phase of the glottal cycle is longer anteriorly than posteriorly. The transillumination signal reflects the longer closed phase, and corresponds fairly well to the rise measured in the ANT portion of the film measures.

This correspondence is again illustrated in Figure 4 for speaker BW. In the photoglottographic signal, the lower portion of the trace begins to rise at about the same time as the trace in the ANT film record. Unlike speaker KH, opening of the glottis occurs slightly earlier in the anterior and posterior portions than in the medial. For this speaker, the anterior part of the glottis was not visible. The film image suggested that the opening was occurring earlier in the anterior portion than was reflected in the film measures. However, both speakers are alike in that glottal closure again occurs almost simultaneously along the entire length of the folds as shown across all of these measures.

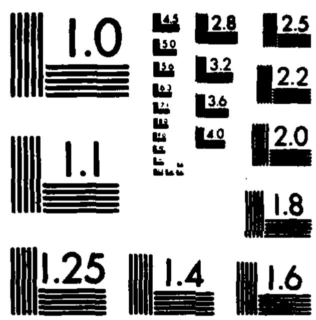
DISCUSSION

The results concerning the reliability of transillumination confirm that the PGG and film measures give essentially the same information about peak glottal opening and glottal closure in normal phonation. We also confirm the observations of other investigators, in that there is more uncertainty about the moment of glottal opening, and this uncertainty appears to arise from the fact that glottal opening is more gradual than glottal closure. It is well

AD-A129 713 STATUS REPORT ON SPEECH RESEARCH A REPORT ON THE STATUS AND PROGRESS OF S..(U) HASKINS LABS INC NEW HAVEN CT M STUDDERT-KENNEDY ET AL. 1983 SR-73(1983) 4/4

UNCLASSIFIED N00014-83-C-0083 F/G 6/16 NL

END
DATE
FILMED
7 83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

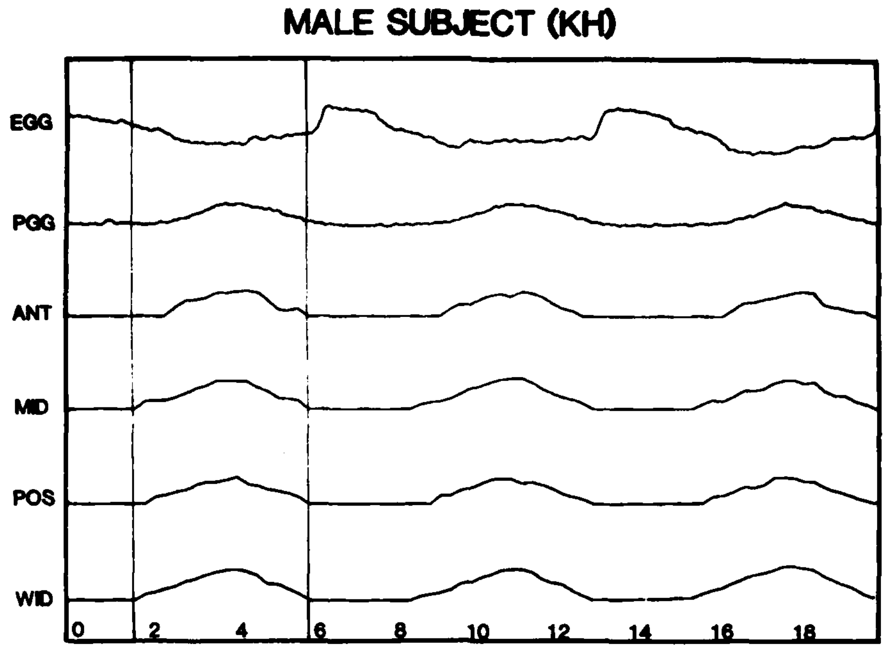


Figure 3. Subject KH, comparison between glottograms and glottal opening measured at different points along the glottis. The curves represent, from top to bottom, EGG, PGG, ANT, MID, POS, WID.

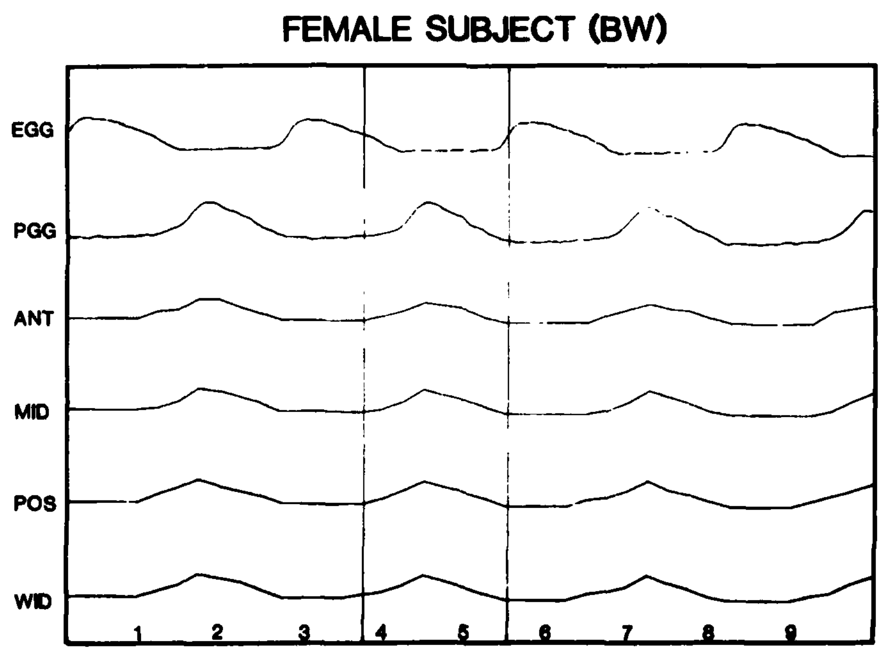


Figure 4. Subject BW, comparison between glottograms and glottal opening measured at different points along the glottis. Curves as in Figure 3.

known that the depth of glottal closure is quite small just prior to opening, while it becomes quite large immediately after closure. There also tend to be greater horizontal phase differences during opening than closing. "Opening" therefore occurs at different times along the anterior-posterior extent of the glottis.

Concerning the relationship between photoglottography and high-speed film measurements, it appears that the PGG signal can be thought of as representing a weighted sum of the widths along the length of the glottis. The weighting function depends on the location of both the light source and sensor with respect to the glottis. When the weights are high near the portion of the glottis that opens first, the agreement is better than when the weights are relatively low. We believe that the weighting functions in our experiment differed for the two subjects. Thus for subject KH, the PGG signal was in agreement with the opening measured at the anterior portion of the glottis. For subject BW, on the other hand, the PGG appeared to be relatively insensitive to the opening movement at the anterior and posterior ends, and the slope of the PGG signal thus increases after the mid portion of the glottis opens.

Considering the EGG signal, its correspondence with other measures of glottal activity appears to confirm its validity as an indicator of vocal-fold contact. Although it is not possible to obtain independent measures of vocal fold contact area, it is plausible that the EGG represents a measure of this quantity. The EGG signal reaches peak amplitude at about the moment of glottal closure indicated by the other measures, suggesting that the depth of glottal contact is maximum at this time. The rate of deflection of the EGG signal just prior to this maximum is very sharp, and it occurs over an interval that is comparable to the interval between film frames (cf. Childers et al., in press). This aspect of the EGG signal agrees with the interpretation that glottal closure is quite abrupt and demonstrates small horizontal phase differences. The EGG signal is also consistent with the notion that glottal opening is more gradual in both the vertical and horizontal dimensions. For the female subject, glottal opening cannot be clearly identified in the EGG waveform; for the male subject, it corresponds only to a mild increase in the rate-of-fall of the curve. A more gradual opening of the glottis for female subjects has also been reported by Kitzing and Sonesson (1974).

In conclusion, glottographic signals appear to be capable of supplying much of the significant information available in high-speed films. In comparison, films not only provide measures of glottal area, but also the distribution of width along the glottis. However, filming procedures are prohibitively difficult and the introduction of the laryngeal mirror for this procedure may have some effect on the phonations that are produced. While the glottographic techniques we have employed cannot detect the distribution of width along the glottis, they can be used to detect the presence of horizontal phase differences during opening and closing and can be used under nearly natural speaking conditions. It appears, therefore, that simultaneous photo- and electroglottographic signals can be used to great advantage in studies of voice production for monitoring the patterns of laryngeal vibrations.

REFERENCES

Baer, T., Titze, I. R., & Yoshioka, H. Multiple simultaneous measures of vocal fold activity. Proceedings of the Vocal Fold Physiology Conference, Madison, Wisconsin, in press.

Childers, D. G., Smith, A. M., & Moore, G. P. Relationships between electroglottograph, speech and vocal cord contact. Folia Phoniatrica, in press.

Coleman, R., & Wendahl, R. On the validity of laryngeal photosensor monitoring. Journal of the Acoustical Society of America, 1968, 44, 1733-1735.

Fant, G., Ondráčková, J., Lindqvist, J., & Sonesson, B. Electrical glottography. Quarterly Status and Progress Report (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm), 1966, 4/1966, 15-21.

Fourcin, A. Laryngographic examination of vocal fold vibration. In B. Wyke (Ed.), Ventilatory and phonatory control systems. London: Oxford University Press, 1974, 315-333.

Harden, R. J. Comparison of glottal area changes as measured from ultrahigh-speed photographs and photoelectric glottographs. Journal of Speech and Hearing Research, 1975, 18, 728-738.

Hutters, B. Problems in the use of the photoelectric glottograph. Annual Report, Institute of Phonology (University of Copenhagen), 10, 1976, 275-312.

Kitzing, P. Methode zur kombinierten photo- und elektroglottographischen Registrierung von Stimmlippenschwingungen. Folia Phoniatrica, 1977, 29, 249-260.

Kitzing, P., & Sonesson, B. A photoglottographic study of the female vocal folds during phonation. Folia Phoniatrica, 1974, 26, 138-149.

Lecluse, F. L. E., Brocaar, M. P., & Verschurre, J. The electroglottography and its relation to glottal activity. Folia Phoniatrica, 1975, 27, 215-224.

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

Löfqvist, A., & Yoshioka, H. Interarticulator programming in obstruent production. Phonetica, 1981, 38, 21-34.

McGarr, N. S., & Löfqvist, A. Obstruent production by hearing-impaired speakers. Journal of the Acoustical Society of America, 1982, 42, 34-72.

Metz, D. E., Whitehead, R., & Peterson, D. An optical-illumination system for high-speed laryngeal cinematography. Journal of the Acoustical Society of America, 1980, 67, 719-720.

Rothenberg, M. Some relations between glottal air flow and vocal contact area. In C. L. Ludlow & M. O. Hart (Eds.), ASHA Reports: Proceedings of the Conference on the Assessment of Vocal Pathology, 1981, 11, 88-94.

Smith, S. Research on the principle of electroglottography. Folia Phoniatrica, 1981, 33, 105-114.

Sonesson, B. On the anatomy and vibratory pattern of the human vocal folds. Acta Oto-laryngologica Supplement, 1960, 156, 1-80.

Yoshioka, H., Löfqvist, A., & Hirose, H. Laryngeal adjustments in the production of consonant clusters and geminates in American English. Journal of the Acoustical Society of America, 1981, 70, 1615-1623.

"COMPENSATORY ARTICULATION" IN HEARING IMPAIRED SPEAKERS: A CINEFLUOROGRAPHIC STUDY*

N. Tye,+ G. N. Zimmermann,+ and J. A. Scott Kelso++

Abstract. Data from three hearing-impaired subjects were compared to data from three hearing subjects to study the effect of constraining the jaw during speech on tongue shape and position for the vowels /i/, /ae/, and /u/. The results showed that although the three hearing-impaired speakers produced more variable tongue shapes and positions in both bite-block and nonbite-block conditions, the bite block had little effect in altering the areas of maximum constriction between the tongue dorsum and maxilla associated with the vowels studied. Two of the hearing-impaired speakers showed less differentiation in tongue shape and position for the vowels /u/ and /ae/ in both jaw-fixed and jaw-free conditions. A third hearing-impaired speaker differentiated the vowels, but the tongue positions observed were different from those of normal hearing speakers. The bite block was shown to have no systematic effect on intelligibility for any of the hearing-impaired speakers. These findings are interpreted in terms of current thinking on sensorimotor integration and movement control with particular reference to "target-based" theories.

INTRODUCTION

A case can be made that the absence or loss of auditory information produces effects on specific articulators and kinematic parameters during speech production. In a recent study of movement kinematics, Zimmermann and Rettaliata (1981) found that an adventitiously deaf speaker showed less distinctive tongue shapes for vowels than expected, when articulatory patterns were viewed relative to a mandibular reference. These findings suggested that the loss of auditory information may lead to a breakdown in the coordination of the tongue dorsum with other structures, and in the timing relations

---

*A similar version will appear in Journal of Phonetics.

+University of Iowa.
++Also University of Connecticut.

Acknowledgment. This research was supported by PHS Grant 07555-10 to the University of Iowa, and NINCDS Grant NS-13617 and BRSG Grant RR-05596 to Haskins Laboratories. We thank Nancy McGarr, Betty Tuller, and especially Katherine Harris for comments on an earlier version, and Pamela Mueller for the preparation of this manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

between voicing and movement onset in a vowel-consonant gesture. Results consistent with these conclusions have been reported by Monsen (1967), Hudgins and Numbers (1942), and McGarr and Harris (1980; see also Osberger & McGarr, 1982, for review). Emerging from such work is a theme that the deaf, who may be deficient in tongue dorsum positioning, rely more heavily on jaw displacement to distinguish between vowels than do normal hearing speakers who display greater flexibility in tongue shaping and movement. If the hearing impaired do not (or cannot) distinguish between vowels on the basis of tongue shapes or movements, but do rely on the jaw for their attempts at vowel production, then it is possible that constraining the jaw, say, by a bite block, would lead to differences in vocal tract shapes and deficits in vowel intelligibility compared to conditions in which the jaw is free to vary.

The study of bite-block speech in the hearing impaired that we undertake here not only allows a test of the foregoing hypothesis, but also may have significant import with regard to recent theorizing in the area of speech production. For example, a principal assumption of contemporary models is that articulatory goals are defined in terms of "targets" of some sort. Though the exact nature of the "targets" has been left vague in most discussions of speech production for a variety of reasons,[1] there is increasing consensus that targets have an auditory basis. For example, Ladefoged, DeClerk, Lindau, and Papcun (1972) suggest that a speaker "...may be able to use an auditory image to arrive at a suitable tongue position" (p. 73). More recently, MacNeilage (1980) has also opted for the auditory nature of "targets," mainly because the acoustic properties of sound are "obviously primary" sources of goals for acquisition of speech sounds. Finally, Gay, Lindblom, and Lubker (1981), following an X-ray examination of bite-block vowels, define the "neurophysiological representation of a vowel target...*in terms of area function related information...specified with respect to the acoustically most significant area function features, the points of constriction along the length of the tract*" (p. 809, italics theirs). According to Gay et al. (1981), their results support a kind of "indirect auditory targeting."

Few would argue the importance of auditory information for speech production, particularly at the acquisition stage (see Pick, Siegal, & Garber, 1982, for review). We ask, however, whether auditory targets (direct or not) are a *necessary* requirement for a talker's ability to adjust to novel contextual conditions. Note that this is not the same question that has been addressed regarding the role of auditory information in the ongoing *control* of articulators. That talkers can adjust the articulators almost immediately, as revealed in normal formant patterns at the first glottal pitch pulse, seems to negate a short-term auditory regulatory role (e.g., Lindblom & Sundberg, 1971). The issue we address here, however, is whether the "target" itself must be auditory in nature.

In the present study we examine, via cinefluorographic and perceptual analysis, the production of vowels in one congenitally and two adventitiously deaf speakers. Overall, we show not only that the hearing impaired "compensate" under the novel conditions created by a bite block but also that intelligibility is relatively unaffected. These data suggest that "auditory representations" of the kind recently proposed in the literature are not a necessary condition for immediate adjustment. Nor, we suspect, are "auditory targets" a sufficient explanation for the phenomenon because they ignore the problem of how a group of muscles might actually attain the so-called "target" positions or points of maximal constriction along the vocal tract. We take these data to offer an alternative proposal that draws on recently emerging concepts in the motor control literature. The latter recognize natural, dynamic properties such as damping and stiffness that are inherent in neuromuscular control systems. Typically, muscle-joint linkages are viewed as dynamically similar to a (nonlinear) mass-spring with controllable equilibrium states. The central idea, promoted by a number of authors (e.g., Bizzi, Dev, Morasso, & Polit, 1978; Fel'dman, 1966, 1980; Fel'dman & Latash, 1982; Kelso, 1977; Kelso & Holt, 1980), is that a system of muscles whose equilibrium lengths are specifiable will achieve and maintain desired configurations when the muscle-generated torques sum to zero. Such a system exhibits the characteristic of equifinality (von Bertalanffy, 1973) in that desired "targets" may be reached from different initial conditions and in spite of unforeseen perturbations encountered during the movement trajectory (cf. Kelso, Holt, Kugler, & Turvey, 1980, for review). This view leads to an interesting, but opposite prediction from the one based on earlier kinematic work on the hearing impaired (Zimmermann & Rettaliata, 1981); namely, that the tongue dorsum will reach similar final configurations *regardless of whether* the jaw is constrained by a bite block or not.

METHODS

Subjects

A 35-year-old, adventitiously deaf male (S1), a 24-year-old congenitally deaf female (S2), and a 34-year-old adventitiously deaf male (S3) served as subjects. S1 was diagnosed as having a profound, bilateral, sensorineural hearing impairment. He had suffered a progressive hearing loss beginning at age 12. S2 was diagnosed as having a bilateral, congenital, sensorineural hearing loss. She has a moderate-to-severe loss at 250 Hz and a profound loss at 500-8000 Hz. A hearing deficit for S3 was first reported when he was 18 months old. He has since been diagnosed as having a profound, bilateral, sensorineural loss at 250-8000 Hz.

Three hearing adults, two males (N1 and N2) and one female (N3) also served as subjects. These subjects served in an earlier collaborative study. Preliminary data have been reported by Kent, Netsell, and Abbs (Note 1).[2]

Speech Task

S1 was tested approximately one year before S2 and S3. Two different speech samples were obtained. S1 uttered the vowels (/i,u,ae/) embedded in the context /h_d/ or /h_t). S2 and S3 uttered the vowels (/i,u,ae/) in isolation.[3] The subjects were instructed to read the sample at a normal conversational rate. S1 read the sample a total of three times, two readings with no bite block and one reading with the bite block. S2 and S3 each made two readings with the bite block and two without it. The hearing subjects, N1, N2, and N3, read the sentence "You heap my hay high happy." Each subject read this sentence twice in each condition.

Apparatus

Cinefluorography was used to measure articulatory positions. The procedures are described in detail by Kent and Moll (1969). The cinefluorographic film rate was 100 frames per second. Hemispherical radiopaque markers, 3.5 mm diameter at the base, were placed on the tongue tip, tongue dorsum, and lower lip. The subjects were allowed to adapt to the markers by speaking and counting prior to filming.

Bite Blocks

Before filming for the hearing-impaired subjects, a bite block was molded from dental acrylic so that the edges of the upper and lower incisors were separated by 10 mm. Care was taken to prevent the bite block from contacting the lateral aspect of the tongue. The subjects were instructed not to speak with the bite block in position until initiation of the filming procedures. Spontaneous speech produced after filming with the bite block in place was not judged to be adversely affected by three phonetically trained observers. The normal hearing controls spoke with three sizes of bite block, but only the data from the 16 mm condition will be presented here.

Analysis of Cinefluorographic Data

Tracings of vocal tract shapes from frames of interest were made from the cinefluorographic films. A vowel "target" was considered achieved when the articulators stayed at the same position for at least three consecutive frames (i.e., 30 msec). The tracings included the outline of the tongue, maxilla, and mandible. Tongue positions were analyzed relative to maxillary and mandibular reference planes (see Kuehn & Moll, 1976; Zimmermann & Rettaliata, 1981). The maxillary framework gives information about changes in tongue position, but does not provide a distinction between changes due to tongue movement and those due to jaw movement. A mandibular reference plane gives information about tongue displacement independent of jaw displacement.

Perceptual Analysis

Tape recordings of utterances of 11 CVCs embedded in carrier phrases produced by the hearing-impaired speakers were presented to eight phonetically trained listeners. The listeners were instructed to rate each speaker on "overall intelligibility" from 1 to 10 (1 being most intelligible). The carriers for S1 differed from those of S2 and S3.[4] The eight listeners also heard and transcribed two productions of /i/, /ae/, and /u/ produced in isolation with and without the bite block. These were randomly presented to the listeners in a free field in a quiet room.[5]

RESULTS

Vocal Tract Shapes

Figures 1a and 1b show the tongue shapes referenced to a maxillary plane for the hearing-impaired (Figure 1b) and normal (Figure 1a) hearing subjects in the bite-block and nonbite-block conditions. The hearing subjects (N1, N2, N3) show more consistency, between and within conditions, in achieving tongue-jaw positions associated with the production of /i/, /u/ and /ae/.

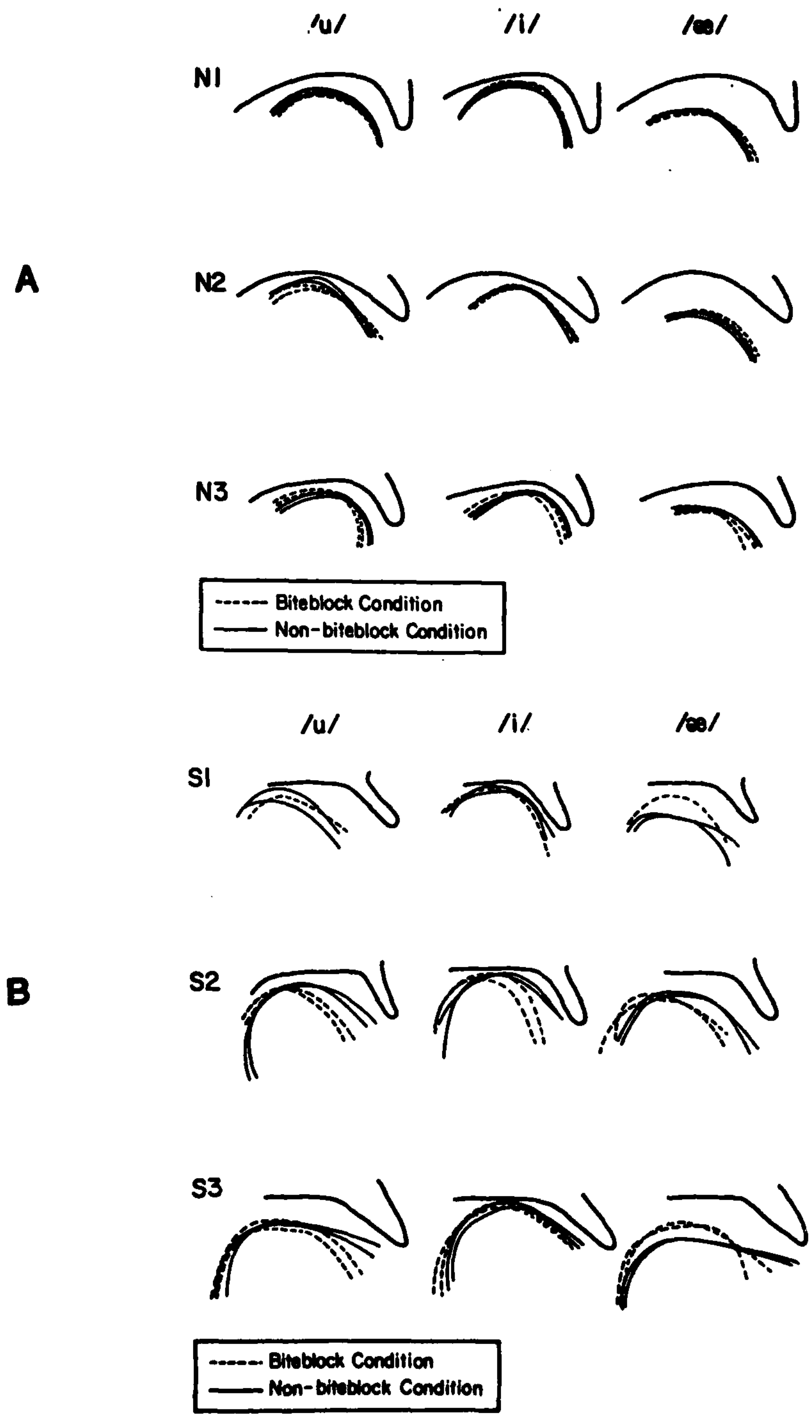


Figure 1. Tongue contours and positions relative to a maxillary reference for /u/, /i/ and /ae/ in the bite-block and nonbite-block conditions. (a) normal hearing speakers, (b) hearing-impaired speakers.

In spite of the variability in tongue shape and positions, the hearing-impaired speakers are, for the most part, as consistent across conditions as they are within conditions in terms of the area of maximum constriction between the tongue dorsum and maxilla. This finding, at least for the vowels /u/ and /i/, suggests that they were able to produce similar vocal tract shapes with and without the bite block. For the production of /ae/ in two of the hearing-impaired subjects (S1 and S3), the distances between the tongue dorsum and maxilla at the region of maximum constriction are different in the bite-block and nonbite-block conditions. The increased distance in the bite-block condition reflects a larger jaw opening without a coincident increased upward displacement of the tongue.

Although the outlines for the hearing-impaired are clearly more variable than for the normal speakers, they nevertheless show a consistent (though not constant) overlap in area of maximum constriction across conditions. Figure 2 shows the vocal tract cross-dimensions (in a manner similar to that employed by Lindblom & Sundberg, 1971) for S2 and N3 in the bite-block and nonbite-block conditions for the production of /i/, /u/, and /ae/. It is clear that the minimum deviations occur at and near the points of maximum constriction, a finding also reported by Gay et al. (1981). Cross-dimension deviation increases with an increase in distance away from the points of maximum constriction, particularly anterior to these points. It is obvious that the cross dimension deviations between conditions are greater for the hearing-impaired speaker than the normal speaker, suggesting differences in the control of the anterior portions of the tongue during vowel production. Also, it should be noted that the region of major constriction appears slightly posterior in the hearing-impaired speaker. The vocal tract shapes in Figures 1a and b lend support to these findings.

Differentiation of tongue shapes and positions among vowels for the bite-block and nonbite-block conditions are shown in Figures 3a (hearing speakers) and 3b (hearing-impaired speakers). This figure shows the composite plots of tongue shapes for /i/, /ae/, and /u/ referred to a maxillary plane. For the /i/ production in both constrained and unconstrained conditions, S2 and S3 show vocal tract shapes that are distinct from those associated with the production of /ae/ and /u/. In fact, they show more differentiation than do hearing subjects. However, while the normal hearing speakers show a definite distinction between the tongue positions for /ae/ versus those for /i/ and /u/, S2 and S3 show more overlap between the shapes associated with /ae/ and /u/. This is evident in the overlap of tongue contours for S2 in both conditions and S3 in the bite-block condition.

The results displayed in Figures 4a and 4b and Figures 5a and 5b show that the distinctions in tongue position evident in Figures 3a and 3b can be accounted for by changes in the displacements of the tongue in relation to the jaw, and are not due solely to changes in jaw displacement. For example, in the bite-block condition for S1 and S3 the tongue position for /i/ is shown to be distinct from those for /ae/ and /u/ (Figure 3b). These contours, with respect to the mandibular reference, indicate the tongue was displaced more for /i/ than for the other vowels (Figure 4b). The increased displacement of the tongue in the bite-block condition compared to the nonbite-block condition, combined with the results in Figure 3b for S3's production of /i/, suggest that increased tongue displacement was associated with an increase in

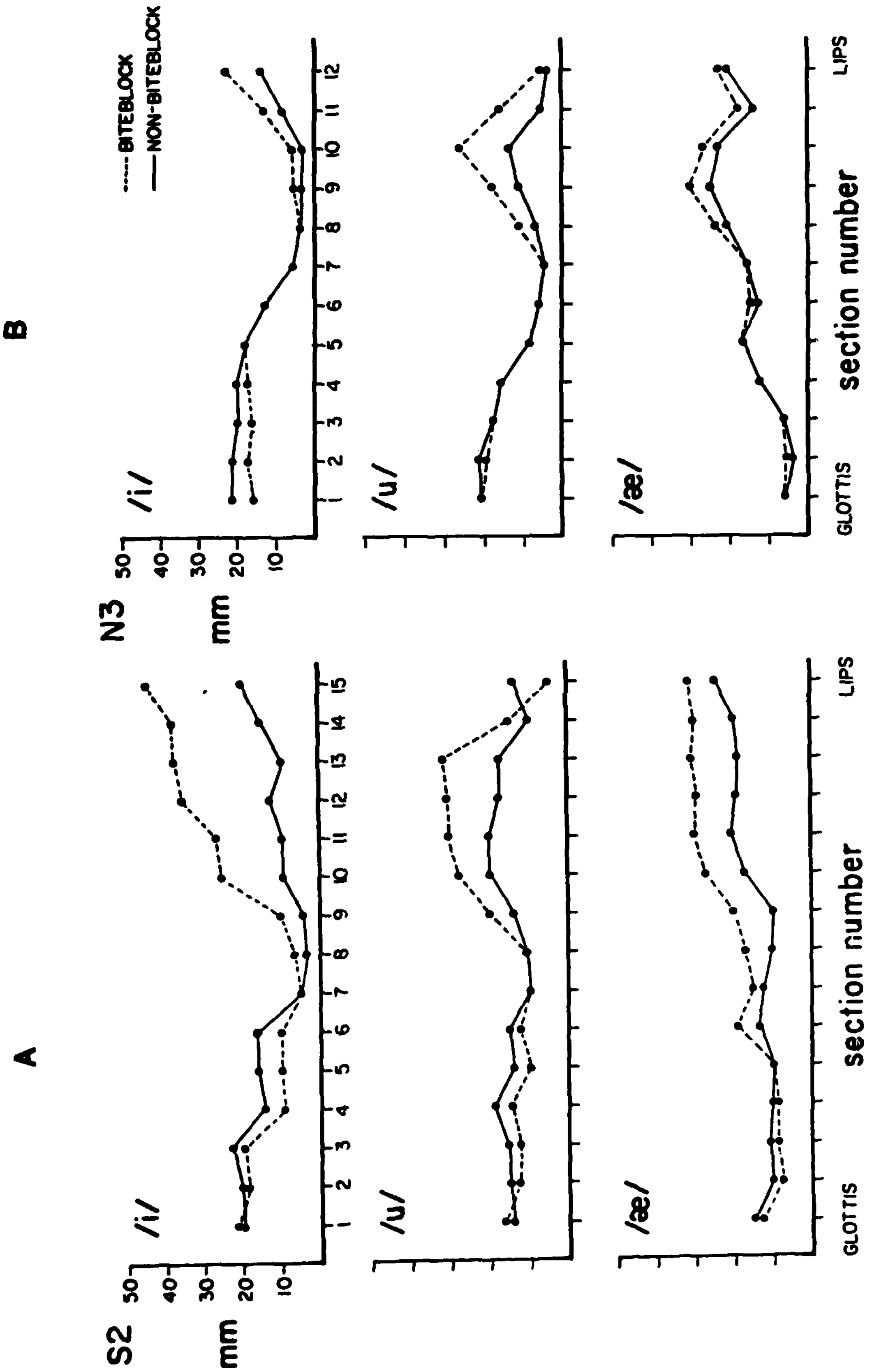


Figure 2. Vocal tract cross dimensions for (a) hearing-impaired speaker (S2), and (b) for a normal hearing speaker (N3) producing /i/, /u/, and /ae/. The data reflect measures taken for one utterance for each speaker.

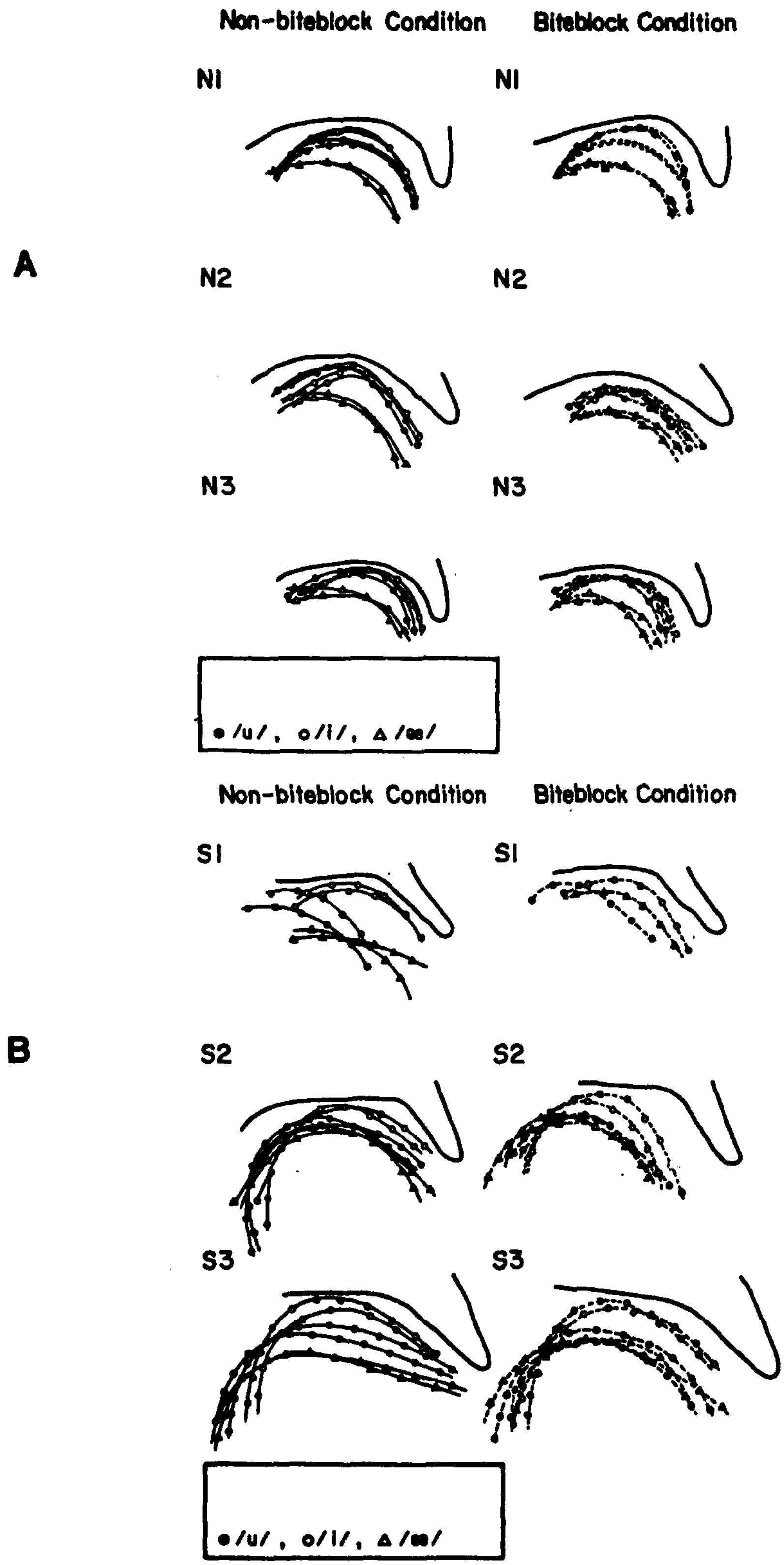


Figure 3. Differentiation between tongue contours and positions relative to maxillary reference for /u/, /i/, /ae/ for the bite-block and nonbite-block conditions. (a) normal hearing speakers, (b) hearing-impaired speakers.

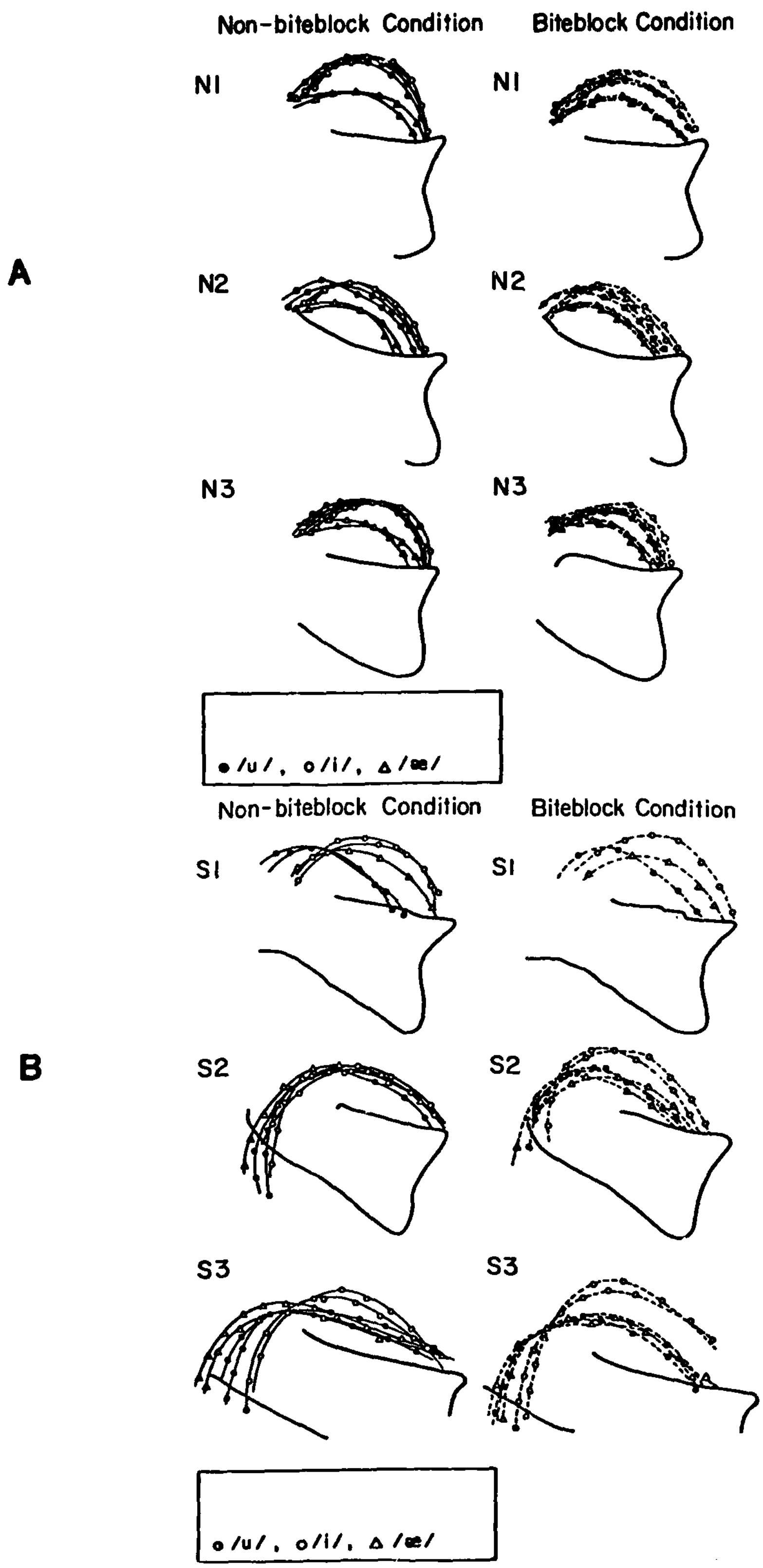


Figure 4. Differentiation between tongue contours and positions relative to mandibular reference for /u/, /i/ and /ae/ in the bite-block and nonbite-block conditions. (a) normal hearing speakers, (b) hearing-impaired speakers.

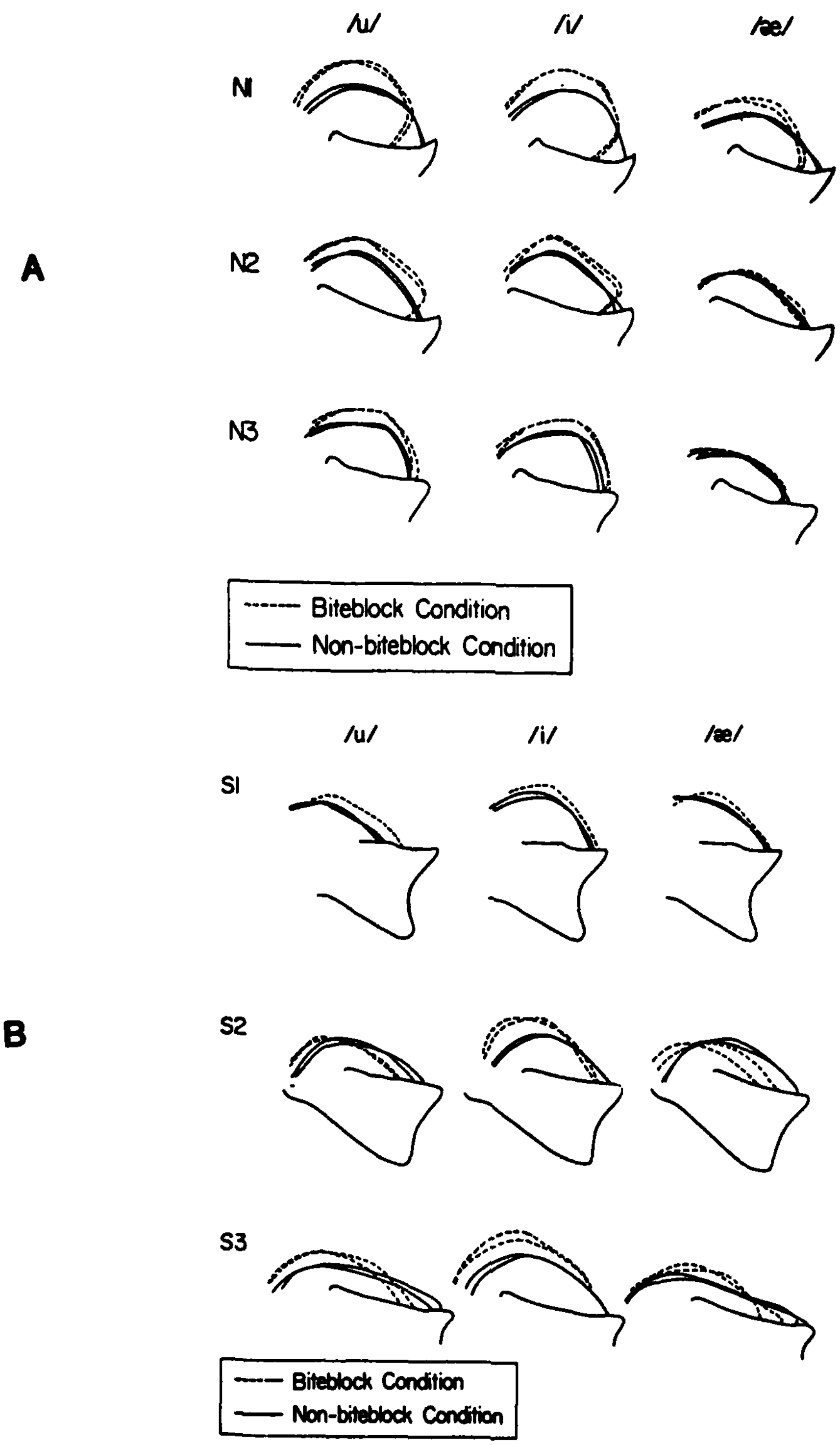


Figure 5. Tongue contours and positions relative to mandibular reference for /u/, /i/ and /ae/ for the bite-block and nonbite-block conditions. (a) normal hearing speakers, (b) hearing-impaired speakers.

jaw opening for the bite-block condition. Figures 5a and 5b also show that there were systematic adjustments in tongue displacement for both hearing-impaired and normal hearing speakers when the jaw was constrained.

Perceptual Results

Each of the eight phonetically-trained listeners ranked the intelligibility of the hearing-impaired speakers in an order that corresponded identically with the judgments of the experimenters: S1 was consistently judged most intelligible, followed by S2 and S3. The results of the vowel transcriptions for S2 and S3 are shown in Table 1. Since S1 did not produce vowels in isolation, so his data are not shown in Table 1. There was no difference in the percent judged errors in vowel production between the bite-block and nonbite-block conditions for either S2 (33% and 35%) or S3 (54% and 52%). The vowels were often judged to be neutralized in both conditions for deaf speakers. The transcription data also showed tongue backing was prevalent in the bite-block condition for the hearing-impaired speakers (e.g., /ae/ was often perceived as /a/).

"Searching" or Oscillatory Behavior

In order to evaluate "searching" or oscillatory movement that may be associated with error correction processes, and to see if there were effects of practice in achieving observed tongue movement patterns, the kinematic trajectories for the first word, "eat" in the carrier, were traced for the first, third, and fifth utterances in the bite-block condition for S2 and S3. Neither the vocal tract shapes associated with /i/ nor the trajectories of movement of the tongue dorsum and jaw to this position were different across trials. Also, the movements to these "vowel" positions were direct and did not display any oscillatory behavior that could be interpreted as "searching" or error correction.[5] However, this is not to suggest that the kinematic patterns of the hearing-impaired speakers were identical to those of the normal hearing speakers (see previous results section).

DISCUSSION

The most interesting result of the present experiment was that the hearing-impaired exhibited so-called "compensatory" movements of the tongue dorsum in the bite-block condition and that these movements generally resulted in the preservation of areas of maximum constriction between the dorsum and the maxilla that were similar for both constrained and unconstrained conditions.

Although the hearing-impaired displayed similar "compensatory" patterns to hearing subjects reported here and elsewhere (Gay et al., 1981; Lindblom & Sundberg, 1971), differences in tongue posturing were nevertheless apparent. In both conditions, the hearing-impaired showed more variable tongue shaping and positioning than the normal hearing subjects. Furthermore, in spite of considerable overlap in regions of maximum constriction of the tongue dorsum in both groups, the positioning of portions of the tongue anterior to the region of maximum constriction differed between conditions for the hearing-impaired subjects, but not for hearing subjects.

TABLE I. Contingency tables for vowels produced by and perceived for S2 and S3 for bite-block (BB) and nonbite-block (NBB) conditions.

S2

NBB					BB				
Produced	Perceived				Produced	Perceived			
	i	æ	u	other		i	æ	u	other
i	8		3	5	i	14			2
æ		14		2	æ		7		9
u	1		10	5	u	1		10	5

Judged errors = 33% Judged errors = 35%

S3

NBB					BB				
Produced	Perceived				Produced	Perceived			
	i	æ	u	other		i	æ	u	other
i	13			3	i	13			3
æ		7		9	æ		7		9
u			2	14	u			3	13

Judged errors = 54% Judged errors = 52%

Two of the hearing-impaired speakers showed less differentiation in tongue shape and position between the productions of /u/ and /ae/ than the hearing speakers in both bite-block and unconstrained conditions. The other speaker (S1), described elsewhere (Zimmermann & Rettaliata, 1981), showed clearly differentiated tongue positions for the vowels /i/, /ae/ and /u/, which may well be related to the better intelligibility for S1 than the other hearing-impaired subjects. Even so, the tongue positioning observed for S1 was markedly different from that of the hearing subjects.

The finding that all three hearing-impaired subjects showed relatively normal tongue contours for the production of /i/ in both experimental conditions, and that the contours for /i/ were the most dissociated from the other vowels, is in accord with the findings of Zimmermann and Rettaliata (1981). The position for the front vowel /i/ may be easiest to learn in the absence of auditory information because it entails primarily a maximum displacement of the tongue dorsum to the palate. That is, the speaker has only to learn to move the dorsum to its greatest extent.

The present data certainly support the acoustic results of Lindblom and Sundberg (1971), and Lindblom, Lubker, and Gay (1979) that indicate auditory information is not critical to the "compensatory" changes in tongue behavior observed when the jaw is constrained. But more important, our results also suggest that "auditory representations" (Gay et al., 1981; Ladefoged et al., 1972) of vowels are not necessarily required to achieve vocal tract configuration associated with /i/, /ae/, and /u/ with the jaw fixed. One presumes that at least the congenitally deaf speaker lacks auditory representations of "vowel targets." Of course, our results do not preclude the existence of some form of "auditory representation" of the target sounds in normal hearing speakers, nor, for that matter, do they negate the importance of audition in the development and maintenance of articulatory patterns.

As we noted in the introduction to the present article, "target-based" theories emphasize the representational aspects of the localization problem (e.g., as auditory or space-coordinate maps) but are mute on how a system of muscles might be so organized as to exhibit targeting behavior. Recent work on other motor activities indicates that learned limb positions can be achieved when afferent information is completely removed. This is the case even when the limb is perturbed during its trajectory to the target or when initial conditions are changed (for relevant animal work see Bizzi et al., 1978; Polit & Bizzi, 1978; for human work see Kelso, 1977; Kelso & Holt, 1980; Kelso, Holt, & Flatt, 1980). These data have been interpreted to suggest that the limbs behave dynamically similar to a nonlinear oscillatory system (Kelso et al., 1980a, 1980b; Fel'dman & Latash, 1982). Extrapolating from this framework to that of speech (see Fowler et al., 1980; Kelso et al., 1980b), achievement of a given vowel target or vocal tract shape may be accomplished by specification of an equilibrium state between the component muscles of the tongue dorsum-jaw system; an equilibrium state being established at a point at which the forces in the muscles summate to zero (Fel'dman, 1966; Kelso & Holt, 1980). Introduction of a bite block may be viewed as altering the balance of forces among articulatory muscles. However, the equilibrium achieved by the tongue dorsum-jaw system during constrained production (i.e., with the jaw fixed) could be achieved by changes in the length-tension ratios of the synergistic muscles involved. That is, a number of combinations of articulatory kinematics (e.g., tongue-jaw positions) may allow for the achievement of the specified equilibrium configuration. The specification of the system's equilibrium state is thought to be determined at higher levels while the details for accomplishment are attributed to lower level, peripheral interactions among the muscles involved. Such muscle groups have been termed functional synergies or coordinative structures to connote a functionally specific set of muscles and joints constrained to act as a single unit (Bernstein, 1967; Boylls, 1975; Greene, 1972; Fowler, 1977; Fowler, Rubin, Remez, & Turvey, 1980; Kelso, Southard, & Goodman, 1979; Saltzman, 1979; Turvey, 1977).

In terms of the present results we suggest that for both hearing-impaired and normal hearing subjects the achievement of similar points of tongue dorsum-maxillary constriction with and without a bite block may be an example of the same dynamical principles derived from other motor activities that involve targeting behavior. That is, even when the jaw is constrained by a bite block, similar regions of maximum constriction or final positions are achieved. While this effect has been termed "compensatory behavior" (Folkins & Abbs, 1977; Lindblom et al., 1979; Lindblom & Sundberg, 1971), the framework offered suggests that the "compensation" is accomplished not through changes in central programs (Lindblom et al., 1979) or through error correction processes based on afferent feedback (Lindblom & Sundberg, 1971; MacNeilage, 1970). Instead, it may be accomplished by a process in which an equilibrium configuration is achieved by virtue of the dynamic characteristics of the muscle-joint system.

The observation that the hearing-impaired display different and more variable tongue positions and shapes than hearing speakers in both jaw-fixed and jaw-free conditions is not inconsistent with the framework that we have elaborated here. Hearing-impaired individuals are likely to have learned different tongue posturing behaviors and different strategies for achieving them because of a lack of available auditory information. The fact that there were changes in tongue contours for certain vowels between conditions although the place of the tongue dorsum-maxillary constriction was held relatively constant in the two conditions suggests that the hearing-impaired have learned to achieve a given point or range of points around the region of maximum constriction for each vowel. The changes in contours for the hearing-impaired, especially the congenitally deaf subject, may suggest that auditory information is used in the learning process to allow fewer degrees of freedom in vocal tract control. That is, in hearing speakers tongue contours may be maintained relatively constant while tongue position is adjusted to distinguish among vowels (Kent, 1970).

The effects of loss of audition on speech kinematics are consistent with Fel'dman's (1974) work. He suggested that removal of afferent information will result in an alteration of the dynamic properties of the muscle groups involved and hence alter the nature of transitional processes without necessarily affecting the achievement of final position. Although much work remains to be done in order to illuminate the processes underlying the control and coordination of speech articulators, we suggest that the theoretical framework referred to here and elaborated in more detail elsewhere (e.g., Fowler et al., 1980; Kelso et al., 1980b; Kelso, Tuller, & Harris, 1983; Kugler, Kelso, & Turvey, 1980) may provide the beginnings of an explanation for the equifinality phenomenon common to many, if not all, motor systems including speech.

REFERENCE NOTE

1. Kent, R. D., Netsell, R., & Abbs, J. A cineflourographic study of bite-block speech. Paper presented at Conference on motor control. Madison, Wi., 1980.

REFERENCES

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

Bertalanffy, L. von. General system theory. Harmondsworth, England: Penguin University Press, 1973.

Bizzi, E., Dev, P., Morasso, P., and Polit, A. Effects of load disturbance during centrally initiated movements. Journal of Neurophysiology, 1978, 41, 542-556.

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

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

Fel'dman, A. G. Functional tuning of the nervous system with control of movement or maintenance of a steady posture. III. Controllable parameters of the muscles. Biophysics, 1966, 4, 565-578.

Fel'dman, A. G. Control in the length of a muscle. Biophysics, 1974 19, 766-711.

Fel'dman, A. G. Superposition of motor programs. I. Rhythmic forearm movements in man. Neurosciences, 1980, 5, 81-90.

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

Folkins, J., & Abbs, J. Motor impairment during inferior alveolar nerve blockade. Journal of Speech and Hearing Research, 1977, 20, 816.

Fowler, C. Timing control in speech production. Unpublished doctoral dissertation Storrs, Ct.: University of Connecticut, 1977.

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

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

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

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

Hudgins, C. V., & Numbers, F. C. An investigation of the intelligibility of the speech of the deaf. Genetic and Psychological Monographs, 1942, 25, 289-392.

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

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

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

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

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

Kelso, J. A. S., & Tuller, B. H. "Compensatory articulation" under conditions of reduced afferent information: A dynamic formulation. Journal of Speech and Hearing Research, in press.

Kelso, J. A. S., Tuller, B. H., & Harris, K. S. A 'dynamic pattern' perspective on the control and coordination of movement. In P. MacNeilage (Ed.), The production of speech. New York: Springer-Verlag, 1983, 137-173.

Kent, R. D. A cinefluorographic-spectrographic investigation of the component gestures in lingual articulation. Unpublished doctoral dissertation, University of Iowa, 1970.

Kent, R. D., & Moll, K. L. Vocal tract characteristics of the stop consonants. Journal of the Acoustical Society of America, 1969, 46, 1549-1555.

Kuehn, D., & Moll, K. A cineradiographic study of VC and CV articulatory velocities. Journal of Phonetics, 1976, 4, 303-320.

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

Ladefoged, P., DeClerk, J., Lindau, M., & Papcun, G. An auditory-motor theory of speech production, UCLA Working Papers in Phonetics. 1972, 22, 48-75.

Lindblom, B., Lubker, J., & Gay, T. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. Journal of Phonetics, 1979, 7, 147-161.

Lindblom, B., & Sundberg, J. Neurophysiological representation of speech sounds. Papers from the Institute of Linguistics, University of Stockholm, 1971, 2, 16-21.

Lindblom, B., & Sundberg, J. Acoustical consequences of lip, tongue, jaw, and larynx movement. Journal of the Acoustical Society of America, 1971, 50, 1166-1179.

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

MacNeilage, P. F. Distinctive properties of speech control. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North Holland, 1980.

McGarr, N. S., & Harris, K. S. Tongue-lip coordination in speech production in a deaf speaker. In J. J. Wolf & D. H. Klatt (Eds.), Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America. New York: Acoustical Society of America, 1980.

Monsen, R. B. Normal and reduced-phonological space: The production of English vowels by deaf adolescents. Journal of Phonetics, 1976, 4, 189-198.
Osberger, M. J., & McGarr, N. S. Speech production characteristics of the hearing impaired. Haskins Laboratories Status Report on Speech Research, 1982, SR-69, 227-288.
Pick, H. L., Siegal, G. M., & Garber, S. R. Development of speech production as a perceptual-motor task. In J. A.S. Kelso & J. E. Clark (Eds.), The development of movement control & coordination. New York: Chicester, 1982.
Polit, A., & Bizzi, E. Processes controlling arm movements in monkeys. Science, 1978, 201, 1235-1237.
Saltzman, E. Levels of sensorimotor representation. Journal of Mathematical Psychology, 1979, 20, 91-163.
Turvey, M. T. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting, and knowing: Toward an ecological psychology. Hillsdale, N.J.: Erlbaum, 1977.
Zimmermann, G., & Rettaliata, P. Articulatory patterns of an adventitiously deaf speaker: Implications for the role of auditory information in speech production. Journal of Speech and Hearing Research, 1981, 24, 169-178.

FOOTNOTES

[1]A dominant reason is "its apparent lack of testability" (MacNeilage, 1980, p. 615).

[2] The data from this previous study were used so we would not expose more subjects to radiation. Note that two hearing-impaired subjects produced isolated vowels. The normal hearing subjects produced vowels in a sentence. It was felt that the different contexts would not significantly affect the results or conclusions, particularly since the major comparison was between bite-block and nonbite-block conditions (within subjects) and not between subjects or groups. Elsewhere it has been shown that the acoustic results of bite-block speech for vowels produced in isolation and vowels produced in a dynamic speech context are near-identical (Kelso & Tuller, in press).

[3]S1 had been part of an earlier study (see Footnote 2). Plots for the normal speakers are for the 16 mm bite-block condition. For the smaller bite-block condition (8 mm) the jaw displacement was not increased over the nonbite-block condition.

[4]Since S1 was part of an earlier study, his sentences differed from those of S2 and S3. S1 produced CVCs in the carrier "eat that ..." while S2 and S3 produced CVCs in the carrier "that's a ...".

[5]Spectrographic analysis was not completed because of the small sample of utterances and the difficulty with reliably measuring the spectrograms of hearing-impaired speakers.

REVIEW*

<u>Pierre Delattre: Studies in comparative phonetics</u>. Edited by B. Malmberg. Heidelberg: Julius Groos Verlag, 1981.

Arthur S. Abramson+

Reviewing a posthumously published book imposes a special obligation on the reviewer to take great care in interpreting the author. While feeling the burden of such a responsibility, I take it to be important that archival journals in our field call the attention of the reading public to what will surely be the last collection of papers by the late distinguished scholar Pierre Delattre. This is my conviction even though my friendship with Delattre and my intellectual debt to him would surely have prevented me from accepting such a task in his lifetime.

The editor of this book, Bertil Malmberg, has carefully chosen four previously published papers, two with co-authors, for reprinting, and he has provided a very interesting introduction of his own. Although Malmberg does say that the papers have appeared previously, he does not give the sources. This is an omission that I shall remedy in my comments on each of the papers. In fact, all of them appeared in the <u>International Review of Applied Linguistics</u> in the period 1968-71. The fact that this is a journal not regularly followed by most phoneticians and other workers in speech research, makes this collection all the more useful. I found the original sources by consulting the bibliography of Delattre's works in the book published in his memory (Valdman, 1972).

It is important here to give some attention to Malmberg's introduction, "Pierre Delattre and Modern Phonetics," since it was written by a person whose views on the man and his scientific setting must be taken very seriously. Although the reader will find this introduction stimulating and informative, he, along with me, may be puzzled and even distressed by Malmberg's insistence that Delattre, in spite of earlier skepticism, had become "convinced of the necessity of the two principles of economy and binarism." He goes on to make much of a "fruitful and intimate collaboration" between Delattre and the late Roman Jakobson. It is true that the two men knew each other and no doubt had much respect for each other, as evidenced by the section entitled "To the Memory of Pierre Delattre" in the recent book by Jakobson and Linda Waugh (1979). In that passage (p. 81), Jakobson's three-day visit to Delattre in Santa Barbara, California is said to have yielded "a plan for a joint, systematic outline of the psychoacoustic correlates of the system of distinc-

*Also <u>Phonetica</u>, 1983, <u>40</u>, in press.
+Also University of Connecticut.
<u>Acknowledgment</u>. This work was supported in part by NICHD Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-73 (1983)]

tive features." Such hearsay reports of private conversations and unrecorded public statements notwithstanding, familiarity with Delattre's publications, especially those within the covers of this volume, would not lead a dispassionate uncommitted reader to the belief that Delattre's attachment to the notion of binary distinctive features was anything more than a willingess not to dismiss such arguments out of hand. That is, when he speaks of, for example, "spread" or "back-rounded" vowels in French in the book under review (p. 82), one might bend over backwards to see binarism lurking between the lines, but the more obvious reading yields merely a traditional phonetic descriptive label.

Jakobson and Waugh (1979, p. 81) tell us that Delattre advocated the slogan "economize and binarize" in his invited paper at the 1967 Sixth International Congress of Phonetic Sciences in Prague. Having been present for this paper, I do recall that Delattre presented his talk with his usual charming flair for the dramatic that made his detailed studies of acoustic cues so much more palatable. Frankly, I cannot recall whether he made such a statement in his oral paper, but in neither the English-language published version of the paper (Delattre, 1968) nor in the proceedings of the congress (Delattre, 1970) does such a sentiment appear! Instead, for this reader at least, the message seems to be that anyone playing the phonological game of distinctive features must be phonetically sophisticated enough to understand that a posited distinctive feature is not likely to be revealed either by the articulatory behavior of the speaker or by his acoustic output. Underlying any such distinctive feature is considerable physical complexity. Summing up the problem, he says (1970, p. 46), "...si les traits pertinents sont des signaux perceptuels qu'on ne peut pressentir qu'indirectement à travers leurs corrélatifs acoustiques et articulatoires, et que les corrélatifs articulatoires ne peuvent être *spécifiés qu'une fois accompli* l'isolement des corrélatifs acoustiques, il n'est peut-être pas possible de toucher les traits pertinents qu'en arrivant à une connaissance suffisante de ce qui est distinctif dans les signaux acoustiques." It is very tempting to interpret this as a warning to the phonologist to make claims about distinctive features only after having found what features of the speech carry the communicative burden.

I shall now make brief mention of the four papers one by one. Since these papers have all appeared before, it may be enough just to give some highlights and a few critical remarks. Without easy access at this time to IRAL, I shall depend on Valdman (1972) to provide bibliographical information on the original publications.

The first paper, written with Michel Monnot (Delattre & Monnot, 1968), is "The Role of Duration in the Identification of French Nasal Vowels." This is an intriguing experimental study of a trading relation between acoustic cues: nasal resonance vs. vowel duration. In French, as is well known, the system of oral vowels is classically described as containing a small subset of vowels minimally distinguished from non-oral counterparts by the simple phonetic feature of nasality. In this paper we find strong analytic support for earlier observations that concomitant with nasality is greater vowel duration. Indeed, experiments with speech synthesis show that this difference in duration is a sufficient acoustic cue to the distinction. Short variants of synthetic vowels with weak simulation of nasal resonance were heard as oral, and long variants, as nasal. The authors speculate in an interesting way about the future of the distinction in French.

The second paper, written with Margaret Hohenberg (Delattre & Hohenberg, 1968), is "Duration as a Cue to the Tense/Lax Distinction in German Unstressed Vowels." Traditionally, it has been observed that the German vowel system contains two sets of vowels, exemplified by such word-pairs as biete/bitte and Kehle/Kelle, said to be distinguished by relative length, although, at least for some of the minimal pairs, there is also a discernible difference in quality. Wishing to avoid assigning phonemic responsibility to either feature, the authors use the terms "tense" and "lax" as cover terms but, at the outset (p. 41, fn. 2), warn the reader that no implication about muscular tension is intended. Anyway, it seems from the sources cited, that dissatisfaction with the status of vowel duration as a satisfactory basis for the distinction arose from the conviction that it was not present in unstressed vowels. The research reported here, however, shows that even in unstressed German vowels, a duration ratio of roughly 3:2 is to be found between the two categories; furthermore, listening tests with synthetic speech, in which vowel durations and vowel formant frequencies, as well as the durations of postvocalic consonant constrictions, were experimentally manipulated, easily demonstrated the overwhelming importance of vowel duration as a perceptual cue to the distinction. Regrettably, the authors appear to contradict themselves (p. 60) by saying, under result number 3, that the two cues of vowel length and vowel color contribute equally well to the distinction in unstressed position, and then, under result number 4, by showing how much more striking and reliable is the duration of the vocalic stretch! That is, the other variables in question certainly have an effect, but they are rather easily overridden by vowel length. A more forthright conclusion to this paper might have insisted on the dominance of duration as a physical underpinning to this feature of German phonology. Indeed, with such results in hand, the authors could have avoided the terms "tense" and "lax" in the title of their paper. After all, it is commonly found in the phonetic literature that clear-cut situations of distinctive vowel length by and large show concomitant differences of vowel color in at least part of the vowel system. It seems very likely, as a matter of fact, that any phonemic distinction closely examined by the experimentalist would reveal that even if a single phonetic dimension, perhaps the one singled out by the phonologist, is dominant, others will also carry perceptually useful information.

The third paper (Delattre, 1969) is "An Acoustic and Articulatory Study of Vowel Reduction in Four Languages." Acoustic and articulatory data are presented for medial vowels under weak stress in English, German, Spanish, and French. This interesting study is marred by a failure to point out a major difference between English and the other three languages. In such word-pairs as disable/disability and abolish/abolition, orthographic a and o in the second members of the pairs represent schwa, that is, reduction of the vowels, if you will, of the first members of the pairs and loss of contrast. The dialect recorded is not mentioned, so it is possible that for at least some of the unstressed English vowels in the sample, "full" vowels are used. It is not surprising, of course, that the plots of formant frequencies and x-ray profiles show much more vowel reduction for English than for the other languages. The results include some interesting differences across these languages in the nature of the vowel reduction observed. It is, by the way, misleading to say at the bottom of page 74 that the IPA charts show only tongue height and fronting; rounding is also a dimension of the charts, whether one uses the old separate charts of primary and secondary Cardinal Vowels or merges them conveniently into one three-dimensional chart.

The final paper in the book, printed as Part I and Part II (Delattre, 1971), is "Consonant Gemination in Four Languages: An Acoustic, Perceptual, and Radiographic Study." As implied by the title, this study, which draws upon German, English, French, and Spanish for its material, is methodologically very elaborate. It examines gemination both at word boundaries and within words. The latter condition, word-internal gemination, is not found in English, and in the other three languages it applies only to /r/. (Of course, in German, as in Beharrung/Behaarung, it should have been pointed out, with a reference to the second paper in this book, that this gemination might best--or at least conventionally—be viewed as part of the vowel-length distinction, although in the other languages of concern here, differences in vowel duration predictably co-occur with phonologically relevant consonant-length distinctions.) The choice of languages having only /r/ for word-interior gemination complicates the matter, since, as shown in this paper, not only relative duration but also other articulatory differences play a role in a way that might not be found in a language like Italian where gemination within the word is found in consonants in which apparently a closure or constriction can simply be held longer. If, however, one makes allowances for phonologically confusing statements here and there, it is possible to derive much enlightening information about the production and perception of this contrast.

Bertil Malmberg and Julius Groos Verlag are to be complimented for their efforts in compiling and publishing this book. Had Pierre Delattre been alive to edit it himself, even with the provocative essay by Malmberg included, no doubt he would have wanted to clarify not only the points I have raised but also many more that he himself would have wished to reconsider in retrospect. This handy collection of some of his last research studies should certainly be on the reading list of all students of experimental phonetics.

REFERENCES

Delattre, P. C. From acoustic cues to distinctive features. Phonetica, 1968, 18, 198-230.

Delattre, P. C. An acoustic and articulatory study of vowel reduction in four languages. International Review of Applied Linguistics, 1969, 7, 295-325.

Delattre, P. C. Des indices acoustiques aux traits pertinents. Proceedings of the Sixth International Congress of Phonetic Science, 1967, Prague, 1970.

Delattre, P. C. Consonant gemination in four languages: An acoustic, perceptual, and radiographic study. Parts I and II. International Review of Applied Linguistics, 1971, 9, 31-52, 97-113.

Delattre, P. C., & Hohenberg, M. Duration as a cue to the tense/lax distinction in German unstressed vowels. International Review of Applied Linguistics, 1968, 6, 367-390.

Delattre, P. C., & Monnot, M. The role of duration in the identification of French nasal vowels. International Review of Applied Linguistics, 1968, 6, 267-288.

Jakobson, R., & Waugh, L. The sound shape of language. Bloomington, Ind.: Indiana University Press, 1979.

Valdman, A. (Ed.) Papers in linguistics and phonetics to the memory of Pierre Delattre. The Hague: Mouton, 1972.

II. PUBLICATIONS

III. APPENDIX

PUBLICATIONS

Abramson, A. S. Review (Pierre Delattre: Studies in Comparative Phonetics.) Phonetica, 1983, 40, in press.

Baer, T., Löfqvist, A., & McGarr, N. S. Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques. Journal of the Acoustical Society of America, 1983, 73, 1304-1308.

Bentin, S., & Carmon, A. Verbal information processing and hemispheric ERP asymmetry. In R. Karrer, J. Cohen, & P. Tueting (Eds.), Brain and information: Event related potentials. New York: New York Academy of Science, in press.

Bentin, S., & Feinsod, M. Hemispheric asymmetry for word perception: Behavioral and ERP evidence. Psychophysiology, in press.

Collier, R., Bell-Berti, F., & Raphael, L. J. Some acoustic and physiological observations of diphthongs. Language and Speech, in press.

Davis, W. E., & Kelso, J. A. S. Analysis of invariant characteristics in the motor control of Down's Syndrome and normal subjects. Journal of Motor Behavior, 1982, 14, 194-212.

Feldman, L. B., & Turvey, M. T. Word recogntion in Serbo-Croatian is phonologically analytic. Journal of Experimental Psychology: Human Perception and Performance, 1983, 9, 288-298.

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

Harris, K. S. Review (Phonological Intervention: Concepts and Procedures.) Applied Psycholinguistics, in press.

Katz, L., & Feldman, L. B. Relation between pronunciation and recognition of printed words in deep and shallow orthographies. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1983, 9, 157-166.

Kelso, J. A. S. Old problems and new directions in motor behavior. Contemporary Psychology, 1983, 28, 8-11.

Kelso, J. A. S., & Saltzman, E. L. Motor control: Which themes do we orchestrate? The Behavioral and Brain Sciences, 1982, 5, 554-557.

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

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

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

Kewley-Port, D., Pisoni, D. B., & Studdert-Kennedy, M. Perception of static and dynamic cues to place of articulation in initial stop consonants. Journal of the Acoustical Society of America, in press.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), The development of movement control and coordination. Chichester: John Wiley, 1982, 5-78.

Lukatela, G., Kostić, A., Feldman, L. B., & Turvey, M. T. Grammatical priming of inflected nouns. Memory & Cognition, 1983, 11, 59-63.

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

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

Niimi, S., Bell-Berti, F., & Harris, K. S. Dynamic aspects of velopharyngeal closure. Folia Phoniatrica, 1982, 34, 246-257.

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

Remez, R. E., & Rubin, P. E. The stream of speech. Scandinavian Journal of Psychology, 1983, 24, 63-66.

Repp, B. H., Milburn, C., & Ashkenas, J. Duplex perception: Confirmation of fusion. Perception & Psychophysics, in press.

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

Studdert-Kennedy, M. On the dissociation of auditory and phonetic perception. In R. Carlson & B. Granström (Eds.), The representation of speech in the peripheral auditory system. Amsterdam: Elsevier Biomedical Press, 1982, 3-10.

Turvey, M. T., Feldman, L. B., & Lukatela, G. The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. In L. Henderson (Ed.), Orthographies and reading. London: Erlbaum, in press.

Whalen, D. H. Vowel information in postvocalic fricative noises. Language and Speech, in press.

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	

Information on ordering any of these issues may be found on the following page.

**DTIC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Language Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Haskins Laboratories
270 Crown Street
New Haven, CT 06510

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
N/A

3. REPORT TITLE
Haskins Laboratories Status Report on Speech Research, SR-73, January - March, 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
March, 1983

7a. TOTAL NO. OF PAGES
332

7b. NO. OF REFS
582

8a. CONTRACT OR GRANT NO.

HD-01994	BNS-8111470
HD-16591	NS13870
N01-HD-1-2420	NS13617
RR-05596	NS18010
PRF-8006144	N00014-83-C-0083

9a. ORIGINATOR'S REPORT NUMBER(S)
SR-73 (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 January-31 March) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

- The influence of subcategorical mismatches on lexical access;
- The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy;
- Grammatical priming effects between pronouns and inflected verb forms;
- Misreadings by beginning readers of Serbo-Croatian;
- Bi-alphabetism and word recognition;
- Orthographic and phonemic coding for word identification: Evidence from Hebrew;
- Stress and vowel duration effects on syllable recognition;
- Phonetic and auditory trading relations between acoustic cues in speech perception: Further results;
- Linguistic coding by deaf children in relation to beginning reading success;
- Determinants of spelling ability in deaf and hearing adults: Access to linguistic structure;
- A dynamical basis for action systems;
- On the space-time structure of human interlimb coordination;
- Some acoustic and physiological observations on diphthongs;
- Relationship between pitch control and vowel articulation;
- Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques;
- Compensatory articulation in hearing impaired speakers: A cinefluorographic study; and
- Review (Pierre Delattre: Studies in comparative phonetics.)

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

UNCLASSIFIED

Security Classification

A-31408

*This document contains no information not freely available to the general public. It is distributed primarily for library use.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Perception: phonetic, categories, lexical auditory, trading relation, acoustic cue Speech Articulation: diphthongs, acoustic, physiologic pitch, vowel, control, relationship larynx, vibration, film, glottography compensation, deaf, speakers, cinefluorography Reading: orthography, phonology, Serbo-Croatian grammar, priming, pronouns, inflection misreadings, beginning reader bi-alphabetic, recognition, words graphemes, phoneme, Hebrew coding. linguistic, deaf, children spelling, structure Motor Control: action, system, dynamic limb, coordination, interlimb time, space Automatic Speech Recognition: stress, vowel, duration, syllable						