DOCUMENT RESUME

ED 094 444	CS 500 814
TITLE	Status Report on Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications, July 1-December 31, 1973. Report No. SR-35/36(1973).
INSTITUTION	Haskins Labs., New Haven, Conn.
REPORT NO	SR-35/36(1973)
PUB DATE	Dec 73
NOTE	220p.
EDRS PRICE	MF-$0.75 HC-$10.20 PLUS POSTAGE
DESCRIPTORS	Articulation (Speech); *Conference Reports; *Educational Research; Higher Education; *Language Development; Language Skills; Psychomotor Skills; *Speech; Speech Skills; *Verbal Communication

ABSTRACT

This report, covering the period of July 1 to December 31, 1973, 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. The materials included in this report are "The Motor Organization of Some Speech Gestures," "Effect of Speaking Rate of Labial Consonant-Vowel Articulation," "Stress and Syllable Duration Change," "Parallel Processing of Auditory and Phonetic Information in Speech Perception," "Perception of Speech and Nonspeech, with Speech-Relevant and Speech-Irrelevant Transitions," "On the Identification of Place and Voicing Features in Synthetic Stop Consonants," "The Lag Effect in Dichotic Speech Perception," "State-of-the-Art Report on Language Processing," "Audible Outputs of Reading Machines for the Blind," and "An Examination of Hemispheric Asymmetry in the Visual Processing of Linguistic Items." (RB)

SR-35/36 (1973)

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 July - 31 December 1973

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

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

ACKNOWLEDGMENTS

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

Information Systems Branch, Office of Naval Research
Contract N00014-67-A-0129-0001 and -0002

National Institute of Dental Research
Grant DE-01774

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

Research and Development Division of the Prosthetic and Sensory Aids Service, Veterans Administration
Contract V101(134)P-71

National Science Foundation
Grant GS-28354

National Institutes of Health
General Research Support Grant RR-5596

National Institute of Child Health and Human Development
Contract NIH-71-2420

The Seeing Eye, Inc.
Equipment Grant

HASKINS LABORATORIES

Personnel in Speech Research

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

Investigators

Arthur S. Abramson* +
Fredericka Bell-Berti*
Gloria J. Borden*
Rene Collier++
James E. Cutting*
Ruth S. Day*
Michael F. Dorman*
Jane H. Gaitenby
Thomas J. Gay*
Katherine S. Harris*
Philip Lieberman*
Leigh Lisker*
Ignatius G. Mattingly*
Paul Mermelstein
Patrick W. Nye
Lawrence J. Raphael*
Donald P. Shankweiler*
George N. Sholes
Michael Studdert-Kennedy*
Michael T. Turvey* +++
Tatsujiro Ushijima++++

Technical and Support Staff

Eric L. Andreasson
Elizabeth P. Clark
Pamela N. Cockey*
Donald S. Hailey
Diane Kewley-Port*
Sabina D. Koroluk
Christina R. LaColla
Roderick M. McGuire
Agnes M. McKeon
Susan C. Polgar*
William P. Scully
Richard S. Sharkany
Edward R. Wiley
David Zeichner

Students

James Bartlett**
Susan Brady*
Joanne Croog**
Nina de Jongh**
Susan Lea Donald*
G. Campbell Ellison*
F. William Fischer**
Christopher F. Foard*
Carol A. Fowler**
Frances J. Freeman*
Gary M. Kuhn*
Elizabeth Miehl**
Terrance M. Nearey*
Barbara R. Pick*
Barbara Pober**
Timothy C. Rand*
Philip E. Rubin*
Helen Simon*
Elaine E. Thompson*

---

*Part-time
**Volunteer
+On leave-of-absence to Ramkhamhaeng University, Bangkok, Thailand.
++Visiting from University of Louvain, Belgium.
+++On leave-of-absence to University of Sussex, Brighton, England.
++++Visiting from University of Tokyo, Japan.

CONTENTS

I. Manuscripts and Extended Reports

The Motor Organization of Some Speech Gestures -- Fredericka Bell-Berti and Katherine Safford Harris 1

Effect of Speaking Rate on Labial Consonant-Vowel Articulation -- Thomas Gay, Tatsujiro Ushijima, Hajime Hirose, and Franklin S. Cooper. . . . 7

Stress and Syllable Duration Change -- Katherine S. Harris 31

Parallel Processing of Auditory and Phonetic Information in Speech Perception -- Charles C. Wood. 39

Perception of Speech and Nonspeech, with Speech-Relevant and Speech-Irrelevant Transitions -- James E. Cutting 55

On the Identification of Place and Voicing Features in Synthetic Stop Consonants -- David B. Pisoni and James R. Sawusch 65

The Lag Effect in Dichotic Speech Perception -- Emily F. Kirstein. 81

State-of-the-Art Report on Language Processing -- A. M. Liberman 107

Audible Outputs of Reading Machines for the Blind -- Franklin S. Cooper, Jane H. Gaitenby, Ignatius G. Mattingly, Patrick W. Nye, and George N. Sholes 117

ABSTRACT: An Examination of Hemispheric Asymmetry in the Visual Processing of Linguistic Items -- Claire Farley Michaels. 121

II. Thesis Supplement

Levels of Processing in Speech Perception: Neurophysiological and Information-Processing Analyses -- Charles C. Wood . . . .[paged separately] 123

III. Publications and Reports 127

IV. Appendix: DDC and ERIC numbers (SR-21/22 - SR-34) 132

V. Cumulative Index (SR-29/30 - SR-35/36) 135

I. MANUSCRIPTS AND EXTENDED REPORTS

The Motor Organization of Some Speech Gestures*

Fredericka Bell-Berti[+] and Katherine Safford Harris[++]
Haskins Laboratories, New Haven, Conn.

A body of speech research has been concerned with describing the phenomena and postulating the rules for the reorganization of speech gestures. We have inspected electromyographic (EMG) data from three speakers of American English and one speaker of Swedish in an effort to discover the manifestations of, and limits on, reorganization at the motor command level. We have defined this reorganization as the merging of EMG activity peaks for sequences of speech sounds. Our EMG data were recorded from several extrinsic lingual muscles--the mylohyoid, genioglossus, and palatoglossus--and were processed using the Haskins Laboratories' EMG data system. We have called the muscles we studied "closers" of the vocal tract, because they are active for articulations requiring tongue raising. The experimental utterances were polysyllabic nonsense words. The muscles studied were active for both vowels and consonants included in the experimental utterances. Stop and fricative consonants, having inherently different degrees of vocal tract closure, were included in the utterances. The figures that follow show EMG signals for short sequences of speech sounds occurring within longer strings.

The vertical line at time zero represents the point in the acoustic signal that was used to line up the individual tokens for averaging. Representative acoustic signals appear above each graph, and are aligned with the zero reference point.

Results

The palatoglossus muscle is active for subject BG for /u/ and /m/ (Figure 1). The EMG curves for the /-um-/ and /-mu-/ sequences of the utterances are different: there is one peak for the /-um-/ sequence of /fumpup/, while there are two separate peaks for the /-mu-/ sequence of /fupmup/.

The mylohyoid muscle is active for subjects LJR and KSH for /i/, /s/, and /k/ (Figure 2). One peak occurs for the /-ik-/ sequence of /ə'pikupə/ (Figure 2a)

---

*Paper presented at the 86th meeting of the Acoustical Society of America, Los Angeles, Calif., November 1973.

[+] Also Montclair State College, Upper Montclair, N. J.

[++] Also City College of the City University of New York.

[HASKINS LABORATORIES: Status Report on Speech Research SR-35/36 (1973)]

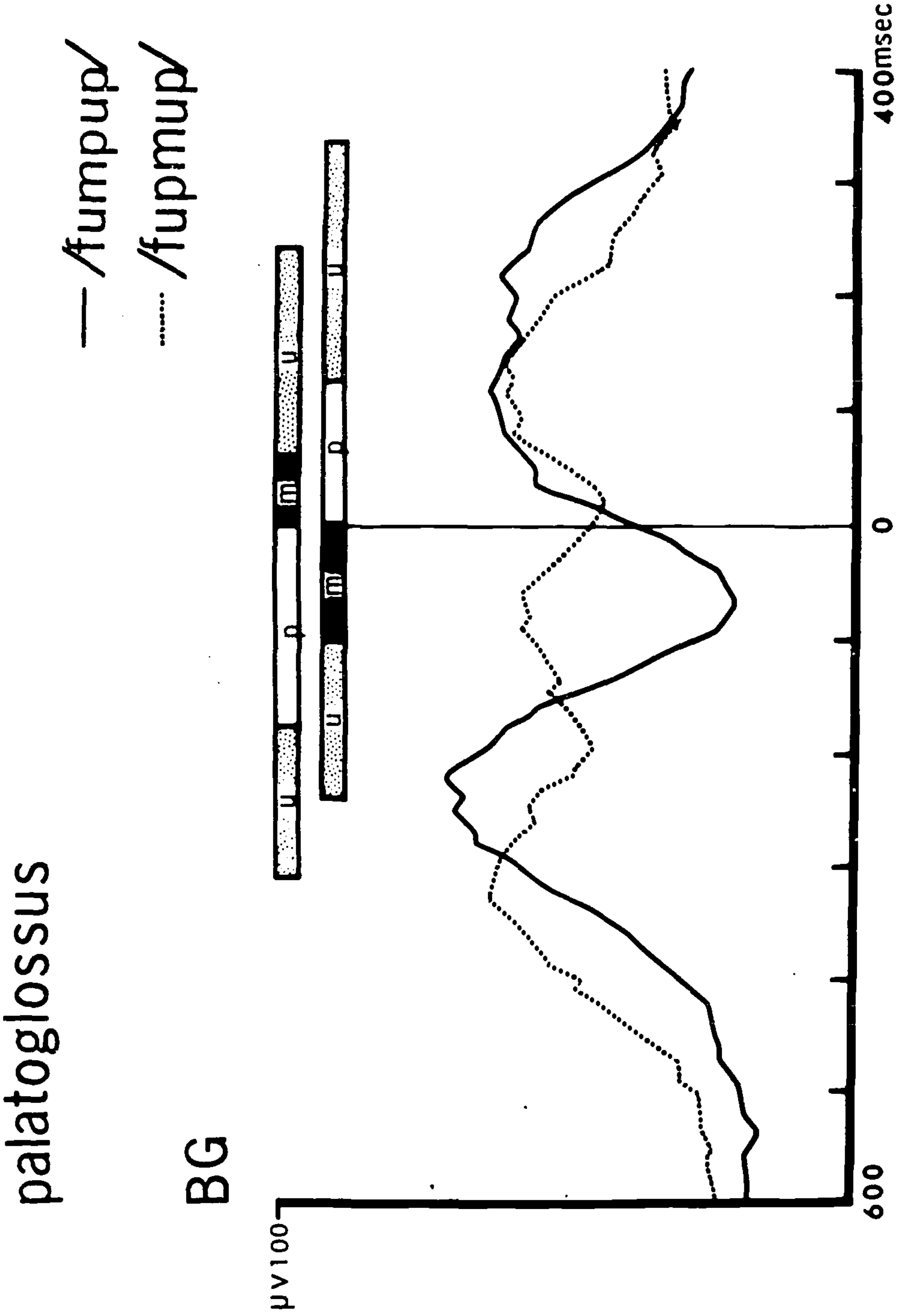

palatoglossus
BG
— /fumpup/
····· /fupmup/
u
p
m
u
u
m
p
u
μv100
600
0
400msec

Figure I

MYLOHYOID

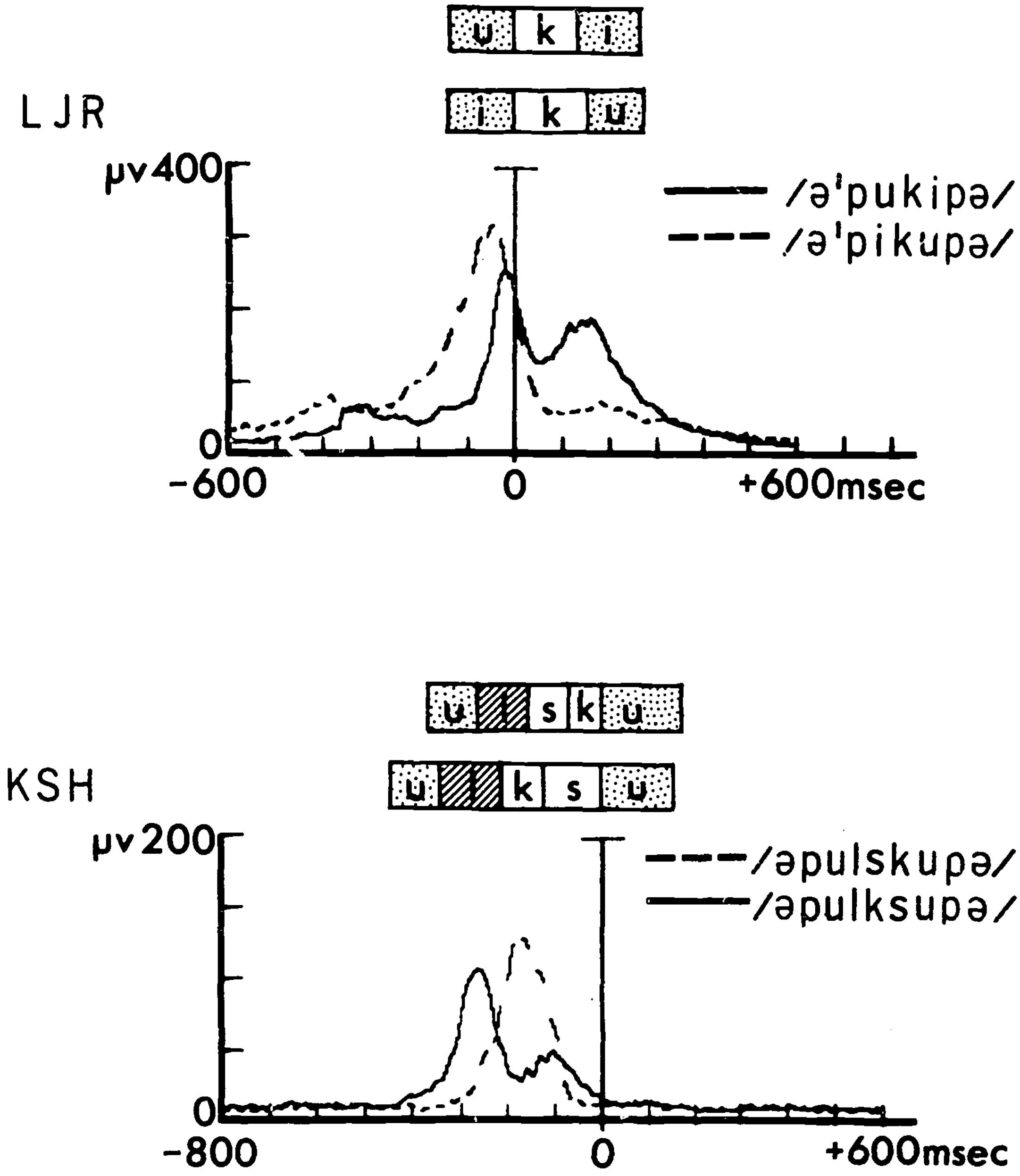


Figure 2

GENIOGLOSSUS

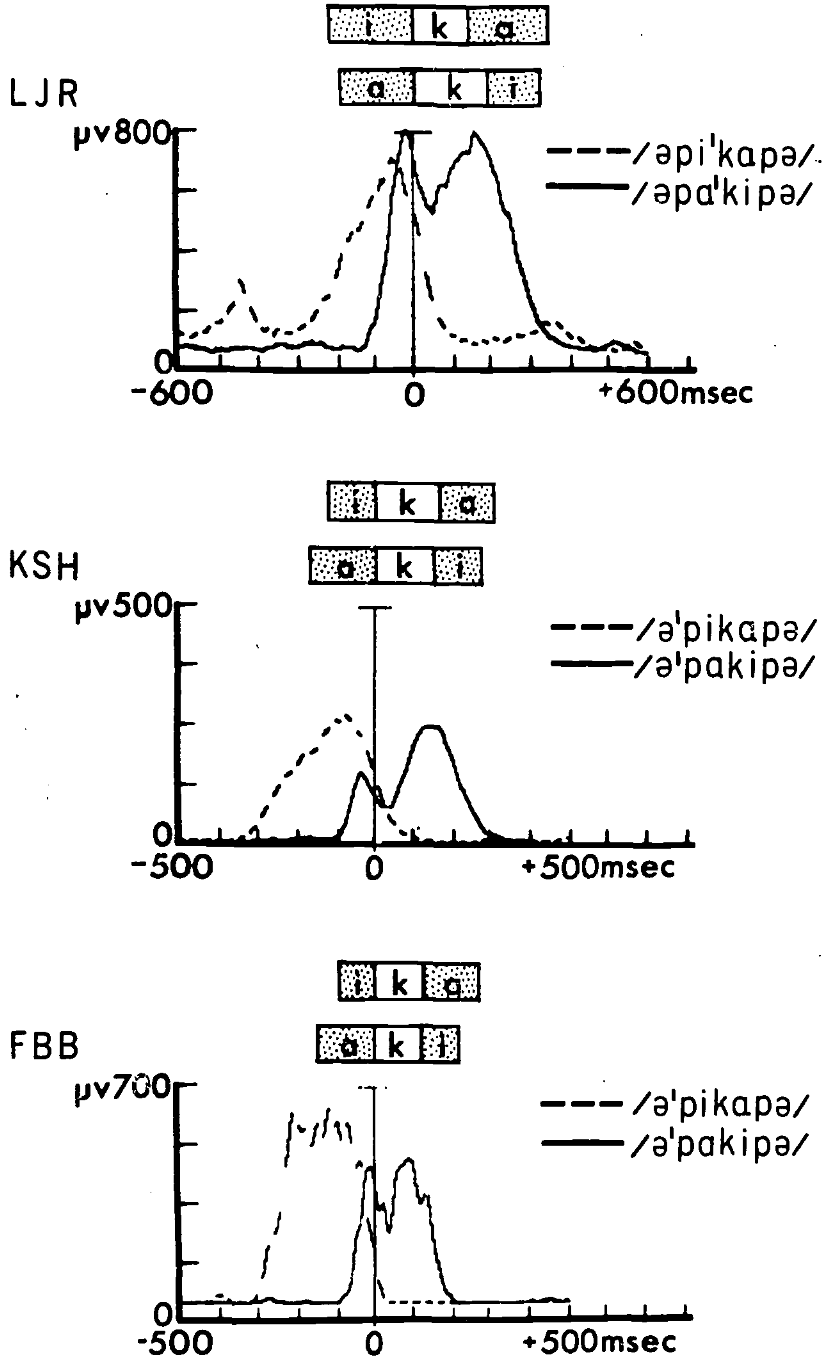


Figure 3

and two peaks occur for the /-ki-/ sequence of /ə'pukipə/ (Figure 2b). In the lower figure we see one peak of activity for the /-sk-/ sequence of /əpul'skupə/ and two peaks of activity for the /-ks-/ sequence of /əpulk'supə/.

The genioglossus muscle is active for subjects LJR, KSH, and FBB in these utterances for /i/ and /k/ (Figure 3). The primary syllable stress varies in these utterances between the second and third syllable. In each case, there is one peak of activity for the /-ik-/ sequence and there are two peaks of activity for the /-ki-/ sequence.

Discussion

We may consider the reorganization of motor commands to be coarticulatory events. One type of reorganization may be viewed as anticipatory, where a future gesture influences the present one. We wish to discover when the commands for a gesture will merge with commands for a future gesture (anticipatory coarticulation at the motor command level).

We find in all these data that there is one peak of activity for a sequence of speech sounds beginning with a more open vocal tract and ending with a less open vocal tract, and there are two peaks of activity for a sequence of speech sounds beginning with a less open vocal tract and ending with a more open vocal tract, regardless of the position of the perceived syllable boundary or primary syllable stress. [The finding of coarticulation occurring across syllable boundaries has previously been reported by Daniloff and Moll (1968), although they were discussing a different type of phenomenon.]

All of our data to this point were collected from muscles that we have called vocal tract closers because their activity increases tongue height (or, for the palatoglossus, for nasal gestures, brings the velum down toward the tongue). In each case, the motor commands for sequences of speech sounds were reorganized (or merged) when the sequence began with a relatively more open vocal tract and ended with a more closed vocal tract, /-ik-/, /-sk-/, /-um-/. Reorganization was not observed when the sequence began with a more closed vocal tract and terminated with a more open vocal tract, /-ki-/, /-ks-/, /-mu-/.

We hypothesize, then, that muscles that are vocal tract closers coarticulate only to the greatest constriction. We presume, by analogy, that muscles that are vocal tract openers coarticulate to the greatest opening. Viewing any speech sequence, closers exhibit anticipatory coarticulation when the vocal tract is being closed and openers exhibit anticipatory coarticulation when the vocal tract is being opened. Only one set of these muscles, either the openers or the closers, may anticipate future gestures within any sequence.

Summary

In summary, we hypothesize that muscles that close the vocal tract coarticulate only to the greatest constriction. And we presume, by analogy, that muscles that open the vocal tract coarticulate only to the greatest opening.

REFERENCE

Daniloff, Raymond G. and Kenneth Moll. (1968) Coarticulation of lip rounding. J. Speech Hearing Res. 11, 707-721.

Effect of Speaking Rate on Labial Consonant-Vowel Articulation*

Thomas Gay,[+] Tatsujiro Ushijima,[++] Hajime Hirose,[+++] and Franklin S. Cooper
Haskins Laboratories, New Haven, Conn.

ABSTRACT

The purpose of this experiment was to study the effect of speaking rate on the articulation of the consonants /p/ and /w/ in combination with the vowels /i/, /a/, and /u/. Two subjects read a list of nonsense syllables containing /p/ and /w/ in all possible vowel-consonant-vowel combinations with /i/, /a/, and /u/ at both moderate and fast speaking rates. Electromyographic recordings from muscles that control movements of the lips, tongue, and jaw were recorded simultaneously with high-speed, lateral-view X-ray films of the tongue and jaw, and with high-speed, full-face motion pictures of the lips. For labial consonant production, an increase in speaking rate is accompanied by an increase in the activity level of the muscle (orbicularis oris) and slightly faster rates of lip movement (both closing and opening). Vowel production, however, shows opposite effects: an increase in speaking rate is accompanied by a decrease in the activity level of the genioglossus muscle and, as shown by the X-ray films, evidence of target undershoot. Jaw movement data show more variable, context-dependent effects on speaking rate. Observed differences are explained in terms of the muscle systems involved.

---

*Expanded version of a paper presented at the 86th meeting of the Acoustical Society of America, Los Angeles, Calif., November 1973.

[+] Also Department of Oral Biology, University of Connecticut Health Center, Farmington.

[++] Faculty of Medicine, University of Tokyo; visiting researcher, Haskins Laboratories, 1970-1972.

[+++] On leave from the Faculty of Medicine, University of Tokyo.

Acknowledgment: The authors wish to thank Dr. J. Daniel Subtelny, head of the Department of Orthodontics, Eastman Dental Center, Rochester, N. Y., for the use of his cinefluorographic facilities; and Mrs. Constance Christian for her help in analyzing the motion picture data.

[HASKINS LABORATORIES: Status Report on Speech Research SR-35/36 (1973)]

INTRODUCTION

The way a speaker produces a given string of phones will show a good deal of variability depending upon, among other things, the suprasegmental features of stress and speaking rate. The control of speaking rate is a good illustration of the complex nature of these allophonic variations. For example, it is commonly known that during faster speech a vowel tends to change color toward the neutral schwa (Lindblom, 1963, 1964). Lindblom's original model proposes that this neutralization is a consequence of the shorter duration of the vowel and is caused by a temporal overlap of motor commands to the articulators. In other words, the articulators fail to reach, or undershoot, their targets because the next set of motor commands deflects them to the following target before the first target is reached. This phenomenon implies further that both the rate of movement of the articulators (specifically the tongue) and the activity levels of the muscles either remain unchanged or are decreased during faster speech. Although similar undershoot effects have been observed for other phones (Gay, 1968; Kent, 1970), a general model of speaking rate control based on timing changes alone is too simple. For example, in a recent study of labial consonant production (Gay and Hirose, 1973) it was shown that an increase in speaking rate is accompanied by both an increase in the rate of movement of the lips and an increase in the activity levels of the muscles. Although changes in the timing of commands to the muscles do occur, the production of labial consonants during faster speech is characterized primarily by an increase in articulatory effort.

The implication that more than one mechanism operates to control speaking rate might be expected. Whereas vowel production involves a movement toward a spatial target, the production of most consonants involves a movement towards constrictive or occlusal targets. Thus, in a strict sense, the concept of undershoot itself cannot be easily applied to consonant production. Of course, too, the phenomena described above are based on a number of different experiments; it is quite conceivable that some of the differences observed are individual ones.

This paper represents an attempt to describe these phenomena further by studying the effect of speaking rate on the articulation of labial consonants on both preceding and following vowels. The specific purpose of the experiment was to study the effect of speaking rate on the coordination of lip, tongue, and jaw movements during the production of the labial consonants /p/ and /w/ in combination with the vowels /i/, /a/, and /u/. The experiment utilized the combined techniques of electromyography (EMG); cinefluorography; and direct, high-speed motion picture photography.

METHOD

Subjects and Speech Material

Speakers were two adult males, both native speakers of American English. The speech material consisted of the consonants /p/ and /w/ and the vowels /i/, /a/, and /u/ in a trisyllable nonsense word of the form /k V_1 C V_2 p/ where V_1 and V_2 were all possible combinations of /i/, /a/, and /u/ and C was either /p/ or /w/. An additional set of trisyllables of the form /kut V pa/ (V = /i/, /a/, /u/) was also constructed. These stimuli were incorporated into the EMG part of the experiment to obtain lip-rounding data for /u/ in a nonlabial consonant environment. The utterance types were randomized into a master list. The carrier phrase, "It's a...," preceded each utterance. Two speaking rates were studied: slow (normal) and fast. Each speaking rate was based on the subject's own appraisal of comfortable slow and fast rates. A brief practice session preceded each run. The subjects were also instructed to produce the first two syllables with equal stress and the final syllable unstressed. The subjects' performances were monitored continuously throughout the run.

Electromyography

For both subjects, conventional hooked-wire electrodes were inserted into muscles that control movements of the lips, tongue, and jaw. These muscles are listed in Table 1. Although all muscle locations showed adequate firing levels at the time of electrode insertion, some locations deteriorated at one time or another during the run. The extent to which this occurred is also indicated in Table 1.

TABLE 1: EMG electrode locations.

Subject FSC	Subject TG
Orbicularis Oris (OO)	Orbicularis Oris (OO)
Genioglossus (GG)	Genioglossus (GG)[2]
Internal Pterygoid (IP)[1]	Superior Longitudinal (SL)[2]
Anterior Belly Digastric (AD)	Internal Pterygoid (IP)[2]
	Anterior Belly Digastric (AD)

[1] Analyzed for combined run only.

[2] Not usable.

The basic procedure was to collect EMG data for a number of tokens of a given utterance and, using a digital computer, average the integrated EMG signals at each electrode position. The EMG data were recorded on a 14-channel instrumentation tape recorder together with the acoustic signal and a sequence of digital code pulses (octal format). These pulses are used to identify each utterance for the computer during processing. A more detailed description of the various aspects of the experimental procedure can be found elsewhere (Hirose, 1971; Kewley-Port, 1971, 1973).

Cinefluorography

Lateral-view X-ray films were recorded with a 16 mm cine camera set to run at 64 fps. The X-ray generator delivered 1 msec pulses to a 9 inch image intensifier tube. The subject was seated with his head positioned in a standard headholder. A barium sulfate paste was used as a contrast medium on the tongue, and tantalum was applied along the midline of the nose, lips, and jaw to outline those structures. The X-ray film records were synchronized with the other records by a pulse train generated by the camera and recorded on the data tape.

High-Speed Motion Picture Photography

High-speed motion pictures of lip movements were recorded with a 16 mm Milliken camera set to run at 128 fps. Because of space constraints the full-face motion pictures of the lips were recorded through a mirror. The motion picture and EMG data were synchronized by an annotation system that displayed the octal code pulses on an LED device placed in the path of the camera. This display was also driven by a signal from the camera to count individual frames between octal codes. Before the run, white reference dots were painted on the subjects' lips at the midline. A scale was fixed to the mirror for calibration of lip movement measurements. A block diagram of the recording system is shown in Figure 1.

Data Recording and Analysis

The combined EMG/cinefluorographic/high-speed motion picture data were recorded at the beginning of the run, after which the EMG part of the experiment continued. For the second segment of the run (EMG only), the word list was repeated ten times at each of the two speaking rates. The EMG data from the combined-techniques part of the run were processed separately from the remainder of the run. This allowed comparisons to be made between the individual (combined run) tokens and the averaged data.

The X-ray films were analyzed by frame-by-frame tracings to obtain the outline of the surface of the tongue as well as a direct measurement of jaw displacement (vertical distance between the incisors). The direct-view motion picture films were analyzed by frame-by-frame measurements of vertical lip opening at the midline. All film measurements were made on a Perceptoscope film analyzer.

Duration measurements were made from the Visicorder tracings. The mean durations of the utterances (token + carrier) were 980 msec and 650 msec for Subject FSC, and 1,030 msec and 670 msec for Subject TG, for the slow and fast speaking rates, respectively.

RESULTS

Lip Movement

Results of the electromyographic analyses for both subjects are summarized in Table 2. This table shows the peak activity levels of the orbicularis oris muscle for all utterances at both speaking rates. The orbicularis oris muscle is largely rssponsible for a closing gesture of the lips and its activity, as shown here, is associated with lip closure during the production of the consonant. In Table 2, C_1 represents the first consonant (either /p/ or /w/) and C_2 represents the final /p/ in the utterance.

Generally speaking, the data summarized in this table show that the peak muscle activity levels of the orbicularis oris are greater during the fast speaking rate condition than during the slow speaking rate condition. These differences,

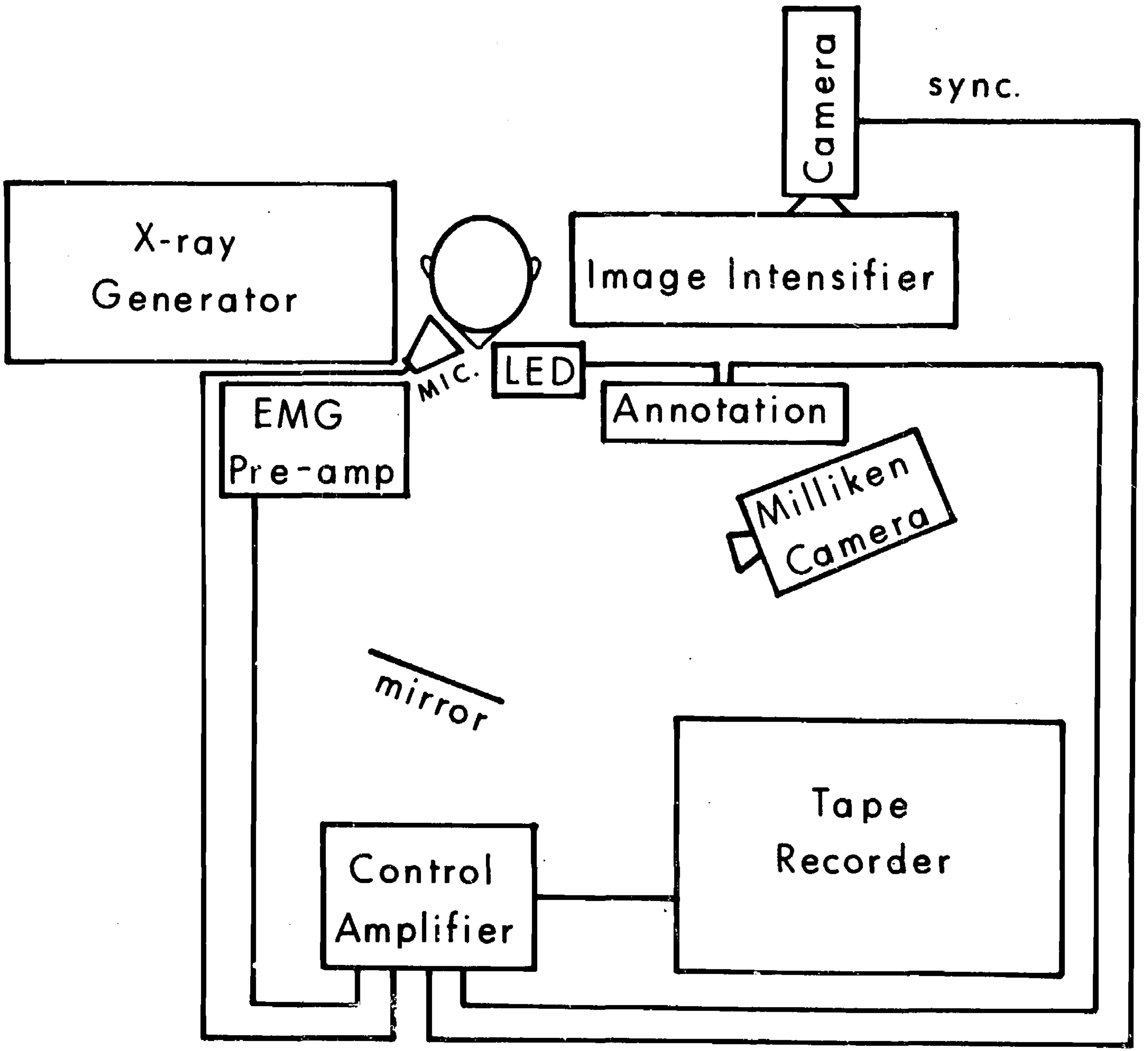


Figure 1: Block diagram of the electromyographic, cinefluorographic, and direct, high-speed motion picture recording systems.

TABLE 2: Averaged and single token (in parentheses) peak EMG values (μv) for the orbicularis oris muscle. Values for the slow speaking rate are in the left column and values for the fast speaking rate are in the right column of each cell. An asterisk (*) indicates higher values for the slow speaking rate condition.

	Subject FSC		Subject TG	
	S C1 F	S C2 F	S C1 F	S C2 F
ipip	120-185 (110-195)	130-140 (90-110)	205-280 (220-270)	200-210 (180-240)
ipap	150-205 (115-120)	125-165 (80-90)	225-325 (210-285)	240-300 (225-290)
ipup	175-215 (150-175)	135-150 (145-170)	215-345 (190-305)	185-345 (165-315)
apip	140-220 (130-145)	140-145 (80-85)	235-270 (235-265)	215-225 (220-235)
apap	145-225 (120-125)	130-155 (120-190)	220-270 (215-295)	245-260 (240-255)
apup	170-205 (140-150)	145-145 (120-200)	220-270 (205-310)	175-255 (175-250)
upip	130-185 (120-180)	130-145 (100-160)	135-290 (155-240)	210-260 (200-250)
upap	150-240 (95-200)	120-150 (90-95)	155-245 (145-230)	100-210 (105-195)
upup	110-265 (115-205)	140-170 (95-95)	165-245 (145-250)	180-195 (180-195)
iwip	150-230 (115-175)	130-170 (115-95)*	175-325 (180-300)	225-235 (230-230)
iwap	150-240 (150-160)	130-165 (95-95)	195-315 (175-295)	250-215* (240-230)*
iwup	200-275 (130-220)	165-170 (95-110)	160-315 (160-295)	200-290 (190-290)
awip	155-195 (130-225)	125-155 (95-70)*	13[illegible]-245 (140-240)	220-205* (235-215)*
awap	175-205 (165-175)	125-155 (120-85)*	135-240 (120-235)	230-190* (235-225)*
awup	190-220 (160-200)	165-165 (155-155)	155-275 (165-230)	225-255 (225-240)
uwip	155-190 (120-240)	125-140 (95-105)	105-205 (95-180)	215-205* (205-190)*
uwap	150-245 (150-265)	125-170 (105-90)*	100-235 (115-220)	240-215* (255-200)*
uwup	160-180 (160-220)	140-150 (95-85)*	85-240 (110-205)	190-270 (200-250)

with only a few exceptions,[1] hold for both C_1 and C_2 and for the single tokens (from the combined run) as well as the averaged tokens. The magnitude of these increases, however, varies from nil to over 100 percent; the only consistent trend is for C_1 differences to be greater than C_2 differences. This, of course, is a probable stress effect. On the whole, these data, like those from a previous study (Gay and Hirose, 1973), demonstrate that the major effect of an increase in speaking rate for labial consonant production is an increase in articulatory effort.

A compatible result was obtained from the frame-by-frame analysis of the direct-view, high-speed motion pictures. Figure 2 shows the vertical lip opening measurements for the /a/ series of utterances. These data show that the rates of lip closing and opening for the consonant are slightly faster for the faster speaking rate condition. These effects occur for both C_1 and C_2 and for all vowels, with the exception of the /upu/ and /uwu/ series for Subject TG. The increase in articulatory speed for the consonant is also carried over to the adjacent vowel as greater lip opening. Lip closure duration is somewhat variable across changes in speaking rate (Table 3), although in most cases it decreases with an increase in speaking rate.

TABLE 3: Lip closure durations (msec) for p_1, rounded to the nearest 5 msec.

	Subject FSC		Subject TG	
	Slow	Fast	Slow	Fast
ipi	80	60	80	60
ipa	100	100	100	80
ipu	90	60	100	70
api	100	70	60	60
apa	70	70	80	80
apu	110	80	80	70
upi	120	80	110	70
upa	110	90	100	80
upu	80	70	90	60

The effects that occur for /p/ also occur for /w/, i.e., an increase in speaking rate is accompanied by both an increase in the activity level of the muscle and an increase in the speed of movement of the lips. Further, the target configuration of the lips (minimum lip opening) remains constant across both

---

[1] The instances where an increase in muscle activity level was associated with the slower speaking rate condition occurred for 10 utterances out of a total of 144. Reversals of the expected result were generally small and occurred only for the final /p/ in the /w/ utterances.

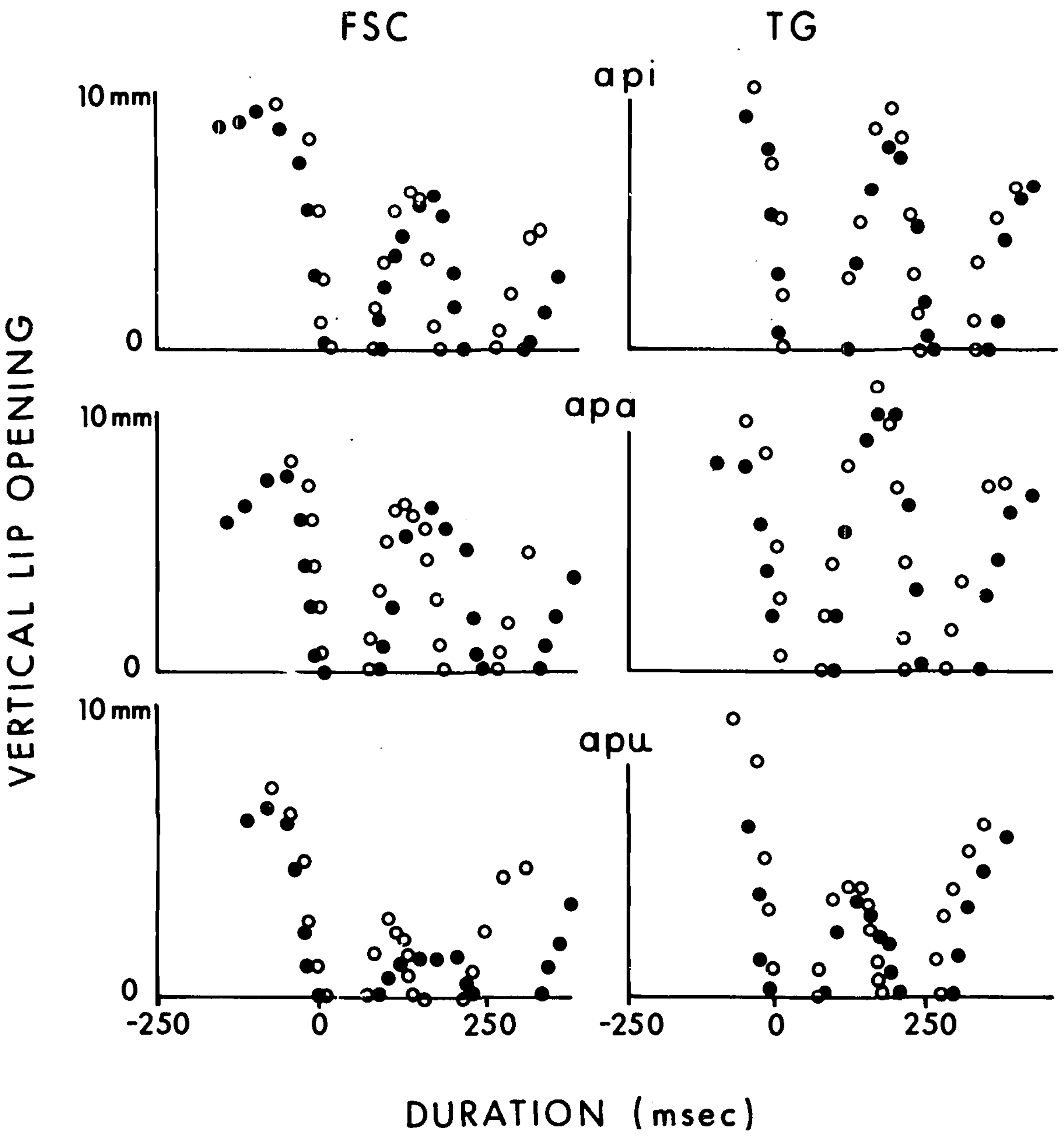


Figure 2: Vertical lip opening measurements. '0' on the abscissa represents the time of lip closure for /p/. The slow speaking rate is represented by filled circles and the fast speaking rate by unfilled circles.

speaking rates. In other words, the lips do not undershoot the /w/ target during faster speech. An example of the /w/ curves is shown in Figure 3.

The EMG data also show both contextual and individual differences in labial consonant production. For those utterances where both consonants are /p/, the EMG data for Subject FSC show higher peaks for p_1 than p_2; however, for Subject TG, the opposite is usually the case: p_2 peaks are usually (although not always) higher than p_1 peaks. For both subjects, muscle activity differences are conditioned more by speaking rate than by position. In other words, the differences in muscle activity levels as a function of speaking rate are greater than differences in muscle activity levels associated with position (C_1 vs C_2). Further, these data do not show any consistent vowel effects on muscle activity levels for consonant closure, although displacement differences were evident in the films (lip opening was greatest for /a/, less for /i/, and least for /u/). This, of course, might be attributed to a trade-off between displacement (/a/ and /i/) and degree of rounding (/u/), or to a trade-off between jaw opening and lip opening.

In summary, both the EMG and motion picture data show that the major effect of an increase in speaking rate on the production of a labial consonant is an increase in articulatory effort and a corresponding increase in the speed of articulatory movement. Both effects imply a reorganization of the commands to the articulators as well as a change in the timing of those commands.

Tongue Movement

The EMG data for the genioglossus muscle of Subject FSC are summarized in Table 4.[2] The genioglossus muscle, which makes up the bulk of the tongue body, is responsible for both protruding and bunching movements of the tongue.[3] The data in this table show that the activity levels of the genioglossus muscle decrease during faster speech. This decrease occurs for all utterances (except for /a/, where the genioglossus muscle shows only resting potentials). The magnitude of these differences is clearly vowel dependent, with the greatest differences occurring for /i/, less for /u/, and, of course, none for /a/. The activity patterns of the muscle at each speaking rate show large and consistent vowel effects in the same directions. This latter finding is a rather common pattern that has been shown before (Smith, 1970; Harris, 1971). Both the vowel and speaking rate effects occur systematically and hold up across /p/ and /w/ as well as across both the first and second vowels.

The X-ray films clearly reflect the decrease in muscle activity during faster speech. Figure 4 shows the position of the tongue (Subject FSC) at the V_1

---

[2] This electrode location was unusable for Subject TG.

[3] The genioglossus muscle is classically divided into two muscle groups. The anterior fibers fan from the mandibular tubercle to subsurface points along the length of the tongue. The posterior fibers run longitudinally from the mandibular tubercle to the hyoid bone. Depending upon the precise location of the recording electrodes, different response patterns can be observed for different features. Based on the specific patterns of activity observed for this location, it is assumed that the electrodes were placed in the posterior or lower anterior fibers of the muscle.

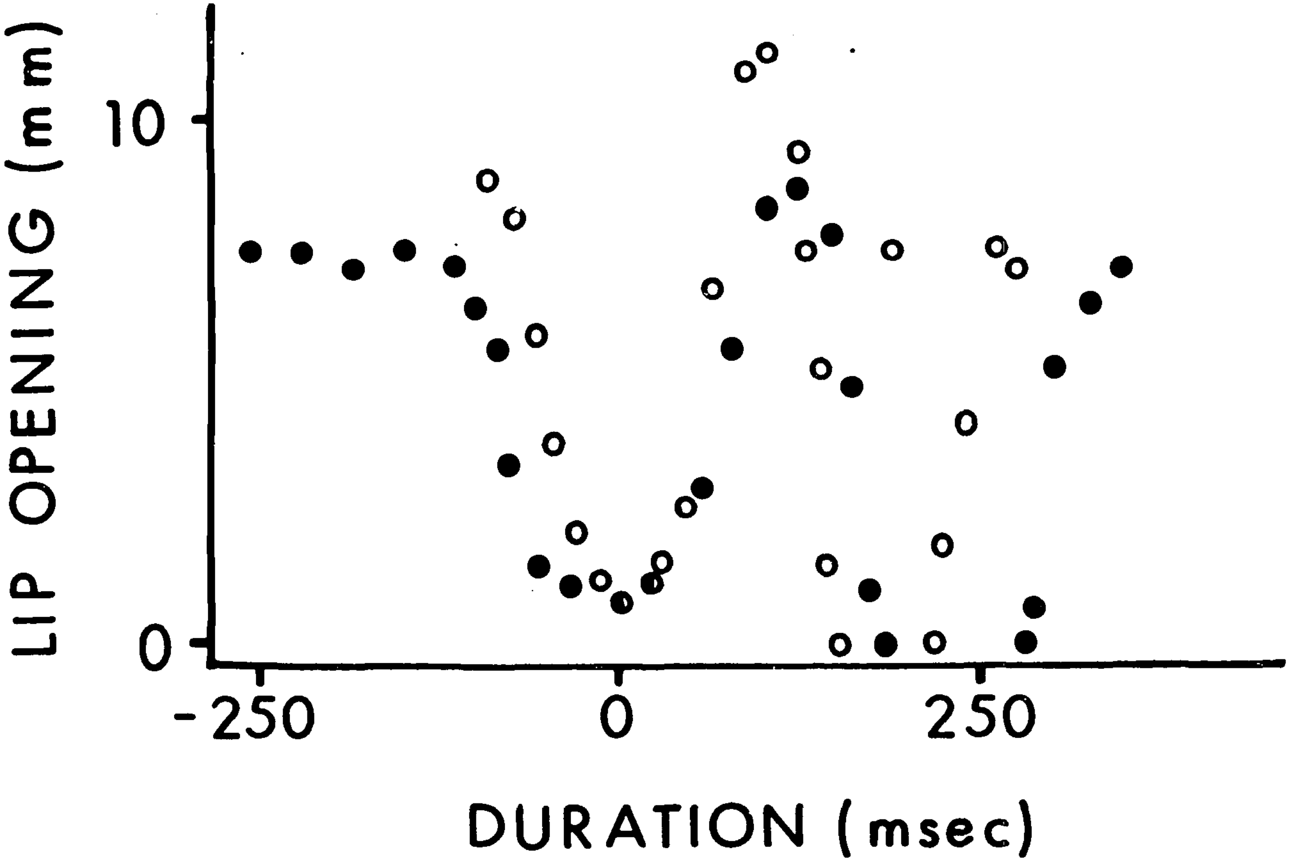


Figure 3: Vertical lip opening measurements for /iwap/ (Subject TG). 'O' on the abscissa equals the *point of minimum lip opening*. The slow rate is represented by filled circles and the fast rate by unfilled circles.

TABLE 4: Averaged and single token (in parentheses) peak EMG values (μv) for the genioglossus muscle for subject F.S.C. Values for the slow speaking rate condition are in the left column and values for the fast speaking rate condition are in the right column of each cell.

	Subject FSC	
	S V1 F	S V2 F
ipip	355-205	335-275
	(450-405)	(490-370)
ipap	325-105	-
	(440-195)	-
ipup	360-180	155-125
	(495-285)	(310-200)
apip	-	350-265
	-	(495-480)
apap	-	-
	-	-
apup	-	160-130
	-	(220-210)
upip	230-125	315-300
	(275-180)	(425-310)
upap	185-105	-
	(240-130)	-
upup	195-105	165-165
	(205-150)	(200-195)
iwip	295-180	305-285
	(380-340)	(370-360)
iwap	295-110	-
	(360-210)	-
iwup	305-175	165-145
	(410-390)	(265-230)
awip	-	310-300
	-	(490-350)
awap	-	-
	-	-
awup	-	165-165
	-	(420-400)
uwip	130-90	350-315
	(195-175)	(485-460)
uwap	140-110	-
	(160-150)	-
uwup	125-85	200-150
	(175-145)	(245-190)

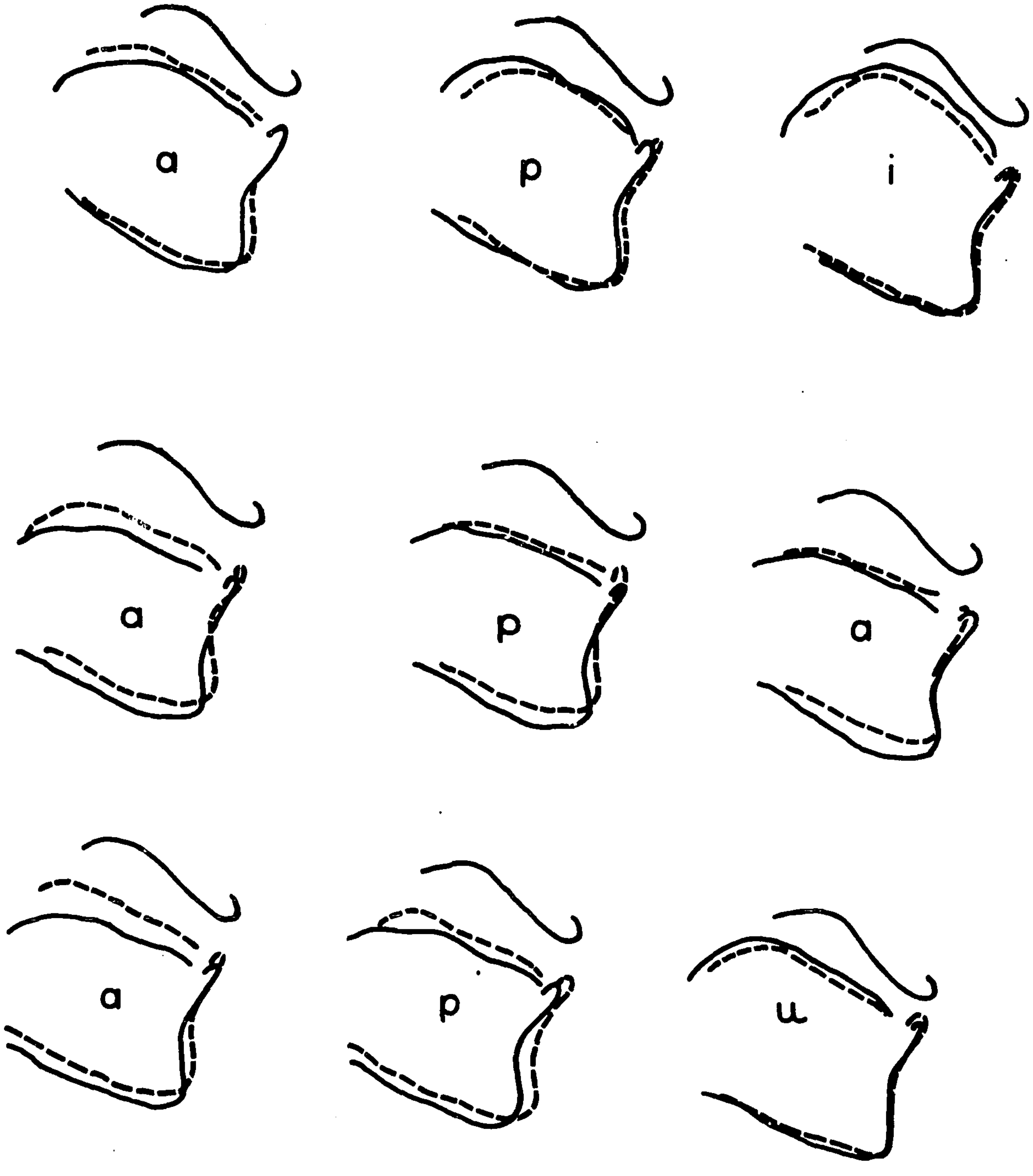


Figure 4: X-ray tracings (Subject FSC) for V_1 target, time of lip closure for /p/, and V_2 target, for slow (solid line) and fast (dashed line) speaking rates.

target, the time of lip closure for /p/, and the V_2 target for the /a/ series of utterances at both speaking rates. It is evident from these tracings that the tongue does not extend as far for the vowel during fast speech as it does during slow speech; it clearly undershoots its target. Tongue undershoot cocurs consistently for both subjects (although with a greater magnitude for FSC) and for both V_1 and V_2. This shortened course of movement is also reflected in the position of the tongue at the time of /p/ closure. During fast speech the tongue is in a lagging (or more neutral) position compared to the tongue during slow speech. These positional relationships are most obvious for the /i/-/a/ contrasts, apparently because of their more directly opposite target positions.

Although it is apparent that during faster speech the tongue follows a shorter, more restricted course from vowel to vowel, the film data could not be used to quantify articulatory rates of movement because of the measurement technique used. However, coupled with the EMG data for Subject FSC, the film data suggest that articulatory rates of movement might very well remain constant across changes in speaking rate. This assumes, of course, that the decrease in EMG activity reflects only a decrease in articulator displacement and not a concurrent decrease in articulator velocity. This suggestion is in agreement with the findings of both Lindblom (1964) and Kent (1970), whose data show little effect of speaking rate on articulatory velocity. However, such a statement is not necessarily universal. For example, Kuehn (1973) has shown that different speakers seem to use different strategies in the control of speaking rate. His data show that some speakers increase speaking rate by increasing articulatory velocity (with a corresponding decrease in amount of undershoot), and others, by decreasing articulatory displacement (with a corresponding decrease in articulatory velocity).

In this experiment, the data for tongue movement differ markedly from those for lip movement. Whereas the tongue shows a decrease in muscle activity and target undershoot during faster speech, lip movement is characterized by an increase in muscle activity and an increase in articulatory speed. The question then arises as to whether such differences are phoneme related or muscle system related. The EMG data on lip rounding for /u/ bear directly on this question. Table 5 shows the averaged peak muscle activity levels of the orbicularis oris during lip rounding for /u/. These figures show that, for both subjects and all utterances, the lip rounding gesture for /u/ is characterized by consistently higher peaks of muscle activity during the faster speaking rate condition. Thus, it would appear that for the phoneme sequences studied here, the observed vowel-consonant differences are, in fact, tongue-lip differences.

TABLE 5: Averaged peak EMG values (μv) of the orbicularis oris muscle for lip rounding of /u/.

	Subject FSC		Subject TG	
	Slow	Fast	Slow	Fast
uti	120	150	80	145
uta	115	165	80	150
utu	145	185	75	140

The X-ray tracings can also serve to illustrate the precision with which vowel targets are attained, irrespective of preceding or following vowels. Figure 5 shows vowel target positions of the tongue for the /a/ series of utterances. Although the routes taken by the tongue toward these targets vary considerably, the final target position is consistently stable. This finding, which is in agreement with whose of Houde (1967) and MacNeilage and DeClerk (1969), characterizes the V_1 and V_2 targets for /p/ as well as for /w/.

Jaw Movement

The EMG data for the anterior belly of the digastric muscle are summarized in Table 6.[4] The anterior belly of the digastric acts to open the jaw; thus, the values shown in this table are associated with jaw opening for V_1 and V_2.

The muscle activity levels of the anterior belly of the digastric show an increase during fast speech. This increase is consistent (small in magnitude for Subject FSC, large in magnitude for Subject TG), but occurs only for the first vowel. For both subjects, this muscle does not show much more than resting potentials for V_2. It is possible, of course, that these peaks reflect, at least in part, a stabilizing gesture for /k/. However, the timing of the peaks is compatible with the vowel, and further, the data for TG show a consistent vowel effect, i.e., large peaks and greater speaking rate differences for /a/-/u/-/i/, increasing in that order.

Figure 6 illustrates the effect of speaking rate on jaw movement. For the vowel, jaw displacement is usually greater during slow speech than during fast speech, but for the consonant, jaw displacement is usually less during fast speech. In other words, jaw movement during fast speech mirrors jaw movement during slow speech, but with lower absolute displacement values. These data are not consistent with recent findings of Abbs (1973), whose measurements for /pip/ and /pæp/ show little effect of speaking rate on jaw displacement.

Figure 6 also shows that for the opening segment of the consonant, jaw opening consistently leads lip opening (lip opening is shown by the vertical lines on each graph). This lead effect, which is evident for both subjects, will be described more fully in the following section. Finally, this figure illustrates that the effect of speaking rate on jaw velocity varies with context. Although a trend seems to exist for rate of movement to increase during faster speech for /a/, and sometimes for /i/, no such trend is apparent for /u/. The effects for /a/ occur most consistently while effects for /i/ are more variable. The absence of any effect for /u/ is obviously because jaw movement for /u/ is miminal at both speaking rates.

Jaw movement is also subject to certain anticipatory coarticulation effects. Figure 7 shows the jaw displacement curves for all /p/ utterances at the slow speaking rate. These curves are arranged so that the data for the first vowel

---

[4] The internal pterygoid muscle was also studied for both subjects. However, for Subject TG this muscle did not show any activity during speech (although it was active for clenching), and for Subject FSC activity was present only during closure for initial /k/.

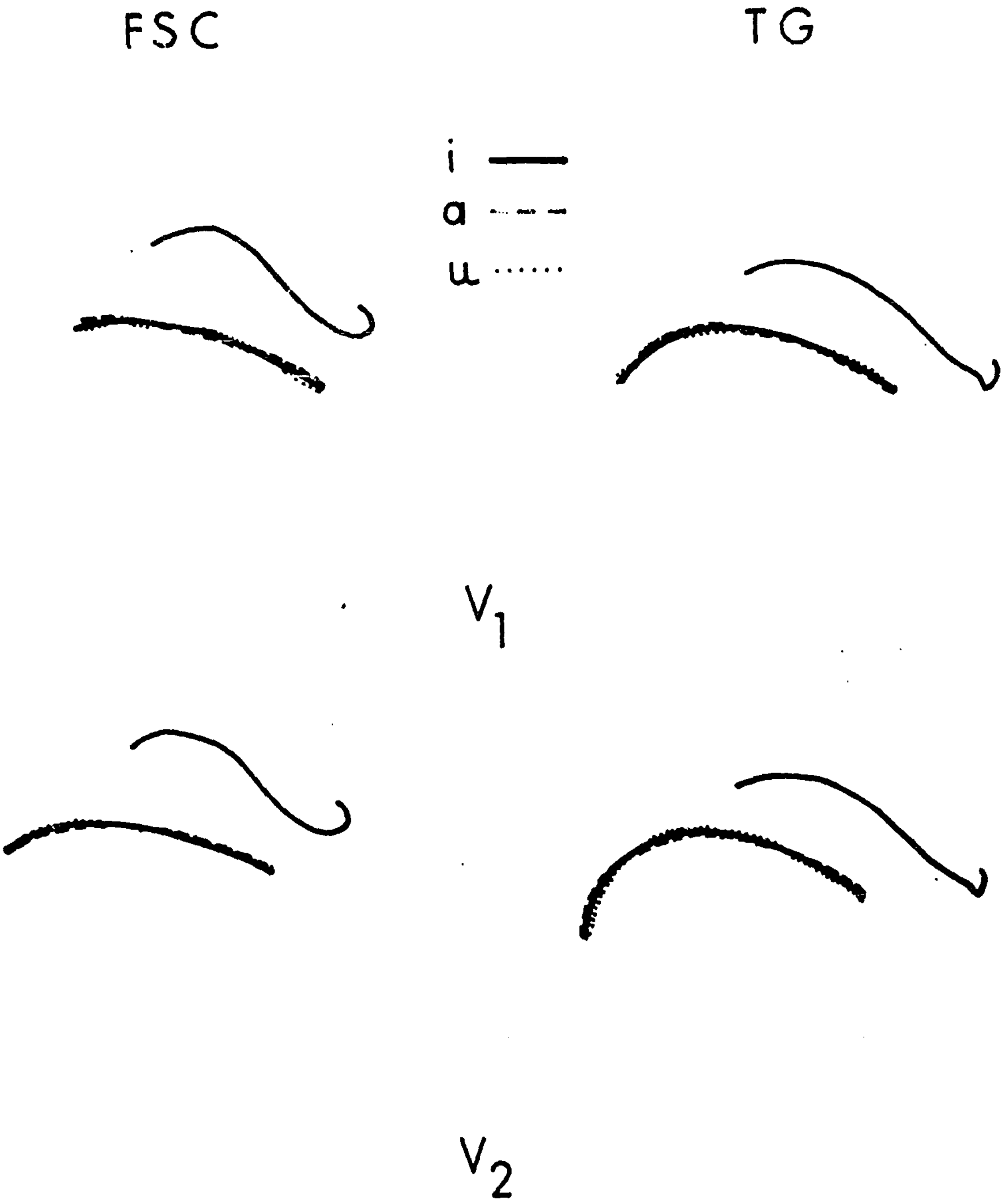


Figure 5: X-ray tracings for the /a/ series of utterances. V_1 represents /apV/ and V_2 represents /Vpa/, both at slow speaking rates.

TABLE 6: Averaged and single token (in parentheses) peak EMG values (μv) for the anterior belly of the digastric muscle. Values for the slow speaking rate condition are in the left column and values for the fast speaking rate condition are in the right column of each cell. An asterisk (*) indicates higher values for the slow speaking rate condition.

	Subject FSC		Subject TG	
	S V1 F	S V2 F	S V1 F	S V2 F
ipip	110-135	-	120-280	15-15
	(90-135)	-	(110-130)	(30-40)
ipap	115-120	-	120-240	30-75
	(130-135)	-	(120-260)	(25-30)
ipup	110-110	-	115-265	25-35
	(125-180)	-	(100-265)	(30-35)
apip	105-170	-	165-400	30-45
	(145-220)	-	(170-420)	(25-25)
apap	135-155	-	140-375	25-35
	(120-165)	-	(130-370)	(20-30)
apup	120-130	-	200-385	30-30
	(95-180)	-	(185-375)	(30-35)
upip	115-125	-	130-290	25-95
	(115-125)	-	(125-265)	(30-40)
upap	120-30*	-	120-305	30-85
	(105-25)*	-	(115-275)	(30-75)
upup	115-140	-	125-285	30-50
	(125-140)	-	(125-295)	(25-30)
iwip	125-145	-	135-255	30-60
	(110-140)	-	(110-230)	(30-45)
iwap	120-125	-	115-250	35-70
	(125-200)	-	(140-230)	(30-60)
iwup	110-125	-	125-260	30-35
	(115-160)	-	(145-255)	(45-50)
awip	125-125	-	190-390	30-55
	(135-140)	-	(170-370)	(25-60)
awap	115-140	-	160-385	25-45
	(125-130)	-	(135-350)	(25-40)
awup	115-110*	-	175-390	25-35
	(120-150)	-	(170-410)	(35-40)
uwip	105-130	-	135-375	25-95
	(115-115)	-	(125-410)	(25-90)
uwap	110-115	-	135-300	25-70
	(95-115)	-	(140-300)	(25-90)
uwup	105-135	-	140-330	25-50
	(100-110)	-	(165-335)	(35-40)

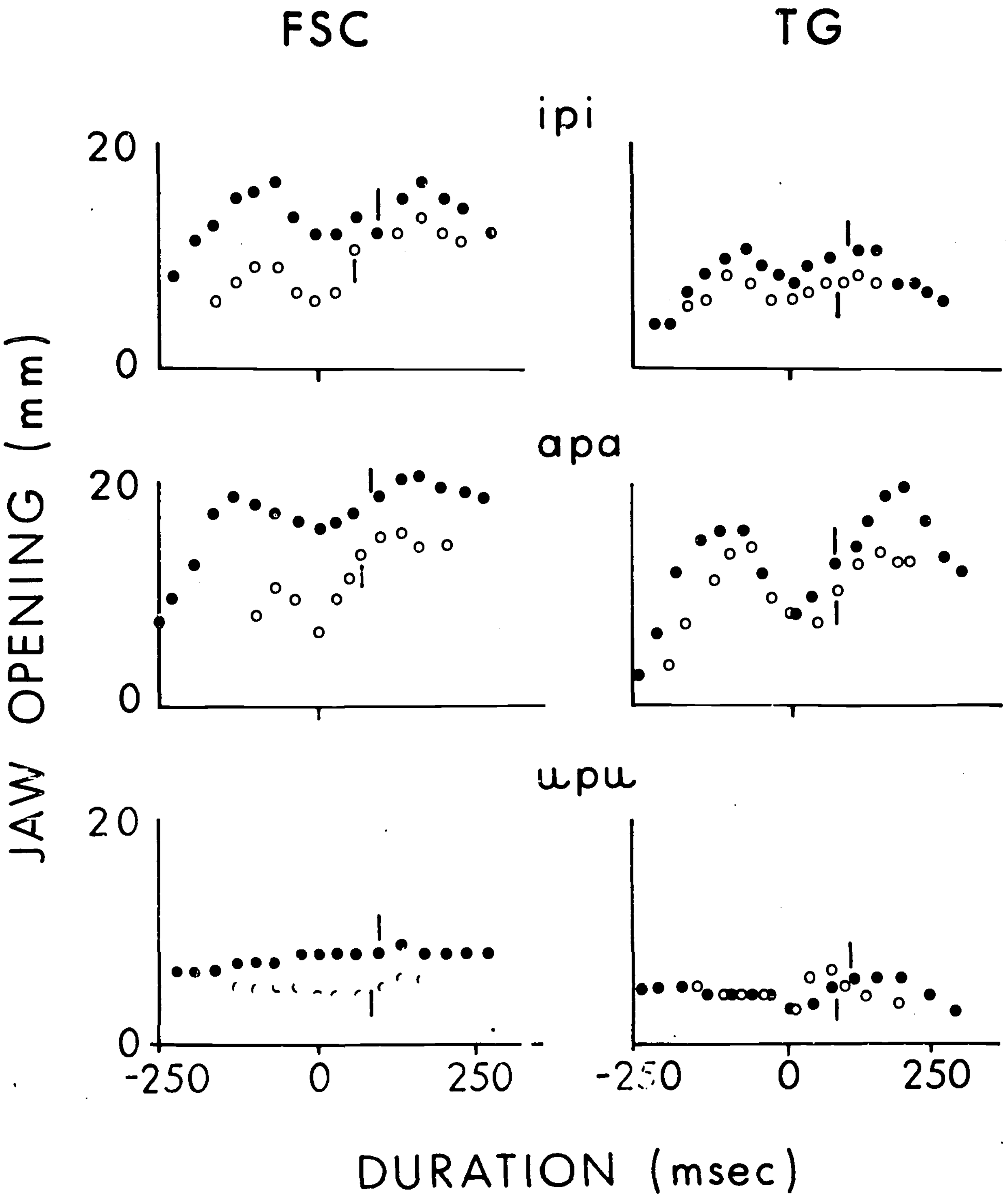


Figure 6: Jaw displacement measurements for slow (filled circles) and fast (unfilled circles) speaking rates. '0' on the abscissa refers to time of lip closure for /p/ and the short vertical lines indicate time of lip opening for /p/.

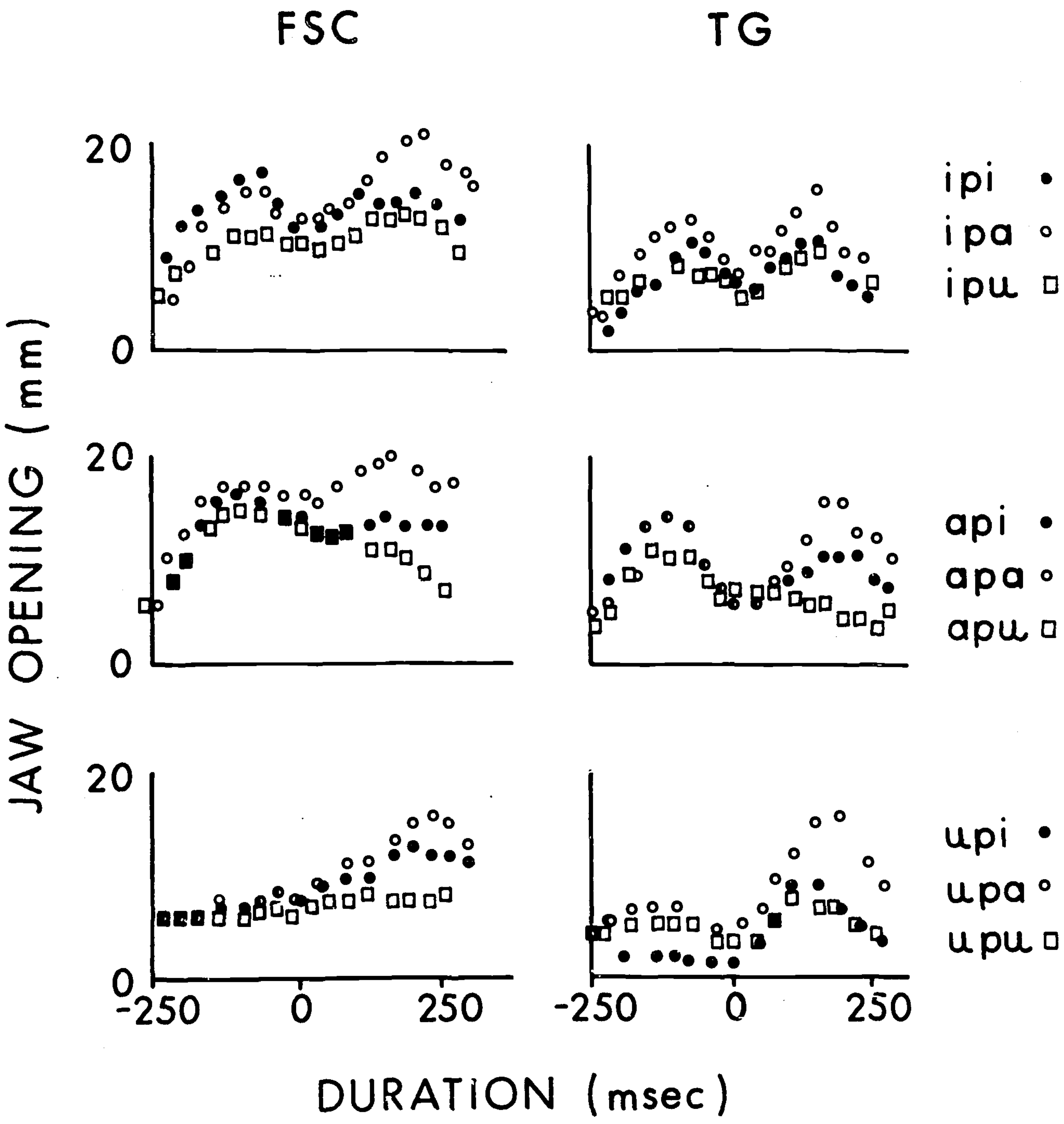


Figure 7: Jaw displacement measurements for all /p/ utterances at the slow speaking rate. '0' on the abscissa indicates time of lip closure for /p/.

in the VCV sequence can be compared to differences in the second vowel. Except for the /u/ series for Subject TG, jaw displacement for V_1 is least when V_2 is /u/. Also, in most cases, the first vowel has relatively little effect on jaw position at the time of /p/ closure for Subject FSC, and almost no effect for Subject TG. Displacement differences for the vowel do not occur consistently between /i/ and /a/, although there is a trend for greater jaw displacement for V_1 when V_2 is /a/ rather than /i/.

Figure 8 shows these data replotted so that second-vowel comparisons can be made, i.e., the effect of a different first vowel on the displacement for the same second vowel. For these comparisons, individual differences are evident. For Subject FSC, large /u/ effects on jaw displacement for the second vowel occur consistently, i.e., jaw displacement for the second vowel is least when the first vowel is /u/. These curves also show an effect of the first vowel on jaw closure for /p/. Here systematic differences in degree of jaw closing for the consonant occur, with greater closure for /u/-/i/-/a/, increasing in that order. These effects are not so apparent for the second vowel (the comparisons in Figure 6). For Subject TG, on the other hand, there are essentially no effects of the first vowel on jaw displacement for the second vowel. Also, there do not seem to be any anticipatory effects for the second vowel at the time of /p/ closure. The jaw movement data for /w/ are essentially the same as for /p/, i.e., they show reduced displacement for the vowel and increased closing for the consonant during faster speech, and decreased displacement for a vowel when either proceeded or followed by /u/.

In summary then, it would seem that jaw movement during fast speech is characterized by a pattern similar to that for slow speech, but with a decrease in overall displacement. Changes in the velocity of jaw movement, when they occurred, did so primarily for /a/. The movement data were only partially supported by the EMG data (for Subject TG, where greater activity levels for the anterior belly of the digastric muscle correlated with the increase in velocity of the jaw for /a/). Of course, too, this experiment sampled only two of the muscles involved in movements of the jaw; this in itself, plus the absence of any activity for the second vowel in the test utterances, clearly indicates that a more complete muscle inventory is needed for an adequate description.

Coordination of Lip, Tongue, and Jaw Movements

Table 7 summarizes timing information for movements of the tongue, jaw, and lips during the closing, closed, and opening segments of the consonant. The data in this table are the relative onset times (lip closure for /p/ or minimum lip opening for /w/ = 0) of tongue movement from the first vowel, and lip and jaw opening and closing for the consonant, at both speaking rates.

For both subjects and both speaking rates, the onset of jaw closing for /p/ lags behind the onset of both tongue movement and lip closing. The onset time of jaw closing from /a/ is earlier than the onset times of either /i/ or /u/. This is probably because jaw displacement for /a/ is greater than that for the other two vowels. The lips begin to close approximately 75 msec ahead of tongue movement when the vowel preceding the consonant is /u/. This is apparently part of the lip-rounding gesture for /u/.

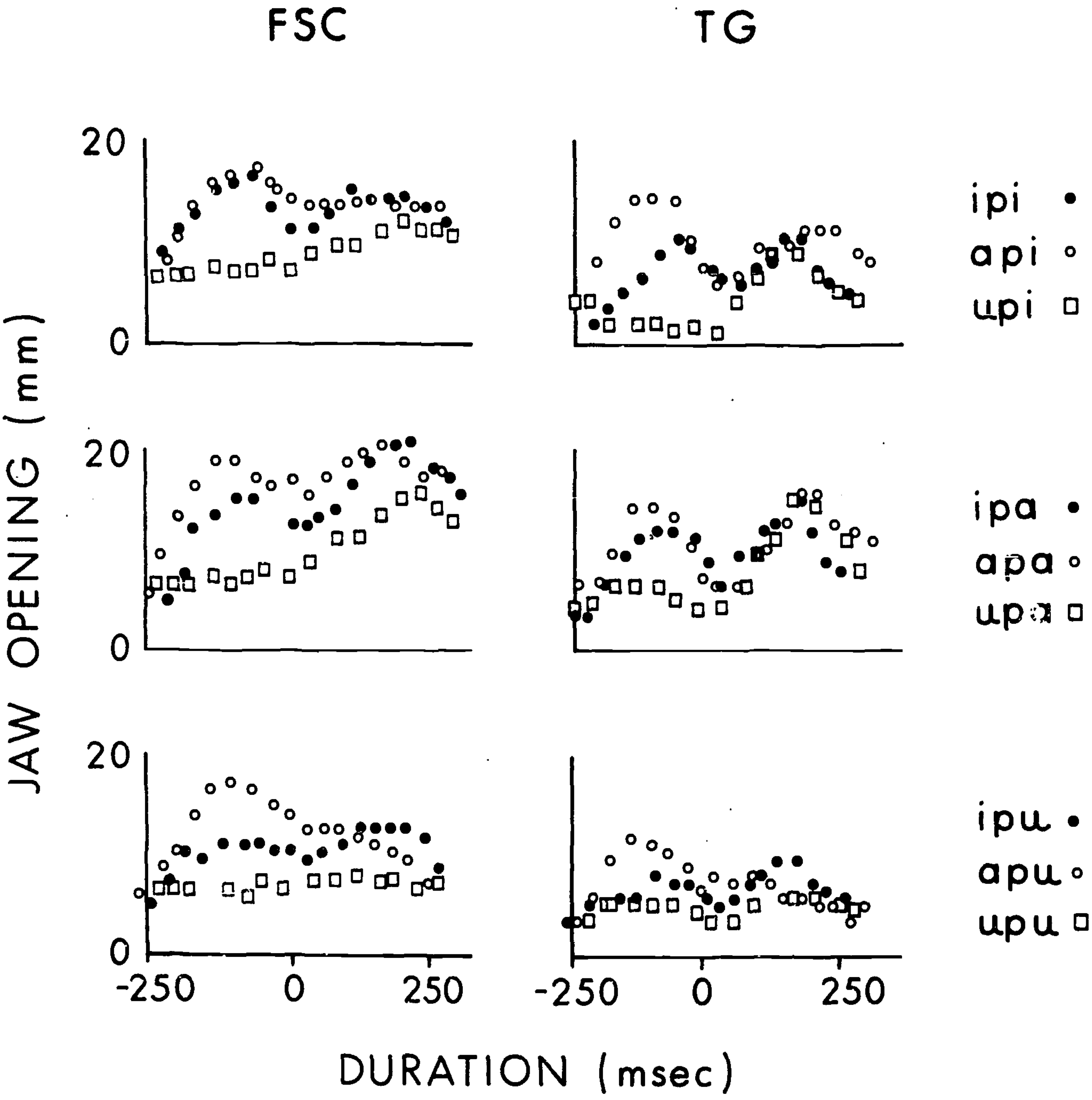


Figure 8: Jaw displacement curves for /p/, replotted for second vowel comparisons.

TABLE 7: Relative onset times (in msec) of tongue, lip, and jaw movements (lip closure for /p/ or minimum opening for /w/ = 0) for the closing, closed, and opening segments of the intervocalic consonant. Values for the fast speaking rate appear below those for the slow speaking rate. All values are averaged over the second vowel.

	Subject FSC								Subject TG						
	Closing			Closed		Opening			Closing			Closed		Opening	
	Tongue	Lips	Jaw	Lips	Jaw	Lips	Jaw		Tongue	Lips	Jaw	Lips	Jaw	Lips	Jaw
ip	- 80	- 80	- 45	0	15	90	35	ip	-100	- 95	-55	0	10	95	20
	- 50	- 75	- 35	0	0	75	20		- 60	- 70	-45	0	0	70	20
ap	- 80	- 95	- 85	0	20	95	45	ap	- 80	- 75	-70	0	10	75	20
	- 40	- 70	- 50	0	10	70	25		- 55	- 65	-50	0	10	70	20
up	- 70	-165	- 30	0	0	100	20	up	- 95	-175	-25	0	0	100	30
	- 50	-135	- 30	0	0	80	20		- 45	-100	-20	0	0	75	20
iw	-155	-135	-100	0	-10	45	40	iw	-105	-125	-80	0	-20	45	35
	-110	- 95	- 75	0	-10	30	35		-100	-100	-75	0	-10	20	25
aw	-120	-155	-120	0	0	70	55	aw	-125	-135	-95	0	0	40	45
	- 65	-110	- 90	0	0	55	45		- 75	- 90	-80	0	-10	30	30
uw	-110	-210	- 95	0	0	120	60	uw	-120	-195	-75	0	0	45	40
	- 75	-140	- 80	0	0	80	40		- 75	-130	-65	0	0	25	30

Jaw closing is usually completed at or slightly after lip closure for /p/, while jaw opening for the following vowel precedes lip opening by approximately 50 msec. Except for shorter lead and lag times, similar patterns emerge for fast speech.

Segment durations are somewhat longer for /w/ than /p/ (50 msec for Subject FSC, 25 msec for Subject TG). Also, the onset times of lip movement for /w/ (with two exceptions) are earlier than those for tongue movement. Jaw closing is completed at or slightly ahead of minimum lip opening, while jaw opening begins at about the same time as lip opening. Closure duration is shorter for /w/ than /p/.

The absence of an anticipatory movement of jaw opening during closure for /w/ contradicts recent data of Gay and Hirose (1973), who showed that for /w/ jaw movement was independent of lip movement, anticipating a following vowel by opening for it during lip closing. This effect was not evident, of course, in the present X-ray data, the examination of which easily explains the discrepancy. In their experiment, Gay and Hirose measured jaw displacement by the movement of a marker painted on the chin. The present X-ray films, however, show that flesh points directly over the mandible can move independently from, and even in opposite directions to, the mandible itself. Thus, indirect measurements such as the above can produce seemingly accurate, but in fact erroneous data. It is also possible that similar errors are inherent in strain gage measurements, especially if the transducer is positioned at the level of the mandibular protuberance.

The major results of this experiment can be summarized as follows. For lip movements associated with either labial consonant production or rounding for the vowel, an increase in speaking rate is accompanied by an increase in the activity level of the muscle and by slightly faster rates of movement. For tongue movement during vowel production, an increase in speaking rate has the opposite effect: a decrease in the activity level of the muscle and a decrease in articulatory displacement. For jaw movement, the major effect of an increase in speaking rate is a decrease in the displacement of the jaw throughout the utterance, i.e., for both the vowel and consonant. Jaw movement is also more sensitive than either lip or tongue movements to changes in phonetic context and shows a lag effect for consonant closing and a lead effect for vowel opening.

DISCUSSION

The results of this experiment show that the control of speaking rate cannot be accounted for by one simple mechanism. As was shown in a previous experiment (Gay and Hirose, 1973), and confirmed here, the major effects of an increase in speaking rate on the production of a labial consonant is an increase in the activity level of the muscles and an increase in the rate of movement of the lips. Both of these effects are apparent consequences of an increase in articulatory effort. As was also mentioned before, a strategy of this type could be expected for a consonant gesture that involves an occlusal target. The articulators must approximate for a stop consonant.[5] Thus, it is reasonable to assume that under the constraints of an increase in speaking rate, they would do so somewhat faster.

---

[5] For /w/, the articulators apparently must reach an invariant target position in order to produce an acoustic steady state.

The data for the tongue (and to some extent the jaw), however, cannot be explained in the same way. For tongue movement, an increase in speaking rate is accompanied by a decrease in displacement (undershoot) and a decrease in the activity level of the muscle. The decrease in the activity level of the genioglossus muscle shows that undershoot is not, as Lindblom (1963) originally suggested, a consequence of an overlap in the timing of commands to the muscle.[6] Although the decrease in muscle activity for the tongue during faster speech might reflect only the decrease in overall displacement of the tongue (and not any changes in its speed of movement), the decrease itself is indicative of some reorganization taking place at the level of the muscle commands.

The EMG data for the tongue indicate that vowels are characterized by different targets for slow and fast speech. In other words, a vowel target cannot be operationally defined by a set of invariant spatial coordinates. Rather, a vowel target must be defined either by a multiple coordinate system or by an articulatory field (with limits). For slow speech, and perhaps for stressed vowels, one given set of coordinates is aimed for, while for fast speech, where articulatory expediency or the constraints of decreased jaw displacement place additional demands on the mechanism, a different set of coordinates is aimed for.

As mentioned above, Kuehn (1973) has shown that different speakers use different strategies to increase speaking rate, i.e., by trade-offs between displacement and velocity. These differences might also be explained in terms of a field or multiple coordinate system. Some individuals might be able to produce a given vowel with a greater degree of freedom than others. That is, the acoustical properties of a given vocal tract might be such that a wider range of formants can produce the same perceptual result. Other tracts might not have these characteristics and, thus, the articulators must (by increasing velocity) attain a more strictly defined set of target coordinates. However, irrespective of both the strategy employed and the effect observed, the crucial point is that the speaking rate control mechanism involves changes in both the timing and organization of commands to the muscles.

One final point which requires reiteration is the contrast between the variability of target position across changes in speaking rate versus the precision in attaining that same target in varied phonetic contexts. The larger speaking rate effects can be interpreted to mean that the greatest challenge to the system lies not in the organization of a target-directed movement, but rather, in the rapid sequencing of such movements.

In summary, the data of this experiment show that speaking rate is controlled by a mechanism that involves more than a simple reordering of the timing of motor commands. Reorganization of the gesture for fast speech involves changes in both the duration and size of the muscle contraction. However, an adequate description of the mechanism requires additional information about how the tongue handles vowel sequences that are separated by both lingual and velar consonants, along with a more detailed description of the variability of target position as a function of a greater number of speaking rate variations.

---

[6] Lindblom's original model implies that only the timing, not the size, of the EMG signal would change during faster speech.

REFERENCES

Abbs, J. (1973) The influence of the gamma motor system on jaw movements during speech: A theoretical framework and some preliminary observations. J. Speech Hear. Res. 16, 175-200.

Gay, T. (1968) Effect of speaking rate on diphthong formant movements. J. Acoust. Soc. Amer. 44, 1570-1573.

Gay, T. and H. Hirose. (1973) Effect of speaking rate on labial consonant production. Phonetica 27, 44-56.

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

Hirose, H. (1971) Electromyography of the articulatory muscles: Current instrumentation and technique. Haskins Laboratories Status Report on Speech Research SR-25/26, 73-85.

Houde, R. A. (1967) A study of tongue body motion during selected speech sounds. Ph.D. thesis, University of Michigan.

Kent, R. D. (1970) A cinefluorographic - spectrographic investigation of the component gestures in lingual articulation. Ph.D. thesis, University of Iowa.

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

Kewley-Port, D. (1973) Computer processing of EMG signals at Haskins Laboratories. Haskins Laboratories Status Report on Speech Research SR-33, 173-184.

Kuehn, D. P. (1973) A cinefluorographic investigation of articulatory velocities. Ph.D. thesis, University of Iowa.

Lindblom, B. (1963) Spectrographic study of vowel reduction. J. Acoust. Soc. Amer. 34, 1773-1781

Lindblom, B. (1964) Articulatory activity in vowels. Quarterly Progress Status Report (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden) STL/RIT 2, 1-5.

MacNeilage, P. F. and J. L. DeClerk. (1969) On the motor control of coarticulation in CVC monosyllables. J. Acoust. Soc. Amer. 45, 1217-1233.

Smith, T. (1970) A phonetic study of the function of the extrinsic tongue muscles. Ph.D. thesis, University of California, Los Angeles.

Stress and Syllable Duration Change*

Katherine S. Harris[+]
Haskins Laboratories, New Haven, Conn.

It is generally assumed that underlying each phoneme there is an invariant articulatory target. At a surface level, this statement is, of course, untrue. There is no moment when the articulators assume a position for a given speech sound--a position that is invariant over changes in the phonemic context and suprasegmental structure. Much of the effort of traditional articulatory phonetics was directed towards writing rules to describe observed differences in articulatory target as a change in allophone selection. Modern physiological research searches for simple rewrite rules to derive observed positional variants from some presumed underlying single articulatory target.

A carefully worked out theory of this sort is Lindblom's (1963) theory of vowel reduction, which was developed to account for the changes in vowel color that accompany changes in stress. If a vowel is destressed, it will tend to be of shorter duration, and to move in vowel color towards the neutral schwa; the latter phenomenon is called vowel neutralization. Lindblom's proposal is that the neutralization is a consequence of the accompanying shortening. Briefly, in a consonant-vowel-consonant (CVC) sequence, although the signals sent to the articulators are constant, the response of the articulators is sluggish. If signals arrive at the muscles too fast, the articulators will start towards the vowel target but will be deflected towards the subsequent consonant target--that is, there will be undershoot. Lindblom tested his theory by having subjects produce sentences containing CVC monosyllables. The effect of rearranging the sentences was to change the stress on one "word" and consequently to change the vowel duration. He made careful measurements of the most extreme positions of the first and second formants, as a function of the vowel length. He found that as vowels lengthened, the formants tended towards asymptotic values which could be described as targets. Equations could be written describing the relation of vowel duration to the departure of formant position from target.

Lindblom's theory seems to be elegant and testable, if one substitutes for "signals" the more specific "muscle contractions." A reformulation in electromyographic (EMG) terms would then perhaps be: "Under conditions of changing stress the EMG signals associated with a CVC sequence will remain constant in amplitude. Only the relative timing of vowel and consonant signals will change."

---

*Presented to the American Association of Phonetic Sciences, Detroit, Mich., 11 October 1973.

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

[HASKINS LABORATORIES: Status Report on Speech Research SR-35/36 (1973)]

Lindblom goes further than this. He assumes that changes in duration operate in the same manner whether they are due to stress or to speaking rate. Gay (1973) has been investigating some aspects of this formulation with respect to speaking rate. I will talk about stress.

As our model for changing stress, we constructed some two-syllable nonsense words, of the form /pVCVp/. The two middle vowels were always /i/, /a/, or /u/, and were always different. The first and last consonants were always /p/, and the middle consonant was /p/ or /k/. There was a neutral carrier on each end. Using this format, we examined EMG signals from several muscles involved in the articulation.

The genioglossus muscle will be discussed first. This is a large muscle making up a great part of the body of the tongue, acting to bunch it. Consequently, activity is always seen for the vowel /i/, usually is seen for /u/, and none is seen for /a/. No activity is seen for the consonants.

The effect of the change in stress is shown in Figure 1. (Time runs along the abscissa; each unit is 100 msec. Averaged energy is on the ordinate. The vertical line shows the point when the voicing for the first syllable ends.) The two curves in each graph show the two conditions of stress. In each graph the thin line shows the utterance with the first syllable stressed and the thick line shows the utterance with the second syllable stressed. The line rows show two different electrode insertions, both into the genioglossus. The curves show the two effects of stress usually found. First, there is a small difference in peak height between stressed and unstressed vowels (Harris, 1971).

The second effect is the "lineup" effect. Notice in the left column, that when the /i/ is stressed, the genioglossus curve begins earlier, but ends at the same time with respect to the offset of voicing. The effect is almost symmetrical when the /i/ is in the second syllable. In other words, the vowel activity lengthens, but dies off in a constant relationship to the offset of voicing.

Figure 2 illustrates the activity of the orbicularis oris muscle, which shows a burst of activity for the first, middle, and last /p/ in these utterances since it acts to close the lips. Note that the middle /p/ peaks over the offset of voicing, indicated by the vertical line. Again, the first /p/ moves leftwards, as the vowel lengthens, when the first syllable is stressed. The last /p/ moves rightwards, as the vowel lengthens, when the last syllable is stressed.

What does this mean? First, the activity for the vowel lengthens. Second, the time between consonant peaks changes systematically. Combining these facts, we get a picture of stress change illustrated by Figure 3.

This figure shows the relationship between orbicularis oris and genioglossus activity, for four disyllables. In all cases, the vowel activity begins as the initial consonant activity wanes. If the vowel is stressed, its activity continues for a longer period than if it is not stressed. The middle or terminal consonant begins as the vowel activity wanes. The vowel seems to lengthen literally--that is, associated muscle activity lasts longer. The temporal relationship of consonant and vowel activity seems to be fixed.

Lindblom's model, then, is wrong on two counts--it posits, first, that under conditions of changing stress the signals to the muscles will remain constant,

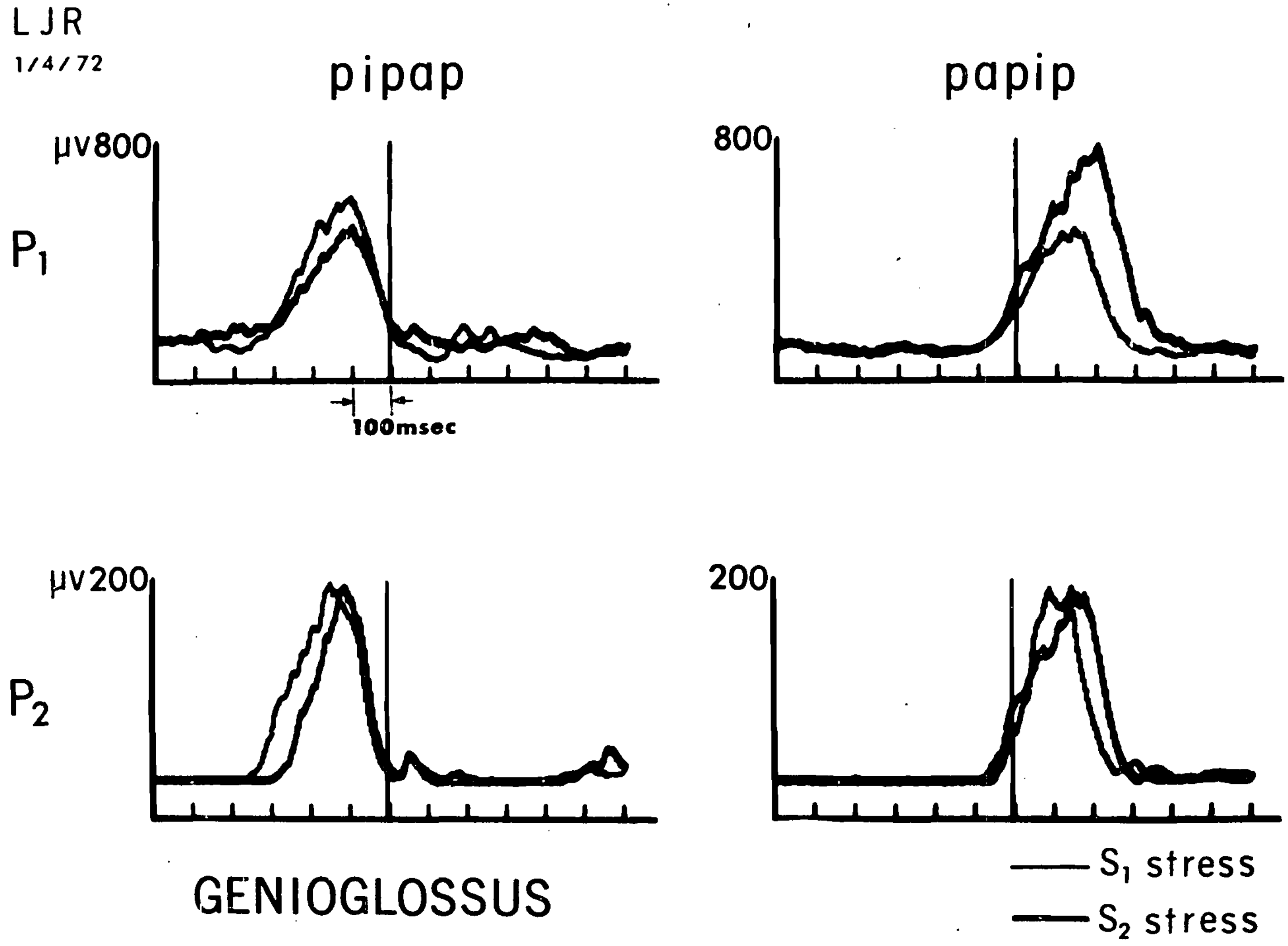
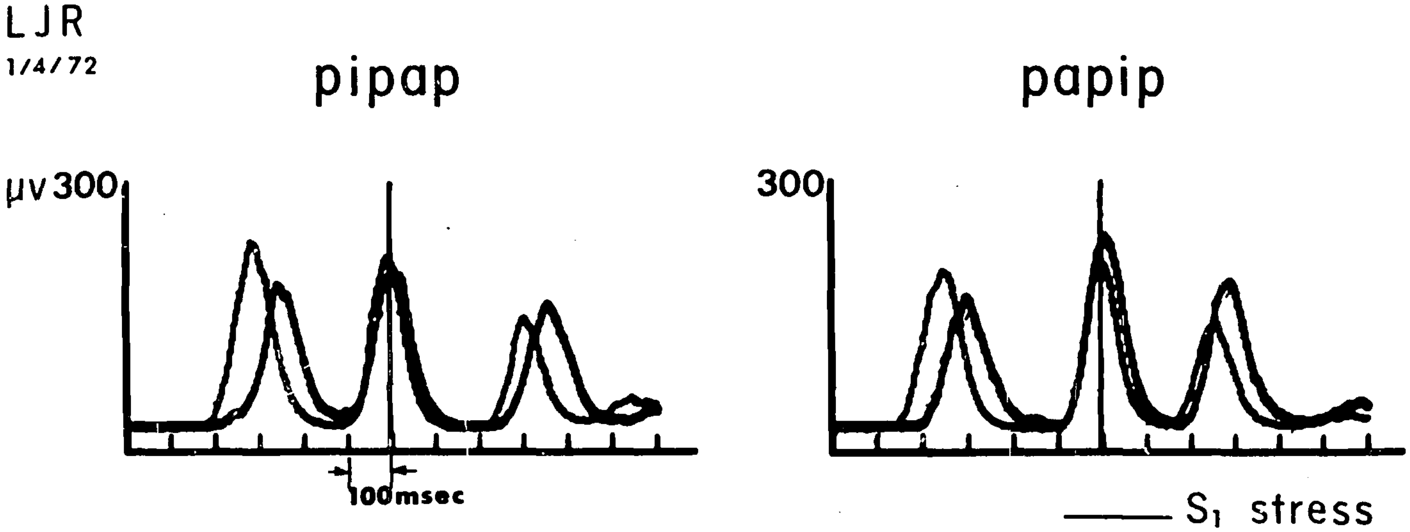

LJR
1/4/72
pipap
papip
μv800
800
P$_1$
100msec
μv200
200
P$_2$
GENIOGLOSSUS
S$_1$ stress
S$_2$ stress

Figure I



ORBICULARIS ORIS

Figure 2

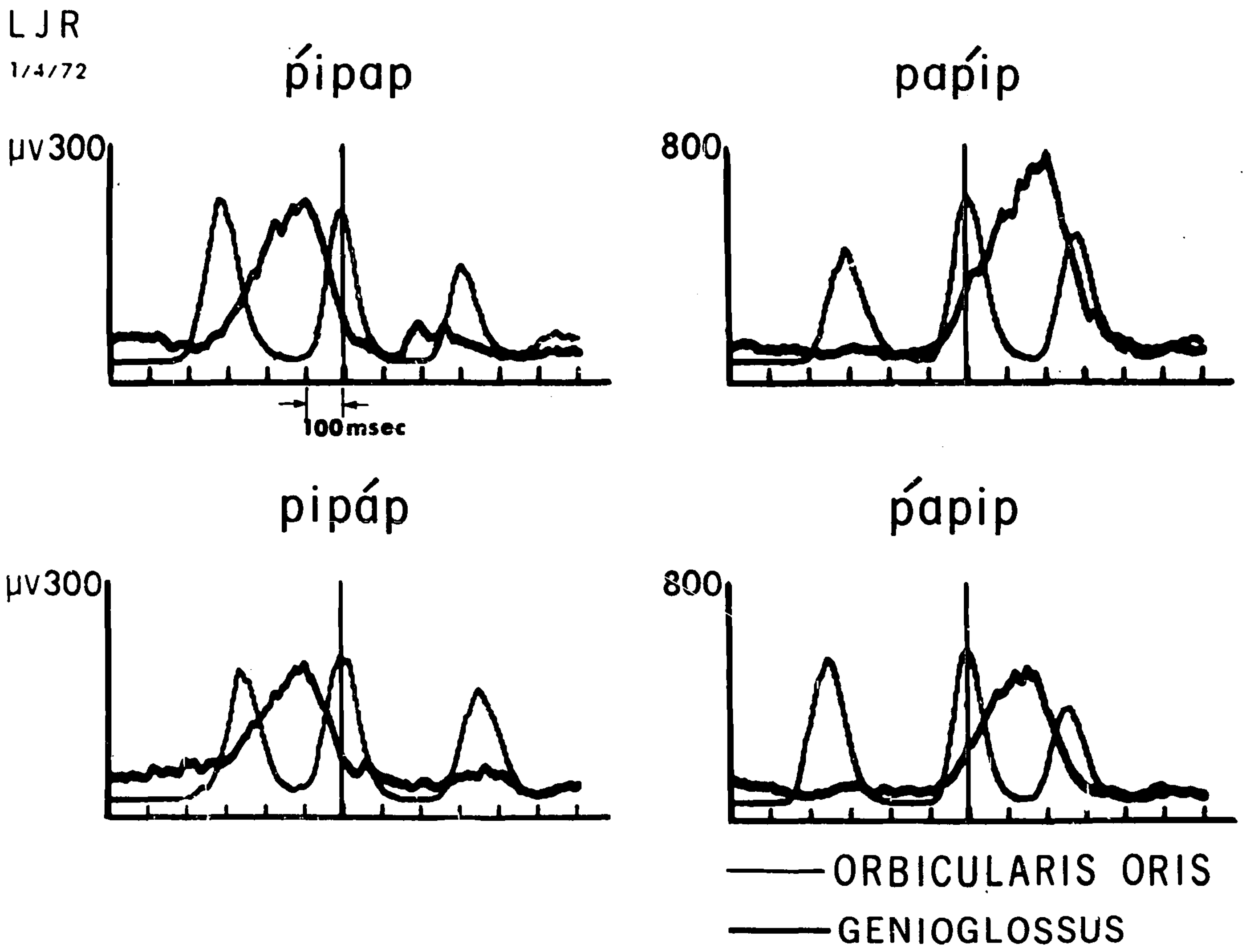

LJR
1/4/72
ṕipap
paṕip
µv300
800
100msec
pipáp
ṕapip
µv300
800
ORBICULARIS ORIS
GENIOGLOSSUS

Figure 3

and second, that the temporal relationship between consonant and vowel signals will change. What is found is a difference in the peak amplitude of the signals to the relevant muscles while the relationship between consonant and vowel signals remains constant.

A closely parallel observation has been made by Kent and Netsell (1971) in a cinefluorographic study of stress contrast effects on various aspects of upper articulator movement in noun and verb forms of various words. They note that in words like "escort," articulator adjustments for the second vowel occur at the same time relative to the intervocalic consonant adjustments regardless of lexical stress. Our observations show that the EMG signals underlying the articulator movement are organic in a way that parallels the output articulator movement.

A second proposal for the stress contrast mechanism is Ohman's (1967) suggestion that stress is manifested by "extra energy" of articulation of the stressed member of a contrasting pair. The results described above are consonant with such a description. However, Ohman's proposal is obviously incomplete. If all muscles reacted to extra stress with more vigorous activity, the effects on antagonistic muscles would cancel each other. There must be uneven effects of stress on various muscles.

The left side of Figure 4 shows some examples from another subject. As before, genioglossus activity clearly lengthens with vowel lengthening as stress changes. There is also somewhat more activity with stress. However, the difference is not huge. The right side shows sample records taken simultaneously from the geniohyoid muscle. The activity appears to be correlated with jaw opening, at least in part. The important point is that stress effects here are very much larger. Apparently, then, the effects of stress are not evenly distributed to all muscles. Perhaps the effects of stress change are greater on jaw muscles than on tongue muscles; I should say, however, that we have a poor understanding of the opening and closing movements of the jaw.

What about Lindblom's (1963) hypothesis about the homogeneity of all duration change mechanisms? He discusses only stress and speaking rate contrasts specifically. A third type of vowel duration change is the well-known effect of voicing status in the terminal consonant. We have scattered information on all three problems. Gay, Ushijima, Hirose, and Cooper (1973) have shown that the effects of speaking rate change are not uniform on all articulatory components. Some gestures show more forceful articulation when rate increases, while others show less activity.

Raphael (1970) has collected some data on terminal consonant effects on vowel duration. The data are similar to what we have just seen--that is, the vowel gesture lengthens before the voiced consonant, but the timing relationship between consonant and vowel is fixed.

Let me conclude by summarizing. First, interrelationship of the consonant and vowel is surprisingly constant over stress lengthening, a finding not predicted by a previous model of the process. Second, stress effects can be considered as "more energetic" enunciation, but the effects of increased energy are distributed unevenly, according to a pattern we do not now understand, over the relevant articulatory muscles. Third, the mechanism of duration change is not uniform for all suprasegmental manipulations.

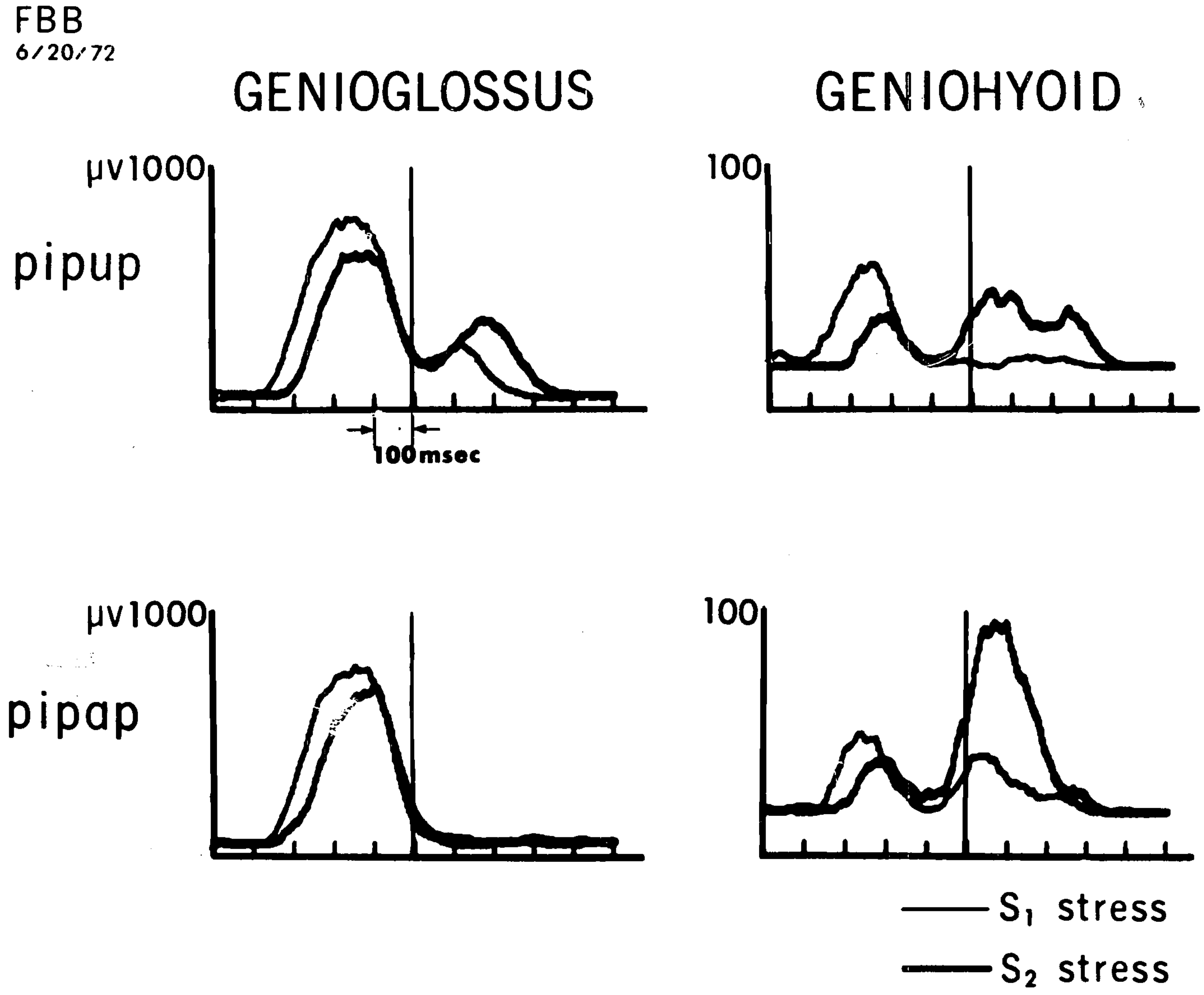

FBB
6/20/72
GENIOGLOSSUS
GENIOHYOID
μv1000
100
pipup
100msec
μv1000
100
pipap
S₁ stress
S₂ stress

Figure 4

REFERENCES

Gay, T., T. Ushijima, H. Hirose, and F. S. Cooper. (1973) The effect of speaking rate on labial consonant-vowel articulation. Haskins Laboratories Status Report on Speech Research SR-35/36 (this issue).

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

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

Lindblom, Bjorn E. F. (1963) Spectrographic study of vowel reduction. J. Acoust. Soc. Amer. 35, 1773-1781.

Ohman, Sven E. G. (1967) Word and sentence intonation: A quantitative model. (Stockholm: Royal Institute of Technology) STR/QPSR 2/3, 20-34.

Raphael, L. (1970) Electromyographic investigation of the articulatory basis of vowel duration differences before voiced and voiceless consonants. J. Acoust. Soc. Amer. 40, 120(A).

Parallel Processing of Auditory and Phonetic Information in Speech Perception*

Charles C. Wood[+]

Recent experiments using a variety of techniques have suggested that speech perception involves separate auditory and phonetic levels of processing. Two models of auditory and phonetic processing appear to be consistent with existing data: a) a strict serial model in which auditory information would be processed at one level followed by the processing of phonetic information at a subsequent level; and b) a parallel model in which at least some portion of auditory and phonetic processing could proceed simultaneously. The present experiment attempted to distinguish empirically between these two models. Subjects identified either an auditory dimension (fundamental frequency) or a phonetic dimension (place of articulation of the consonant) of synthetic consonant-vowel syllables. When the two dimensions varied in a completely correlated manner, reaction times were significantly shorter than when either dimension varied alone. This "redundancy gain" could not be attributed to speed-accuracy trades, selective serial processing, or differential transfer between conditions. These results are consistent only with a model in which auditory and phonetic information can be processed in parallel.

Auditory and Phonetic Levels of Processing in Speech Perception

Current theories of speech perception generally view the process by which linguistic information is extracted from an acoustic speech signal as a hierarchy of logically distinct levels or stages (see, for example, Fry, 1956; Fant, 1967; Stevens and Halle, 1967; Stevens and House, 1972; Studdert-Kennedy, in press). Recently we began a series of experiments designed to investigate possible levels of processing involved in the perception of isolated consonant-

---

*Expanded version of a paper presented at the 86th meeting of the Acoustical Society of America, Los Angeles, Calif., 30 October 1973.

[+] Neuropsychology Laboratory, Veterans Administration Hospital, West Haven, Conn., and Yale University, New Haven, Conn. Present address: Division of Neuropsychiatry, Walter Reed Army Institute of Research, Washington, D. C. 20012.

Acknowledgment: This research was supported by National Institute of Mental Health Grant M-05296 and National Science Foundation Grants GB-3919 and GB-5782 to W. R. Goff, and by a National Institute of Child Health and Human Development grant to Haskins Laboratories. The author wishes to thank W. R. Goff, R. S. Day, W. R. Garner, J. R. Jennings, and J. D. Wicke for their advice and comments on the manuscript; F. S. Cooper for making available the facilities of Haskins Laboratories for the generation and analysis of the synthetic speech stimuli; and T. C. Fisher and J. Jasiorkowski for technical assistance.

vowel (CV) syllables (Day and Wood, 1972a, 1972b; Wood, 1973). The results of these experiments, together with those using other procedures (cf. Studdert-Kennedy and Shankweiler, 1970; Studdert-Kennedy, Shankweiler, and Pisoni, 1972; Tash and Pisoni, 1973), provide empirical support for a basic distinction between two levels of processing in phonetic perception: 1) an auditory level, in which an acoustic speech signal is analyzed into a set of corresponding auditory parameters; and 2) a phonetic level, in which abstract phonetic features are extracted from the results of the preliminary auditory analysis.

The basic paradigm used in these experiments was a two-choice speeded classification task similar to that employed by Garner and Felfoldy (1970) to study patterns of interaction between stimulus dimensions. Subjects were presented a sequence of synthetic CV syllables which varied between two levels of a given dimension and were required to identify which level of that dimension occurred on each trial.[1] Reaction time (RT) for the identification of each dimension was measured under two conditions: a) a single-dimension control condition in which only the target dimension to be identified varied in the stimulus sequence, and b) a two-dimension orthogonal condition in which both the target dimension and the irrelevant nontarget dimension varied orthogonally. For a given target dimension the only difference between the control and orthogonal conditions was the presence or absence of irrelevant variation in the nontarget dimension. Therefore, a comparison of the RTs from these two conditions indicates the degree to which each dimension may be processed independently of irrelevant variation in the other dimension.

The initial experiments in this series analyzed the interactions between the following pairs of dimensions: a) an auditory and a phonetic dimension, b) two auditory dimensions, and c) two phonetic dimensions. Day and Wood (1972a) and Wood (1973, Experiment 1) compared the auditory dimension--fundamental frequency--with the phonetic dimension--place of articulation of voiced stop consonants. For convenience, these dimensions will be referred to as Pitch and Place, respectively. In both experiments irrelevant variation in Pitch significantly interfered with the processing of Place, but irrelevant variation in Place produced minimal interference with the processing of Pitch. Thus, the processing of the auditory dimension appeared to be independent of the phonetic dimension but not the reverse.

A different pattern of results was obtained in experiments where the two dimensions were from the same class--both auditory or both phonetic. Wood (1973, Experiment 2) compared the same levels on the Pitch dimension used in the initial experiments with another auditory dimension, overall Intensity. Neither of these auditory dimensions could be processed independently of the other; that is, irrelevant variation in either dimension produced substantial interference with the identification of the other. Finally, for the case of two phonetic dimensions Day and Wood (1972b) compared Place with formant positions of the vowels in the CV syllables. The pattern of interaction between these phonetic dimensions was identical to that obtained for two auditory dimensions: neither phonetic dimension could be processed independently of irrelevant variation in the other.

---

[1] The term "dimension" is used in this paper to refer to aspects or properties of stimuli which are varied in a given experiment. It should be emphasized that this term explicitly does not imply that the stimulus property in question is singular or unitary in a perceptual sense. The latter is an empirical question (cf. Garner, 1970, 1973, in press; Garner and Felfoldy, 1970).

These results may be summarized in the following way. When both dimensions were members of the same class--both auditory or both phonetic--the interaction between them in this paradigm was a mutual or symmetric interaction. This result is typical of that obtained for "integral" stimulus dimensions in the experiments of Garner and Felfoldy (1970), and is what would be expected if the two dimensions were extracted by a single perceptual process or by multiple processes that are strongly dependent upon each other. In contrast, the interaction between a phonetic dimension and an auditory dimension was a unidirectional or asymmetric interaction. This kind of interaction is evidence a) that the two dimensions are not extracted by a single perceptual process, and b) that the component processes for the phonetic dimension are in some way dependent upon those for the auditory dimension.

Serial Versus Parallel Organization of Auditory and Phonetic Levels

Two process models appear to be consistent with the unidirectional interaction between Place and Pitch: a) a strict serial or sequential model in which auditory information would be processed at one level followed by the processing of phonetic information at a subsequent level, and b) a parallel model in which at least some portion of auditory and phonetic processing could proceed simultaneously.[2] Various forms of both serial and parallel processing have been incorporated in most theories of speech perception (cf. Stevens and House, 1972; Studdert-Kennedy, in press).

A serial organization of auditory and phonetic levels could account for the unidirectional interaction between Place and Pitch by the position of each component process in the sequence. According to this model, a response could be made based on the output of the auditory level alone without processing by the subsequent phonetic level, while access to the phonetic level could occur only after prior auditory processing. In contrast, a parallel model could not provide such a direct explanation for the unidirectional interference between Place and Pitch. However, the parallel model would be more consistent than a serial model with the finding of the initial experiments that RTs for Place and Pitch were not significantly different in the control conditions.[3]

---

[2] Nickerson (1971) pointed out an important limitation of the terms "simultaneous" and "parallel" as used in current information-processing experiments: "What may appear to be simultaneous activities at one level of analysis may prove to be the result of an efficient switching process when the analysis is carried to a more microtemporal level" (p. 276). Following Nickerson's analysis, the terms simultaneous and parallel will be used in the present paper to indicate: "... that the processes in question are proceeding concurrently relative to the time scale on which they are measured, *which is to admit the possibility of inter*mittent switching of attention between one process and another on a more microtemporal scale" (p. 277).

[3] According to a strict serial model any task which requires a response based on information about Place would also require prior processing of information about Pitch, therefore resulting in longer processing times for Place than for Pitch (for data which seem to satisfy such a model, see Posner and Mitchell, 1967; Posner, 1969). As described above, RTs in the orthogonal condition of the initial experiments were indeed longer for Place than for Pitch. However, (continued next page)

The purpose of the present experiment was to investigate more directly the degree to which the processing of auditory and phonetic dimensions occurs serially or in parallel. An efficient way of distinguishing between serial and parallel processing in the context of the RT experiments described above is to include, in addition to the control and orthogonal conditions, a third condition in which the two dimensions are completely correlated. In such a <u>correlated</u> condition both the target dimension and the redundant nontarget dimension provide sufficient information for a correct response. The outcome of interest is simply whether the additional information provided by the redundant dimension can be used by the subject to facilitate performance (i.e., to decrease RT relative to the control conditions without sacrificing accuracy).[4]

A strict serial model would predict no decrease in RT in the correlated condition, since according to this model the Pitch dimension would always have to be processed first, followed by the processing of Place at a subsequent level. In contrast, given certain reasonable assumptions, the parallel model would predict a decrease in RT (usually referred to as a "redundancy gain") in the correlated conditions relative to the control conditions for both dimensions. According to this model, the processing of both dimensions would be initiated simultaneously and the response would be based on the component process which was completed first. The decrease in RTs in the correlated conditions would therefore occur statistically from the fact that the distribution that results from selecting the faster of two "competing" distributions is faster than either "competing" distribution alone.[5]

---

[3] (continued)
according to a serial model this result should be true for the control condition as well, since in both conditions identification of Pitch would require only one level of processing, while identification of Place would involve processing at both levels. The control condition RTs were in fact slightly longer for Place than for Pitch in both of the initial experiments (mean differences of 10.0 and 3.4 msec), although in neither case was the difference statistically significant. Thus, the existing data are equivocal about the serial-parallel distinction. These results suggest either a) that processing of auditory and phonetic information can occur in parallel, or b) that processing is serial but the extra processing time required for the phonetic level is extremely small relative to the variability of the RTs in the initial experiments.

[4] Garner (1970; see also Garner and Felfoldy, 1970) has argued that the distinction between "integral" and "separable" stimulus dimensions must logically precede any consideration of serial versus parallel processing, since the serial-parallel question is inappropriate for pairs of integral dimensions. The unidirectional interference between Place and Pitch in the initial experiments is clear evidence that these dimensions are not completely integral, and therefore that the serial-parallel question is appropriate.

[5] For a detailed analysis of the assumptions and predictions of various serial and parallel models in this and other paradigms, see Egeth (1966); Smith (1968); Hawkins (1969); Biederman and Checkosky (1970); Egeth, Jonides, and Wall (1972); Grill (1971); Lockhead (1972); Nickerson (1971); Townsend (1971); Saraga and (continued next page)

METHOD

Subjects

Each of six subjects (five volunteers and the experimenter) served in an experimental session lasting approximately two hours.

Stimuli

The acoustic stimuli were the four synthetic CV syllables /bae/-104 Hz, /bae/-140 Hz, /gae/-104 Hz, and /gae/-140 Hz, corresponding to the two levels on the two dimensions Place and Pitch. These two-formant stimuli were generated by the Haskins Laboratories parallel resonance synthesizer and were prepared to be equal in all acoustic parameters other than the two dimensions explicitly varied for experimental purposes. All four stimuli had identical fundamental frequency contours (falling), intensity contours (falling), duration (300 msec), and formant frequencies appropriate for the vowel /ae/. Pairs of stimuli differing on the Place dimension differed in the direction and extent of the second formant transition (Liberman, Delattre, Cooper, and Gerstman, 1954; Delattre, Liberman, and Cooper, 1955), while pairs of stimuli differing on the Pitch dimension differed in fundamental frequency (initial fundamental frequencies of 104 Hz versus 140 Hz). Spectrograms of the four stimuli are shown in Figure 1.

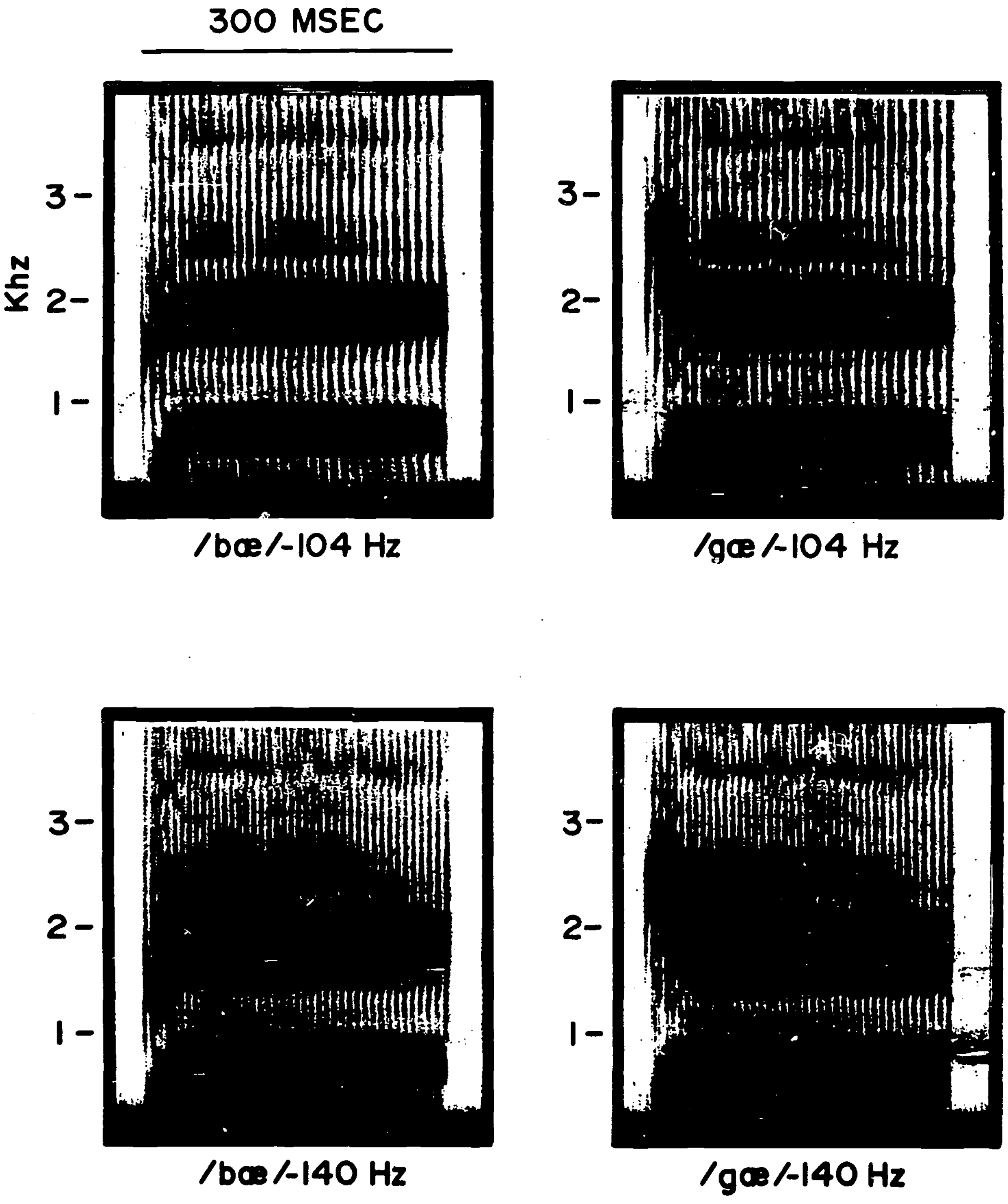
Identification Task, Dimensions, and Conditions

Subjects listened to blocks of trials in which the Place dimension, the Pitch dimension, or both dimensions could vary within a block. One dimension was specified as the target dimension for each block of trials and subjects were required to indicate as rapidly as possible which of the two levels on that target dimension occurred on each trial. Subjects made their responses by pressing one of two response buttons with either the index or middle finger of the preferred hand. Each button was assigned to the same level on each dimension throughout the experiment.

Place and Pitch were each specified as the target dimension in three different conditions, with each condition presented in a separate block of 64 trials. For the target dimension these three conditions were identical; that is, in all three conditions the target dimension varied randomly between its two levels and subjects had to identify which level on that dimension occurred on each trial. The only difference between conditions was the status of the nontarget dimension. In the control condition, the nontarget dimension was held constant at one of its two levels throughout the entire block of trials. For half the subjects the nontarget dimension was held constant at one level and for the remaining half it was held constant at the other level. In the orthogonal condition, the target dimension again varied randomly but in this case the nontarget dimension varied orthogonally. Thus all four stimuli occurred randomly in the orthogonal condition. The control and orthogonal conditions were therefore identical to those

---

[5] (continued)
Shallice (1973). Biederman and Checkosky (1970) and Lockhead (1972) present particularly clear discussions of the statistical rationale for the prediction of a parallel model that RTs in the correlation condition should be faster than those of the single-dimension control conditions.



/bæ/-104 Hz

/gæ/-104 Hz

/bæ/-140 Hz

/gæ/-140 Hz

Figure 1: Spectrograms of the four synthetic syllables. Stimuli differing in Place (/bæ/ versus /gæ/) differed in the direction and extent of the F2 transition (left versus right half of the figure), while stimuli differing in Pitch (104 Hz versus 140 Hz) differed in fundamental frequency (upper versus lower half). The four stimuli were identical in all acoustic parameters other than Place and Pitch.

employed in the previous RT experiments described above (Day and Wood, 1972a, 1972b; Wood, 1973). Finally, in the correlated condition both dimensions again varied but in this case were completely correlated. The stimuli for this condition were /bae/-104 Hz and /gae/-140 Hz, regardless of which dimension was specified as the target dimension.

Apparatus

The stimuli were presented binaurally through Koss Pro-4AA earphones from a Precision Instrument FM tape recorder through computer-controlled relays and a Grason-Stadler electronic switch. A 64-item series of each of the four stimuli was recorded on a separate channel of the stimulus tape, synchronized to 60 μsec accuracy under computer control. On a given trial any of the four possible stimuli could be presented to the subject by closing a relay between the appropriate stimulus channel and the subject's earphones. For each block of trials a LINC computer read a pseudorandom sequence of predetermined stimulus codes from paper tape and closed the appropriate relays in that specified sequence. In this way the same four-channel stimulus tape was used for all six conditions.

The subject's identification responses and RTs were recorded to the nearest msec by an external clock (Beckman-Berkley Universal Counter-Timer). The clock was triggered simultaneously with stimulus onset and was halted by the subject's response. Following the response on each trial the LINC read the RT from the counter and punched the stimulus code, response, and the RT for that trial on paper tape for later analysis.

Procedure

Each subject received one block of 64 trials in each of the six conditions (two dimensions x three conditions per dimension) in an order specified by a balanced latin square. Over the group of six subjects each of the six conditions occurred once in each sequential position and preceded and followed every other condition once.

At the beginning of the experimental session, each subject was informed of the general nature of the experiment, the stimuli and dimensions to be presented, and the identification tasks to be required. Both speed and accuracy were strongly emphasized in all conditions. For the control condition subjects were instructed that only the target dimension would vary in that blocks of trials, and on each trial they were to identify which level on that dimension had occurred. For the orthogonal condition they were told that one dimension would be the target dimension but that the other dimension would also vary. They were instructed to identify the target dimension and to ignore variations in the irrelevant nontarget dimension. Finally, in the correlated condition subjects were told that the two dimensions would be completely correlated, with the levels on both dimensions specifying the same response. In this condition they were again instructed to identify the target dimension as rapidly and accurately as possible, and to use the extra information provided by the redundant dimension if possible.

Prior to the block of trials for each condition, subjects received at least eight practice trials under conditions identical to those they would receive in the following block. These practice trials were designed to stabilize RT performance and allow subjects to become familiar with the stimulus set and identification task they would received in that block.

Data Analysis

As in the initial experiments, subjects made very few errors, averaging less than 3 percent over the entire experiment. Since there were no significant differences in the number of errors among any of the six conditions (Mann-Whitney U tests), error scores will not be considered in detail in the analysis below.

For statistical analysis of the RT data, a complete four-way factorial analysis of variance was computed (Subjects x Conditions x Dimensions x Within Cell). The data entered into this analysis were the untransformed RTs with the single exception that all values greater than 1 sec were set equal to 1 sec. This procedure eliminated the few very long RTs due to subjects' failure to press the response buttons sufficiently to make electrical contact, etc. Subsequent individual comparisons among main effect and interaction means were made according to the Scheffe procedure (Winer, 1962).

RESULTS

Before examining the correlated conditions for evidence of serial or parallel processing, it is important to establish that the results obtained in the control and orthogonal conditions of the present experiment were similar to those in the corresponding conditions of the initial experiments. These data are presented in Table 1. For Place, there was an increase in RT of 57.3 msec from the control to the orthogonal condition, while the increase between conditions for Pitch was only 3.8 msec. In the analysis of variance, the effects of Conditions, Dimensions, and the Condition x Dimension interaction were significant, F (2,126) = 154.3, $p < .001$; F(1,63) = 53.73, $p < .001$; and F(2,126) = 26.12, $p < .001$, respectively. According to the results of the Scheffe analysis on the Condition x Dimension interaction means in Table 1, a difference ≥ 29.7 msec was significant at $p < .001$. Thus, these data are consistent with the unidirectional interference between Place and Pitch obtained in the initial experiments, and therefore indicate that analysis of the correlated conditions for evidence of parallel processing is appropriate for these dimensions.

TABLE 1: Mean reaction time (in msec) for each dimension and condition.

Dimension	Condition		
	Control	Correlated	Orthogonal
Place	386.8	342.8	444.1
Pitch	381.4	346.1	385.2

(Note: According to the Scheffe method for individual comparisons, a difference between any pair of means ≥ 29.7 msec is significant at $p < .001$.)

In the correlated conditions there were substantial decreases in RT for both dimensions (Table 1): 43.9 msec for Place and 35.3 msec for Pitch

($p < .001$). These significant redundancy gains are clearly in accord with the predictions of a parallel model.

However, several alternative explanations for the redundancy gain must be ruled out before concluding that Place and Pitch were actually processed in parallel. First, as mentioned above there were no significant differences in errors among any conditions of the experiment, therefore eliminating the possibility that speed-accuracy trades could be responsible for the observed reduction in RT in the correlated conditions.

A second way in which a redundancy gain could be obtained in absence of parallel processing is the strategy of "selective serial processing" (SSP, cf. Garner, 1969; Morton, 1969; Biederman and Checkosky, 1970; Garner and Felfoldy, 1970; Felfoldy and Garner, 1971). According to the SSP strategy the underlying mode of processing is strictly serial but subjects are presumed to have the ability to select which dimension they actually process in tasks with redundant dimensions. An apparent redundancy gain could therefore be produced if each subject performed the correlated task based on the faster of the two single dimensions, regardless of which was specified as target dimension by the instructions.

In an explicit analysis of the SSP strategy, Felfoldy and Garner (1971) suggested that either of the following two conditions must be met in order for SSP to be effective: a) that RTs for one dimension are significantly faster than the other across all subjects (as would be the case if the two dimensions differed greatly in discriminability); or b) that each subject is able to identify one dimension significantly faster than the other, but the faster dimension is different for different subjects. In either case the SSP strategy could produce decreased RTs in the correlated condition relative to the control condition, but without true parallel processing. The first of these two conditions for SSP was clearly not met in the present experiment, since mean RT for Place and Pitch in the control conditions was not significantly different (Table 1). However, there was a significant interaction of Subjects x Dimensions, $F(5,315) = 25.24$, $p < .001$, indicating that there were reliable differences in RTs between the two dimensions for individual subjects. Therefore, it is logically possible that SSP could have produced the redundancy gain.

Whether or not subjects actually used the SSP strategy can be evaluated directly by comparing the RTs from the correlated conditions to RTs from the faster of the two control conditions for each subject (Garner, 1969; Morton, 1969; Biederman and Checkosky, 1970; Garner and Felfoldy, 1970; Felfoldy and Garner, 1971). Optimal performance for each subject under the SSP model would result in the RT of the correlated conditions being equal to the RT of that subject's faster control condition. In contrast, the parallel model predicts that RTs in the correlated conditions will be faster than the control, even after correcting for the possible use of SSP. In the present experiment the mean of the faster control RTs for each subject was 369.4 msec. In contrast, mean RTs in the correlated conditions were 342.8 msec for Place and 346.1 msec for Pitch (Table 1). A separate analysis of variance and subsequent individual comparisons among these means (Scheffe method) showed that RTs in both correlated conditions were significantly faster than each subject's faster control RTs, $F(2,126) = 14.05$, $p < .001$. Therefore, the obtained redundancy gain cannot be attributed to the SSP strategy.

A third source of an apparent redundancy gain in absence of parallel processing is the possibility of differential transfer between control and correlated conditions. As pointed out by Biederman and Checkosky (1970), greater positive transfer between the two correlated conditions than between the two control conditions might tend to reduce artifically the RTs in the correlated condition relative to the control condition. This would be true both for parallel processing and for SSP, since in either case the two correlated conditions would actually be repetitions of the same task while the two control conditions would always be different. To examine this possibility of differential transfer, the control conditions received first and second in sequence by each subject were compared to the correlated conditions received first and second by each subject, without regard to the actual target dimension in each condition. For the control conditions mean RT for the second block was 1.3 msec faster than the first, while for the correlated conditions the second block was 2.9 msec slower. Thus, there was minimal transfer between the two blocks of trials within each condition, and the direction of the obtained differences favored the control and not the correlated conditions.

Finally, the question of explicit instructions to the subjects regarding the correlation between dimensions should be considered. It might be argued that such instructions could bias subjects toward responding faster in the correlated conditions, leading to a false conclusion about a redundancy gain. In their analysis of the conditions in which the SSP strategy is effective, Felfoldy and Garner (1971) compared the effects of explicit instructions about the correlation between dimensions with implicit instructions consisting of exposure to only the control and correlated stimulus sequences. Using stimulus dimensions which do not produce a redundancy gain under conditions of neutral instructions (Garner and Felfoldy, 1970), Felfoldy and Garner (1971) showed that the explicit instructions produced extensive use of SSP by all subjects, but no evidence of a redundancy gain beyond that attributable to SSP alone. That is, RTs in the correlated conditions were significantly faster than in the control conditions, but were not significantly different from each subject's faster control condition.

The explicit instructions were therefore used in the present experiment to maximize the ability to discriminate between a redundancy gain due to parallel processing and one produced by SSP. Indeed, a result demonstrating selective serial processing of auditory and phonetic information would have implications for models of speech perception as important as those of parallel processing. The correction for SSP described above rules out possible biased reduction of RTs in the correlated condition *per se* since it demonstrates that correlated RTs were faster than would be possible according to a serial model. However, this correction does not completely eliminate the possibility that the explicit instructions produced artifically inflated RTs in the control conditions, thereby producing faster RTs in the correlated conditions than in the control. This possibility is unlikely since: a) the control RTs for both dimensions in the present experiment averaged 25 msec faster than those of extremely well practiced subjects in the experiment of Wood (1973, Experiment 1) under similar conditions, and b) the differences in RT between control and orthogonal conditions in this experiment were virtually identical to those of Wood (1973, Experiment 1). Thus, if subjects inflated the control RTs they would have also had to inflate the orthogonal RTs by a precisely equal amount. These observations make the biased inflation of the control RTs extremely unlikely.

DISCUSSION

The results of this experiment have implications both for models of speech perception in a narrow sense, and for the broader question of how multidimensional stimuli of any kind are perceived. The relation of the present results to both these problem areas is discussed below.

Relation to Multidimensional Information Processing

The way in which humans process multidimensional stimuli has been the subject of considerable theoretical and experimental effort (cf. Garner, 1962, 1970, 1973, in press; Posner, 1964; Egeth, 1966, 1967; Lockhead, 1966, 1970, 1972; Egeth and Pachella, 1969; Nickerson, 1971). The large number of experiments devoted to this question can be classified into two distinct approaches or patterns of major emphasis: a) those which emphasize stimulus concepts and focus upon the nature of the stimulus dimensions which make up the multidimensional stimuli, and b) those which emphasize processing concepts and focus upon the nature of the processes by which the multidimensional stimuli are perceived.

Major emphasis upon stimulus concepts has come from the distinction between integral and separable stimulus dimensions (Garner and Felfoldy, 1970; Garner, 1970, 1973, in press). Expanding upon previous suggestions by Torgerson (1958), Attneave (1962), Shepard (1964), Lockhead (1966), and Hyman and Well (1968), Garner and Felfoldy (1970) argued that the concept of integral dimensions could best be defined by converging experimental operations (Garner, Hake, and Eriksen, 1956): "Integral dimensions are those which lead to a Euclidean metric in direct distance scaling, produce a redundancy gain when the dimensions are correlated and some measure of speed of classification is used, and produce interference in speed of classification when selective attention is required with orthogonal stimulus dimensions" (p. 238). Added to this list of converging operations are the recent data of Handel and Imai (1972) which distinguish between integral and separable dimensions using free classification tasks. Since the present experiment employed two of the converging operations listed above, conditions directly analogous to those of Garner and Felfoldy (1970), the present results may be compared to those expected of integral and separable dimensions.

The results for the Place dimension correspond exactly to those of one member of an integral pair of dimensions: a significant redundancy gain in the correlated condition and significant interference in the orthogonal condition. However, the results for the Pitch dimension showed a different pattern, one which is consistent with neither integral nor separable dimensions. In this case there was again a significant redundancy gain in the correlated condition, but there was minimal interference in the orthogonal condition.

These results pose two problems for a strict distinction between integral and separable dimensions. First, the interference between Place and Pitch in the orthogonal conditions was unidirectional or asymmetric. Previously the integral-separable distinction has always been considered to be symmetric, with either equal interference or no interference between a pair of dimensions. However, as discussed in detail by Garner (1973), recent data suggest that integrality and separability may be more accurately considered as the two ends of a continuum rather than as a strict dichotomy. The second problem for the integral-separable distinction posed by the present results is that the processing of the Place and Pitch dimensions was affected differentially by the correlated and orthogonal conditions. That is, instead of strictly obeying the converging operations notion of integrality and separability presented above, Place and Pitch produced one pattern of results in the correlated conditions and another pattern in the orthogonal conditions. These results indicate that subjects appear to have some option in the way they process Place and Pitch: a) they can process the two dimensions in parallel, as demonstrated by the redundancy gain for both dimensions in the correlated conditions; or b) they can process the Pitch dimension selectively, as demonstrated by the minimal interference produced by irrelavant variation in Place in the orthogonal condition.

The processing options for Place and Pitch also pose problems for existing process models. A number of authors have attempted to develop a single serial or parallel model which could account for a wide variety of information-processing tasks (see discussions by Egeth, 1966; Smith, 1968; Hawkins, 1969; Sternberg, 1969; Biederman and Checkosky, 1970; Grill, 1971; Nickerson, 1971; Townsend, 1971; Saraga and Shallice, 1973). Basic serial and parallel models have been modified to include distinctions between exhaustive and self-terminating processes, fixed or random orders of search, fixed or variable durations for each component process, and a number of others. In addition, Lockhead (1972) presented a "blob" or holistic processing model in which a multidimensional stimulus is first processed holistically, with subsequent serial processes as required by the task. The present results are inconsistent with any model that specifies a particular mandatory process--serial, parallel, or holistic. These data provide further evidence for the suggestions made by Grill (1971), Nickerson (1971), Townsend (1971), and Garner (in press) that no single stimulus distinction or process model may be universally appropriate. Rather, under different conditions subjects may employ various strategies, including serial processing, parallel processing, holistic processing, and combinations thereof, depending upon the constraints of the stimuli and perceptual tasks involved. Instead of performing "critical experiments" to establish the validity of various models, a more appropriate strategy would appear to be the empirical distinction between mandatory and optional processes, and the investigation of stimulus properties and task conditions related to each (Grill, 1971; Nickerson, 1971; Townsend, 1971; Garner, in press).

Relation to Models of Speech Perception

The idea that speech perception may involve some form of parallel processing has been suggested on logical grounds by a number of investigators (cf. Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Stevens and House, 1972; Studdert-Kennedy, in press). The results of the present experiment provide clear evidence that auditory and phonetic information can be processed in parallel, and they provide a starting point for the investigation of possible parallel processing of other kinds of linguistic and nonlinguistic information in speech perception.

However, despite the strong evidence in the present experiment for parallel processing of auditory and phonetic dimensions, such a conclusion contradicts the intuitively reasonable idea that auditory and linguistic processes are organized serially, with linguistic processes dependent upon those performed by the general auditory system. This idea is clearly stated by Stevens and House (1972): "All acoustic signals undergo some common peripheral processing, and up to a certain point in the auditory system the nature of this early processing is the same whether the signal is speech or is not speech" (p. 9). The suggestion that linguistic processes are dependent upon the peripheral auditory system must certainly be correct in some form, since all acoustic signals must be transduced by the receptor apparatus. Thus, the "common peripheral processing" for speech and nonspeech undoubtedly includes the receptor apparatus, and presumably includes much of the subcortical auditory system. However, the exact extent to which this "common peripheral processing" extends anatomically into the auditory system remains to be determined, and the locus at which processing peculiar to speech is initiated continues to be one of the most interesting questions in speech perception.

The general organization of auditory and phonetic processes that seems necessary to account for these observations would require at least three components: a) a common peripheral component for the transduction and preliminary analysis of all acoustic signals, b) a "central" auditory component for the additional processing of nonlinguistic auditory information, and c) a "central" phonetic component for the extraction of phonetic features from the results of the preliminary auditory analysis. The two central components would be capable of functioning in parallel, but both would be dependent upon the output of the prior peripheral processing. From this point of view, the common peripheral stage could not be directly manipulated in RT experiments, since it would be mandatory for all auditory processing tasks. As a working hypothesis this "hybrid" organization would be consistent with the parallel processing of auditory and phonetic information demonstrated by the present experiment, and with the previous experiments which suggest that the perception of speech involves specialized neural mechanisms not required for the perception of nonspeech (cf. Liberman et al., 1967; Studdert-Kennedy and Shankweiler, 1970; Mattingly, Liberman, Syrdal, and Halwes, 1971; Wood, Goff, and Day, 1971; Studdert-Kennedy et al., 1972; Wood, 1973).

REFERENCES

Attneave, F. (1962) Perception and related areas. In Psychology: A Study of a Science, Vol. 4, ed. by S. Koch. (New York: McGraw-Hill) 619-659.

Biederman, I. and S. F. Checkosky. (1970) Processing redundant information. J. Exp. Psychol. 83, 486-490.

Day, R. S. and C. C. Wood. (1972a) Interactions between linguistic and nonlinguistic processing. J. Acoust. Soc. Amer. 51, 79(A).

Day, R. S. and C. C. Wood. (1972b) Mutual interference between two linguistic dimensions of the same stimuli. J. Acoust. Soc. Amer. 52, 175(A).

Delattre, P. C., A. M. Liberman, and F. S. Cooper. (1955) Acoustic loci and transitional cues for consonants. J. Acoust. Soc. Amer. 27, 769-773.

Egeth, H. E. (1966) Parallel versus serial processes in multidimensional stimulus discrimination. Percept. Psychophys. 1, 245-252.

Egeth, H. E. (1967) Selective attention. Psychol. Bull. 67, 41-57.

Egeth, H. E., J. Jonides, and S. Wall. (1972) Parallel processing of multielement displays. Cog. Psychol. 3, 674-698.

Egeth, H. E. and R. Pachella. (1969) Multidimensional stimulus identification. Percept. Psychophys. 5, 341-346.

Fant, G. (1967) Auditory patterns of speech. In Models for the Perception of Speech and Visual Form, ed. by W. Wathen-Dunn. (Cambridge, Mass.: MIT Press) 111-125.

Felfoldy, G. L. and W. R. Garner. (1971) The effects on speeded classification of implicit and explicit instructions regarding redundant dimensions. Percept. Psychophys. 9, 289-292.

Fry, D. B. (1956) Perception and recognition in speech. In *For Roman Jakobson*, ed. by M. Halle, H. G. Lunt, and C. H. van Schooneveld. (The Hague: Mouton) 169-173.

Garner, W. R. (1962) *Uncertainty and Structure as Psychological Concepts*. (New York: Wiley).

Garner, W. R. (1969) Speed of discrimination with redundant stimulus attributes. Percept. Psychophys. *6*, 221-224.

Garner, W. R. (1970) The stimulus in information processing. Amer. Psychol. *25*, 350-358.

Garner, W. R. (1973) Paul M. Fitts memorial lectures. Presented at the University of Michigan, Ann Arbor.

Garner, W. R. (in press) Attention: The processing of multiple sources of information. In *Handbook of Perception*, ed. by E. C. Carterette and M. P. Friedman.

Garner, W. R. and G. L. Felfoldy. (1970) Integrality of stimulus dimensions in various types of information processing. Cog. Psychol. *1*, 225-241.

Garner, W. R., H. W. Hake, and C. W. Ericksen. (1956) Operationism and the concept of perception. Psychol. Rev. *63*, 149-159.

Grill, D. P. (1971) Variables influencing the mode of processing of complex stimuli. Percept. Psychophys. *10*, 51-57.

Handel, S. and S. Imai. (1972) The free classification of analyzable and unanalyzable stimuli. Percept. Psychophys. *12*, 108-116.

Hawkins, H. L. (1969) Parallel processing in complex visual discrimination. Percept. Psychophys. *5*, 56-64.

Hyman, R. and A. Well. (1968) Perceptual separability and spatial models. Percept. Psychophys. *3*, 161-165.

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

Liberman, A. M., P. C. Delattre, F. S. Cooper, and L. J. Gerstman. (1954) The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychol. Monogr., Whole No. *68*.

Lockhead, G. R. (1966) Effects of dimensional redundancy on visual discrimination. J. Exp. Psychol. *72*, 95-104.

Lockhead, G. R. (1970) Identification and the form of multidimensional discrimination space. J. Exp. Psychol. *85*, 1-10.

Lockhead, G. R. (1972) Processing dimensional stimuli: A note. Psychol. Rev. *79*, 410-419.

Mattingly, I. G., A. M. Liberman, A. K. Syrdal, and T. Halwes. (1971) Discrimination in speech and nonspeech modes. Cog. Psychol. *2*, 131-157.

Morton, J. (1969) The use of correlated stimulus information in card sorting. Percept. Psychophys. *5*, 374-376.

Nickerson, R. S. (1971) Binary-classification reaction time: A review of some studies of human information-processing capabilities. Psychon. Monogr. Suppl. *4*(17) (Whole No. *65*).

Posner, M. I. (1964) Information reduction in the analysis of sequential tasks. Psychol. Rev. *71*, 491-504.

Posner, M. I. (1969) Abstraction and the process of recognition. In *Psychology of Learning and Motivation*, Vol. 3, ed. by G. Bower and J. T. Spence. (New York: Academic Press).

Posner, M. I. and R. F. Mitchell. (1967) Chronometric analysis of classification. Psychol. Rev. *74*, 392-409.

Saraga, E. and T. Shallice. (1973) Parallel processing of the attributes of single stimuli. Percept. Psychophys. *13*, 261-270.

Shepard, R. N. (1964) Attention and the metric structure of the stimulus space. J. Math. Psychol. 1, 54-87.

Smith, E. E. (1968) Choice reaction time: An analysis of the major theoretical positions. Psychol. Bull. 69, 77-110.

Sternberg, S. (1969) The discovery of processing stages: Extensions of Donders' method. In Attention and Performance II, ed. by W. G. Koster. (Amsterdam: North Holland).

Stevens, K. N. and M. Halle. (1967) Remarks on analysis by synthesis and distinctive features. In Models for the Perception of Speech and Visual Form, ed. by W. Wathen-Dunn. (Cambridge, Mass.: MIT Press) 88-102.

Stevens, K. N. and A. S. House. (1972) Speech perception. In Foundations of Modern Auditory Theory, ed. by J. V. Tobias. (New York: Academic Press) 1-62.

Studdert-Kennedy, M. (in press) The perception of speech. In Current Trends in Linguistics, Vol. XII, ed. by T. A. Sebeok. (The Hague: Mouton). [Also in Haskins Laboratories Status Report on Speech Research SR-23 (1970) 15-48.]

Studdert-Kennedy, M. and D. Shankweiler. (1970) Hemispheric specialization for speech perception. J. Acoust. Soc. Amer. 48, 579-594.

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

Tash, J. and D. B. Pisoni. (1973) Auditory and phonetic levels of processing as revealed by reaction time. Paper presented at the 85th meeting of the Acoustical Society of America, Boston, Mass.

Torgerson, W. S. (1958) Theory and Methods of Scaling. (New York: Wiley).

Townsend, J. T. (1971) A note on the identifiability of parallel and serial processes. Percept. Psychophys. 10, 161-163.

Winer, B. J. (1962) Statistical Principles in Experimental Design. (New York: McGraw-Hill).

Wood, C. C. (1973) Levels of processing in speech perception: Neurophysiological and information-processing analyses. Unpublished doctoral dissertation, Yale University. [Also in Haskins Laboratories Status Report on Speech Research SR-35/36 (this issue).]

Wood, C. C., W. R. Goff, and R. S. Day. (1971) Auditory evoked potentials during speech perception. Science 173, 1248-1251.

Perception of Speech and Nonspeech, with Speech-Relevant and Speech-Irrelevant Transitions

James E. Cutting[+]
Haskins Laboratories, New Haven, Conn.

The course of speech processing, from registration of the acoustic signal to phonetic interpretation of that signal, may be represented as a number of operations engaging different mechanisms. Since speech perception is primarily confined to the left hemisphere of the brain, many of these operations may be assumed to take place there. The results of the present study, together with results of previous studies, indicate that along with a general speech processor in the left hemisphere there is a purely auditory device suited to analyze transitions, whether those transitions are part of a speech signal or not.

Differences in the perception of speech and nonspeech have been well documented. Perhaps the best information about the processing of these two types of stimuli stems from dichotic listening. In general, speech tasks, whether identification or temporal-order judgment, yield right-ear advantages (Kimura, 1961; Day and Cutting, 1971). On the other hand, nonspeech tasks, whether identification or temporal-order judgment, generally yield left-ear advantages (Chaney and Webster, 1966; Day and Cutting, 1971). The results of a recent study (Cutting, 1973) are in accord with these findings. In a temporal-order judgment task, speech stimuli yielded a large right-ear advantage and nonspeech stimuli yielded a small advantage to the left ear. Although the ear advantage for nonspeech stimuli was not significant, it was significantly different from the speech stimuli. The nonspeech stimuli, however, were quite different from those used in many studies. They were sine-wave analogs to the speech stimuli: that is, three pure tones were synthesized to correspond to the middle frequency of each of the three formants of each of the synthetic speech stimuli. As in previous studies, processing appeared to be different for the two types of stimuli, each type of stimuli requiring different general amounts of processing in each hemisphere, even though the stimuli shared many acoustic characteristics.

Since speech/nonspeech differences have been given so much attention in the past (Kimura, 1967; Semmes, 1968; Mattingly, Liberman, Syrdal, and Halwes, 1971; Day, in press), consider the other, and perhaps more interesting, general result of the Cutting study. Stimuli with transitions were perceived differently than stimuli without transitions regardless of whether they were speech or nonspeech: a stimulus which contained rapid frequency transitions required special processing in the left hemisphere, and this processing was not dependent on the stimulus having been classified as speech.

[+] Also Yale University, New Haven, Conn.

[HASKINS LABORATORIES: Status Report on Speech Research SR-35/36 (1973)]

The processing capabilities of the left hemisphere appear to be fundamentally different from those of the right hemisphere. Most of these differences have been thought to be related to language, and range through all levels from phonetics to semantics. Since speech and language processing is primarily confined to one hemisphere, it would be advantageous to have certain subsidiary systems in that hemisphere to assist in the demodulation of the incoming speech signal. A subsystem which tracks rapidly changing formant frequencies would not typically be needed in a hemisphere geared for nonlinguistic analysis since such changing frequencies often do not appear to be pertinent to the perception of nonspeech sounds. The notion of an auditory analyzer residing only in the left hemisphere is in keeping with Semmes' (1968) views of hemisphere differences in processing capabilities. She states that left-hemisphere function is characterized by "focal" organization, whereas the right hemisphere is characterized by a more "diffuse" organization. Certainly the analysis of rapid pitch modulations is a "focal" task requiring very precise detectors.

It seems reasonable to suppose that the analysis of transitions is independent of whether the stimulus is speech or nonspeech. Otherwise, the system must necessarily make an early decision determining whether or not a stimulus is speech, before starting to analyze its transitions. Such a process would be unnecessarily cumbersome, if not untenable. Moreover, the notion of an independent transition analyzer in the left hemisphere is congruent with the results of Darwin (1971). He found that fricatives with formant transitions yielded a right-ear advantage, while the same fricatives without transitions yielded no advantage.

Cutting (1973) employed two variables in his stimuli: they could be speech or nonspeech, with transitions or without transitions. Speech processing is primarily a left-hemisphere task, and it appears that the analysis of transitions is also a left-hemisphere task. The two processes appear to be independent of one another, and also additive. Consonant-vowel (*CV*) stimuli, both [+ speech] and [+ transition], yielded a large right-ear/left-hemisphere advantage since the two variables favor left-hemisphere processing. Steady-state vowels (*V*) and sine-wave analogs of the *CV* stimuli (*CVsw*), however, had only one positive value on the two dimensions and thus yielded smaller ear difference scores: *V* stimuli were [+ speech] and [- transition] and *CVsw* stimuli were [- speech] and [+ transition]. Sine-wave analogs of the vowel stimuli (*Vsw*) were both [- speech] and [- transition], and consequently yielded a left-ear advantage. It should be noted that speech/nonspeech was a more potent dimension than transition/nontransition: *V* stimuli, for example, yielded a larger right-ear advantage than did *CVsw* stimuli.

These results, however, do not necessarily support the notion of an *auditory* transition analyzer in the left hemisphere. It remains possible that this secondary mechanism is, in reality, a language-based subsystem. Perhaps, because of their resemblance to speech syllables, Cutting's *CVsw* stimuli triggered this processor into analyzing the transitions appropriate for the stop consonants. The present study was designed to test whether stimuli with transitions not corresponding to phoneme segments yield equal ear advantages to those stimuli that have transitions corresponding to phoneme segments. An affirmative result would support the notion that this device is an auditory feature detector, while a negative result would indicate that it is a linguistic feature detector.

Method

Stimuli. CV and CVsw stimuli were used again as in Cutting (1973). They were, or corresponded to, the CV syllables [bi, gi, bae, gae, bɔ, gɔ]. Stimuli containing the same vowel, or sine waves which corresponded to the same vowel, (such as [bae, gae]) were identical in all respects except for the second-formant transition. In all cases, the second-formant transition of [b] rose for 50 msec to the resting frequency of the formant, while the transition of [g] fell to that frequency. First- and third-formant transitions were both always upgliding. Two other sets of stimuli were synthesized. One set was similar to the CV stimuli in that its members contained formants and formant transitions typically found in speech. The particular array of formant transitions, however, could never have been produced by a human vocal tract. As in the CV stimuli the second-formant transition could be either upgliding or downgliding in frequency. The first- and the third-formant transitions, on the other hand, were always downgliding. The extent of the two transitions was the same, only the direction was changed. Since these stimuli had transitions which did not correspond to any consonant speech segment they are designated C'V stimuli. The fourth stimulus set consisted of sine-wave analogs of the C'V stimuli, and are designated C'Vsw stimuli. Four stimuli, one from each class, are displayed in Figure 1. Thick bars indicate formants in the speech and speech-like stimuli, while narrow lines indicate sine waves in the nonspeech stimuli. The actual bandwidth of the formants was 60, 90, and 120 Hz for the first, second, and third formants, respectively. The bandwidth of the sine waves was essentially zero.

Subjects. Ten Yale University undergraduates participated in Task 1, and 16 others participated in Task 2. All were right-handed native American English speakers with no history of hearing difficulty. None had previous experience listening to synthetic speech or to dichotic stimuli.

Task 1: Diotic Identification

A brief identification task was run to be sure that the sine-wave stimuli were not identifiable as speech.

Procedure. Only CV and CVsw stimuli were selected to be identified. Subjects (Ss) listened to a tape of 120 items presented in random order one at a time: (2 classes of stimuli) x (6 stimuli per class) x (10 observations per stimulus). They were instructed to write down their responses as BEE, DEE, GEE, BAA, DAA, GAA, BAW, DAW, or GAW for each item. In this manner Ss were forced to try labeling the sine-wave stimuli as speech. Note that there were, in fact, no stimuli which began with the phoneme [d]. No practice or training was given.

Results. CV stimuli were correctly identified on 82 percent of all trials, while CVsw stimuli were correctly identified on only 19 percent of all trials, only a few percentage points above chance. Responses were parsed into their component segments, and performance was measured for both consonants and vowels. In the CV stimuli, consonants were correctly identified 82 percent of the time and vowels 97 percent of the time. The corresponding scores for the CVsw stimuli were 38 and 45 percent correct.

These results strongly suggest that the formant stimuli were processed as speech and the sine-wave stimuli as nonspeech. Although Ss gave consistently correct identifications for the CV stimuli, they scattered their identifications over the nine possible responses for the CVsw stimuli.

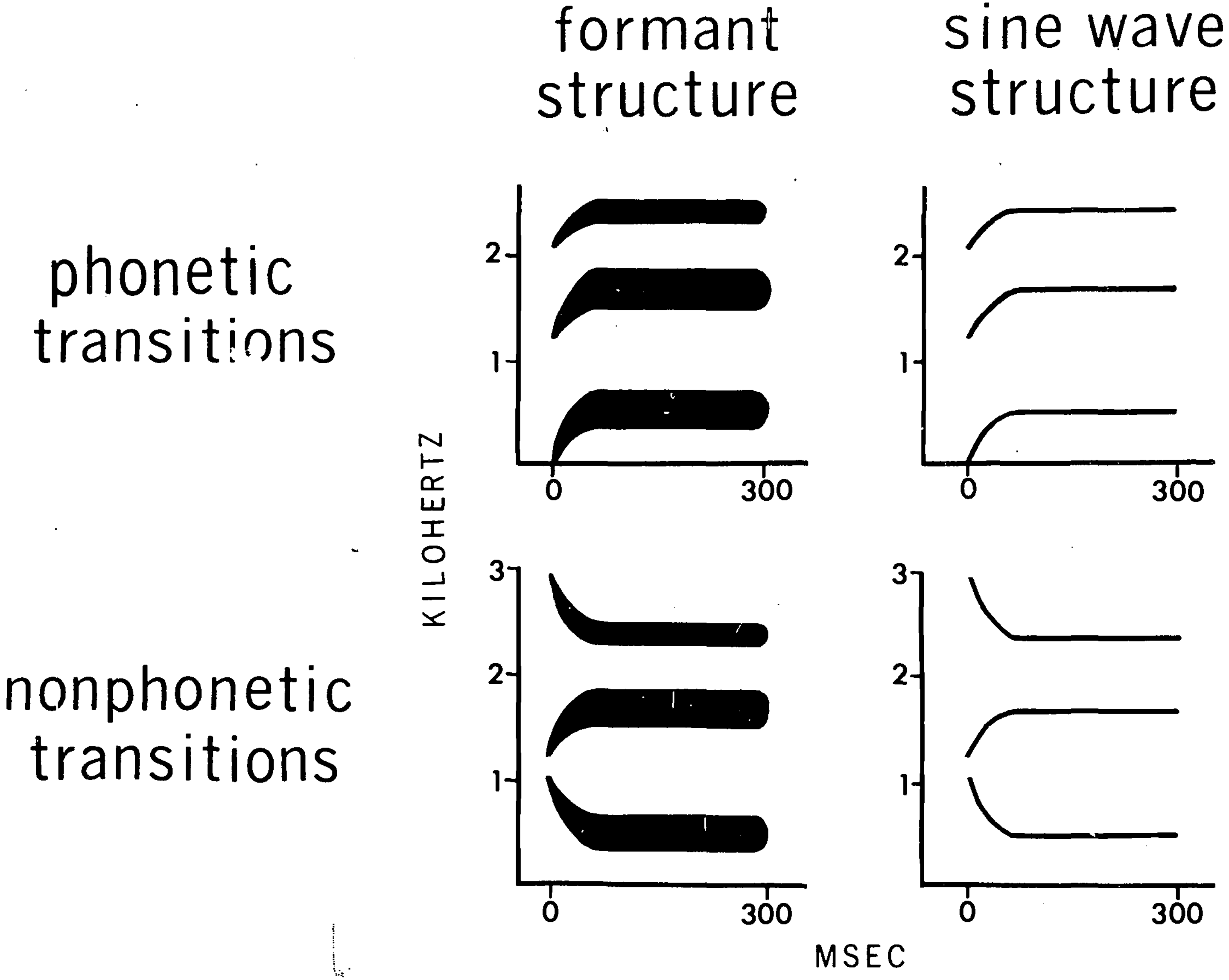


Figure 1: Schematic spectrograms of speech and nonspeech stimuli, with phonetic and nonphonetic transitions.

Task 2: Dichotic Temporal-Order Judgment

A dichotic listening task was devised to assess the role of transitions in the perception of speech and nonspeech signals. The paradigm was a temporal-order judgment task not requiring Ss to give verbal labels to the stimuli, and was identical to that of Cutting (1973).

Tapes and procedure. Trials for the dichotic temporal-order judgment task were constructed so that Ss recognized the leading stimulus in a given dichotic pair by means of a subsequent probe stimulus. A trial consisted of a dichotic pair with a temporal onset asynchrony of 50 msec, followed by 1 sec of silence, followed by a diotic stimulus which was one of the members of the dichotic pair. Ss were instructed to regard the diotic stimulus as a probe which asked the question: "Is this the stimulus which began first?" Figure 2 shows a schematic representation of two such trials. Consider sample trial 1, where Stimulus 1 begins before Stimulus 2 by 50 msec, and the probe stimulus is Stimulus 1. Since the probe is identical to the stimulus that began first, the correct response is yes. In sample trial 2, the dichotic pair is the same as in trial 1, but the probe stimulus is different. Since Stimulus 2 did not begin before Stimulus 1, the correct response for trial 2 is no. The 50 msec onset asynchrony was chosen because it is a particularly difficult interval at which to judge stimulus order (see Day, Cutting, and Copeland, 1971; Day and Cutting, 1971; Day and Vigorito, 1973).

Four tapes were constructed, one for each class of stimuli. Each tape consisted of 48 trials: (6 possible pairs) x (2 channel arrangements) x (2 possible probes) x (2 observations per pair). The stimuli used in each trial were always selected from the same class of stimuli. CV trials, for example, were constructed out of CV stimuli that shared neither the same vowel nor the same consonant: thus, for example, /bi/ was paired with /gae/ or /gɔ/. CVsw, C'V, and C'Vsw trials were constructed using the same rules applied to CV stimuli. Stimuli in the dichotic pair were counterbalanced for leading and lagging position. The probe stimulus chosen for each trial and the channel assignments of the stimuli in the dichotic pair were also counterbalanced in the random sequence of trials.

Ss listened to each tape twice, reversing the earphones after one pass through the tape. The order of channel assignments was counterbalanced across Ss. Each group listened to the four tapes in a different order, determined by a balanced latin square design. Ss listened to a total of 384 trials consisting of a dichotic pair and a diotic probe, writing Y for yes or N for no for each trial. Four practice trials were given before each stimulus class in order to familiarize Ss with the stimuli.

Results

In general the task was quite difficult: overall performance for all trials and all types of stimuli was 60 percent correct. Performance for each of the four types of stimuli was comparable: the average score for each was between 59 and 61 percent, with no significant differences among them. Because of the comparability of performance levels no phi coefficient analysis was necessary (see Kuhn, in press).

Figure 2

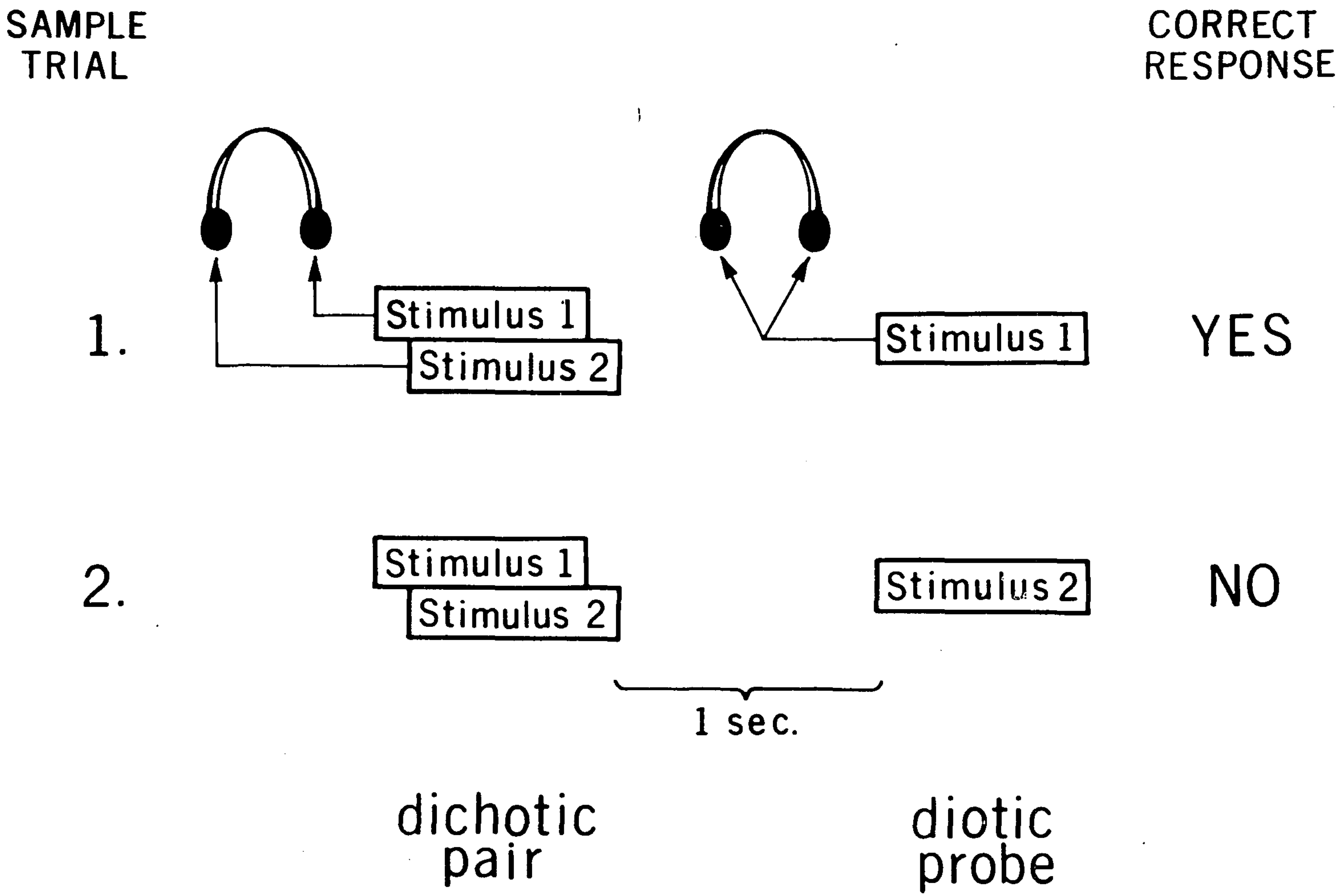


Figure 2: A paradigmatic display of two dichotic temporal-order judgment trials with a diotic probe.

The pattern of ear advantages is quite interesting, but it is necessary first to note how the results were scored. Consider again the sample trials in Figure 2. The correct response for sample trial 1 is yes, while the correct response for trial 2 is no. If in the dichotic pair Stimulus 1 was presented to the right ear and Stimulus 2 to the left ear, and if the S responded yes for the first trial and no for the second, he would have been correct on both. This would be scored as two correct responses for the right-ear leading stimulus. On the other hand, if the S had responded no and yes, respectively, for the sample trials both would be wrong and his score for the right-ear leading stimulus would be docked for two incorrect responses. (Of course, if the channels had been reversed with Stimulus 1 presented to the left ear and Stimulus 2 to the right the logic would be entirely reversed.)

CV and C'V trials. There was a large significant ear difference for the formant stimuli. Ss were 65 percent correct in responding to the probe stimulus when the leading stimulus was presented to the right ear, and 54 percent correct when it was presented to the left ear, yielding a net 11 percent difference [$F(1,15) = 5.38$, $p < .05$]. There was no significant difference between CV and C'V stimuli.

CVsw and C'Vsw trials. No significant ear advantage was found for sine-wave stimuli. Ss were 61 and 60 percent correct for right-ear and left-ear leading trials, respectively--a net 1 percent right-ear advantage.

The results of each of the four conditions are shown in Figure 3. The effect of phonetic versus nonphonetic transitions was not an important factor in either formant or sine-wave stimulus perception.

Speech vs. nonspeech. Although the C'V stimuli did not have phonetic transitions, their vowel portions were speech-like enough to make the stimuli sound like syllables with a garbled onset, good enough to rate the label "speech." Sine-wave stimuli, regardless of the nature of their transitions, always sounded like nonspeech. Accepting this redefinition of the speech/nonspeech distinction, there was a significant difference in the ear advantages for the two types of stimuli [$F(1,15) = 16.41$, $p < .001$]. Moreover, the magnitude of the speech/nonspeech difference in the present study was comparable to that found in Cutting's (1973) previous study using this paradigm.

Discussion

The results of the present study suggest that the transition analyzer in the left hemisphere is a language-independent device. Stimuli with transitions which do not correspond to any phoneme segment yield results nearly identical to stimuli which have phonetically appropriate transitions. This result is consonant with that of Kimura and Folb (1968), who found a right-ear advantage for the perception of speech played backwards. Backwards speech, like C'V stimuli, has transitions that are often inappropriate for the perception of any specific speech segments, but it is nevertheless heard and processed as speech.

The existence of a purely auditory analyzer in the left hemisphere, in addition to the usual speech processor, may be able to explain the variation in ear advantages found within specific phonemes. Indeed, Studdert-Kennedy and Shankweiler (1970) reported differential ear advantages for various stop consonants: although the differences were not significant, [b, g, p, k] yielded

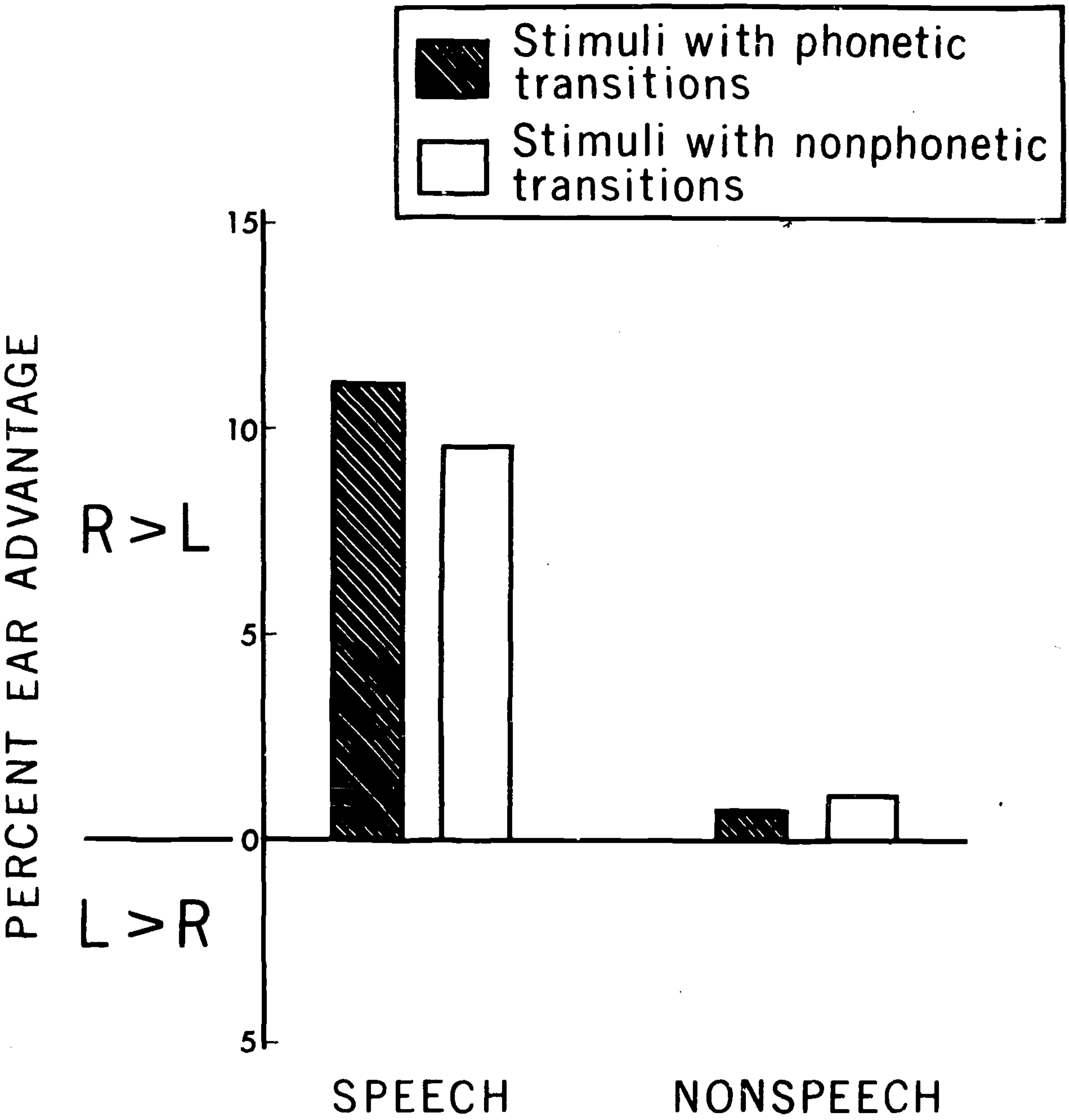


Figure 3: Percent ear advantages for speech and nonspeech stimuli, with phonetic and nonphonetic transitions.

larger right-ear advantages than [d, t] in many different vowel contexts. In general, the labial [b, p] and the velar [g, k] stops have second-formant transitions of greater extent than do the alveolar stops [d, t]. Perhaps these systematic differences in ear advantages stem from differential engagements of the auditory analyzer. A possible effect of transitions can be seen not only in the consonants but also in the vowels. Studdert-Kennedy and Shankweiler reported differential ear advantages for the several vowels that they used. Not surprisingly [i], the vowel which has the most extreme second-formant position and which typically undergoes much context-conditioned variation in CVC syllables, yielded the largest right-ear advantage. Furthermore, it had the only ear advantage for a vowel that was significant.

Increased right-ear advantages appear to result from increased acoustic variation in both nonspeech and speech contexts. Apart from the results of the present study, Halperin, Nachshon, and Carmon (1973) found that the introduction of acoustic variation within nonspeech sounds could change a left-ear advantage into a right-ear advantage. Furthermore, the more the acoustic variation, the greater the tendency for the right-ear/left-hemisphere system to excel over its counterpart.

The results of the present study, taken in conjunction with those of Cutting (1973), Darwin (1971), and Halperin et al. (1973), provide strong evidence that the left hemisphere is not only specialized for processing speech, but is specialized for processing certain purely auditory events as well. It seems ecologically parsimonious for man to have developed both systems within the same hemisphere of the brain. Certainly most speech utterances are characterized by a myriad of formant transitions which must necessarily be analyzed before phonetic decisions can be made about them. That the appropriate auditory analyses and the phonetic decisions about them appear to occur in relatively close proximity is an example of an elegant way to build a system.

REFERENCES

Chaney, R. B. and J. C. Webster. (1966) Information in certain multidimensional sounds. J. Acoust. Soc. Amer. 40, 2, 447-455.

Cutting, J. E. (1973) Perception of speech and nonspeech, with and without transitions. Haskins Laboratories Status Report on Speech Research SR-33, 37-46.

Darwin, C. J. (1971) Ear differences in the recall of fricatives and vowels. Quart. J. Exp. Psychol. 23, 46-62.

Day, R. S. (in press) Engaging and disengaging the speech processor. In *Hemispheric Asymmetry of Function*, ed. by Marcel Kinsbourne. (London: Tavistock).

Day, R. S. and J. E. Cutting. (1971) What constitutes perceptual competition in dichotic listening? Paper presented at the Eastern Psychological Association meeting, New York, April.

Day, R. S., J. E. Cutting, and P. M. Copeland. (1971) Perception of linguistic and nonlinguistic dimensions of dichotic stimuli. Paper presented at the annual meeting of the Psychonomic Society, St. Louis, Minn., November. (Also in Haskins Laboratories Status Report on Speech Research SR-27, 193-197.)

Day, R. S. and J. M. Vigorito. (1973) A parallel between encodedness and the ear advantage: Evidence from a temporal-order judgment task. J. Acoust. Soc. Amer. 53, 358(A).

Halperin, Y., I. Nachshon, and A. Carmon. (1973) Shift in ear superiority in dichotic listening to temporal pattern nonverbal stimuli. J. Acoust. Soc. Amer. 53, 46-50.

Kimura, D. (1961) Cerebral dominance and the perception of verbal stimuli. Canad. J. Psychol. 15, 166-171.

Kimura, D. (1967) Dual functional asymmetry of brain in visual perception. Neuropsychol. 4, 275-285.

Kimura, D. and S. Folb. (1968) Neural processing of backwards-speech sounds. Science 161, 395-396.

Kuhn, G. M. (in press) The phi coefficient as an index of ear differencss in dichotic listening. Cortex.

Mattingly, I. G., A. M. Liberman, A. K. Syrdal, and T. Halwes. (1971) Discrimination in speech and nonspeech modes. Cog. Psychol. 2, 131-157.

Semmes, J. (1968) Hemispheric specialization: A possible clue to mechanism. Neuropsychol. 5, 11-26.

Studdert-Kennedy, M. and D. P. Shankweiler. (1970) Hemispheric specialization for speech perception. J. Acoust. Soc. Amer. 48, 579-594.

On the Identification of Place and Voicing Features in Synthetic Stop Consonants

David B. Pisoni[+] and James R. Sawusch[++]

Two models of the interaction of phonetic features in speech perception were used to predict subjects' identification functions for a bidimensional series of synthetic consonant-vowel syllables. The stimuli varied systematically in terms of the acoustic cues underlying the phonetic features of place of articulation and voicing. Model I assumed that phonetic features are additive and are processed independently in perception. Model II assumed that the phonetic features interact and are not processed independently. The fit of Model II to the bidimensional series data was better than the fit of Model I, suggesting that the phonetic features of place and voicing in stop consonants are not processed independently but rather show a mutual dependency.

Theoretical accounts of speech sound perception have frequently proposed some type of articulatory-motor involvement during perceptual processing (Liberman, 1957; Stevens, 1960; Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Stevens and Halle, 1967). One reason for this may be that research on speech sound perception has drawn its descriptive categories from the account of speech production offered by phoneticians. Thus, the articulatory dimensions that distinguished different classes of speech sounds in production served as the basis for uncovering the acoustic cues that distinguish different speech sounds in perception. Spectrographic analysis and perceptual experiments revealed that the sounds of speech were not arrayed along a single complex dimension but could be specified in terms of a few simple and independent dimensions (Gerstman, 1957; Liberman, 1957). Acoustic dimensions were found in early experiments with synthetic speech to provide distinctions in perception corresponding to the articulatory dimensions of speech production, suggesting that perceptual and articulatory dimensions of speech may be intimately linked (Delattre, 1951; Liberman, 1957).

[+] Indiana University, Bloomington, and Haskins Laboratories, New Haven, Conn.

[++] Indiana University, Bloomington.

Acknowledgment: This research was supported in part by a PHS Biomedical Sciences Grant (S05 RR 7031) and PHS Training Grant (T01 MH 11219-04) to Indiana University and in part by a grant from the National Institute of Child Health and Human Development to Haskins Laboratories. We wish to thank M. Studdert-Kennedy and R. M. Shiffrin for comments and suggestions on the manuscript.

Two articulatory features to receive a great deal of attention in the description of stop consonant production are place of articulation and voicing. Both these features have fairly well defined acoustic properties which presumably mirror the differences in production (Delattre, 1951). For example, the feature of place of production refers to the point of constriction in the vocal tract where closure occurs. The acoustic cues that underlie the place feature in consonant-vowel (CV) syllables are reflected in the formant transitions into the following vowel, particularly the direction and extent of the second and third formant transitions (Liberman et al., 1967). In contrast, the voicing feature is related to the presence or absence of periodic vibration of the vocal chords. The acoustic cues that underlie the voicing feature in stop consonants in initial position are reflected in terms of the relative onset of the first formant transition (i.e., F1 "cutback") and the presence of aspiration in the higher formants (Liberman, Delattre, and Cooper, 1958). This compound acoustic cue has been called "voice-onset time" (VOT) by Lisker and Abramson (1964) and corresponds to the time interval between the release from stop closure and the onset of laryngeal pulsing.

Figure 1 presents schematized spectrographic patterns which show the acoustic cues for place and voicing features for the CV syllables /ba/, /da/, /pa/, and /ta/. There is a relatively simple relation between articulatory features of place and voicing and their respective acoustic cues when the vowel is held constant (see also Liberman, 1970). Consonants within a particular row share voicing; /ba/ and /da/ are voiced, /pa/ and /ta/ are voiceless. The major acoustic cue for voicing in these syllables is the cutback or elimination of the initial portion of the first formant. Consonants within a particular column share place of production: /ba/ and /pa/ are bilabial stops, /da/ and /ta/ are alveolar stops. The primary acoustic cue for place is the direction and extent of the second and third formant transitions.

Several perceptual experiments employing stop consonant-vowel syllables have concluded that the features of place and voicing are processed independently of each other. For example, Miller and Nicely (1955) analyzed the perceptual confusions among 16 CV syllables presented to listeners under various signal-to-noise ratios and filtering conditions. They computed the sum of the information transmitted by the features separately and in combination. Since the two values were approximately equal, they concluded that the features used in their analysis were mutually independent. Among these features were place and voicing. As a part of a larger investigation of dichotic listening, Studdert-Kennedy and Shankweiler (1970) reached the same conclusion by a similar analysis of place and voicing confusions among stop consonants.

These studies imply that features are extracted separately during early perceptual processing and are later recombined in response. Figure 2 represents a simplified block diagram of this process. The output of the auditory analysis is a set of acoustic cues $\{c_i\}$. These cues are combined and from them a set of phonetic features $\{f_j\}$ is recognized. Finally, the phonetic features are combined to yield the perception of the phonetic segment. Together, stages 2 and 3 form what Studdert-Kennedy (1973) has described as the "phonetic" stage of processing. We assume that phonetic features are recognized or identified in short-term memory (STM) when the auditory patterns derived from the acoustic cues have made contact with some representation generated from synthesis rules residing in long-term memory. We assume that abstract phonetic features have an articulatory rather than acoustic reality in STM although we will not try to justify this assumption at present.

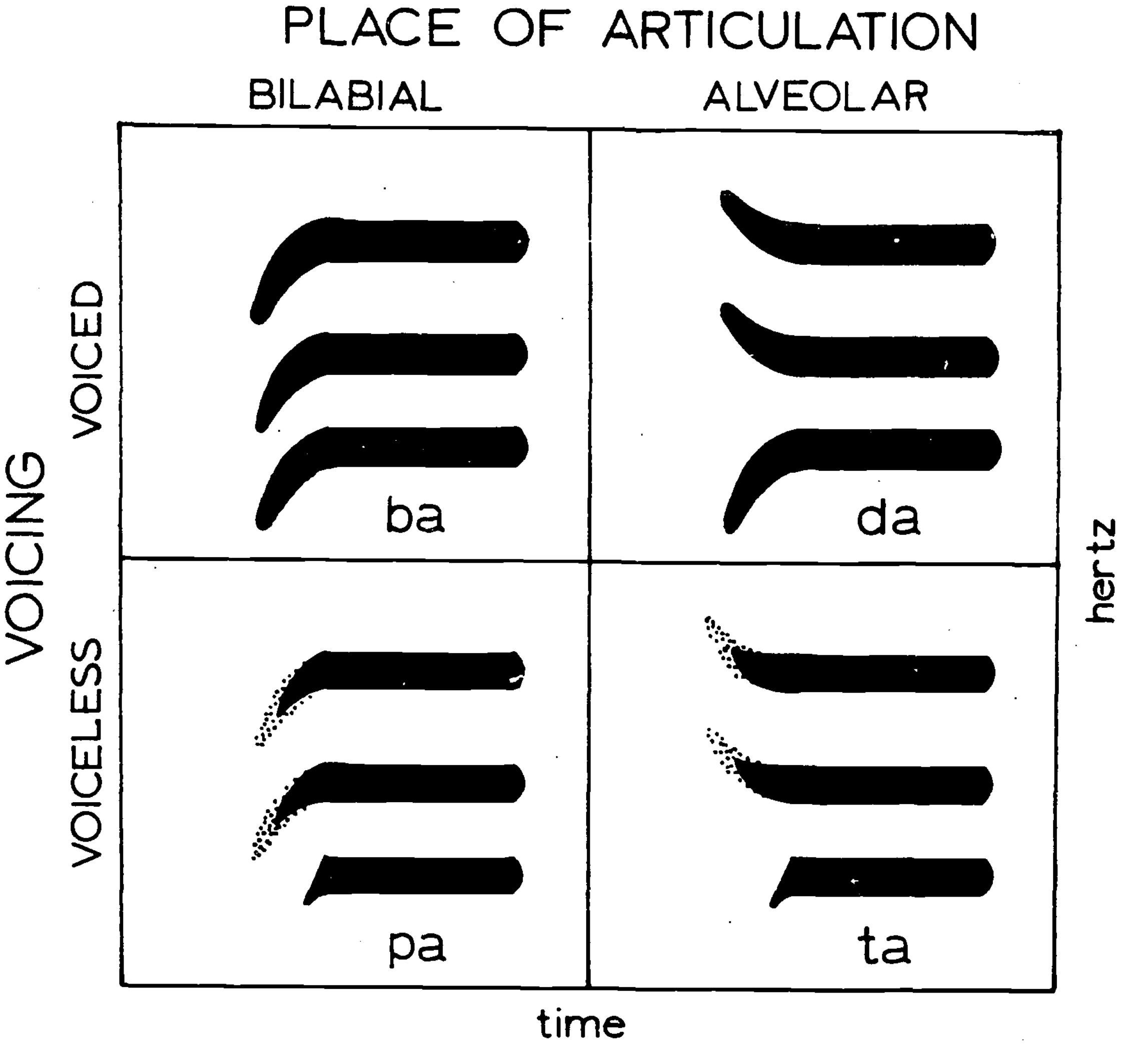


Figure 1: A schematized sound spectrogram of the syllables /ba/, /da/, /pa/, and /ta/ as used in the present experiment.

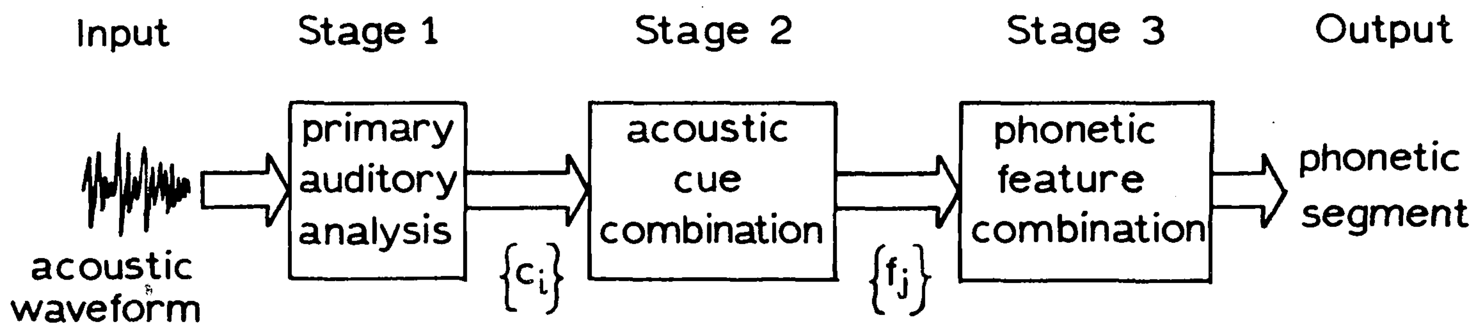


Figure 2: A block diagram of stages of perceptual analysis for phonetic segments. The input is the acoustic waveform and the output is a sequence of phonetic segments.

To say that phonetic features are independent implies independence of processing at all three stages of Figure 2. That is, the acoustic cues are extracted from the acoustic waveform separately and independently of each other in stage 1. Then the phonetic features are extracted separately and independently from the acoustic cues in stage 2. Finally, the phonetic features are combined separately and independently of each other in stage 3, resulting in a particular phonetic segment.

The independence of phonetic feature processing in stage 3 may be described quantitatively by a simple linear or additive model. The phonetic features of place (f_1) and voicing (f_2) that are output from stage 2 are weighted separately and then added together in stage 3. Equation 1 expresses this concept algebraically:

$$X = a_1f_1 + a_2f_2 \qquad (1)$$

Here, f_1 is the amount of the place feature of stimulus X output from stage 2, and a_1 is its associated weight. Similarly, f_2 is the amount of the voicing feature of X output from stage 2, and a_2 is its associated weight. Since these two features are sufficient to distinguish among the four stops b, d, p, and t, we will ignore other phonetic features as being redundant and nondistinctive.

However, evidence for nonindependence of phonetic features, in particular the features of place and voicing, has also been presented by several investigators. This nonindependence could come at any of the three levels mentioned. For example, Haggard (1970) put the dependence relationship of the features of place and voicing in the second stage, where phonetic features are extracted. In the model of Haggard (1970), the listener's decision on the voicing feature is partly determined by his prior decision on the place feature.

Lisker and Abramson (1964, 1967) reached the corresponding conclusion upon examination of their production data for stop consonants in initial position. The voicing feature as reflected in VOT depends on the feature of place of production; the VOT lag at the boundary between voiced and voiceless stops increases as place of production moves further back in the vocal tract (i.e., from /ba/ to /da/ to /ga/). Given the anatomical and physiological constraints on speech production, this position is a priori more plausible.

This particular concept of nonindependence, where the feature of voicing partly depends on the feature of place, may also be expressed algebraically. We will again assume independence of processing in stages 1 and 2. The nonindependence in stage 3 may be expressed as:

$$X = a_1f_1 + a_2f_2 + b(1-f_1)f_2 \qquad (2)$$

Here, a_1, f_1, a_2, and f_2 are the same as in equation 1. However, the constant b represents the weight given to the interaction term of place and voicing $[(1-f_1)f_2]$.

The purpose of the present experiment was to reexamine the identification of place and voicing features in stop consonants and to determine by means of a new experimental paradigm whether these two phonetic features (i.e., place and voicing) are combined additively or nonadditively in stage 3 as shown in Figure 2.

Stimuli for the experiment were three sets of synthetic speech sounds that varied systematically in the acoustic cues underlying the two phonetic features of place and voicing. One series of stimuli varied in the acoustic cues that underlie the phonetic features of place, while holding the voicing feature constant (/ba/ to /da/ with VOT at 0 msec). A second series varied the acoustic cues underlying the phonetic feature of voicing, while holding the place feature constant (/ba/ to /pa/ with F2 and F3 always rising). The final series varied the acoustic cues underlying both phonetic features simultaneously (/ba/ to /ta/). These three sets of speech sounds were presented separately to listeners for identification into the categories /ba/-/da/, /ba/-/pa/, and /ba/-/ta/, respectively. The use of synthetically produced stimuli made it possible to control experimentally the correlation between place of production and voicing that Lisker and Abramson (1964, 1967) had found in natural speech.

Our principle aim was to determine whether the probabilities of identification along the bidimensional continuum (/ba/ to /ta/) could be predicted from some combination of the probabilities along the separate unidimensional series. We consider below two possible models of ways these separate features might be combined in the bidimensional case. Both Model I and Model II are concerned with the manner in which phonetic features are combined in phonetic perception. All processing up to stage 3 of Figure 2 is assumed to be independent according to the definition of independence for these stages given previously. We also assume that processing in stages 1 and 2 takes place in parallel and is automatic in the sense that Ss do not have control over these stages of perceptual processing. (See also Shiffrin and Geisler, 1973, and Shiffrin, Pisoni, and Castenada-Mendez, in press.)

Model I: Linear Combination of Phonetic Features

Hereafter, if a S identified a stimulus as /ba/ it will be denoted B and likewise, /da/ as D, /pa/ as P, and /ta/ as T. In the /ba/ to /da/ series only the acoustic cues underlying the phonetic feature of place of articulation were varied. Since processing in stage 2 is assumed to be independent (i.e., separate for different phonetic features), the only variation in the output of stage 2 on the /ba/ to /da/ (place) series should be in feature f_1, the phonetic feature of place of articulation. Accordingly, since the only variation in the input to stage 3 is in f_1, the output of stage 3 (a phonetic segment) is assumed to vary directly with the input (f_1) and thus accurately reflect f_1. However, due to noise in the acoustic waveform and in the first two stages of processing, the outputs of stage 2 are assumed to be probabilistic in nature. Thus, Ss' judgments of the stimuli from the/ba/ to /da/ series (the probability of responding D to a stimulus, Pr[D]) may then be construed as accurately reflecting the input (F_1) to stage 3. Similarly, Pr[P] from the /ba/ to /pa/ (voicing) series may be construed as accurately reflecting the input of the voicing feature (f_2) to stage 3. Now, we can represent f_1 and f_2 from equations 1 and 2 as follows:

$$f_1 = \Pr[D] \text{ on the /ba/-/da/ series (PLACE)} \qquad (3)$$

$$f_2 = \Pr[P] \text{ on the /ba/-/pa/ series (VOICING)} \qquad (4)$$

Substituting equations 3 and 4 into equation 1 we obtain equation 5:

$$\Pr[T] = a_1 \Pr[D] + a_2 \Pr[P] \qquad (5)$$

Pr[T] in equation 5 represents the probability of a T response on the bidimensional /ba/ to /ta/ series.

One additional assumption will be made. This is shown in equation 6:

$$a_2 = 1 - a_1 \text{ where } 0 \leq a_1 \leq 1 \qquad (6)$$

This constraint is placed on a_1 and a_2 so that Pr[T] will equal one when both Pr[D] and Pr[P] are equal to one. Since only one parameter is being used, we delete the subscript from the constant _a_.

If we now combine equations 5 and 6 and delete the subscript on the constant _a_ we obtain equation 7:

$$\Pr[T] = a \Pr[D] + (1-a) \Pr[P] \qquad (7)$$

Equation 7 represents Model I. This model assumes independence of the features of voicing and place. If we estimate parameter _a_ from the data by the method of least squares, then Model I can be used to predict the bidimensional /ba/ to /ta/ identification function based on the unidimensional /ba/ to /da/ and /ba/ to /pa/ data.

Model II: Nonlinear Combination of Phonetic Features

A development similar to that given for Model I may be applied to equation 2. If we combine equations 2, 3, 4, and 6, we obtain equation 8:

$$\Pr[T] = a' \Pr[D] + (1-a') \Pr[P] - b(1-\Pr[D]) \Pr[P] \qquad (8)$$

Here, _a'_ is used to distinguish this parameter from the parameter _a_ of Model I. A major disadvantage of equation 8 is that it requires two different parameters, _a'_ and _b_ to be estimated from the data.

Equation 8 assumes that _S_s employ information about both phonetic features, place and voicing, to make their decision on the /ba/ to /ta/ series. However, either of these features alone may be sufficient for a _S_ to distinguish between /ba/ and /ta/ when only two response categories are permitted. For example, a _S_ could identify /ba/ and /ta/ on voicing alone (i.e., if voiced, respond /ba/; if voiceless, respond /ta/) or on place alone (i.e., if bilabial, respond /ba/; if alveolar, respond /ta/). Since these stimuli differ in both voicing and place, _S_s may use only one of these features in their decision. However, it is also possible that a particular decision on one feature necessarily entails a particular decision on the other. This is even quite likely considering the constraints on production. In production, a shift in place of articulation entails a shift in VOT, but not vice versa. On the other hand, in perception, the shift in VOT may serve as a cue to a shift in place of articulation.

Previous investigators have found that decisions based on the voicing feature are more consistent and, in some sense, easier than decisions based on other features, including place (Miller and Nicely, 1955; Studdert-Kennedy and Shankweiler, 1970: Shepard, 1972). One reason for this finding may be the multiplicity of cues to the voicing feature (Liberman et al., 1958; Lisker and Abramson, 1964; Summerfield and Haggard, 1972), as compared with the relatively restricted number of cues to the place feature. If subjects were to use only one feature, it seems likely that they would use the feature of voicing for the /ba/ to /ta/ series. We can operationalize this assumption by setting a' from equation 8 to zero. This means that $1-a'$ will be 1 and that the probability of a /ta/ response ($Pr[T]$) will be the result of the amount of the voicing feature present minus the interaction of place and voicing. This is summarized in equation 9, which represents Model II:

$$Pr[T] = Pr[P] - b(1-Pr[D])\,Pr[P] \qquad (9)$$

The term representing the dependence of voicing on place $(1-Pr[D])\,Pr[P]$ has been operationalized this way to insure that $Pr[T]$ does not become negative or greater than one.

Model II can be used to predict the bidimensional /ba/ to /ta/ series from the /ba/ to /da/ and /ba/ to /pa/ data by estimating the parameter b with the method of least squares. By setting parameter a' to zero, Model II assumes that *S*s categorize the stimulus as either /ba/ or /ta/ on the basis of the voicing feature alone. Thus, parameter b may be used as an estimate of how much a *S*'s decision on the voicing feature depends upon the place information in the stimulus.

METHOD

Subjects. Subjects were nine students in introductory psychology, participating as a part of the course requirement. Each *S* was a native American speaker of English, right-handed, and reported no history of a speech or hearing disorder.

Stimuli. The three synthetic speech syllable series were /ba/ to /da/, /ba/ to /pa/, and /ba/ to /ta/. Each series contained 11 stimuli. The /ba/ to /da/ series varied in the initial frequencies of the second and third formant transitions. The second formant varied from an initial value of 1,859 Hz (/ba/) to an initial value of 3,530 Hz (/da/) in ten equal steps. The /ba/ to /pa/ series varied in VOT from 0 msec VOT (/ba/) to a +50 msec VOT (/pa/) in 5 msec steps. Aspiration replaced the harmonics in the second and third formant transitions for the duration of the F1 cutback. The /ba/ to /ta/ series combined the two component changes in a one-to-one fashion, resulting in the third 11-stimuli sequence. All stimuli were of 300 msec duration with a 50 msec transitional period followed by a 250 msec steady-state vowel (/a/). The three series of synthetic stimuli were prepared on the speech synthesizer at Haskins Laboratories and recorded on magnetic tape.

Procedure. The experimental tapes were reproduced on a high quality tape recorder (Ampex AG-500) and were presented binaurally through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape recorder playback was adjusted to give a voltage across the headphones equivalent to 80 db SPL re 0.0002 dyn/cm^2 for the steady-state calibration vowel /a/.

On each tape *S*s heard 10 presentations of each of the 11 stimuli in random order with 4 sec between stimuli. *S*s were run in two groups, 5 *S*s in the first group and 4 *S*s in the second. Each group heard each tape three times, resulting in 30 judgments of each stimulus for each *S*. In addition, the /ba/ to /ta/ tape was presented twice more with a different set of instructions. The order of tape presentation was randomized with one group hearing the /ba/ to /da/ tape first.

For each of the tapes Ss were told that they would hear synthetic speech syllables and they were to identify them as /ba/ or /da/, /ba/ or /pa/, /ba/ or /ta/. Ss were told to record their identification judgment of each stimulus by writing down the initial stop consonant in prepared response booklets.

RESULTS AND DISCUSSION

The identification probabilities for the /ba/ to /da/ (place) and /ba/ to /pa/ (voicing) series were in accord with previous experiments. All the stimuli at one end of the series were consistently categorized one way and all the stimuli at the other end were consistently categorized the other way. There were a few transition stimuli (generally one or two in the middle of the series) which were categorized both ways at a near chance (.5) level. Data from two Ss were eliminated from subsequent analyses since they responded to the /ba/ to /da/ series at a chance level throughout. (One S came from each of the groups.) Identification functions from these two series for a typical S (S number 1) are shown in Figures 3A and 3B. The /ba/ to /ta/ identification function for the same S is shown in Figure 3C.

In order to estimate the weighting factor (a) from Model I for each S, a was allowed to vary from 0.0 to 1.0 in increments of .02. The squared error between the predicted and observed identification functions was then calculated for each value of a. The value which resulted in the minimum squared error for each S was chosen as the best estimate of a. These values of parameter a and their associated squared errors are shown in Table 1. In six out of seven Ss the proportion of the variance accounted for by the predicted values exceeded 86 percent. The mean proportion of variance accounted for over all Ss by Model I was 89.7 percent.

TABLE 1: Variance accounted for by Model I.

Subject	Constants		Minimum	Percent of
	a	1-a	Squared Error	Variance Accounted For
1	0.66	0.34	.1898	94.3
2	1.00	0.00	.3848	86.2
3	0.30	0.70	.0633	97.8
4	0.00	1.00	.1965	92.6
5	1.00	0.00	1.1814	62.9
6	0.00	1.00	.0301	99.6
7	0.64	0.36	.1148	94.7

The data were analyzed a second time for Model II. Parameter a' had been set to zero, in accord with the assumption that Ss would use only the voicing feature in making their judgment. Parameter b, the weight of the interaction term in Model II, was allowed to vary from 0.0 to 1.0 in increments of .02. The squared error between predicted and observed identification functions was also computed. The values for each S that resulted in minimum squared error are shown

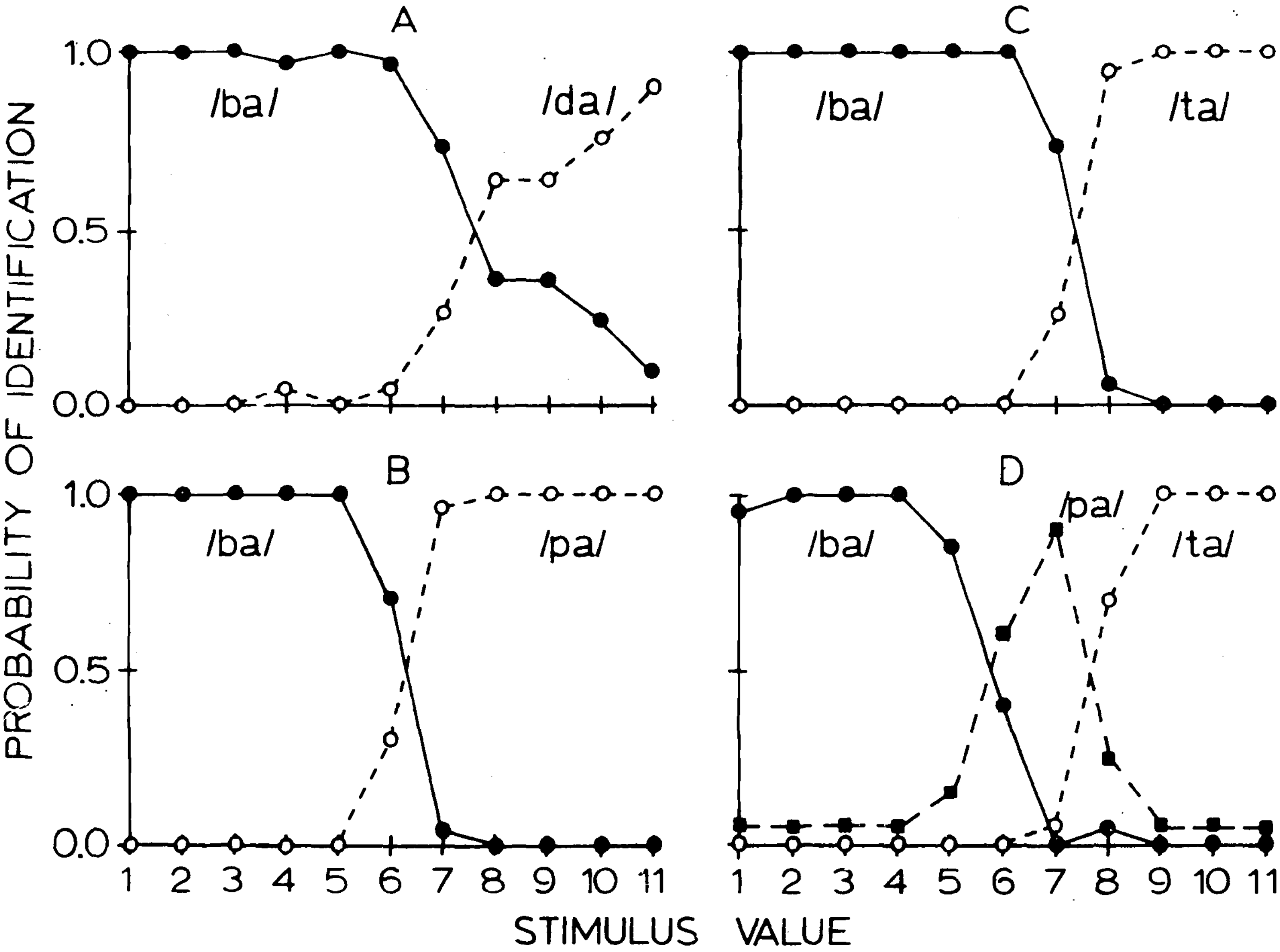


Figure 3: Identification functions for a representative S. Part A is for the /ba/ to /da/ series, part B the /ba/ to /pa/ series, part C the /ba/ to /ta/ series with two response alternatives, and part D is for the /ba/ to /ta/ series with four response alternatives.

in Table 2. The proportion of the variance accounted for by the predicted function was computed and is shown in Table 2. The proportion of the variance accounted for by Model II is greater than or equal to that accounted for by Model I for every S. The overall mean proportion of variance accounted for was 92.9 percent in Model II.

TABLE 2: Variance accounted for by Model II.

Subject	Constants a'	Constants 1-a'	Constants b	Minimum Squared Error	Percent of Variance Accounted For
1	0.00	1.00	0.66	.1831	94.5
2	0.00	1.00	1.00	.1790	93.1
3	0.00	1.00	0.30	.0570	97.8
4	0.00	1.00	1.00	.1064	95.6
5	0.00	1.00	1.00	.6570	74.8
6	0.00	1.00	0.00	.0301	99.6
7	0.00	1.00	0.64	.1154	94.7

Both Model I and Model II predict the identification probabilities along the bidimensional speech series reasonably well. However, predictions from Model II, the interaction model, fit the observed probabilities somewhat better than predictions from the additive model. There was an increase in the proportion of variance accounted for in four out of the seven Ss with Model II. For three of the Ss the variance accounted for by Model II remained the same as in Model I, although the parameter values changed. In fact, the three Ss with the highest proportion of variance accounted for in Model I are the three Ss for whom Model II shows no gain.

We suggested earlier that identification of the bidimensional series /ba/ to /ta/ might be based on the use of only one feature--voicing--since Ss were constrained to only two response categories. Parameter a' in Model II was set at zero on the assumption that the place feature is based entirely on the voicing feature and would not contribute directly to the response decision. The strength of the interaction model, Model II, can be tested by letting parameter a' vary as in equation 8. Accordingly, when the squared error between the predicted and observed probabilities was obtained by equation 8, a' was estimated to be zero for every S. The estimates of parameter b were identical to those obtained with equation 9 where a' was previously set to zero. This suggests that our original assumption was correct. Ss apparently relied more on the voicing feature than the place feature in the two category bidimensional series.

The extent to which place information enters into the voicing decision for each S is reflected in parameter b from Model II (equation 9). This parameter is greater than zero for all Ss except one, indicating that place information does affect the voicing decision, although only in terms of an interaction. Although the fit of the additive model (Model I) is good, the better fit of the interactive model (Model II) and the generally nonzero estimates of the interaction term support the notion that the phonetic features of place and voicing in stop consonants are not combined independently in stage 3.

Probabilities for the second identification function generated by Ss for the /ba/ to /ta/ series with four response alternatives were also computed. Although this condition was included in the experiment almost as an afterthought, the results were not only surprising but consistent among Ss. The identification function for the same representative S as before in this condition is shown in Figure 3D. The high probability of a P response for stimulus 7 and the distribution of P responses around this mode is of special interest. If a S were responding P at random, the Pr[P] in this series should be .25 for all stimuli instead of approximately zero everywhere except for a few stimuli. This same pattern of P responses was found for all Ss tested. The peak probability of a P and the stimulus at which it occurred are shown in Table 3. When the data for each S is broken down by tape presentation, the same results are observed (see Table 3).

TABLE 3: Peak Pr[P] in the second /ba/-/ta/ series.

Subject	Peak Pr[P]	Split-Half Pr[P]		Stimulus Where Peak Occurs
		1	2	
1	.90	1.00	.80	7
2	1.00	1.00	1.00	7
3	.85	.90	.80	6,7
4	.80	.70	.90	7
5	.80	.90	.70	7
6	.85	1.00	.70	8
7	.65	.90	.40	7

In contrast, the Pr[D] was much lower for all Ss except one, and showed greater variability when subject to split-half analysis. These data are shown in Table 4. One S reported only a single /da/ in 220 test trials.

It would appear that the occurrence of /da/ identifications in the /ba/ to /ta/ series with four response categories may be randomly distributed. On the other hand, the occurrence of /pa/ identifications is highly consistent both within and across Ss and the peak probability is never less than .65.

If the phonetic features of place and voicing combined separately and additively in stage 3 as Model I would predict, the identification functions for this second series should resemble the data for the first /ba/ to /ta/ series. This did not occur, as shown in Figure 3D. Ss showed consistent use of the /pa/ response in the second /ba/ to /ta/ series at levels well above chance expectation. The peak Pr[P] in this second bidimensional series occurred at a stimulus whose place value generally corresponded to a high Pr[D] in the /ba/ to /da/ (place) series (see Table 5). Similarly, the peak Pr[P] stimulus in the bidimensional

TABLE 4: Peak Pr[D] in the second /ba/-/ta/ series.

Subject	Peak Pr[D]	Split-Half Pr[D] 1	Split-Half Pr[D] 2	Stimulus Where Peak Occurs
1	.10	0.00	.20	7
2	.80	.80	.80	5
3	.45	.60	.30	4
4	.35	.60	.10	5
5	.50	1.00	0.00	5
6	.05	.10	0.00	10
7	.30	.20	.40	5

TABLE 5: Peak Pr[P] in the second /ba/-/ta/ series; Pr[D] in the /ba/-/da/ series; and Pr[P] in the /ba/-/pa/ series to the corresponding stimulus.

Subject	Peak Pr[P]	Corresponding Pr[D]	Corresponding Pr[P]	Stimulus
1	.90	.267	.967	7
2	1.00	.800	.967	7
3	.85	.500	.967	7
4	.80	.900	.933	7
5	.80	.933	.933	7
6	.85	.733	.833	8
7	.65	.867	1.000	7

series corresponds to a stimulus in the /ba/ to /pa/ (voicing) series which exhibits a high Pr[P] (see Table 5). A model that assumes separate, additive weighting of the features, such as Model I, would predict that the stimulus where the peak Pr[P] occurs in the second bidimensional condition would be categorized as /ta/ and not /pa/.

A model to fit these four-response data was constructed based on equation 8. However, since Model II did not fit the four-response data very well, even when parameter a' was allowed to vary, another term was added in which the place feature is dependent on the voicing feature. This model now reflects an interdependence of these two features on each other. This model, summarized in equation 10 below, has two parameters to be estimated:

$$Pr[T] = a'Pr[D] - b'(1-Pr[P])\,Pr[D] + (1-a')\,Pr[P] + (1-b')\,(1-Pr[D])\,Pr[P] \quad (10)$$

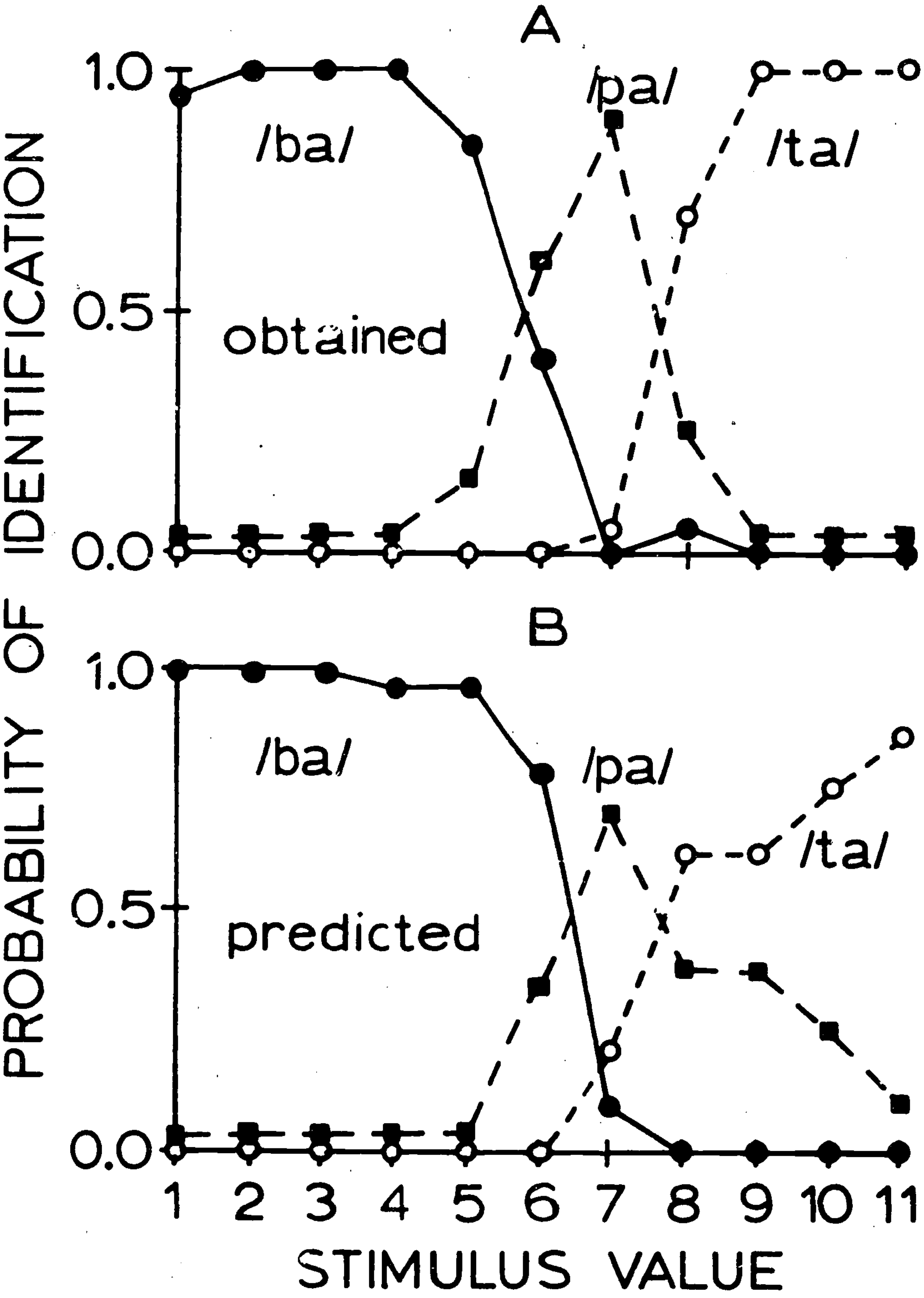


Figure 4: Observed identification function for a representative S in the four-response /ba/ to /ta/ series (part A) and the predicted function for the same S using equation 10 (part B).

Equation 10 generally failed to predict the magnitude of the Pr[P] in the four-response condition, although equation 10 did generally predict a peak at stimulus seven. The fit of equation 10 to one S's data is shown in Figure 4. The obtained identification function is shown in panel A; the predicted function derived by equation 10 is shown in panel B. The response data for all conditions for this same S were shown previously in Figure 3.

The failure of equation 10 to predict accurately the entire set of probabilities for the second /ba/ to /ta/ series may be attributed to two possible factors. First, Ss' identification functions for the two component series (/ba/ to /da/ and /ba/ to /pa/) were somewhat noisy. Second, processing in stages 1 and 2 of Figure 2 may not be independent as we have assumed. Any nonindependence of processing, especially in stage 2 where the phonetic features are extracted, would affect the assumptions made in deriving Model I and Model II.

In summary, an additive model which assumes independence in the processing of phonetic features cannot account for the identification functions when the acoustic cues underlying place and voicing in stop consonants are varied systematically. Rather, it appears that an interaction model handles the data much better and provides additional support for the evidence previously reported by Lisker and Abramson (1964, 1967) and Haggard (1970) with different experimental procedures. The perception of an acoustic cue underlying a particular phonetic feature (e.g., place or voicing) may not be invariant with changes in the acoustic cues underlying other phonetic features. This conclusion is scarcely surprising since covariations in the acoustic cues derive directly from production constraints, and is added evidence of the close link between speech perception and production.

CONCLUSION

An additive model which assumes independence of processing at all stages did a creditable job in predicting the response probabilities along a bidimensional series of synthetic stop consonants when Ss were constrained to two responses. However, a model that does not assume additive processing (i.e., nonindependence) in the stage where phonetic features are combined does even better than the independence (i.e., additive) model. When Ss are given four responses from which to choose, the additive model fails completely. In contrast, the nonadditive model, while not yielding an excellent fit, does predict occurrence of /pa/ identifications on the /ba/ to /ta/ series. Based on these perceptual data with synthetic speech stimuli, we conclude that phonetic features in stop consonants are not combined independently to form phonetic segments.

REFERENCES

Delattre, P. (1951) The physiological interpretation of sound spectrograms. Public. Mod. Lang. Assoc. Amer. LXVI, 864-875.

Gerstman, L. J. (1957) Perceptual dimensions for the friction portions of certain speech sounds. Unpublished doctoral dissertation, New York University.

Haggard, M. P. (1970) The use of voicing information. Speech Synthesis and Perception (University of Cambridge, London) 2, 1-15.

Liberman, A. M. (1957) Some results of research on speech perception. J. Acoust. Soc. Amer. 29, 117-123.

Liberman, A. M. (1970) Some characteristics of perception in the speech mode. In Perception and Its Disorders, Proceedings of A.R.N.M.D., ed. by D. A. Hamburg. (Baltimore, Md.: Williams and Wilkins).

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

Liberman, A. M., P. C. Delattre, and F. S. Cooper. (1958) Some cues for the distinction between voiced and voiceless stops in initial position. Lang. Speech 1, 153-167.

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

Lisker, L. and A. S. Abramson. (1967) Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences (Prague).

Miller, G. A. and P. E. Nicely. (1955) An analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Amer. 27, 338-352.

Shepard, R. N. (1972) Psychological representation of speech sounds. In Human Communication: A Unified View, ed. by E. E. David, Jr. and P. B. Denes. (New York: McGraw-Hill).

Shiffrin, R. M. and W. S. Geisler. (1973) Visual recognition in a theory of information processing. In The Loyola Symposium: Contemporary Viewpoints in Cognitive Psychology, ed. by R. Solso. (Washington, D. C.: Winston).

Shiffrin, R. M., D. B. Pisoni, and K. Castenada-Mendez. (in press) Is attention shared between the ears? Cog. Psychol. (Also in Indiana Mathematical Psychology Program Report 73-3.)

Stevens, K. N. (1960) Toward a model for speech recognition. J. Acoust. Soc. Amer. 32, 47-55.

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

Studdert-Kennedy, M. (1973) The perception of speech. In Current Trends in Linguistics, Vol. XII, ed. by T. A. Sebeok. (The Hague: Mouton).

Studdert-Kennedy, M. and D. Shankweiler. (1970) Hemispheric specialization for speech perception. J. Acoust. Soc. Amer. 48, 579-594.

Summerfield, A. Q. and M. P. Haggard. (1972) Perception of stop voicing. Speech Perception (Queen's University of Belfast, Northern Ireland) 2.

The Lag Effect in Dichotic Speech Perception

Emily F. Kirstein[+]

INTRODUCTION

An important factor in dichotic competition is the temporal alignment of syllable onsets at the two ears. It is well known that if different syllables are presented simultaneously to opposite ears, syllables at the right ear are more accurately identified than those at the left (Shankweiler and Studdert-Kennedy, 1967; Studdert-Kennedy and Shankweiler, 1970). Recently, it was discovered that the size of the right-ear effect can be increased by delaying syllable onsets at the right ear 5 to 120 msec behind the left and, conversely, that the ear advantage can be reduced or even reversed (giving a left-ear superiority) by causing the left-ear syllable to be delayed behind the right (Lowe, Cullen, Thompson, Berlin, Kirkpatrick, and Ryan, 1970; Studdert-Kennedy, Shankweiler, and Schulman, 1970). That is, in general, lagging syllables at either ear have an advantage over leading syllables; this "lag effect" is seen superimposed on the right-ear effect in dichotic experiments.

In control experiments syllables differing in time of arrival by 5 to 120 msec were mixed electronically and the resulting signal was delivered to one ear (monotic) or to both ears (diotic). For these conditions leading syllables were more accurately identified than lagging syllables (Lowe et al., 1970; Studdert-Kennedy et al., 1970; Kirstein, 1971; Porter, 1971a). Studdert-Kennedy et al. (1970) attributed the diotic and monotic lead effect to peripheral masking. They considered the dichotic lag effect to be a higher-level phenomenon involving competition for perceptual processing. They proposed that the lagging syllable is more intelligible than the leading syllable because it interrupts the phonetic analysis of the leading syllable and captures the speech processor.

[+] University of Michigan, Dearborn Campus, Department of Psychology.

Acknowledgment: This paper reports a portion of the research conducted at Haskins Laboratories for my doctoral dissertation in psychology at the University of Connecticut (Kirstein, 1971), under the supervision of Dr. Alvin Liberman. I wrote this paper while a postdoctoral trainee at the MIT Speech Communication Group, supported by a National Institutes of Health training grant to the MIT Research Laboratory of Electronics. I would particularly like to thank Dr. Ignatius Mattingly, Dr. Michael Studdert-Kennedy, and Dr. Kenneth Stevens for their help and encouragement.

[HASKINS LABORATORIES: Status Report on Speech Research SR-35/36 (1973)]

All the experiments referred to above used as stimuli the stop consonant-vowel (CV) syllables /ba/, /da/, /ga/, /pa/, /ta/, /ka/, where consonants differed but vowels were shared. Subsequently, other stimuli have been used to determine precisely the conditions required to elicit the lag effect. Apparently, as long as there are stop consonants in CV syllables contrasting between ears, the lag effect will persist, despite substantial variations in the acoustics of interaural competition. For example, the lag effect has been demonstrated where vowels as well as consonants vary between ears, as in /ba/-/ge/ (Kirstein, 1971), where the fundamental frequency of the syllables varies between ears (Halwes, 1969), and where the duration of the competing syllables has been shortened from the usual 300 msec to only 75 msec (Porter, 1971a). The minor influence of such acoustic variations strengthens the view that a critical condition for producing the lag effect is the "perceptual class" of the stimuli.

It is not certain whether the lag effect is peculiar to speech or whether it is a more general phenomenon of auditory perception. Darwin (1971) asserted that the lag effect was related to the perception of rapidly changing acoustic signals (transitions) whether in speech or nonspeech tasks. He supported this claim by demonstrating a lag effect for perception of pitch changes (rising, falling, level) in the initial 50 msec of a 150 msec steady-state vowel. However, Porter (1971b) found no evidence of a lag advantage in perception of formant transitions isolated from the speech signal itself. Also, in Darwin's study, the use of a vowel as a carrier of the pitch transition makes it difficult to classify the sounds unambiguously as nonspeech. Thus, at present there is no strong evidence to refute the hypothesis of Studdert-Kennedy et al. (1970) that the lag effect is a speech perception phenomenon.

Not all classes of speech sounds are equally effective for eliciting the lag effect. Porter (1971b) compared stop consonants (/b/, /d/, /g/) with sonorant consonants (/l/, /w/, /y/). He found that some subjects had a lag effect for both stops and sonorants while other subjects had the lag effect for stops only. Porter, Shankweiler, and Liberman (1969) presented steady-state vowels dichotically with delays between ears, and they found a slight advantage for leading over lagging vowels. However, Kirstein (1971) found a preference for lagging vowels if the vowels were embedded in CV syllables. She also found a lag effect for isolated steady-state vowels among subjects who had previously taken a dichotic test involving stop consonants. The finding that the lag effect is an extremely robust effect for stops, less robust for sonorants, and marginal for vowels supports the view that the effect is related to special decoding processes in speech perception. The term "encoding" has been used by Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) to refer to the fact that the acoustic cues for perception of a particular phoneme may be greatly affected by the nature of the adjacent phonemes. Liberman et al. (1969) proposed that highly encoded phonemes like stop consonants require special decoding to arrive at phonetic identification, while unencoded phonemes like the isolated steady-state vowels could optionally be identified through purely auditory perception modes. (Sonorant consonants are more highly encoded than vowels, but less encoded than stops.) The finding that the lag effect occurs for vowels under some circumstances but not under others suggests that the lag effect is related to the perceptual mode (speech or nonspeech) rather than to some acoustic feature of the stimulus.

The present research is concerned with the methodology of lag effect experiments, and specifically with the role of attention. In dichotic experiments listeners are generally required to attend simultaneously to both ears. The question examined here is whether leading syllables might be as accurately identified as lagging syllables are if attention were concentrated on leads or lags only, rather than divided between them. It had been claimed by Inglis (1965) and by Treisman and Geffen (1968) that the ear advantage in dichotic tasks can be attributed to systematic biases on the part of subjects in their order of reporting the two ears or in the distribution of attention between ears. To control for these factors in the study of ear effects Kirstein and Shankweiler (1969) introduced the procedure of having listeners concentrate on one ear at a time, reporting only the syllables at the "attended" ear. They found that for dichotically presented stop consonants report was more accurate under right-ear attention than under left-ear attention. They concluded that neither response bias nor the distribution of attention could explain the ear asymmetry. In dichotic experiments where syllables are temporally offset between ears, the distribution of attention and response ordering might also affect the pattern of identification errors. It seemed desirable, therefore, to determine whether the lag advantage would occur with attention directed selectively toward lagging or leading syllables.

METHOD

A dichotic tape consisting of pairs of CV syllables was constructed. Syllables within a pair always contrasted in the initial consonant (/b/, /d/, or /g/) but shared the same vowel. Within a pair, one syllable was always delayed relative to the other by 10, 30, 50, 70, or 90 msec. This tape was presented to subjects under three different task conditions. In the two-response task the listeners were instructed to report both consonants on each trial and to indicate which was the clearer. This is essentially the method of Studdert-Kennedy et al. (1970). In the ear monitoring task subjects were instructed to concentrate their attention on a particular ear and to report only the consonants at the "attended" ear. In the temporal order task listeners were instructed to attend to the order of arrival of the consonants within a pair and report only the leading stops or only the lagging stops according to instructions.

Stimuli. Nine consonant-vowel syllables were synthesized on the Haskins Laboratories' parallel resonance speech synthesizer. These were /ba/, /da/, /ga/, /bɛ/, /dɛ/, /gɛ/, /bɔ/, /dɔ/, /gɔ/. The duration of each syllable was 350 msec. Syllables beginning with /d/ or /g/ started with a 10 msec noise burst, followed by appropriate formant transitions. No burst was needed with /b/. In all cases the formant transitions were completed and the steady-state vowel parameters reached within 70 msec.

The intelligibility of the syllables was assessed by asking four listeners unfamiliar with synthetic speech to identify the consonants in the 180-trial randomization. Ninety-five percent correct identifications were obtained, an adequate level of intelligibility for the dichotic tests.

Waveforms of the stimuli were stored on a computer disc file using the Pulse Code Modulation System (Cooper and Mattingly, 1969). This computer system also controlled the alignment of syllable onsets as the syllables were recorded in pairs in a specified order onto the two-channel dichotic test tape.

In the design of the dichotic tape, care was taken to counterbalance to prevent confounding of ear effects, lag effects, and tape channel imbalance. Each ear received the same number of lag and lead trials, and the same permutations of the syllables. From the nine CV syllables there are 18 possible permutations of two syllables where only the consonants differ. A 180-trial randomization was assembled in which each of the 18 permutations occurred twice at each of the five delay intervals (10, 30, 50, 70, and 90 msec), once with channel-1 delay and once with channel-2 delay. All conditions of offset were randomly ordered on the tape. There was a 6-sec pause between pairs and a 10-sec pause after ten pairs.

Procedure. The dichotic tape was played from a General Radio stereo tape deck into a special amplifier built by D. Zeichner of Haskins Laboratories for group dichotic experiments. As many as six subjects could be tested at one time. The subjects listened to the tape over Grason-Stadler stereo headphones. The tape was presented at a comfortable listening level, and the output intensity from the two channels was equated to within 1 db with the aid of calibration signals on the tape.

As an added control for channel effects, the dichotic tape was always presented twice within a test session with the headphone orientation physically reversed on the second presentation.

Two-response task. The subjects were told that they would receive two different syllables on each trial, one syllable to each ear, and that the syllables would differ between ears in the consonant (/b/, /d/, or /g/), never in the vowel. The instructions were to report both of the consonants, giving two different responses and guessing if necessary. An added aspect of this task was the clarity judgment. The responses were to be ordered on the answer sheet so that the clearer consonant was written in the first column for each pair and the less clear consonant in the second column. If one of the two responses was a guess, it was to be written in the second column.

Each subject had 360 trials in a single 1-hour test session.

Ear monitoring task. The subjects were told that they would receive two different syllables on each trial, one syllable at each ear, and that the syllables would differ between ears in the consonant, (/b/, /d/, or /g/), never in the vowel. The instructions were to attend to a particular ear designated by the experimenter and to write down on each trial the consonant arriving at the "attended" ear. The subjects were required to respond on every trial even if the response was a guess. Each subject had 360 trials under left-ear attention and 360 trials under right-ear attention.

Each subject had two 1-hour (360-trial) test sessions, with each session subdivided into four 90-trial blocks. At the start of a block the subject was told whether to report the right ear or the left, and this instruction was in effect for the entire block. Within a session the order of the blocks was Right-Left-Left-Right or Left-Right-Right-Left; a subject was randomly assigned to one of these orders for his first session and was automatically assigned to the other order for the second session.

Temporal order task. The subjects were told that they would receive two different syllables on each trial, one at each ear; that the syllables would differ between ears in the consonant (/b/, /d/, or /g/); and that the syllables would also differ slightly in onset time, with the leading syllable randomly at the right or left ear. The instructions were to attend to the order of arrival of the two syllables and to report either the lagging or leading syllable, as specified by the experimenter. The subjects were required to respond on every trial even if the response was a guess. Each subject had 360 trials reporting the leading syllables and 360 trials reporting the lagging syllables.

Each subject had two 1-hour (360-trial) test sessions, with each session subdivided into four 90-trial blocks. At the start of a block the subject was told whether to report the leading syllables or the lagging ones, and this instruction was in effect for the entire block. Within a session the blocks were arranged Lags-Leads-Leads-Lags or Leads-Lags-Lags-Leads; a subject was randomly assigned to one of these orders for his first session and was automatically assigned to the other order for the second session.

Subjects. The ear monitoring and temporal order tasks were originally studied together as part of a single experiment. A group of 12 subjects took both tasks, half taking the temporal order first, and half the ear monitoring task. Later, an additional group of 10 subjects was run on the ear monitoring task only, making a total of 22 ear monitoring subjects. (The data analysis revealed no systematic differences between the original group of 12 and the added group on the ear monitoring task. For purposes of analysis sometimes only the original 12 subjects are considered while in other cases data are presented for the entire group of 22.)

The two-response task was performed by 24 subjects, none of whom had previously participated in dichotic experiments.

The subjects, all Introductory Psychology students at the University of Connecticut, received course credit for their participation. They were native speakers of English, were self-classified "right-handers," and had (to their knowledge) normal hearing in both ears.

RESULTS

Two-response task. The percent correct responses on the two-response task is shown in Figure 1, presented according to the various stimulus conditions: lag vs. lead, right vs. left ear, and length of interaural delay. The data replicate the earlier findings: lagging syllables were more accurately identified than leading syllables, and the right ear was more accurate than the left ear. The stops that were lagging and at the right ear were most often correctly identified, next in accuracy were left-ear lags, then right-ear leads, and finally left-ear leads. The only exception to this ordering was at 10 msec, where right-ear leads were slightly better than left-ear lags. An analysis of variance on the two response data is summarized in Table 1. Both the right-ear effect and lag effect were highly significant. Also, averaging over lags and leads and over the left and right ears, there was a statistically significant rise in the number of correct identifications for longer interaural delay intervals. Figure 1 shows that both the lag advantage and ear advantage vary in magnitude depending on the length of interaural delay; however, in the analysis of variance only the interaction of lag effect with delay was significant.

The clarity-judgment instructions were apparently being followed, since errors occurred primarily on second responses: 5 percent of responses in the

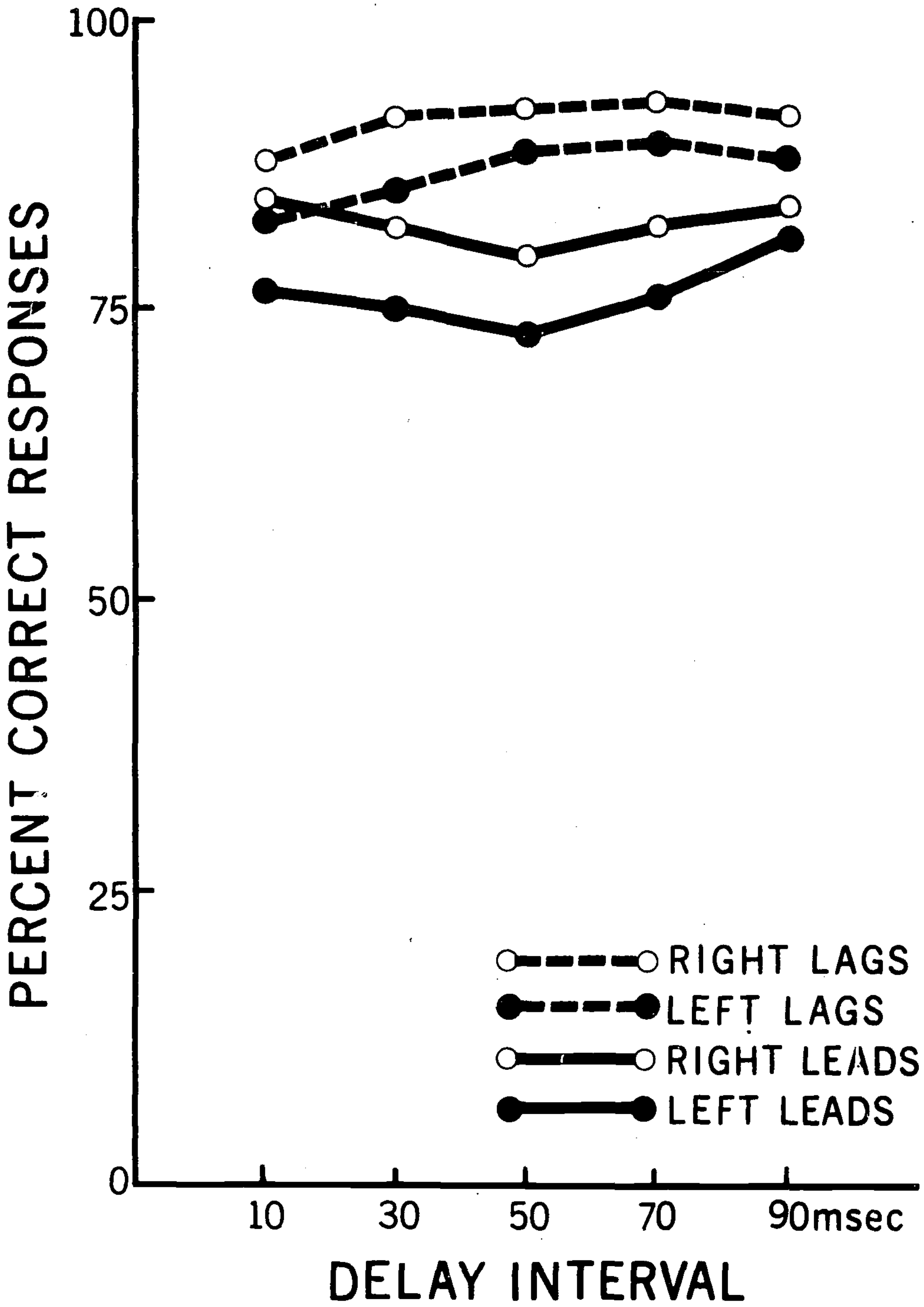


Figure 1: Percent correct responses on two-response task (N=24). Each point is based on 864 trials, 36 trials per subject.

TABLE 1: Analysis of variance summary based on the number of correct responses in each task.

Source of Variation	Two-Responses (N=24)			Ear Monitoring (N=22)			Temporal Order (N=12)		
	F	df	p	F	df	p	F	df	p
Right Ear Effect	21.8	1,23	*	15.1	1,21	*	6.24	1,11	<.05
Lag Effect	28.6	1,23	*	37.5	1,21	*	17.3	1,11	<.005
Length of Delay	6.78	4,92	*	46.8	4,84	*	6.82	4,44	*
Ear Effect x Delay	1.66	4,92	NS	7.88	4,84	*	2.95	4,44	<.05
Lag Effect x Delay	6.90	4,92	*	9.75	4,84	*	7.95	4,44	*
Ear x Lag Effect	2.01	1,23	NS	8.47	1,21	<.01	1.02	1,11	NS
Ear x Lag x Delay	.57	4,92	NS	.94	4,84	NS	.54	4,44	NS

* indicates effect is significant with p<.001 or better.

first column were errors, but 28 percent of second responses were errors. Errors on second rssponses decreased with longer interaural delay intervals, but the first-response error rate was independent of delay interval. An analysis of first responses showed that lagging stops were judged to be clearer than leading and the right ear was judged clearer than the left. (The first-response results are shown in Figure 7, which is discussed later. Of the 24 subjects, 23 favored lags over leads in first responses, and 18 favored the right ear over the left. (Only the direction of preference was tabulated, not the magnitude of the effect.)

Ear monitoring and temporal order tasks. The results of the ear monitoring task are shown in Figure 2 and the temporal order results in Figure 3. Both tasks gave essentially the same results. On the ear monitoring task, listeners were more accurate on right-ear attention than on left-ear attention, and they were more accurate when the attended ear received a lagging syllable than when it received a leading syllable. On the temporal order task, listeners made fewer errors under "report lags" instructions than under "report leads" instructions, and report of either lags or leads was more accurate from the right ear than from the left. The analysis of variance for the selective listening tasks is summarized in Table 1. For both tasks the lag effect and right-ear effect were significant, and both the ear effect and the lag effect showed significant variations in magnitude with the length of interaural delay. Accuracy of report by ear or by temporal order improved significantly with longer delay intervals.

Since 12 subjects took both the temporal order and ear monitoring tasks, it was possible to make a comparison of the consistency of individual ear effects and lag effects across the two tasks. A measure of each subject's lag effect was obtained for the temporal order task by subtracting the number of correct responses under "report leads" from the number of correct responses under "report lags." For the ear monitoring tasks, a measure of the lag effect was obtained by summing the number of correct right-ear and left-ear responses when the attended ear received lagging syllables and subtracting the number of correct responses when the attended ear received leading syllables. Similarly, a measure of the ear effect was determined for each subject on each task. The reliability of individual differences across tasks was assessed by calculating the Pearson product-moment correlation coefficient. The individual lag effect measures gave a correlation coefficient of .85 across tasks, and the ear effect measures gave a correlation coefficient of .95, both of which are highly significant, with $p < .001$. These results indicate that individual lag effect scores and individual ear effect scores are reliable, even when the measures are obtained on different tasks.

The types of errors made on the ear monitoring and temporal order tasks fell into two categories. The response could be identical to the unattended syllable, in which case it is termed an "intrusion" error; or the response could differ from both the attended and unattended syllables, in which case it is termed a "nonintrusion" error. Figure 4 gives a breakdown of all responses as correct responses, intrusion errors, or nonintrusion errors. Intrusions were the primary source of errors under both attention conditions. That is, listeners had difficulty in discriminating between the attended and the unattended syllables. The ability to select the attended syllable improved with longer delay, and Figure 4 shows that this improvement is entirely due to a reduction in the number of intrusions of unattended syllables. Somewhat surprisingly, nonintrusion errors increased slightly with longer interaural delay, whereas they might have been

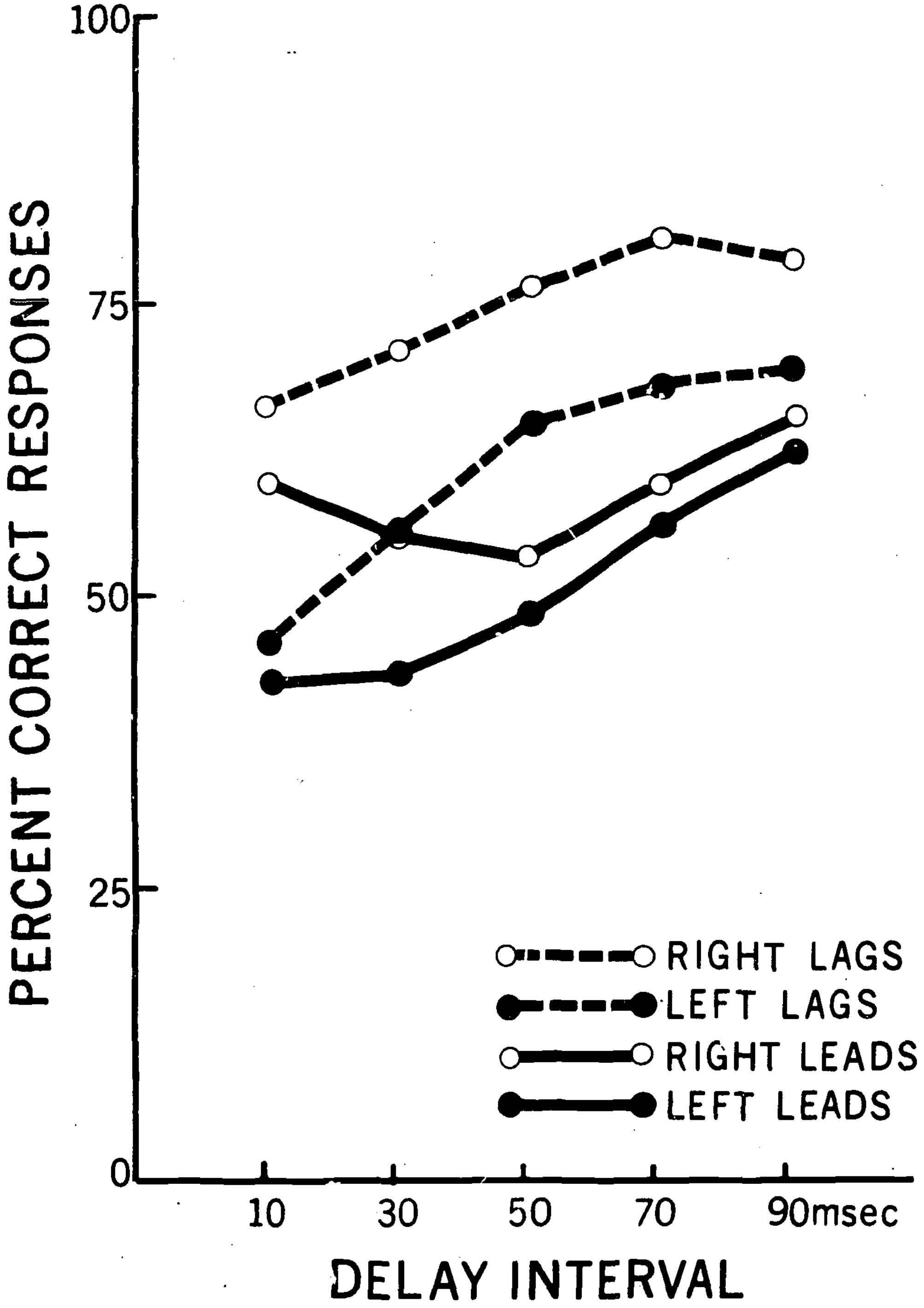


Figure 2: Percent correct responses on ear monitoring task (N=22). Each point is based on 792 trials, 36 trials per subject.

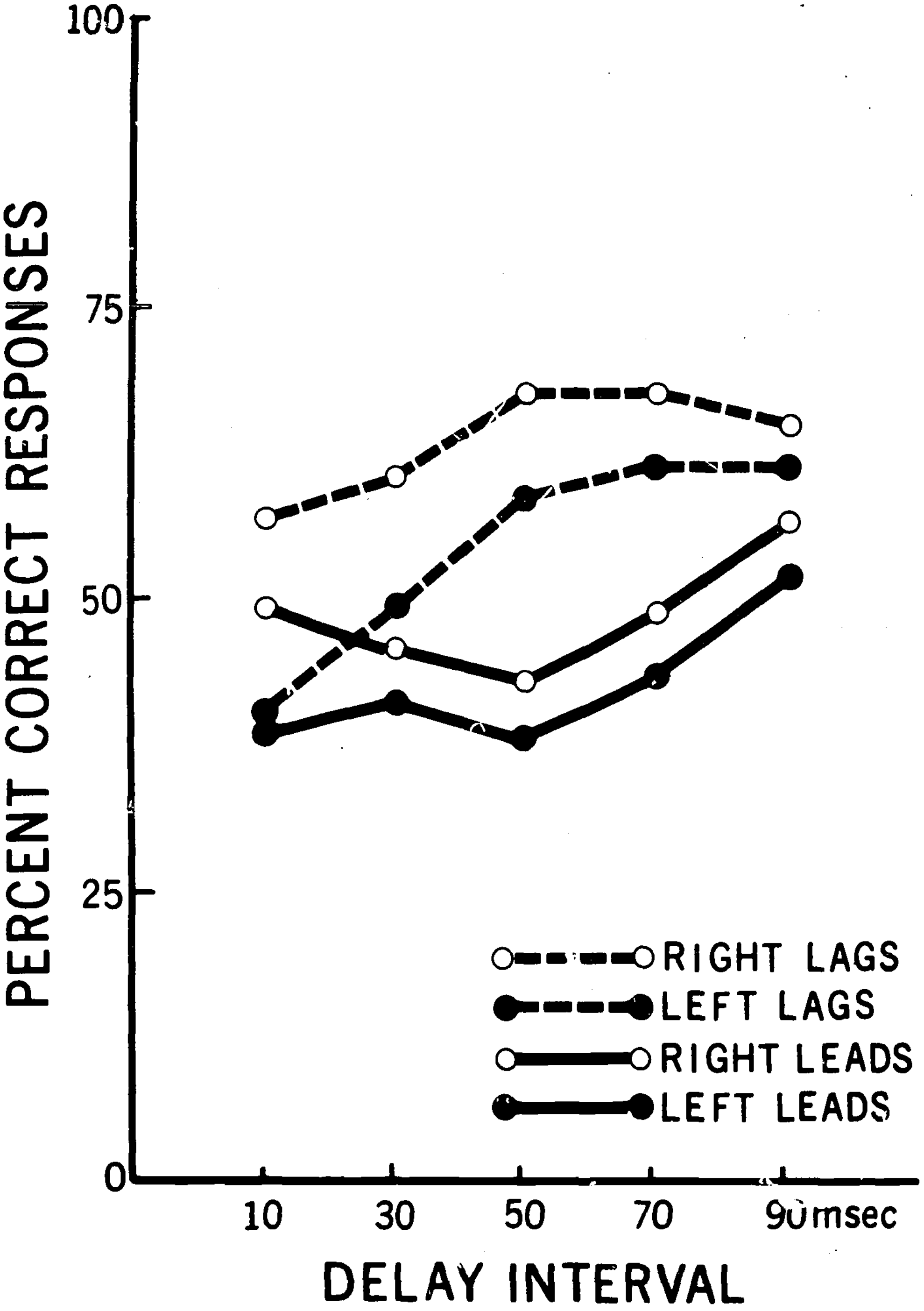


Figure 3: Percent correct responses on temporal order task (N=12). Each point is based on 432 trials, 36 trials per subject.

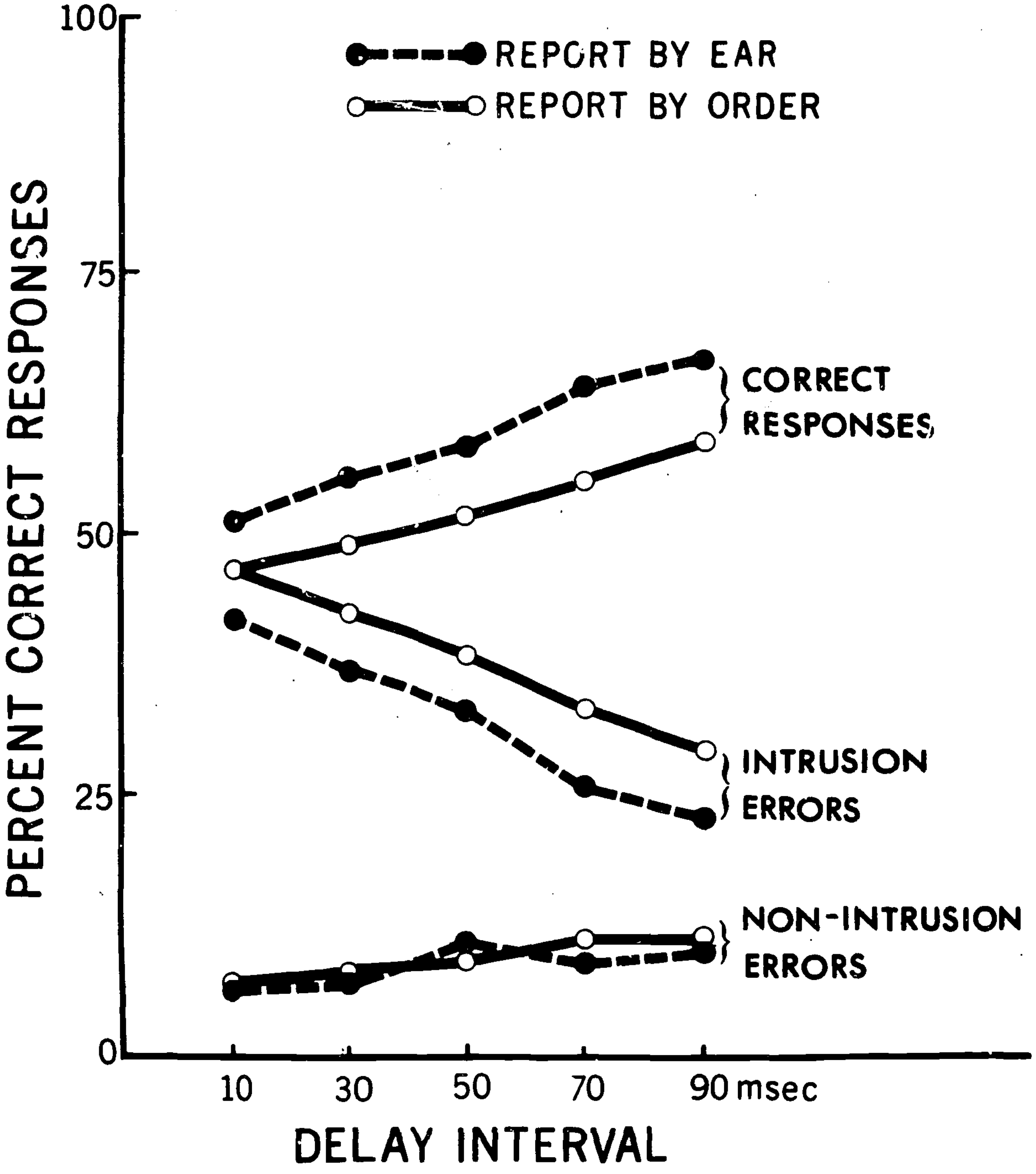


Figure 4: Percent correct responses, intrusion errors, and nonintrusion errors on selective listening tasks (N=12).

expected to decline. For the ear monitoring task, the increase in nonintrusion errors with longer interaural delay proved to be statistically significant [Friedman two-way analysis of variance $\chi_r^2 = 14.8$ $p < .01$ (Siegel, 1956)]. Overall, reporting a particular ear was more accurate than report by order of arrival ($F = 17.8$ df = 1,11 $p < .005$).

Lag effect and ear effect as functions of delay interval. Thus far it has been demonstrated that there is an advantage for lagging syllables and right-ear syllables in all three tasks. The present section examines how the magnitude of lag effect and ear effect varied with length of delay.

The lag effect was treated independently of the right-ear effect by computing the mean percent correct on right- and left-ear lags and subtracting from this the mean percent correct on right- and left-ear leads. Figure 5 displays these lag effect scores at each delay for the three tasks. For the two-response task separate plots are shown for the lag effect in clarity judgments (first responses) and for the lag effect in intelligibility (both responses). The same trend was observed in all tasks: the advantage for lags over leads progressively increased with longer interaural delay up to 50 msec; with still longer delays the lag effect began to diminish. The lag effect showed a maximum at 50 msec for all three tasks, so this location must be considered a reliable finding, at least for this particular set of stimuli.

The finding of a lag effect peak in the 50-70 msec delay range is also in agreement with other observations (Berlin, Lowe-Bell, Cullen, Thompson, and Loovis, 1973; Studdert-Kennedy et al., 1970; Porter, 1971b).

A right-ear effect score at each delay was computed similarly by subtracting the mean percent correct on left-ear leads and lags from the mean percent correct on right-ear leads and lags. These scores are plotted in Figure 6. The right-ear advantage was greatest at short interaural delay intervals and declined with longer delays. Again all three tasks showed the same trend, although for the two-response task the change in ear effect with delay was not statistically significant. From these results it can be inferred that the right-ear advantage would have been maximal with simultaneous onsets.

While the pattern of results was basically the same in all tasks, the magnitude of ear advantage and lag advantage did vary considerably among the tasks, as can be seen in Figures 5 and 6. Statistical factors were considered first in trying to account for these magnitude differences. It can be shown that the maximum possible lag effect or ear effect which can be obtained in the two-response task is only 50 percent whereas in the selective listening and clarity judgment tasks a 100 percent lag effect or ear effect could be obtained.[1] A correction

---

[1] The 50 percent ceiling on the lag effect in the two-response task derives from the fact that the subject must give two different responses on each trial from a set of only three possible responses. Suppose that the listener always heard correctly the lagging consonants but not the leading consonants. His guess for the second response would nevertheless be correct for the leading consonant for half the trials. Thus, the lag effect would have been 100 percent considering only first responses but is automatically reduced by half when both responses are considered. The same argument applies to the ear effect. Using a larger response set would have given a higher ceiling on the magnitude of the effects obtained in the two-response task.

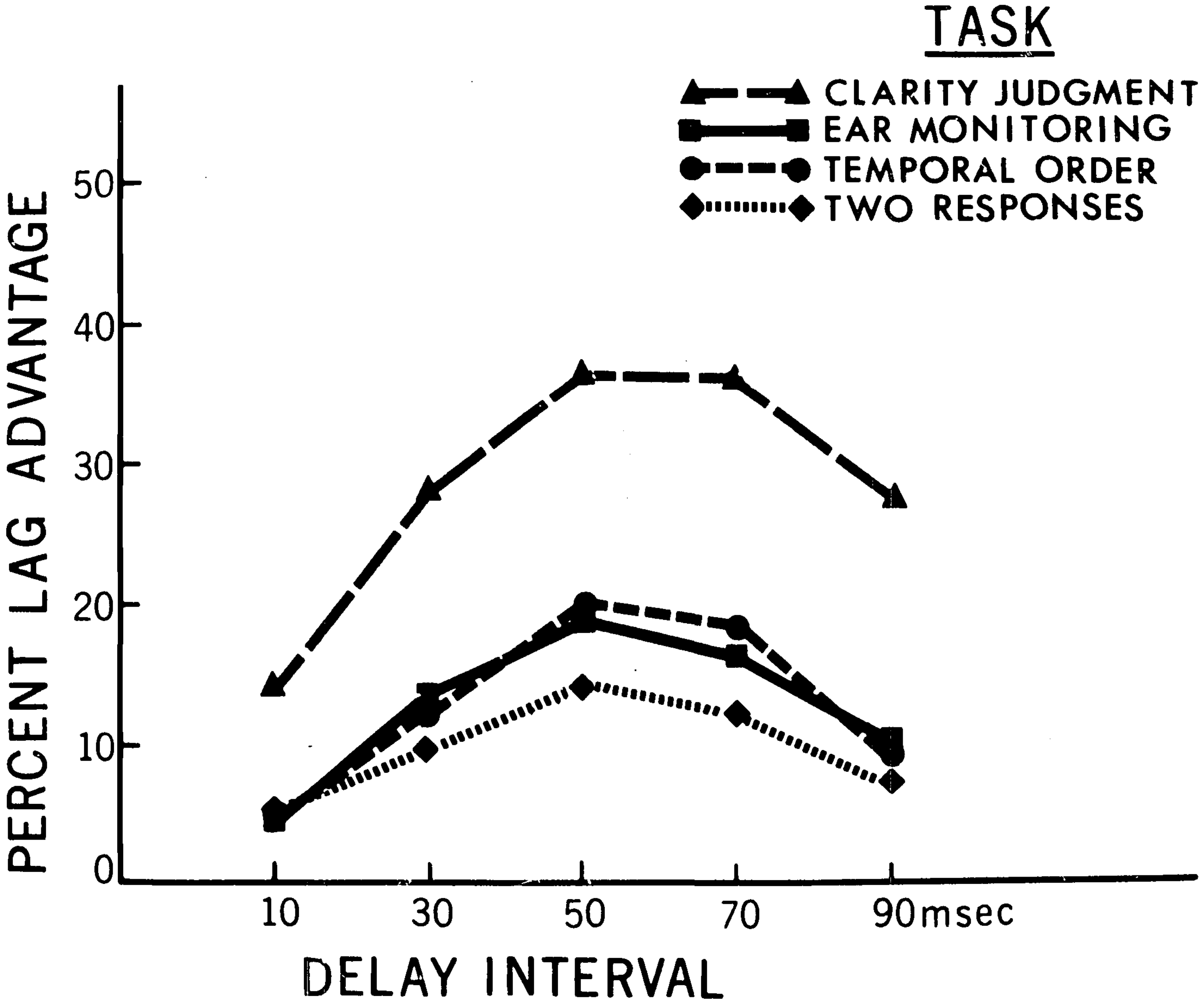


Figure 5: Magnitude of lag effect as a function of interaural delay interval. Comparison of clarity judgments, ear monitoring, temporal order, and two-response tasks.

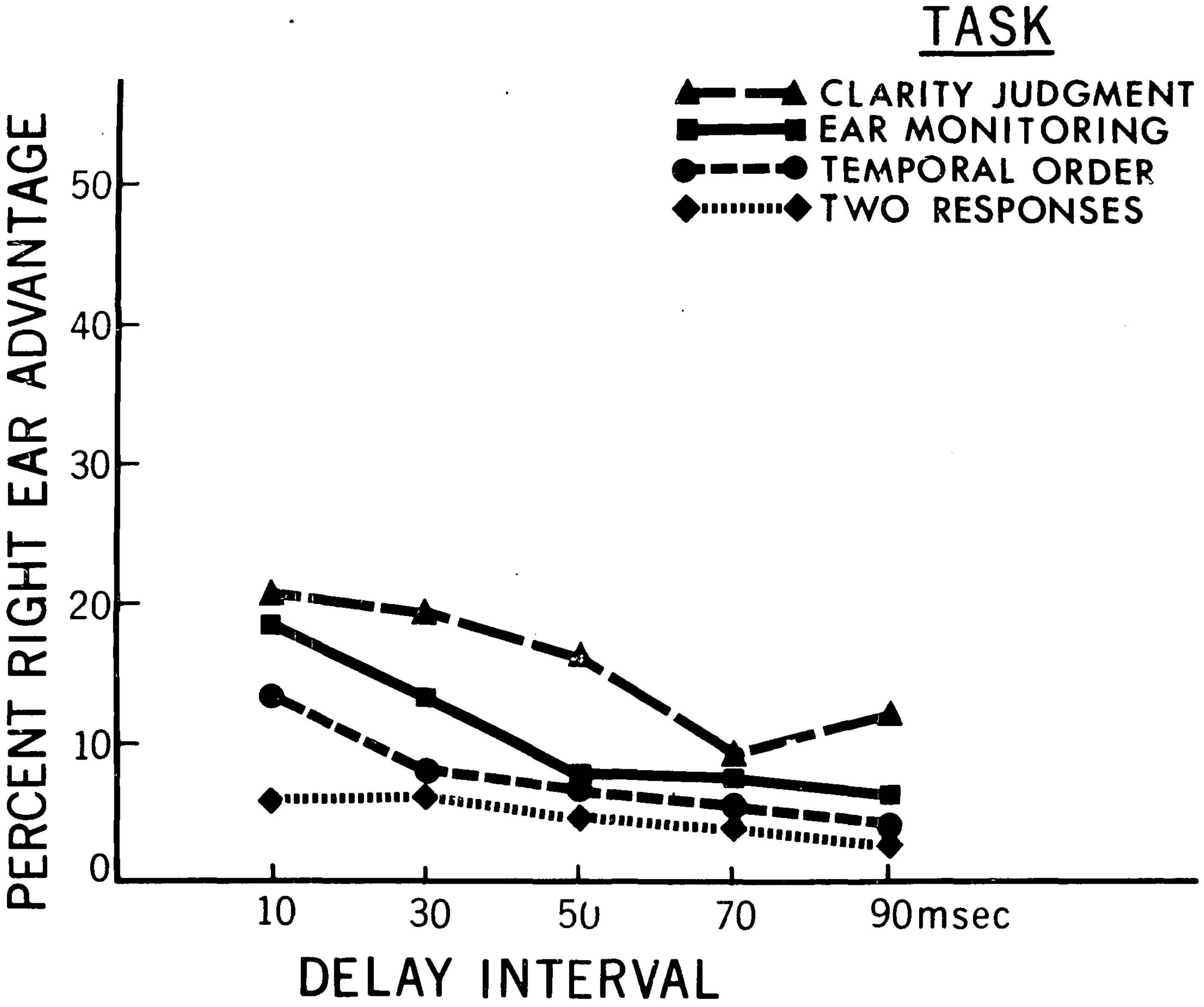


Figure 6: Magnitude of right-ear effect as a function of interaural delay interval. Comparison of clarity judgments, ear monitoring, temporal order, and two-response tasks.

must be applied to the two-response data (multiplying all ear effect and lag effect scores by 2) before comparing the two-response with the other tasks. However, even after this correction has been applied, the magnitude of the effects obtained in clarity judgments is still greater than in any of the other tasks. A possible explanation for this result is that clarity judgments are more sensitive than the other tasks to the effects being studied. The clarity judgment task required listeners to compare the competing syllables qualitatively, while the two-response and selective listening tasks asked subjects to identify the syllables.

The idea that clarity judgments are more sensitive than identification tasks receives support from a more detailed comparison of clarity judgments with responses on the ear monitoring task. Figure 7 compares these two tasks directly, plotting on the same chart the percent "clearer" judgments for the various temporal offset conditions against the percent of trials on which these same syllables were correctly identified under selective listening. If the curves for the two tasks were to coincide, this would indicate that the subjects could correctly identify in ear monitoring only those stops which were independently judged to be the relatively clearer within the pair. Divergence between the ear monitoring and clarity judgment curves gives the percentage of trials on which the "less clear" consonant could be identified when attention was concentrated on that sound. It is evident that very frequently the "less clear" sound could in fact be identified. It is interesting also that at short delays the responses given were not greatly affected by the specific task instructions, while longer delays produced greater divergence between tasks. The greater sensitivity of the clarity judgment over the identification task can be seen in the fact that as we move from the most favored condition (right-ear lags) to the least favored conditions (left-ear leads) the change between conditions is greater for clarity judgments than for identification.

DISCUSSION

All three tasks gave essentially the same patterns of identification errors. The effects observed in all tasks were the lag effect, the right-ear effect, the variation in the magnitude of lag effect and ear effect with delay, the improved performance with longer delay, and the susceptibility to intrusion errors in the selective listening tasks. The consistency in error pattern across tasks is convincing evidence that there are genuine variations in intelligibility of dichotic syllables depending on ear of arrival, order of arrival, and temporal offset. These effects are clearly not under the listeners' voluntary control, and they cannot be explained in terms of attentional strategies or response order.

It is often said that dichotic presentation causes errors because of perceptual competition between stimuli, that is to say, because the perceptual system is unable to process both ears simultaneously. The term "perceptual competition" is usually interpreted to mean that verbal stimuli presented simultaneously to opposite ears must compete for entry to language processing areas. The assumption that only one ear at a time can have access to language areas of the brain underlies much theorizing about the ear effect and lag effect. For example, Kimura (1961, 1967) attributed the right-ear advantage for dichotically presented verbal stimuli to competition between inputs for access to language processing areas of the left hemisphere; the right ear was thought to win the competition because of the greater strength of its neuroanatomical connections to the left hemisphere. The explanation of the lag effect offered by Studdert-Kennedy et al.

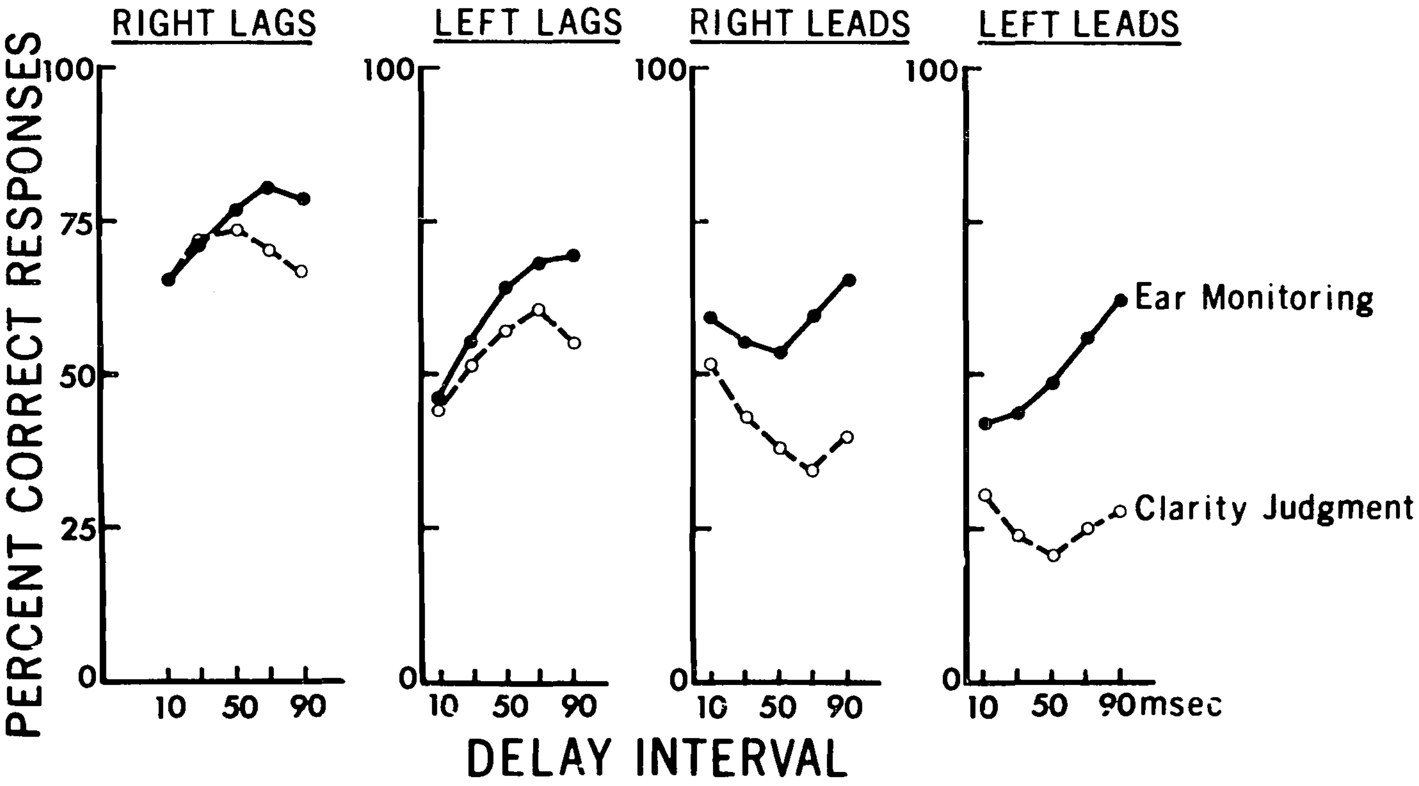


Figure 7: Clarity versus identifiability of dichotic stop consonants. Clarity judgments and ear monitoring scores are compared for all delay conditions.

(1970) also assumes that only one ear at a time can be admitted to the speech processor. They proposed that the lag effect occurs because the leading syllable is ejected from the processor when the lagging syllable arrives on the opposite channel. It will be argued in the discussion that follows that recent data do not support the idea that the ears must compete for entry to the speech processor. An attempt will be made to redefine the notion of dichotic competition and to account for the ear effect and lag effect in light of that definition.

The view that dichotic stimuli compete for perceptual processing was elaborated by Broadbent (1958) in his "filter theory" of selective attention. According to that theory, the flow of sensory data to central processing areas is regulated by a filtering mechanism which blocks peripherally all irrelevant sensory channels. Broadbent treated the two ears as separate sensory channels. He considered that in dichotic listening only one ear could be processed at a time and that the system would have to "switch channels" in order to process both ears. Peripheral blocking of one ear and channel switching were assumed to be under a subject's voluntary control.

A motivation for the filter theory was to explain the finding of Cherry (1953) that when different continuous messages were presented to opposite ears listeners could easily repeat back (shadow) the message at one ear and ignore the other. Moreover, they showed no retention of the unattended message. Broadbent attributed the shadowing results to the filtering out of the unattended message so that only the attended message could have access to linguistic processing. However, in subsequent research using the shadowing technique the unattended ear was not fully suppressed. For example, Treisman (1960) found that words in the unattended messages would occasionally be repeated by the shadower if those words were semantically probable within the context of the attended message. Also, if the experimenter switched messages between ears, subjects would sometimes also switch ears unconsciously, maintaining the continuity of the shadowed message. Such intrusions from the unattended ear relating to the semantics of the messages indicate that the unattended message must have been analyzed linguistically, and subjects' failure to retain the content of the unattended message in these tasks must be attributed to memory rather than perception.

Ear monitoring experiments with dichotic syllables prove even more strikingly that listeners cannot voluntarily select a particular ear for perceptual analysis and filter out the other. When stop-vowel syllables are presented dichotically with simultaneous onsets, subjects frequently confuse ears when instructed to report a particular ear (Kirstein and Shankweiler, 1969; Halwes, 1969) and when instructed to report both syllables by ear (Gerber and Goldman, 1971). The present results expand these earlier findings by showing that confusions between ears persist even when syllable onsets are not precisely simultaneous. Accuracy of monitoring is, however, related to interaural delay interval, and the fact that intrusion frequency depends on the timing relation between ears is strong evidence that listeners cannot voluntarily exclude a particular ear from processing. The results suggest, rather, that selection from a particular ear depends on the physical relation between the dichotic stimuli. The ear monitoring results contradict two assumptions of the filter theory: first, that dichotic inputs are always strongly "tagged" by ear of origin, and second, that listeners can voluntarily turn off an ear in dichotic tasks.

If the unattended ear cannot be voluntarily inhibited peripherally, then how are we to account for the ease of selectively shadowing a particular ear for continuous dichotic messages? Moray (1969) considered this matter and concluded that the mechanism of selection of a particular ear is the same in dichotic listening as in ordinary binaural situations where many messages are arriving simultaneously to both ears--the so-called "cocktail party" effect. A particular message can be selected for attention providing there are sufficient physical cues to establish a distinct spatial origin for the message. In everyday listening situations the two ears do not behave as independent channels. Ordinarily the same signals arrive at both ears, perhaps with slight differences in time, phase, or intensity between ears; these interaural differences provide the physical cues for spatial localization of the sound. Thus, normally inputs at the two ears are integrated to yield a unitary percept and are compared to locate the sound source (Cherry, 1961). Moray proposed that dichotic inputs are handled in essentially the same manner.

If we assume, following Moray, that in dichotic listening the auditory system compares stimuli from the two ears and locates the source based on differences between ears, we can then readily understand why reporting stop consonants from a particular ear is difficult, while shadowing continuous messages is much easier. For ongoing messages the acoustic signals at opposite ears would be generally quite distinct at any point in time, so that by comparing the two inputs the auditory system could establish two distinct sound sources and a subject could shadow the message emanating from a particular location. In stop-vowel syllables the acoustic information distinguishing one stop from another is contained within the first 70 msec or less of any syllable, and the vowels are acoustically identical at the two ears. Thus, if selection between ears presupposes a clear acoustic distinction between simultaneous dichotic signals, it is understandable that selection would be faulty for CV syllables. The role of acoustic similarity between ears in selective listening for dichotic stop consonants was demonstrated convincingly in an experiment by Halwes (1969). He presented CV syllables dichotically with syllable onsets precisely aligned for simultaneity and compared accuracy of selective report for syllable pairs which shared fundamental frequency at the two ears or which differed in fundamental frequency. When fundamental frequency was shared, listeners were unable to distinguish the attended from the unattended ear; performance was significantly improved when fundamental frequencies varied. The present study shows that another physical dimension, interaural asynchrony of onsets, is also important in selection.

Confusion between ears for dichotic stops was attributed by Halwes to an acoustic fusion effect. Halwes claimed that the perceived localization of a syllable presented to one ear was shifted toward the midline when an acoustically similar syllable was delivered at the opposite ear. Often the listeners heard a single syllable localized at the midline or diffusely rather than at a particular ear. Earlier, a similar phenomenon was described by Broadbent and Ladefoged (1957). They presented the first formant of a synthetic vowel to one ear and the second formant to the opposite ear; these fused perceptually into a single vowel.

While listening to the tapes for the present experiments, I observed frequent fusion for short delay (10-30 msec) trials. For these, generally only one of the two stops could be heard, and it often could not be definitely assigned to either ear. For longer delays, two stimuli could usually be detected although often the identity of one was still unclear. These observations accord with the experimental data. On the two-response task, for example, the subjects could identify one stop from each pair, regardless of delay, but correct identification of both stops was facilitated by longer offsets. For ear monitoring, longer delays reduced intrusions; this effect can be explained by the assumption that selection can occur only under conditions where both syllables can be detected.

A somewhat surprising result in the ear monitoring task was that while intrusion errors decreased with longer interaural delay, nonintrusion errors increased. The increase in nonintrusion errors with longer delays may result from the fact that at the shortest delays one nearly always hears one of the two syllables clearly, although that might be the unattended syllable. At longer delays listeners would be better at discriminating the attended from the unattended syllable, but they would often be unable to identify the attended syllable. That is, the increase in nonintrusion errors with longer delay can be related to the temporal alignment condition where the attended syllable can be selected but not identified.

Dichotic fusion phenomena are of interest because they support the hypothesis that there is perceptual integration of dichotic stimuli. Cutting (1972) proposed that dichotic fusion can occur at various levels of perceptual processing. He considered the effects described by Halwes (1969) and by Broadbent and Ladefoged (1957) to be low-level or auditory fusions because both effects depend on a purely acoustic property of the stimuli: the fusion is disrupted if fundamental frequency is varied between ears. Other types of fusion have been discovered in recent experiments where the integration apparently arises at higher perceptual levels and does not depend so critically on acoustic parameters. An example of a higher-level fusion is the phenomenon of "feature blending" (Studdert-Kennedy and Shankweiler, 1970). If stop consonants contrast between ears in two distinctive features, place of articulation and voicing, many responses are "blends" where the voicing feature at one ear is combined with the place feature at the other. Studdert-Kennedy, Shankweiler, and Pisoni (1972) found no change in frequency of blend responses whether vowels are shared between ears, as in /pi/-/di/, or whether vowels vary, as in /pi/-/du/; based on this result they argued that left- and right-ear syllables are not blended acoustically, but that abstract phonetic features become mixed between ears in the course of phonemic identification. Switching of stimulus elements between ears was also reported by Treisman (1970) for dichotic consonant-vowel-consonant syllables, differing between ears in all three phonemes; responses often combined phonemes from opposite ears, for example, "taz" + "geb" → "teb." A final type of dichotic fusion to be considered is the combination of simultaneously presented dichotic consonants to yield a perceived sequence of consonants (Day, 1968). Day presented a word beginning with a stop consonant to one ear and a word beginning with a liquid to the opposite. For example, one ear received "lack" and the other "back." Many subjects (about half) reported hearing a single word beginning with a stop-liquid cluster, "black;" the remaining subjects heard one or both of the actual stimulus items. This effect has been termed "phonological fusion" because the structure of the response is apparently determined by a rule of English phonology which permits stop + liquid clusters syllable initially but probibits liquid + stop. Phonological fusion occurs with CV syllables as well as with more complex words (Cutting, 1973).

The existence of perceptual fusions such as feature blending, phoneme switching between ears, and phonological fusion argues strongly against the hypothesis that only one ear at a time has access to the speech processor. For fusion to occur, stimuli to both ears must enter the processor and undergo phonetic analysis within a single "time frame." Moreover, it is interesting that feature blending was observed originally in a task giving a highly reliable right-ear effect. Apparently, the same stimulus material can give rise to different types of perceptual effects--either fusion effects (where the response contains parts of both stimuli) or suppression effects (where only one of the two stimuli is correctly reported). If there is a mechanism for preventing an overload of sensory data into the speech processor, it is unlikely that such a device would work so sporadically, producing sometimes suppression and sometimes fusion. It is argued here that the right-ear effect, lag effect, and fusion effects all represent outcomes of normal speech perception strategies applied to stimuli arriving at the speech processor from both ears.

A significant observation in connection with the lag effect is that fusion responses for stop-liquid pairs occur frequently even when the stop and liquid are temporally offset between ears by 50-100 msec (Day, 1970; Cutting, 1973). Day found a constant fusion rate whether stop or liquid led and regardless of length of offset. In Cutting's work, stop + liquid cluster responses were more common when the stop led, but still occurred frequently when the liquid led. The occurrence of fusion for stop-liquid pairs contradicts the notion that the arrival of a delayed syllable automatically causes the leading syllable to be ejected from the speech processor. The fusion results suggest that if the lagging syllable causes interruption of the processing of the leading syllable, it does so only after the phonetic class of the lagging syllable (stop or liquid) has been determined. We may infer that both lagging and leading sounds undergo at least rudimentary phonetic analysis.

Why would the lag effect and ear effect occur regularly for stop-stop and liquid-liquid pairs but not for liquid-stop or stop-liquid? Two explanations seem reasonable. Perhaps the notion "perceptual processor" has been too broadly defined. If there are separate processors or "feature detectors" for stops and for liquids, then a stop-liquid pairing may not constitute a condition of perceptual competition. The second approach is to view the situation linguistically. In English phonology stop-liquid consonant clusters are permitted syllable initially, while liquid-stop, stop-stop, and liquid-liquid are prohibited sequences. If the speech processor were attempting to integrate the two inputs into a single syllable, a "correct" response would be available only for stop-liquid pairs. Liquid-stop pairs could be construed as stop-liquid, but for liquid-liquid or stop-stop there is no possible response which could integrate the two ears.

Thus, even if inputs from both ears are admitted to the speech processor, there are still at least two levels of processing at which perceptual competition could rise. First, there is the level of "feature extraction," where distinctive phonetic features like voicing, place of articulation, nasality, etc. are identified. Second, there is a higher level where the percept, consisting of a sequence of segmental phonemes, is formed. The phonetic features extracted from the speech signal provide only part of the relevant information for the decisions made at this second stage. Here the phonological rules of the language play a role. For example, nasality occurring during a vowel would be assigned by the phonological rules of English to the following consonant, but for French, nasality would be assigned to the vowel itself. Also, in running speech, not all words are clearly articulated, and context facilitates understanding. The influence of context on speech perception may operate at this stage.

It is proposed here that the right-ear effect arises at the feature extraction level of processing, but that the lag effect and perceptual fusions both arise at the phonemic decision stage. For this model to work, we must assume that the feature extraction stage involves an independent analysis of acoustic cues for both values of a binary feature. That is, rather than being either voiced or voiceless, the output of voicing analysis for English stops might be both voiced and voiceless, providing there were sufficient cues (aspiration, VOT, f_0 contour, etc.) to support both analyses. The output from feature extraction would include a weighting or probability based on the number of acoustic cues consistent with each feature value. For example, on a dichotic trial with a voicing contrast between ears (+) voicing could have a .80 weighting and (-) voicing a .50 weighting. Under binaural conditions (i.e., without competition) the analysis would favor a particular value more strongly. This approach is consistent with the finding in dichotic studies that listeners often do identify stops at both ears correctly, and that listeners rarely misperceive feature values which are shared between ears (Kirstein and Shankweiler, unpublished data). This view is consistent with the importance of multiple acoustic cues in perception of many distinctive features and with the context sensitive nature of the cues (e.g., VOT for voiceless stops varies with place of articulation).

In their model of the ear effect, Studdert-Kennedy and Shankweiler (1970) attributed the right-ear advantage to degradation in the auditory representation of left-ear syllables due to their more circuitous neural connections to the left hemisphere. They considered that ipsilateral connections were inhibited in dichotic presentation so that left-ear stimuli could reach speech processing areas in the left cerebral hemisphere only by first traveling to the right hemisphere and then crossing to the left via the corpus callosum. If the right ear provides superior sensory data to the speech processor, this would, in the present model, produce higher weightings for features extracted from right-ear syllables. However, there might conceivably be reasons for weighting the right ear more highly than the left besides the quality of their auditory representation. The model claims simply that both ears are admitted to the speech processor but that the syllables originally presented at the right ear emerge from the feature extraction stage of processing with higher weighting than left-ear syllables. This idea is consistent with the finding that stop consonants contrasting between ears in only voicing or place of articulation give smaller ear effects than consonants contrasting in both features (Shankweiler and Studdert-Kennedy, 1967; Studdert-Kennedy and Shankweiler, 1970).

The output from the feature extraction level is also the input to a higher level of processing where the identity and order of segmental phonemes are decided. Because the input to this stage may contain incompatible feature values (e.g., voiced and voiceless, labial and velar), there must be perceptual strategies available to resolve conflicting analyses. It is proposed that the lag effect reflects one such strategy and that feature blending and phonological fusion reflect others. The blending in the response of place and voicing from opposite ears could be a function of weights assigned during the feature analysis. The highest voicing and place values may have been originally from opposite ears. This account agrees with Studdert-Kennedy and Shankweiler (1970) in considering the locus of feature blending to be at a stage of processing subsequent to feature extraction where features are combined to yield segmental phonemes.

How many distinct syllables the listener can hear for a dichotic pair probably depends on acoustic fusion as well as on the phonological analysis. The auditory system might recognize two sound sources or only one, and this conclusion could in turn affect the number of phonemic solutions produced by the phonological analysis. A striking result in dichotic experiments is the wide variation among individuals in frequency of fusion responses for stop-liquid pairs (Day, 1970) and in their accuracy of selecting a particular ear in ear monitoring tasks. Such individual differences might arise in the auditory analysis of spatial localization for stimuli which are physically similar but not identical.

During the phonological analysis a number of possible phonemic solutions might be rejected; the listeners would "hear" only those that are finally accepted. Solutions involving low-weighted features would probably be rejected. It is suggested that the lag effect occurs because as a perceptual strategy leads tend to be rejected more than lags. That is, if the feature extraction produces first one feature value and then another, incompatible with the first, the phonological analysis may strongly favor the second result. The second feature value might be treated as if it were a revision of the first. The claim is being made that both the lagging and leading syllables undergo phonetic analysis but that the leading syllable is subsequently rejected. This would occur for stop-stop pairs or for liquid-liquid pairs. For stop-liquid pairs the leading syllable would not be rejected because stop-liquid is an acceptable phonemic sequence.

It is argued that the lag effect and right-ear effect both involve speech perception, but that within the course of speech processing the two effects are independent. Moreover, it is claimed that the ear effect arises at a level of processing prior to the lag effect. Is there evidence in the present data to support these hypotheses? First, there is the fact that the lag effect and ear effect have completely different temporal parameters: the ear effect is greatest with the shortest offsets while the lag effect increases with longer delays up to 60 msec. Beyond the different time functions for the two effects, there are other aspects of the data which suggest that the effects arise at different levels.

At any delay, the size of the ear advantage was the same whether measured on lags--as percent correct right ear lags-left ear lags--or measured on leads--as percent correct right ear leads-left ear leads.[1] It is puzzling that the size of the ear effect should be the same on lags and leads because the conditions of competition between ears are quite different in the two cases. The onset of a lagging syllable always coincides with some portion of an ongoing stimulus at the other ear but the onset of a leading syllable coincides with silence in the other ear. Since it is clear from the data that the ear effect is enhanced by simultaneous competition, we would expect the ear difference in lags to be much greater than in leads. That is, left-ear leads should suffer only slightly in comparison with right-ear leads, but for lags the left ear should be much poorer

---

[1] For the two-response and temporal order tasks there was no interaction between size of ear effect and lag vs. lead. In ear monitoring there was a significantly smaller ear effect in leads than in lags. However, the analysis presented here was based on correct responses only. In intrusion responses from the unattended ear, the right-ear effect was greater in lags than leads so if all ear monitoring responses are considered, there is no interaction between ear effect and lag effect.

than the right ear. Interestingly, the model just proposed provides a way of accounting for the data. Suppose that the ear effect arises at the feature extraction level of processing and that at this level the ear difference is only on lags: features extracted from left-ear lags have received lower weights than those extracted from right-ear lags. At the phonological decision stage there is a tendency to favor lags over leads. The asymmetry between ears arising initially in lags would then be mirrored in leads as well. Right-ear leads are always paired with left-ear lags; these would be rejected less frequently as a phonemic solution than left-ear leads, which are paired with right-ear lags.

The relation between the right-ear effect and lag effect has always seemed paradoxical because all models of the ear effect imply a temporal advantage for the right ear. This is supported by Springer (1971), who reported a 50 msec right-ear advantage in reaction time for correctly identified dichotic stops. The procedure of temporally offsetting stops at the two ears was proposed originally as a possible method of measuring the right-ear effect (Studdert-Kennedy et al., 1970; Berlin et al., 1973). It was thought that giving the left ear a certain lead time would make left- and right-ear stops equally intelligible. The finding that left-ear lags wash out the ear advantage was, thus, incompatible with our understanding of the ear effect. The present view in which the lag effect and ear effect arise at different levels of perceptual processing offers a solution to this paradox. In this model only left-ear lags suffer from the laterality effect, not left-ear leads. Thus, the use of a left-ear lead to compensate for the right-ear advantage would be effective at the lower level of processing but would be wiped out at the higher level. By assigning the lag effect to a higher level of perceptual analysis than the ear effect, we leave open the possibility that the lag effect may not be specific to speech perception. The lag effect may reflect a general strategy in various sense modalities for handling incompatible inputs in pattern perception.

REFERENCES

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

Broadbent, D. E. (1958) Perception and Communication. (New York: Pergamon Press).

Broadbent, D. E. and P. Ladefoged. (1957) On the fusion of sounds reaching different sense organs. J. Acoust. Soc. Amer. 29, 708-710.

Cherry, E. C. (1953) Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Amer. 40, 447-455.

Cherry, E. C. (1961) Two ears--but one world. In Sensory Communication, ed. by W. A. Rosenblith. (Cambridge, Mass.: MIT Press) 99-117.

Cooper, F. S. and I. G. Mattingly. (1969) Computer-controlled PCM system for investigation of dichotic speech perception. J. Acoust. Soc. Amer. 46, 115(A).

Cutting, J. E. (1972) A preliminary report on six fusions in auditory research. Haskins Laboratories Status Report on Speech Research SR-31/32, 93-107.

Cutting, J. E. (1973) Levels of processing in phonological fusion. Ph.D. dissertation, Yale University. (Also in Haskins Laboratories Status Report on Speech Research SR-34, 1-53.)

Darwin, C. J. (1971) Dichotic backward masking of complex sounds. Quart. J. Exp. Psychol. 23, 386-392.

Day, R. S. (1968) Fusion in dichotic listening. Unpublished Ph.D. dissertation, Stanford University.

Day, R. S. (1970) Temporal order judgments in speech: Are individuals language-bound or stimulus-bound? Haskins Laboratories Status Report on Speech Research SR-21/22, 71-87.

Gerber, S. E. and P. Goldman. (1971) Ear preference for dichotically presented verbal stimuli as a function of report strategies. J. Acoust. Soc. Amer. 49, 1163-1168.

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

Inglis, J. (1965) Dichotic listening and cerebral dominance. Acta Otolaryngol. 60, 231-238.

Kimura, D. (1961) Cerebral dominance and the perception of verbal stimuli. Canad. J. Psychol. 15, 156-165.

Kimura, D. (1967) Functional asymmetry of the brain in dichotic listening. Cortex 3, 163-178.

Kirstein, E. F. (1971) Temporal factors in perception of dichotically presented stop consonants and vowels. Ph.D. thesis, University of Connecticut.

Kirstein, E. and D. Shankweiler. (1969) Selective listening for dichotically presented consonants and vowels. Paper presented at 40th annual meeting of Eastern Psychological Association, Philadelphia, Pa., April.

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

Lowe, S. S., J. K. Cullen, C. Thompson, C. I. Berlin, L. L. Kirkpatrick, and J. T. Ryan. (1970) Dichotic and monotic simultaneous and time-staggered speech. J. Acoust. Soc. Amer. 47, 76(A).

Moray, N. (1969) Listening and Attention. (Baltimore, Md.: Penguin Books).

Porter, R. J. (1971a) The effect of temporal overlap on the perception of dichotically and monotically presented CV syllables. J. Acoust. Soc. Amer. 50, 129(A).

Porter, R. J. (1971b) The effect of delayed channel on the perception of dichotically presented speech and nonspeech sounds. Ph.D. dissertation, University of Connecticut.

Porter, R., D. Shankweiler, and A. M. Liberman. (1969) Differential effects of binaural time differences in perception of stop consonants and vowels. Proceedings of the 77th annual convention of the American Psychological Association.

Shankweiler, D. and M. Studdert-Kennedy. (1967) Identification of consonants and vowels presented to left and right ears. Quart. J. Exp. Psychol. 19, 59-63.

Siegel, S. (1956) Nonparametric Statistics. (New York: McGraw-Hill).

Springer, S. P. (1971) Ear asymmetry in a dichotic detection task. Percept. Psychophys. 10, 239-241.

Studdert-Kennedy, M. and D. Shankweiler. (1970) Hemispheric specialization for speech perception. J. Acoust. Soc. Amer. 48, 579-594.

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

Studdert-Kennedy, M., D. Shankweiler, and S. Schulman. (1970) Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. J. Acoust. Soc. Amer. 48, 599-602.

Treisman, A. M. (1960) Contextual cues in selective listening. Quart. J. Exp. Psychol. 12, 242-248.

Treisman, A. M. (1970) Perception and recall of simultaneous speech stimuli. Acta Psychol. 33, 132-148.

Treisman, A. and G. Geffen. (1968) Selective attention and cerebral dominance in perceiving and responding to speech messages. Quart. J. Exp. Psychol. 20, 139-150.

State-of-the-Art Report on Language Processing*

A. M. Liberman+
Haskins Laboratories, New Haven, Conn.

To provide a framework for our discussion, I will set down in outline form the questions that arise in my mind when I wonder how we might get language into the hearing-impaired child. Some of these questions were raised by Ira Hirsh in his keynote speech the other evening, which reinforces my belief that they will help to organize our discussion this morning.

I should confess at the outset that I know very little about deaf children, even less, indeed, than I know about language processing. It is the more appropriate, therefore, that I should try to make the outline neutral. But I do have views (even biases) that may, in one way or another, influence what I say, so, before proceeding with the outline, I should get them on the record. The most relevant of these concern the function or purpose of grammar. My colleagues and I have written about those views in other places, and at length (Liberman, Mattingly, and Turvey, 1972; Mattingly, 1972a; Liberman, 1973) if only for that reason, I should be as brief as possible.

I believe that grammar--or, more exactly, grammatical recoding--serves to reshape information to enable the speaker-listener to move it efficiently between a nonlinguistic intellect, which has facilities for the processing and long-term storage of ideas, and a transmission system, where sounds are produced and perceived. Without the grammatical reshaping that comes so naturally to all normal human beings, we should have to communicate our ideas by means of sounds that were uniquely and holistically different from each other--one sound pattern, however simple or complex, for each idea. In that case, the number of ideas we could transmit would be limited to the number of sounds we can produce and identify. (Precisely that limitation applies to the normal communication of nonhuman animals if, indeed, it is true that those creatures lack the capacity for grammatical coding.) We do not know exactly how many messages could be transmitted by that kind of "language." But, given the richness of the intellect and the comparative poverty of the transmission system, the scope of such a nongrammatical "language" is orders of magnitude less than that which is afforded by the

---

*Edited transcript of a talk presented at a workshop on Sensory Capabilities of Hearing-Impaired Children held October 26 and 27, 1972, in Baltimore, Md.; to be published in a book of the same title, edited by Rachel E. Stark (Baltimore, Md.: University Park Press, in press).

+Also University of Connecticut, Storrs, and Yale University, New Haven, Conn.

[HASKINS LABORATORIES: Status Report on Speech Research SR-35/36 (1973)]

grammars that are so readily available to human beings and that, in my view, set language apart from other perceptual and cognitive processes (Liberman et al., 1972; Liberman, 1973). All this is to imply what some of my colleagues and I believe about the biology of grammar--that the capacity for grammatical processing evolved as a kind of interface, matching the output of the intellect to the vocal tract and the ear. If that is so, the biological development of those grammatical processes should have been influenced by the possibilities and limitations of the mismatched structures they connect. Natural grammatical processes should, then, reflect those influences. Unfortunately, we do not know how far "up" and "down" the grammatical interface the effects of intellect and transmission system go. I have some guesses, based on the formal resemblances between speech and the rest of grammar (Mattingly and Liberman, 1969), but I see no point in inflicting them on you. I will, however, suggest that the point of view I have expressed here is relevant to our concerns. When congenital deafness prevents the use of the normal transmission system, what are the consequences for grammar? If the deaf child bypasses speech entirely, as in the case of natural sign, must he then use a grammar different from the grammar of spoken languages? If so, what are the differences and, more to the point, how far "up" the system do they extend? If the grammar of spoken languages is not appropriate for the transmission system used in sign, how adequately can the signer adapt it? Or must he contrive a more suitable grammar? If so, how well does this more suitable grammar work, and at what cost in effort?

My views about grammatical recoding take a more specific form the more closely we approach to speech at the transmission end of the system. They also become more directly relevant to our concern, since we must surely try to understand speech if we want to know how the deaf child might cope with it. Thus, I think we should want to understand the function of the phonetic representation so that we can better appreciate the consequences, if any, of the failure to develop it in a proper way. We should also want to understand the relation between the phonetic representation and the sound, for we cannot otherwise see how the deaf child might use prosthetic devices that drastically alter the acoustic signal or, in the extreme case, transform it for delivery to the eye or the skin. (See: Liberman, Cooper, and Studdert-Kennedy, 1968.)

Let us consider first the function of the phonetic representation in the conversion between ideas and sounds. We do not know the shape of ideas in the intellect, but we should doubt that they are strung out in time as they are after they have been transformed into sound. If that doubt is well founded, we should suppose that the meaning of the longer segments of language (for example, sentences) must transcend the meaning of the shorter segments (for example, words) that they comprise. There is, then, a need for a buffer in which the shorter segments can be held until the meaning of the longer segments has been extracted. I suspect that the universal phonetic features became specialized in the evolution of language as appropriate physiological units for storage and processing in that short-term buffer. (See: Liberman et al., 1972.) Since the substitution of sign for speech does not remove the need to spread ideas in time, we must wonder how or how well the need for short-term storage is met. We should also wonder what happens when, instead of bypassing the normal transmission system entirely, as in natural sign, one rather enters directly (if only approximately) at the level of the phonetic representation by finger spelling or by writing (and reading) the letters of the alphabet. For if those substitute signals do not engage the phonetic features, then the deaf child may have to make do with other representations--for example, visual or kinesthetic images. How efficient are these nonphonetic representations for the storage and processing in short-term memory that the perception of language may be presumed to require? (See, for example, Conrad, 1972.)

But suppose that instead of avoiding the sound by representing the phonetic message directly, (as in finger spelling or alphabetic writing) we try, as many have, to present the acoustic signal to the deaf child in a form (for example, spectrographic) suitable for presentation to an organ other than the ear. One of the problems we will then encounter arises from the nature of the relation between the sound and the phonetic message. A great deal of evidence supports the conclusion that speech is not an alphabet or simple substitution cipher on the phonetic representation, but rather a complex and grammatical code (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Liberman et al., 1968). Indeed, if speech were not a complex code it could not work efficiently, for just as the transmission system is not well matched, most broadly, to the intellect, so also is it unable, more narrowly, to deal directly with the phonetic representation. Thus, the rate at which the phonetic message is (or can be) communicated--up to 20 or 30 phonetic segments per second--would far exceed the temporal resolving power of the ear if, as in a simple cipher, each phonetic segment were represented by a unit sound. But there is another, equally important problem we should expect to have if the phonetic representation were transmitted alphabetically: the listener would have great difficulty identifying the order of the segments. Though little is known about the ability of the ear to identify the order of discrete (nonspeech) sound segments, recent work suggests that it fails to meet the normal requirements of speech by a factor of five or more (Warren, Obusek, Farmer, and Warren, 1969). That is, when segments of distinctive nonspeech sounds are arranged in strings of three or four, their order can be correctly identified only when the duration of each segment is five or more times longer than speech sounds normally are.

The complex speech code is a grammatical conversion that nicely evades both those limitations of auditory perception: several segments of the phonetic message are commonly folded into a single segment of sound, which takes care of the problem posed by the temporal resolving power of the ear; and there are context-conditioned variations in the shape of the acoustic cues that provide nontemporal information about the order of segments in the phonetic message, thus getting around the ear's relatively poor ability to deal with order on a temporal basis. (See: Liberman et al., 1967; Liberman et al., 1972.) But for our purposes the important point is that these gains are achieved at the cost of a very complex relation between phonetic message and acoustic signal. We are not normally aware of how complex this relation is because the decoding is ordinarily done by an appropriately complex decoder that speech has easy access to. Unfortunately for the needs of the deaf child, however, that decoder is connected to the auditory system. What happens, then, when we try to present the raw (that is, undecoded) speech signal to some other sense organ, such as the eye? On the basis of what we know about speech we can, I think, understand some of the difficulties that are encountered; we can also, perhaps, see opportunities that might be exploited.

I should like to turn now from a more specific concern with grammatical processes near the transmission end of the system to consider some hypotheses about language that deal with grammatical processes more generally. In speaking of the function of these processes, I have suggested that by appropriately interfacing mismatched structures of intellect and transmission, grammar makes possible the efficient communication of ideas from one person to another. But I believe that an equally important function of a grammar is to enlarge the possibilities for communicating ideas to oneself. By getting ideas out of the inarticulate intellect and down at least part way into the language system, we conceivably achieve a kind of control that we could not otherwise have managed. If so, having a grammar confers on us much the same kind of advantage that a mathematics does. A significant part of normal human cognitive work may then depend in one way or another on grammatical processes. In that case we have reason to be concerned about the consequences that may follow when these processes are tampered with.

I have also spoken of the human intellect as though it were in no sense linguistic--that is, as if all the accommodating to the transmission system had been done by the development of the grammatical interface. That leaves out of account the possibility that in the evolution of language the intellect and the transmission system themselves underwent alterations that tended to reduce the mismatch. In the case of the vocal tract, indeed, there is evidence that such an accommodation did occur. The vocal tract of human beings is different from that of other primates (Lieberman, 1968; Lieberman, Klatt, and Wilson, 1969; Lieberman, Crelin, and Klatt, 1972), and the difference appears to have produced for us a greater ability to transmit the phonetic message, thus easing somewhat the job that the speech grammar has to do. But what of the other end of the system? Was the originally nonlinguistic intellect also altered in the direction of a better fit to the other structures in the linguistic system? We do not know, of course, but if it was, then we should have to suppose that the human intellect is to some extent specifically adapted to normal grammatical processes. Given that possibility, we have another reason for wondering whether alteration of normal grammatical processes might have consequences for intellectual ability.

Throughout this introduction I have spoken of "natural" grammatical recodings, which implies a bias I particularly want to get on the record--namely, that such recodings are not arbitrary inventions or cultural artifacts, but rather the reflections of deeply biological processes. I believe, as do many other people who concern themselves with language, that human beings come equipped with the capacity to develop grammars, including, as I have already emphasized, the grammar of speech that connects the phonetic message to the acoustic signal. To the extent that we force these processes into unnatural channels, we can expect to encounter difficulties. Unnatural grammars will very likely be hard to learn, especially if they are as complex as they may need to be. Indeed, the fact that people do not learn to read spectrograms suggests that we cannot, by learning, acquire a grammar of speech or make the natural grammar work with an organ other than the ear (Liberman et al., 1968).

Now let us turn to the outline I spoke of at the beginning, the one that might help us to organize our discussion. Though the shape of the outline conforms rather well to the views I have just talked about, the outline itself does not prejudge any of the issues it raises, or so I hope. The larger division in the outline is between those methods that would aim at delivering to the hearing-impaired child as close an approximation to the spoken language as possible, and those that would use a different transmission system, such as, for example, the gestures of sign. The first method is further divided between those presenting speech in unencoded form (that is, as a signal from which the phonetic message has not been extracted) and those presenting it in decoded form (that is, for example, as a phonemic or phonetic transcription). With undecoded speech there is, of course, an additional, subordinate choice among modalities: do we present the signal to the ear, the eye, or the skin?

I. COMMUNICATION OF A STANDARD, SPOKEN LANGUAGE

It seems reasonably obvious that we should want, if possible, to develop in the deaf child a reasonable approximation to standard, spoken language. Because the greatest number of natural grammatical processes is then used, the fullest possible development of language becomes a relatively easy matter, and there is the least risk of crippling the kinds of cognitive processes that normal grammatical processes ordinarily serve. Those advantages are, of course, in addition to giving the child access to standard literature of all kinds and the ability to communicate more readily with normal-hearing people. I do not mean to propose that we eschew all other possibilities, since the advantages of trying to give the child an approximation to a standard language can be outweighed by many considerations. Indeed, I do not mean to propose anything here, but only to frame the possibilities.

A. Transmission of the Undecoded Speech Signal

I said in my introductory remarks that there is a complexly encoded relation between the phonetic message and the acoustic signal. The salient characteristic of the speech code is that information about successive segments of the phonetic message is often transmitted simultaneously on the same parameter of the sound. As a consequence, there is, in general, no acoustic criterion by which one can identify segments of sound that correspond in size or number to the segments of the phonetic message, and the acoustic shape of the cues for a phonetic segment will often vary greatly as a function of context. The perception of speech requires, then, a complex decoding operation. In this section we will consider those ways of presenting speech, including even rather elaborately processed speech, in which, no matter how well the speech signal penetrates the person's deafness, the decoding job has yet to be done. But first, by way of introduction, I should say more about the speech code and the speech signal. Thus, I should emphasize that the relation between phonetic message and sound is not always that of a complex grammatical code; there are, intermittently, quite transparent or unencoded stretches. In those parts of the speech signal that carry the phonetic message in encoded form, there is, as I have pointed out, the complication that information about more than one phonetic segment is carried simultaneously on the same acoustic parameter. In the transparent or unencoded stretches, however, there is no such complication: a segment of sound carries information about only one phonetic segment. In slow articulation the vowels and fricatives, for example, are reasonably transparent, as are some aspects of the distinctions among phonetic manner classes. The fact that the phonetic message is sometimes encoded in the speech signal and sometimes not becomes important later in this section of the outline when we consider how to present the speech signal to an organ other than the ear.

I should also emphasize here that there is an aspect of the speech signal that has, in principle, nothing to do with encodedness, but that nevertheless can make speech hard to deal with, especially for the deaf. I refer to the well-known fact that speech is, from an engineering standpoint, a very poor signal.

The acoustic energy is not highly concentrated in the first two or three formants, which carry most of the important linguistic information, but is rather smeared broadly through the spectrum. Moreover, some of the most important acoustic cues are rapid frequency changes of the formants, the so-called formant transitions; such rapid frequency swings are, by their nature, physically indeterminate. In the processing we normally do in speech perception, therefore, we must not only decode the signal so as to recover the phonetic segments which are so complexly encoded in it, but also, apparently, clean up the signal--track the formants, as it were--and deliver to the decoder a clearer parametric description of a still undecoded signal. I know of no evidence that human beings have devices (shall we call them property filters?) to do that job. It is nonetheless relevant to our concerns, however, to know that the linguistically important acoustic cues are poorly represented, and to wonder, then, whether we might help the deaf by altering speech to make it a better signal.

1. Getting the undecoded speech signal in by ear. If we are to deal with the undecoded speech signal, then we should want, if possible, to get it in by ear in order to take advantage of all the physiological equipment, including especially the speech decoder, that is naturally connected to the auditory system. But we must then alter the speech signal in some way that is calculated to evade the condition of deafness. The simplest and most common alteration is amplification. I will not discuss that remedy further, except to say the obvious, that it does not always solve the problem.

I would rather consider other, more complicated alterations in the speech signal. Here I have in mind that, as I said in the introduction to this section, the speech signal may be hard to deal with, not only because of its peculiarly complex relation to the phonetic message, but also because the important cues are not among the most prominent features of the acoustic landscape. By using what we now know about those cues, and by taking advantage of the techniques that enable us to manipulate them in convenient and flexible ways, I should think we might be able to make speech significantly more intelligible to the deaf. We should want first, for this and for other more general purposes, to extend our knowledge about the acoustic cues by discovering exactly which ones deaf people can and cannot hear. Then we should explore the possibility of producing a more effective signal by putting the acoustic energy where it counts, and by specifically reinforcing certain cues. Of course, many of the alterations that might, on a common-sense basis, be expected to help could only be managed with totally synthetic speech, since it is beyond our present technological capabilities to process "real" speech as to produce those patterns that are likely to prove most effective. But it is nonetheless worthwhile, I think, to see how much better we can do with even the most extreme, synthetic departures from normal speech. We all know that what is technologically not feasible today is child's play tomorrow, so if we find that certain kinds of synthetic speech can be got through to the deaf better than natural speech, we can look forward realistically to the possibility of someday being able to produce such signals from "real" speech. But there might also be an immediate application. I have in mind the problems of the congenitally deaf child and the possibility that the development of his linguistic system might be promoted--or, more exactly, not held up--if speech could more effectively be got through to him. Of course, if we could provide him only with exposure to appropriately tailored synthetic speech, he could not interact with it in the normal way. Still, he might, like the chaffinch, gain something important if his normal language mechanisms had proper data to work on.

There are other possibilities for alterations in the speech signal that might also increase intelligibility for the deaf. In that connection I should like to take particular note of some work done recently by Timothy Rand (1973). That work is the more relevant because a member of our conference, Dr. Pickett, has results that are related to those of Rand, and Dr. Pickett will, I believe, describe those results for us at this session. Rand has found that when the formants are split between the ears the two higher formants are, to a significant extent, released from the masking effects of the lowest one. More specifically, the procedure and the findings are as follows. Using synthetic speech to have the stimulus control he needs, Rand presents binaurally the syllables [ba], [da], and [ga], which are distinguished only by the transitions of the second and third formants. He then determines by what amount he must reduce the intensities of the second and third formants to bring the subjects' accuracy of identification down from nearly 100 percent, where it is before the intensity reduction, to a level just slightly above chance. In another condition, he carries out exactly the same procedure, but this time with dichotic rather than binaural presentation. In the dichotic condition the first formant is presented to one ear, the second and third formants to the other. The first thing to be said about the results is that, as had been known before, the listener fuses the two inputs quite readily and hears an intelligible utterance. But, for our purposes, the more important result is that, in order to produce a reduction in intelligibility equal to that of the binaural condition, Rand must, in the dichotic condition, reduce the intensities of the second and third formants by an additional 15 db. That is, in the dichotic condition the transition cues for the stop consonants can, other things equal, be heard (and used) by the subjects at a level 15 db lower than that required in the normal binaural condition. Thus, it is as if the dichotic presentation produced a 15 db release from masking. I should emphasize that Rand's work has been done with normal-hearing subjects, and the degradation in the speech has so far been only in the form of intensity reduction. Still, we might want to consider the implications that Rand's work could have for improving speech intelligibility with the deaf. Perhaps Dr. Pickett will do that.

2. Getting the undecoded signal in through a nonauditory modality. Over the years, and especially in the recent past, attempts have been made to help the deaf by presenting the speech signal to the eye or the skin. Those attempts were very adequately reviewed by Dr. Pickett at the 1970 meeting in Easton, Maryland. As our contribution to an earlier meeting at Gallaudet, Franklin Cooper, Michael Studdert-Kennedy, and I undertook to describe the difficulties facing anyone who tries to decode the acoustic stream of speech without the aid of the physiological decoder that normally does it for him (Liberman et al., 1968). Indeed, the source of those difficulties should be apparent on the basis of what I have said here today about the complexly encoded nature of the relation between the acoustic signal and the phonetic message. If the sounds of speech were an alphabet on the phones--that is, if there were a discrete acoustic segment for each phonetic segment, or if the segments were merely linked as in cursive writing--then it should be no more difficult to read spectrograms than than to read print. (Of course we should still have to contend with the fact that signal-to-noise ratio of speech would be poorer by far than that of print; that would, however, pose no very serious problem.) But, as I have said already, the relation of the speech signal to the message it carries is not that simple. Though the speech code matches the requirements of the phonetic representation to the particular limitations of the transmission system, thus permitting these two structures to work well together, it does so at a price; to extract the

phonetic message from the acoustic signal requires a special and complex decoder. Such a decoding mechanism is apparently quite readily available to all human beings, but, unfortunately for our present purposes, it is connected to the auditory system, and experience in trying to learn to read spectrograms suggests that it cannot be transferred to the eye (or the skin).

Given what we know about the speech code and the way it is normally perceived, we have reason to be pessimistic, I think, about the possibility that the eye or the skin can ever be a wholly adequate substitute for the ear as a pathway for speech sounds or even as an alternative entry to the speech decoder. It does not follow, however, that no useful information about the speech signal can be transmitted through nonauditory channels. There are, as I have pointed out, relatively transparent or unencoded stretches of speech in which the relation between acoustic signal and phonetic message is quite straightforward. Since these stretches are not in need of complex decoding, they might be more readily "understood" when transmitted through the eye or the skin.

At all events, I would suggest that in the design of prosthetic aids for the deaf we take into account what we now know (or could, by further research, learn) about the speech code. We should then more clearly see both the difficult problems and the promising possibilities.

B. Transmission of the Decoded Speech Signal

In an alphabetically written language there is a fairly straightforward relation--a rather simple substitution cipher, indeed--between the segmented optical shapes and the phonetic or phonemic segments they represent. We might suppose, therefore, that in presenting language to the eye of the deaf child it would be the better part of wisdom not to offer the raw speech signal, which requires decoding, but rather an alphabetic representation, which does not. Indeed, this seems the more reasonable because we know that while normal-hearing people have not learned to read spectrograms, some have learned to read language in an alphabetically written form.

But the matter is not that simple. There is abundant evidence that reading is a secondary linguistic activity in the sense that it is grafted onto a spoken-language base (Mattingly, 1972b). Thus, reading came late in the history of our race. Moreover, an alphabet, which represents the decoded phonetic segments, is the most recently invented orthography, and it is significant that it has been invented only once. Most relevant of all, of course, is the fact that among normal-hearing children many who speak and perceive speech perfectly well nevertheless cannot learn to read.

We should not be surprised, then, to discover that congenitally deaf children, having had little or no chance to master the primary spoken language, find it exceptionally difficult to acquire a secondary, written form of it. Indeed, the fact that such children have more than the normal amount of trouble learning to read, and that they do not normally attain so high a final level of achievement, is itself strong evidence for the essentially secondary nature of reading. It seems intuitively reasonable to me that a child (or anyone else) should have difficulty mastering the grammatical (as opposed to the lexical) elements of language if his initial and only exposure is to the written forms, but I don't know how to talk about that in any intelligent way. I will only say, therefore, that it is surely important to us that reading is significantly harder for those who do not speak--that it is, in effect, difficult to acquire the language by eye.

How much do we know about this and what else should we try to learn? Is the deaf child's success in reading related to his ability to deal, by whatever means, with the spoken language? If so, what is the nature of the relation? Is there some kind of threshold effect--that is, is some certain amount of competence with the spoken language enough to enable the child to break through and acquire the rest of the language by reading? Can we discover whether experience with particular aspects of the spoken language is more important than experience with some others? And what does it mean, precisely, to say that a congenitally deaf child reads poorly? What kinds of errors does he make, for example, and how do those compare with the errors made by normal-hearing children? Are the deaf child's errors spread evenly (or randomly) over all aspects of language, or do the difficulties pattern in ways that make sense in terms of anything we know about language? Is there any factual support for my intuition that the deaf child might have more trouble with the grammatical items than with the lexical ones? Is that what is reflected in the comment I heard from one of the participants at this conference, that teachers sometimes refer to the performance of deaf children in reading as "noun calling?" If, as I suggested earlier, the phonetic representation normally provides an efficient vehicle for storage and processing in short-term memory, what kinds of alternative representations are available to the deaf child, and how well do they work for the same purpose?

Our outline would be incomplete if we omitted another method of communicating decoded speech to the deaf child, though in this case the decoding is not complete and only some aspects of speech are communicated at all. I refer to "lip reading." The gestures of articulation occur at a stage just prior to the one where much of the most severe encoding occurs. Though the gestures do not thereby escape as many complications as my colleagues and I had once supposed, still they are, by contrast with the acoustic signal, more simply related to the phonetic message. To the extent that the deaf child can see at least some of the articulatory gestures, he has access to a reasonably straightforward representation of the phonetic message. Conceivably, we will want to consider today what we now know or ought to try to learn about lip reading. We may also want to wonder whether there are greater possibilities with that method than have yet been realized.

II. COMMUNICATION BY AN OTHER-THAN-SPOKEN LANGUAGE

Given the problems that the deaf child has with speech, we must consider alternative means of communication. Surely the most obvious and important of these is sign language. Unfortunately for our purposes, and for me, I know almost nothing about sign, so I will not presume to talk about it. All that I can do is to include it in our outline as a subject that you may want to discuss, and, more presumptuously, raise a few questions that my own biases lead me to ask.

Seeing grammar as a kind of interface, I assumed in my introductory remarks that it might bear the marks of the several structures, intellect and transmission system, that it connects. On that basis I raised questions about the consequences of using a different transmission system. In sign the transmission system is very different, involving neither the vocal tract nor the ear. I should ask, then, as I did earlier, whether the grammar of sign is different from that of any spoken language, and if so, exactly how different? (Apart from its relevance to our understanding of the deaf, an answer to that question should be of interest to students of language, because it tells us something about how far up the grammatical interface the effects of the transmission system extend.) If the grammar of sign is very different, is there a price to be paid, either in effort or in efficiency, for not being able to use, as the normal-speaking person does, those grammatical processes that presumably evolved with language and are now a part of our physiology? You probably know more than I do about research on sign, including, for example, the work of Stokoe (1960) or that of Bellugi and Fischer (1972). If so, I hope you will include sign in our discussion. In any case, it is time for me to stop talking and, instead, to invite from you the comments that are the principal purpose of this meeting.

REFERENCES

Bellugi, U. and S. Fishcer. (1973) A comparison of sign language and spoken language. Cognition 1, 173-200.

Conrad, R. (1972) Speech and reading. In Language by Ear and by Eye, ed. by J. F. Kavanagh and I. G. Mattingly. (Cambridge, Mass.: MIT Press) 205-240.

Liberman, A. M. (1973) The specialization of the language hemisphere. In The Neurosciences: Third Study Program, ed. by F. O. Schmitt and F. G. Worden. (Cambridge, Mass.: MIT Press) 43-56.

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

Liberman, A. M., F. S. Cooper, and M. Studdert-Kennedy. (1968) Why are spectrograms hard to read? Amer. Ann. Deaf 113, 127-133.

Liberman, A. M., I. G. Mattingly, and M. T. Turvey. (1972) Language codes and memory codes. In Coding Processes in Human Memory, ed. by A. W. Melton and E. Martin. (Washington, D. C.: V. H. Winston) 307-334.

Lieberman, P. (1968) Primate vocalizations and human linguistic ability. J. Acoust. Soc. Amer. 44, 1574-1584.

Lieberman, P., E. S. Crelin, and D. H. Klatt. (1972) Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and chimpanzee. Amer. Anthropol. 74, 287-307.

Lieberman, P., D. H. Klatt, and W. A. Wilson. (1969) Vocal tract limitations of the vocal repertoires of rhesus monkey and other nonhuman primates. Science 164, 1185-1187.

Mattingly, I. G. (1972a) Speech cues and sign stimuli. Amer. Scient. 60, 327-337.

Mattingly, I. G. (1972b) Reading, the linguistic process, and linguistic awareness. In Language by Ear and by Eye, ed. by J. F. Kavanagh and I. G. Mattingly. (Cambridge, Mass.: MIT Press) 133-147.

Mattingly, I. G. and A. M. Liberman. (1969) The speech code and the physiology of language. In Information Processing in the Nervous System, ed. by K. N. Leibovic. (New York: Springer Verlag) 97-117.

Rand, T. (1973) Dichotic release from masking for speech. Haskins Laboratories Status Report on Speech Research SR-33, 47-55.

Stokoe, W. (1960) Sign language structure. Studies in Linguistics, Occasional Paper 8 (Buffalo, N. Y.: University of Buffalo).

Warren, R. M., C. J. Obusek, R. M. Farmer, and R. T. Warren. (1969) Auditory sequence: Confusion of patterns other than speech or music. Science 164, 586-587.

Audible Outputs of Reading Machines for the Blind*

Franklin S. Cooper, Jane H. Gaitenby, Ignatius G. Mattingly,[+] Patrick W. Nye,
and George N. Sholes
Haskins Laboratories, New Haven, Conn.

Introduction

A frequent topic of complaint from the blind and visually handicapped concerns the long delays that often occur in receiving recordings of spoken texts. The alleviation of these delays by use of a High Performance Reading Machine which can provide supplementary reading services to blind people is the goal of the research being carried out at Haskins Laboratories. From a technical standpoint, the results of this research indicate that the automatic production of spoken text from print is entirely feasible. Thus, a reading machine of this kind, installed in a major library, could respond to requests from individual blind subscribers by providing direct listening access to, or recordings of, clearly intelligible synthetic speech from ordinary printed texts. These recordings can be made at rates much faster than a human speaker can produce them. Hence, the availability of a fast, library-based reading machine service could make a substantial contribution toward meeting the educational, recreational, and vocational needs of blind people.

Status of the Research

A prototype reading system has been constructed at Haskins Laboratories and has been in operation for nearly a year. Continuing efforts are being made to improve the performance of the machine at different levels. High on the list of activities during recent months have been the introduction of improvements in the quality of the speech and the incorporation of an optical character recognition (OCR) machine into the system to provide for the input of typewritten texts.

Looking ahead to the eventual deployment of a reading machine system, a collateral study which has gathered momentum during the past six months has focused attention on the intelligibility, comprehensibility, and acceptability of synthetic speech. The data from these tests are intended to show where efforts on speech improvement should be concentrated, and to test the reliability

---

*Summary prepared for the <u>Bulletin of Prosthetics Research</u>, BPR 10-20, Fall 1973.

[+] Also University of Connecticut, Storrs.

[HASKINS LABORATORIES: Status Report on Speech Research SR-35/36 (1973)]

of the reading machine system. If, as expected, the results of these studies confirm the feasibility and utility of an automated reading system from both technical and user standpoints, the resources might then be found to build a Pilot Reading Service Center. This Center would provide an experimental service to the blind community in its area and would act as a model on which other regional centers could later be based.

Installation of the OCR Equipment

The Cognitronics System/70 optical character recognition equipment, which the Laboratories purchased with funds from The Seeing Eye Foundation, was put into operation in mid-March 1973. Since then, as a first step, an output program has been developed to punch a paper tape copy of the typewritten pages automatically scanned by the optical reader. This tape is then read by the DDP-224 computer which performs the remainder of the processing required to generate synthetic speech. As described in more detail in earlier reports,[1] the DDP-224 computer--using a phonetic dictionary--converts the orthographically spelled text received via the tape reader into phonetic text. During the conversion of the text into phonetic form, stress and intonation markers are introduced in readiness for speech synthesis. If all the words contained in the original text have been found in the dictionary and if the punctuation available in the original text has provided an adequate guide to the insertion of intonation marks, synthesis proceeds automatically. However, editorial intervention is sometimes required and provision has therefore been made, just prior to synthesis, for an editor to check the dictionary output. New words are continually being added to the dictionary which now contains over 150,000 entries.

The use of the paper tape medium to convey texts from the optical reader to the main computer has been adopted merely as an interim measure. Work is in hand on the design and implementation of a direct electrical connection between the Cognitronics reader and the DDP-224 computer. This connection will permit rapid conversion of the fairly large volumes of text required for evaluation purposes--particularly for those requiring acceptability and tolerability judgments. One such evaluation project (for which the system is currently being readied) involves the regular conversion of articles from a New Haven daily newspaper into synthetic speech and the subsequent appraisal by blind veterans at the Veterans Administration Hospital in West Haven, Conn. The New Haven Register provides the Laboratories with punched paper tapes of an article. The PDP-8 computer (which is an integral part of the OCR reader) is then used to recode the text so that the DDP-224 computer can read it and perform the speech synthesis.

Synthetic Speech Evaluation

In the area of evaluation, recent activity has concentrated on an analysis of the data obtained from a closed response version of the Fairbank's Rhyme Test and on the administration and analysis of a new test procedure using meaningless sentences. (The absence of meaning makes the recognition of words in continuous speech much more difficult.)

---

[1] See Haskins Laboratories Status Report on Speech Research SR-29/30 (1972) 91-95, 97-104; SR-31/32 (1972) 177-182.

The Modified Rhyme Test, described in an earlier report,[2] was administered in synthetic speech to thirty inexperienced sighted students and six blind students at the University of Connecticut. As a control test, the words were presented in natural speech generated by a single speaker. Three hundred monosyllabic words were used in six different orders of presentation. The overall intelligibility scores were found to be 92.5% for synthetic speech and 97.3% for natural speech--the former indicating needed synthesis improvements and the latter agreeing well with the data obtained by other workers. Initial /v/ and final /r/ in particular--as well as the labial, labiodental, and dental fricatives in general--were isolated as the least intelligible phones. However, an intrinsic limitation of the Modified Rhyme Test is that individual consonants are presented an unequal number of times, in unequal vowel environments, and in an imbalanced proportion of initial versus final syllable positions. The subjects' ability to recognize words in synthetic speech was shown to improve consistently over the course of the tests; thus it is possible that the low occurrence of some phones may have contributed to their low intelligibility scores. The finding that a listener's performance with synthetic speech improves with experience is consistent with the observations of many other workers. Customarily, the best scores are obtained if the "training period" with synthetic speech extends over several hours. However, this period is very short compared with the learning time demanded by nonspeech reading aids. In view of the fact that it is intended to be used on large volumes of reading matter, the modest amount of learning required in no way lessens the potential usefulness of synthetic speech in a Reading Service Center application.

The latest test conducted in the evaluation program utilized 126 nouns, 63 adjectives, and 63 past-tense verbs--all monosyllables selected from the 2,000 most frequently used words in English. Words from each category were randomly selected to create 200 meaningless sentences of the grammatical form exemplified in this sentence: "The gold rain led the wing." These sentences were recorded in both naturally spoken and synthesized speech in batches of 50 sentences with a 10-sec interval between each sentence. During that interval the 32 sighted test subjects were required to write down the sentence they had heard in ordinary English orthography. Lacking semantic context cues, the test proved to be the most difficult yet administered. A full phoneme-by-phoneme analysis of the natural speech and synthetic speech errors made by each subject has been undertaken to discover not only the most common confusions made but also the phonetic environments in which the errors occurred. A large volume of data has been obtained and the concluding phase of the analysis is still in progress. Discussion of the results will appear in the next report.

Experiments in Alphabetic-to-Phonetic Conversion

While the reading machine output is being evaluated with a view to early deployment, research efforts are continuing toward the improvement of the speech output. By deliberate choice, the current methods of assigning and modifying stress in the phonetic string are simple and direct. The results, however, while clearly superior to what might be expected if stress and intonation were

---

[2] See Haskins Laboratories Status Report on Speech Research SR-33 (1973) 77-91.

totally absent, are not entirely natural. The problem of improving the intonation patterns in the speech output has two parts. One part involves the observation of natural speech patterns and the determination of rules relating these patterns to the syntactic and lexical content of the sentences. A second part involves the development of a flexible Experimental Synthesis Program in which the rules governing acceptable stress and intonation may be examined.

Comparison of the current synthesized output with samples of natural speech have recently led to new experiments involving the use of increased vowel duration as a supplemental cue for stress. In the Experimental Synthesis Program now being written, when a syllable is chosen to be stressed its lexical vowel is mapped into the phonetic output string as a diphthong or occasionally as a triphthong. This increase in formant excursion is applied in addition to the usual pitch excursion. When, on the other hand, syllables are marked for low stress, the vowels are in general mapped into the single vowel shwa. All syllables moving from the phonetic dictionary to the output are additionally marked according to whether they occur in phrase-final position or not (a phrase in this sense being indicated by an intonation contour symbol). Thus, in the last phrase before a final end-of-intonation pause, trailing resonant phones as well as central vocalic phones are protracted. This gives a partially filled-pause effect which, together with the normal distinctive pitch excursion, highlights the conclusion of the phrase.

With these methods, certain prosodic features observed in natural speech are emphasized in the synthetic output in sharp phonetic relief. This appears to increase the intelligibility of sentences, although at the expense of naturalness. Further investigation will be required to obtain a satisfactory balance among the various cues that convey acceptable stress and phrasing within a synthesized sentence.

Speech Synthesis

A new OVE II cascade formant synthesizer has recently been installed as the output stage of the reading machine system. The OVE replaces a parallel resonance synthesizer that was built at the Laboratories several years ago. While in many respects less flexible as a general speech research tool, the new synthesizer has its formant filters connected in series. This arrangement is better suited to a reading machine application since it establishes automatically the correct relative formant energy levels and reduces significantly the amount of calculation performed by the computer during the production of vowels. Synthesis programs designed for the OVE have been in operation since February 1973 and the device is already producing speech which seems better than that from the older model.

However, one of the most striking deficiencies of the OVE synthesizer is its limited performance on the production of nasals. The OVE has one parallel nasal resonator available, but spectrographic analyses of natural speech suggest that additional resonances and antiresonances may be needed. At present, an investigation is in progress to discover the extent to which the perception of nasality can be enhanced within the limitations of the existing hardware. A search is also being actively pursued to find ways in which additional components, designed to generate the appropriate spectra, can be added. The results of this enquiry promise to provide a substantial improvement in voice quality.

ABSTRACT

An Examination of Hemispheric Asymmetry in the Visual Processing of Linguistic Items*

Claire Farley Michaels+

The two cerebral hemispheres differ in their linguistic capabilities; in the processing of language by ear and by eye the left hemisphere is clearly superior to the right. Two aspects of the left-hemisphere/right-visual-field superiority for verbal material were examined: the earliest processing stage at which the superiority appears and the characteristics of verbal material that are relevant to that superiority.

One approach sought to determine whether hemispheric differences were to be found in the latencies required for the detection of phonological lawfulness and/or the detection of lexical membership. These two processes are assumed to underlie performance in the word/nonword classification task. In three experiments subjects were presented lateralized letter trigrams for classification as "word" or "nonword." No hemispheric differences were observed. The absence of asymmetry in this task could be attributed to apparent changes in processing mode during the course of an experiment. In addition, the spatial (right hemisphere) nature of the manual response signaling the classification may have obscured any left-hemisphere superiority.

Another approach to the aspects of left-hemisphere superiority took advantage of visual masking. Dichoptic masking functions trace out a J-shaped relation between target identifiability and the time elapsed between target and mask onset. The descending portion of such functions can be said to describe masking arising in an icon-construction stage while the ascending part describes masking originating in an icon-identifying stage. A left-hemisphere superiority was detected only in the latter case and this superiority held for both words and nonwords. A more detailed analysis revealed that the left hemisphere responded more than the right to a meaning distinction (e.g., CAT vs. CAG) but less than the right to a number-of-syllables distinction (e.g., CAG vs. CKG). These observations were discussed in the context of current conceptualizations of the relation between language and brain.

---

*A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, University of Connecticut, Storrs, 1973.

+Department of Psychology, Lake Forest College, Lake Forest, Ill. 60045.

Acknowledgment: This research was done at Haskins Laboratories, under the direction of Michael Turvey, with grant support to Haskins Laboratories from the National Institutes of Health and the National Institute of Child Health and Human Development. Support was also received from a University of Connecticut Predoctoral Fellowship.

II. THESIS SUPPLEMENT

LEVELS OF PROCESSING IN SPEECH PERCEPTION:
NEUROPHYSIOLOGICAL AND INFORMATION-PROCESSING ANALYSES

by

Charles C. Wood[+]

A dissertation* presented to the faculty of the
Graduate School of Yale University
in candidacy for the degree of
Doctor of Philosophy

1973

[+] Present address:
Division of Neuropsychiatry
The Walter Reed Army Institute of Research
Washington, D. C. 20012

*Abridged--see Table of Contents

[HASKINS LABORATORIES: Status Report on Speech Research SR-35/36 (1973)]

ACKNOWLEDGMENTS

The research reported in this dissertation is the product of contributions from a number of different people. To all of them I express my deepest appreciation. In particular, I wish to thank my advisor and chairman W. R. Goff and advisors R. S. Day and W. R. Garner, who encouraged me to pursue this work and provided invaluable advice throughout its course. Dr. Goff provided not only the training, physical facilities, and instrumentation required for this research, but an environment in which it was a true pleasure to work. In addition, I would like to thank the other members of my committee, W. Kessen, F. S. Cooper, and E. W. Rubel, for their helpful suggestions, advice, and criticisms. To Dr. Cooper I am especially grateful for making available the unique facilities of the Haskins Laboratories for the generation and analysis of the synthetic speech stimuli used in these experiments. I also wish to acknowledge the excellent technical assistance of T. C. Fisher and J. Jasiorkowski and the skill of G. D. Goff and J. E. Fritz in drawing the illustrations.

This dissertation was supported in part by the Veterans Administration, Public Health Service National Institute of Mental Health Grant M-05286 and National Science Foundation Grants GB-3919 and GB-5782 to W. R. Goff, and by a National Institute of Child Health and Human Development grant to Haskins Laboratories.

Editor's Note: This dissertation was selected by The Psychology Section of the New York Academy of Science as winner of the annual James McKeen Cattell award for an outstanding doctoral dissertation in psychology.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS .... S-iii

LIST OF TABLES .... S-vi

LIST OF FIGURES .... S-vii

INTRODUCTION S-1

- Auditory and Phonetic Levels of Processing in Speech Perception S-1
- Differences Between the Perception of Speech and Nonspeech Stimuli S-2
- Differences Between the Perception of Auditory and Phonetic Dimensions of the Same Speech Stimuli S-3
- Rationale for the Present Experiments S-5

METHOD S-7

- The Basic Paradigm Common to All Experiments S-7
- Subjects S-8
- Stimuli S-8
- Apparatus S-9
- Procedure S-10
- Data Analysis S-11

EXPERIMENT 1: PLACE OF ARTICULATION S-12

- Stimuli S-12
- Results and Discussion S-15

EXPERIMENT 2: INTENSITY S-19

- Stimuli S-20
- Results and Discussion S-20

EXPERIMENT 3: SECOND FORMANT TRANSITIONS S-24

- Stimuli S-24
- Results and Discussion S-26

TABLE OF CONTENTS (cont'd)

Page

EXPERIMENT 4: PITCH CONTOUR S-30

Stimuli S-30

Results and Discussion S-32

ADDITIONAL ANALYSES OF NEURAL ACTIVITY DURING THE PROCESSING OF AUDITORY AND PHONETIC DIMENSIONS S-36

Spectral Analysis of Background EEG S-38

Method S-38

Results S-39

Gaussian Characteristics of Background EEG: Analysis of EEG Amplitude Histograms S-39

Method S-43

Results S-44

Signal Averaging of Activity During the Intertrial Interval Between Successive Stimuli S-44

Method S-46

Results S-46

Signal Averaging Synchronized to Subjects' Button-Press Responses Instead of Stimulus Onset S-48

Method S-48

Results S-48

GENERAL DISCUSSION S-51

Summary of Experiments S-51

Relative Discriminability of the Stimulus Dimensions in Each Experiment S-52

Individual Differences S-54

Relation of the Present Experiments to the Distinction Between Integral and Nonintegral Stimulus Dimensions S-55

Converging Operations and the Concept of Specialized Neural Mechanisms for Speech Perception S-57

TABLE OF CONTENTS (cont'd)

Page

*Serial Versus Parallel Organization of Auditory and Phonetic Levels [abridged] S-60

SUMMARY AND CONCLUSIONS S-62

REFERENCES S-63

LIST OF TABLES

Table 1: General form of the 2 x 2 paradigm used in each experiment S-7

Table 2: Summary of analysis of variance for Experiment 1 ..... S-17

Table 3: Summary of analysis of variance for Experiment 2 ..... S-22

Table 4: Summary of analysis of variance for Experiment 3 ..... S-28

Table 5: Summary of analysis of variance for Experiment 4 ..... S-34

Table 6: Percentage of EEG segments in which the Gaussian hypothesis was rejected S-45

LIST OF FIGURES

Page

Figure 1: Wide-band spectrograms of the four stimuli used in Experiment 1 .... S-13

Figure 2: Narrow-band spectrograms of the same stimuli shown in Figure 1 .... S-14

Figure 3: Reaction time data for Experiment 1 .... S-16

Figure 4: Average evoked potentials during identification of Place and Pitch in Experiment 1 .... S-18

Figure 5: Reaction time data for Experiment 2 .... S-21

Figure 6: Average evoked potentials during identification of Intensity and Pitch in Experiment 2 .... S-23

Figure 7: An example of the relationship between the CV syllables used in Experiment 1 and the F2 transitions used in Experiment 3 .... S-25

Figure 8: Reaction time data for Experiment 3 .... S-27

Figure 9: Average evoked potentials during identification of F2 transitions and Pitch in Experiment 3 .... S-29

Figure 10: Narrow-band spectrograms of the four stimuli used in Experiment 4 .... S-31

Figure 11: Reaction time data for Experiment 4 .... S-33

Figure 12: Average evoked potentials during identification of Pitch Contour and Pitch in Experiment 4 .... S-35

Figure 13: Frequency spectra from the left hemisphere (C3) during identification of Place S-40

Figure 14: Frequency spectra from the right hemisphere (C4) which occurred simultaneously with those shown in Figure 13 .... S-41

Figure 15: Average frequency spectra across subjects during identification of Place and Pitch S-42

Figure 16: Averaged activity during the intertrial intervals between successive stimuli S-47

Figure 17: Averaged activity synchronized to subjects' button-press responses S-49

INTRODUCTION

Auditory and Phonetic Levels of Processing in Speech Perception

Current theories of speech perception suggest that the process by which a listener understands the linguistic content of a spoken message consists of several distinct conceptual levels or stages. Although different levels have received primary emphasis from different authors, there is general agreement that any satisfactory account of speech perception must include at least auditory, phonetic, phonological, syntactic, and semantic levels (see, for example, Fry, 1956; Fant, 1967; Liberman, 1970; Stevens and House, 1972; Studdert-Kennedy, in press). Typically, these levels are considered to be organized in a hierarchical manner, with the input to one level roughly corresponding to the output of the previous level in the hierarchy. However, since there is ample evidence that constraints at one level can modify operations at both lower and higher levels (cf. Studdert-Kennedy, in press), an adequate description of the speech perception process must also account for these interactions between levels.

The present research was concerned with the auditory and phonetic levels in such a conceptual hierarchy; that is, those perceptual processes which intervene between the acoustic speech signal and a stage roughly corresponding to the identification of phonemes. These two levels can be isolated experimentally from syntactic and semantic levels by investigating the perception of isolated phonetic stimuli in the form of nonsense syllables. While an intuitive distinction between auditory and phonetic levels of processing has been evident in a number of theories of speech perception (Fant, 1967; Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Stevens and Halle, 1967; Stevens and House, 1972), this distinction has been made most explicit in the recent work of Studdert-Kennedy (Studdert-Kennedy and Shankweiler, 1970; Studdert-Kennedy and Hadding, 1971; Studdert-Kennedy, Shankweiler, and Pisoni, 1972; Studdert-Kennedy, in press).

The auditory level is characterized by its direct relation to the acoustic signal. This level is assumed to consist of those neural processes which analyze the acoustic input into a set of auditory parameters. "It is automatic, that is beyond voluntary control; it transforms the acoustic waveform that impinges on the ear to some time-varying pattern of neurological events of which the spectrogram is, at present, our closest symbolic representation" (Studdert-Kennedy, in press). Thus, the auditory level may be characterized as that portion of the speech perception process that is nonlinguistic, and therefore includes those mechanisms that operate on speech and nonspeech signals alike.

In contrast to the direct relationship between the auditory level and the acoustic signal, the phonetic level is best characterized by its abstractness. That is, the phonetic features which are assumed to be the output of the phonetic level are not inherent in the speech signal, but rather are abstract linguistic entities. Instead of a one-to-one mapping of sound to phoneme in the fashion of an alphabet, the phonemes are linked to the acoustic signal by a complex set of transformations which have been called the "speech code" (Liberman et al., 1967). The specific details of this code and the empirical data on which it is based have been thoroughly reviewed elsewhere (Liberman et al., 1967; Mattingly and Liberman, 1969; Liberman, 1970; Cooper, 1971; Liberman, Mattingly, and Turvey, 1971), and will not be considered in detail here. However, it will be useful to consider briefly the most important single characteristic of the speech code: the parallel transmission of phonetic information in the acoustic signal.

Parallel transmission of phonetic information is accomplished in three general ways. 1) A single acoustic segment in the speech signal can carry information simultaneously about two or more successive phonetic segments. Therefore, there is often little correspondence between acoustic segments in the signal and phonetic segments in the message. 2) Information about multiple phonetic features of the same phonetic segment is transmitted simultaneously in time by different portions of the speech signal. 3) The contrast between values of a given phonetic feature may be cued by a number of different acoustic parameters, both simultaneously and successively in time. Thus, information about perceptually distinct phonetic segments is completely merged and intermixed in the acoustic signal. This characteristic "encoding" of phonetic information in the speech signal provides speech communication with its great efficiency, but at the same time requires specialized perceptual mechanisms, in addition to the auditory system, which are capable of decoding the phonetic content from the speech signal (Liberman et al., 1967).

To summarize, the nonlinguistic auditory level is assumed to perform a preliminary analysis of the acoustic speech signal, resulting in a set of auditory parameters that are the neural representation of that signal. This process is not unique to speech but operates on any acoustic signal within the audible range. In contrast, mechanisms at the phonetic level are assumed to perform the linguistic decoding process by which the particular complex of acoustic cues for a given phonetic feature is extracted from the results of the auditory analysis. On logical grounds alone it has been argued that this decoding process requires specialized neural mechanisms in addition to the general auditory system (Liberman et al., 1967; Mattingly and Liberman, 1969; Liberman, 1970; Mattingly, Liberman, Syrdal, and Halwes, 1971), and empirical evidence to be described below is consistent with this suggestion. Thus, the auditory level is assumed to correspond roughly to the general auditory system of man and other primates, while the phonetic decoding process is attributed to the existence in the human brain of additional neural mechanisms specialized for speech perception.

Differences Between the Perception of Speech and Nonspeech Stimuli

Much of the evidence supporting a distinction between auditory and phonetic levels of processing has come from differences in the way speech and nonspeech stimuli are perceived; in particular, 1) tendencies toward "categorical" versus "continuous" perception (for recent reviews see Studdert-Kennedy, Liberman, Harris, and Cooper, 1970a; Pisoni, 1971; Mattingly et al., 1971), and 2) ear advantages in dichotic listening (for a recent review see Studdert-Kennedy and Shankweiler, 1970).

Categorical perception refers to the tendency for the discrimination of certain speech sounds to be a function of phonetic categories rather than a function of physical differences between stimuli. Listeners can easily discriminate between two stimuli selected from different phonetic categories, but they cannot discriminate between two stimuli selected from the same phonetic category even though the physical difference between the two stimuli is equal in both cases.

In other words, for these stimuli, differential discrimination is limited by absolute identification. In contrast, differential discrimination among levels of a nonspeech dimension such as pitch is far better than absolute identification, with well over 1,000 different values discriminated differentially, and absolute identification limited to the familiar seven, plus or minus two. Thus, while the discrimination among levels of a nonspeech dimension is related to the physical stimulus dimension, the discrimination among certain speech stimuli is limited to a large extent by the linguistic knowledge of the listener.

A second important difference between the perception of speech and nonspeech stimuli is the nature of ear advantages in dichotic listening. The basic finding in such experiments is that when pairs of speech stimuli are presented simultaneously to the two ears (i.e., dichotically), those stimuli presented to the right ear are more accurately identified than those presented to the left ear: a "right-ear advantage." In contrast, for nonspeech stimuli the typical result has been either no ear advantage or an advantage in favor of the left ear. Kimura (1961a), who originally reported the right-ear advantage for dichotically presented digits, attributed it to two factors: 1) the prepotency of the auditory pathways from each ear to contralateral auditory cortex demonstrated by neurophysiological experiments; and 2) the predominant lateralization of language function in the left hemisphere demonstrated by the analysis of language disorders following damage to the left and right hemispheres (Milner, 1967; Geschwind, 1970). If this interpretation were correct, then subjects with known language dominance of the right hemisphere should show a left-ear advantage for speech in a comparable experiment. This result was indeed obtained (Kimura, 1961b). Although the initial dichotic listening experiments employed digits and other meaningful verbal stimuli, subsequent experiments have clearly shown that the right-ear advantage is not dependent upon higher level syntactic or semantic processes. Some of the largest and most reliable right-ear advantages have been obtained using isolated consonant-vowel syllables (Studdert-Kennedy and Shankweiler, 1970). A number of alternative explanations to the Kimura model of the right-ear advantage have been suggested, including response bias, attentional bias, and order of report effects. However, the right-ear advantage is still obtained in paradigms that eliminate these alternative explanations.

Differences Between the Perception of Auditory and Phonetic Dimensions of the Same Speech Stimuli

Another approach to the distinction between auditory and phonetic levels of processing has been recently reported by Day and Wood (1972a) and Wood, Goff, and Day (1971). Instead of concentrating upon differences between the perception of speech and nonspeech stimuli, these experiments compared the perception of auditory and phonetic dimensions of the same speech stimuli. Thus, while previous experiments compared auditory and phonetic processing by varying the nature of the acoustic stimuli, the experiments of Day and Wood (1972a) and Wood et al. (1971) presented a single set of speech stimuli and varied the nature of the processing task. Since these initial experiments form the immediate background for the present research, they will be described in some detail.

The experimental paradigm used by Day and Wood (1972a) was a two-choice speeded-classification task similar to that employed by Garner and Felfoldy (1970) to study interactions between stimulus dimensions in information processing. Subjects were presented series of synthetic consonant-vowel (CV) syllables that varied between two levels on a given target dimension, and were required to identify which level of the target dimension occurred on each trial. The two stimulus dimensions compared in this experiment were: 1) a phonetic dimension, place of articulation of voiced stop consonants; and 2) an auditory dimension, fundamental frequency or pitch. For convenience, these two dimensions will be referred to as Place and Pitch, respectively. Reaction time (RT) for the identification of each dimension was measured under two conditions: 1) a single-dimension control condition, in which only the target dimension varied in the stimulus sequence; and 2) a two-dimension orthogonal condition, in which both the target dimension and the irrelevant nontarget dimension varied orthogonally in the stimulus sequence. For each dimension the only difference between the control and orthogonal conditions was the presence or absence of irrelevant variation in the nontarget dimension. Therefore, a comparison of the RTs from these two conditions indicated the degree to which each dimension was processed independently of irrelevant variation in the other dimension.

Three possible interactions between dimensions could be obtained in this paradigm. 1) Irrelevant variation in each dimension could produce interference with the processing of the other. This result is typical of that obtained for "integral" stimulus dimensions by Garner and Felfoldy (1970). Such a result would suggest that the two dimensions were automatically extracted by a single perceptual process, or by multiple processes which are strongly dependent upon each other. 2) A second possibility is that irrelevant variation in neither dimension could interfere with the processing of the other. This result is typical of that for "nonintegral" dimensions in the experiment of Garner and Felfoldy (1970). Such a result would suggest that the perceptual processes for the two dimensions are largely independent. 3) The final possibility is that the interference between dimensions could be unidirectional; that is, irrelevant variation in one dimension could interfere with the processing of the other, but not the reverse. This result would also imply that multiple processes are involved in the extraction of the two dimensions. However, the unidirectional interference would imply a dependence of the processes for one dimension upon those for the other.

The results of Day and Wood (1972a) for the Place and Pitch dimensions followed the third hypothetical pattern described above. For identification of Place, there was a substantial increase in RT from the control to the orthogonal condition, indicating that irrelevant variation in Pitch significantly interfered with the processing of Place. In contrast, there was only a slight increase in RT for Pitch, indicating that subjects could ignore or filter the Place dimension to a considerable degree when required to process Pitch. These results suggest that different levels of processing underlie the identification of auditory and phonetic dimensions of the same speech stimuli. In addition, they suggest that the phonetic level processes are in some way dependent upon those performed by the auditory level.

Neurophysiological evidence for a distinction between auditory and phonetic levels of processing was obtained in a related experiment by Wood et al. (1971). This experiment used the same basic strategy as the RT experiment of Day and Wood (1972a), by comparing neural activity during the identification of auditory and phonetic dimensions of the same speech signal. Averaged evoked potentials produced by the same synthetic CV syllable were recorded over the left and right hemispheres during Place and Pitch identification tasks similar to those of the control condition in the experiment of Day and Wood (1972a). In this way, neural activity that occurred during the Place and Pitch tasks could be compared directly, without differences between tasks in the acoustic speech signal or its presentation probability, in the subjects' motor responses, in RT, or in the recording apparatus. These controls were necessary to eliminate the possibility that obtained evoked potential differences between tasks were produced by factors other than the perceptual processes required for each task.

If the processing of auditory and phonetic dimensions of a speech signal were accomplished by a single perceptual process, then evoked potentials during the Place and Pitch tasks would merely be random samples from a single population and should differ only by sampling fluctuations. In the time interval between the onset of the speech signal and subjects' identification responses, evoked potentials at locations over the right hemisphere were indeed identical for both tasks. However, significant differences in evoked potentials were obtained at left-hemisphere locations during the same time interval. These results indicate that different neural events occur in the left hemisphere during the identification of auditory and phonetic dimensions of the same acoustic signal.

Thus, the neurophysiological data of Wood et al. (1971) provide additional support for the distinction between auditory and phonetic levels of processing demonstrated by the RT experiment of Day and Wood (1972a). Both experiments suggest that the identification of Place involves an additional level of processing that is not required for the identification of the Pitch of the same speech signal. In addition, the evoked potential data relate this additional level of processing to the concept of hemispheric specialization for speech perception, derived from the analysis of language disorders following brain damage and the results of dichotic listening experiments.

Rationale for the Present Experiments

Together with the categorical perception and dichotic listening experiments described above, the experiments of Day and Wood (1972a) and Wood et al. (1971) provide a strong set of converging operations (Garner, Hake, and Eriksen, 1956) upon the distinction between auditory and phonetic levels of processing in speech perception, and upon the idea that the phonetic level processes are performed by specialized neural mechanisms which are lateralized in one cerebral hemisphere. The present investigation was a direct extension of the experiments of Wood et al. (1971) and Day and Wood (1972a) (hereafter called the initial experiments) and had three main purposes:

a) To specify in greater detail the nature of the acoustic stimuli and processing tasks responsible for the neurophysiological and RT results of the initial experiments. Such information would provide further substantiation for the basic distinction between auditory and phonetic levels, allow a more detailed understanding of the specific functional operations performed by each level, and clarify the nature of interactions between levels.

b) To make a stronger test of the convergence of the RT and neurophysiological findings upon the distinction between auditory and phonetic processing. Although it was suggested above that both sets of results reflect the same underlying difference between auditory and phonetic levels of processing, the two initial experiments did not use identical paradigms or stimulus sets. These differences reflect methodological constraints imposed by the distinct backgrounds from which each experiment was derived. If both experimental operations actually do converge upon the single concept of specialized phonetic mechanisms distinct from the general auditory system, then it should be possible to obtain both the neurophysiological and RT findings in a single experiment. In addition, both response measures should be entirely consistent over a range of acoustic stimuli and identification tasks which differ in the aspect of whether or not phonetic processing is required.

c) To obtain more information concerning the characteristics of the neural activity that occurs during phonetic processing. Although Wood et al. (1971) clearly demonstrated that different neural events occur during phonetic and auditory processing, that experiment provided relatively little information concerning the characteristics of those events or their relation to neurophysiological correlates of other perceptual phenomena.

The present investigation consisted of four experiments, each of which compared the perception of two dimensions of the same speech stimuli. In all four experiments a two-choice identification paradigm similar to that of Day and Wood (1972a) combined the methodology of the RT and evoked potential experiments so that both response measures could be obtained in each experiment. One stimulus dimension, Pitch, was identical in all experiments, in order to provide the same auditory processing task as a common baseline in all four experiments. Individual experiments differed from the others only in the second stimulus dimension compared to Pitch. The nature of the information carried by this second dimension and its status as linguistic or nonlinguistic constituted the principal manipulation across experiments.

Experiments 1 and 2 sought to provide further validation of the initial RT and evoked potential experiments which distinguished between auditory and phonetic levels of processing. Experiment 1 was a replication of the comparison between the Place and Pitch dimensions made in the initial experiments, while Experiment 2 was a control experiment comparing Pitch with another auditory dimension, Intensity. Experiments 3 and 4 were designed to provide more specific information about the acoustic stimuli and processing tasks responsible for the differences between phonetic and auditory dimensions. Both experiments investigated dimensions with different degrees of approximation to the phonetic dimension Place used in Experiment 1 and the initial experiments. Experiment 3 analyzed the acoustic cue for the Place distinction, the second formant transition, in isolation rather than in phonetic context as in Experiment 1. Experiment 4 analyzed Pitch Contour, which is basically an auditory dimension, but which under the appropriate conditions can cue a linguistic distinction.

METHOD

The Basic Paradigm Common to All Experiments

Subjects were presented series of trials consisting of synthetic CV syllables which varied along a given target dimension. On each trial subjects were required to identify which of two possible levels on the target dimension had occurred by pressing one of two response buttons as rapidly as possible. Subjects' identification accuracy, identification speed (RT), and averaged evoked potentials time-locked to the acoustic stimulus, were recorded on each trial. Identification of the two dimensions was measured under two conditions: 1) a single-dimension control condition, in which only the target dimension varied in the stimulus sequence; and 2) a two-dimension orthogonal condition in which both the target dimension and the irrelevant nontarget dimension varied orthogonally in the stimulus sequence.

A general form of this 2 x 2 paradigm is shown in Table 1 for the hypothetical stimulus dimensions A and B. Each dimension has two levels, 1 and 2, and the stimuli for each experiment consist of all possible combinations of the two levels of each dimension (i.e., A_1B_1, A_1B_2, A_2B_1, and A_2B_2). The set of possible stimuli for each cell in the paradigm was selected from this complete set of

TABLE 1: General form of the 2 x 2 paradigm used in each experiment.

TARGET DIMENSION	CONDITION	
	CONTROL	ORTHOGONAL
A	A_1B_1 A_2B_1 A_1B_2 A_2B_2	A_1B_1 A_2B_1 A_1B_2 A_2B_2
B	A_1B_1 A_1B_2 A_2B_1 A_2B_2	A_1B_1 A_2B_1 A_2B_2 A_2B_2

four. For example, consider the case in which dimension A is the target dimension and dimension B is the nontarget dimension shown in the top half of Table 1. In the orthogonal condition shown on the right, both dimensions vary orthogonally and the stimulus set for this cell consists of all four stimuli. In the control condition shown on the left, only the target dimension A varies and the nontarget dimension B is held constant at one of its two levels. Notice that there are two ways in which the nontarget dimension can be held constant: either at level 1 shown in the upper half of this cell, or at level 2 shown in the lower half. In order to allow each of the four stimuli to occur an equal number of times in each cell, two separate blocks of trials were presented for each cell. In the control condition the stimulus sets for these two blocks are shown in the upper and lower halves of the control cells in Table 1. In the orthogonal condition there is only one possible stimulus set, and the two blocks of trials for this condition were repetitions of this same set. The lower half of Table 1 shows the corresponding stimulus sets for blocks of trials in which dimension B is the target dimension. By constructing the stimulus sets in this manner, it was possible for each of the four stimuli to occur an equal number of times both within and between cells.

The use of the design shown in Table 1 allowed both the RT and the evoked potential data to be obtained in each experiment. As described above, the degree to which each dimension could be processed independently of irrelevant variation in the other was analyzed by comparing RTs in the control and orthogonal conditions. For the RT data, this paradigm was virtually identical to that of Day and Wood (1972a). The analysis of neural activity comparable to that of Wood et al. (1971) was made by comparing evoked potentials recorded during the control condition for each dimension. As described in detail above, this comparison requires that the evoked potentials for each dimension be equated for factors such as the acoustic stimuli, their presentation probability, subject's motor response, RT, electrode locations, and the recording apparatus. Only under these circumstances may obtained differences in evoked potentials be attributed to differences in the perceptual processes for each dimension.

Subjects

Six male and six female paid volunteers aged 19-24 served as subjects in all four experiments. All subjects were right handed and had no history of hearing difficulty.

Stimuli

The four acoustic stimuli used in each experiment were generated by the Haskins Laboratories' parallel resonance synthesizer and edited under the Haskins' executive system (Mattingly, 1968). In Experiments 1, 2, and 4, the synthetic stimuli were two-formant CV syllables of 300 msec duration. In Experiment 3, the stimuli were portions of the full CV syllables, the second formant transitions. The specific set of stimuli for each experiment will be described below. The four stimuli for each experiment were precisely equated for all acoustic parameters except for the two dimensions that were explicitly varied.

From the synthesizer each stimulus was digitized and stored on a magnetic disc via the Haskins Laboratories' pulse-code modulation system (Cooper and Mattingly, 1969). A five-channel stimulus tape was prepared for each experiment, consisting of four stimulus channels and one channel of trigger pulses. A series of 64 trigger pulses was first recorded on magnetic tape at 5-sec intertrial intervals by a Precision Instrument FM tape recorder (frequency response: ± 0.5 db, DC to 10 kHz at 30 ips). One of the synthetic stimuli was then played into the analog-to-digital converter of a PDP-12 computer, digitized at a sampling rate of 60 μsec, and stored in core memory. The prerecorded trigger pulses were used to trigger the computer, causing the stimulus in memory to be played through the digital-to-analog converter and recorded on another channel of the analog tape. This process was repeated for the remaining three stimuli for that experiment, resulting in a stimulus tape with the four stimuli and the trigger pulses occurring 64 times in parallel on separate channels all synchronized to 60 μsec accuracy. Any of the four possible stimuli could therefore be presented on a given trial by connecting the appropriate channel of the tape recorder to the subject's earphones.

Apparatus

Subjects were seated comfortably in an Industrial Acoustics Corporation sound-attenuating and electrically shielded chamber which was illuminated at moderate intensity. The stimuli were presented binaurally from the tape recorder to G. C. Electronics earplug-type earphones through a Grason-Stadler Model 829D electronic switch at 65 db SL against a constant 30 db SL masking noise (Grason-Stadler Model 701 noise generator). Both of these values were determined individually for each subject prior to the experiment. Intensities of the synthetic stimuli and white noise were separately controlled by Hewlett Packard Model 350B two-channel decade attenuators in series with the tape recorder and white noise generator.

The electroencephalogram (EEG) was recorded using a Grass Model 7 polygraph with Grass Model P511 wide-band AC EEG preamplifiers (system gain: 2×10^4) and was monitored visually throughout each block of trials. Half-amplitude low- and high-frequency settings of the amplifiers were 0.1 Hz and 300 Hz, respectively. Scalp recordings were made using Grass silver disc electrodes from two symmetrically located locations over each hemisphere, each referred to a linked ear reference. The scalp locations were T3 and C3 over the left hemisphere and T4 and C4 over the right hemisphere according to the International 10-20 system (Jasper, 1958). Subjects were grounded through a plate attached to the left wrist. The impedance of all electrodes was monitored carefully throughout each recording session and maintained at less than 4.0 kohms at 10 Hz. Particular care was taken to equalize impedances of the two ear reference electrodes.

The amplified EEG from the four scalp locations (T3, C3, C4, T4) was entered into a LINC computer for on-line analog-to-digital conversion and signal averaging. The LINC sampling epochs were 490 msec long with 256 time points per sampling epoch. Three sampling rates were used in the 490 msec epoch: 1 point every 0.5 msec for the first 60 points, 1 point every 1 msec for the next 66 points, and 1 point every 3 msec for the remaining 130 points. The LINC stored the resulting averaged evoked potentials on digital magnetic tape separately for each stimulus in each block of trials. In addition to this on-line processing, all channels of EEG together with pulses synchronized with stimulus onset and the subjects' identification responses were recorded on a Honeywell Model 8100 FM tape recorder (frequency response: ± 0.25 db, DC to 625 Hz at 3-3/4 ips) for subsequent off-line data analysis.

The LINC also controlled stimulus presentation order and recorded the subject's reaction time. On each trial the LINC read a stimulus code from paper tape and closed a relay to present the stimulus specified by that code to the subject's earphones. Separate paper tapes were made for each cell in the 2 x 2 paradigm shown in Table 1, each tape containing a different constrained random order of the possible stimuli for that cell. The random orders were determined by reference to a random number table with the constraint that each possible stimulus for a given block of 64 trials occur an equal number of times in that block, with no runs of the same stimulus longer than five. No subject ever received the same random order in any cell across all four experiments. Reaction time was recorded to 1 msec accuracy using a counter-timer (Beckman-Berkeley Model 7531R) which was triggered at stimulus onset and halted by the subject's button-press response. The LINC then read and translated the RT from the counter-timer and punched the stimulus code and RT for that trial on paper tape for later analysis.

Procedure

Before beginning the experiments, each subject served in an initial orientation session to become familiar with the apparatus and experimental procedures. The methods of EEG and evoked potential recording were explained in some detail and subjects received practice in all conditions of the first experiment they were scheduled to receive. The practice trials were presented under conditions identical to the actual experiment and allowed subjects to stabilize performance in the identification tasks. In addition, these practice trials allowed the experimenter to assess subjects for two criteria that each subject had to meet in order to participate in the actual experiments: 1) they had to perform the identification tasks accurately with mean RTs of less than 600 msec in the control conditions; and 2) they had to perform the tasks so that stable EEG and evoked potentials could be recorded without muscle or movement artifact. Two potential subjects out of the initial 14 tested did not meet these criteria and were excused. One subject failed to meet the RT criterion, while the other showed a large movement artifact which could not be eliminated. The remaining 12 subjects met both criteria and were continued for the remainder of the experimental procedure.

Since each subject participated in all four experiments, they received the experiments in an order specified by a balanced latin square to control for possible effects of presentation order. For this purpose the 12 subjects were placed in four groups of three, with a different order of the four experiments specified for each group by the latin square. Each experiment was given in a separate session, with at least two days intervening between sessions. To control for possible effects of presentation order within a given experiment, the four cells in the 2 x 2 paradigm were also presented in an order specified by a balanced latin square. Each subgroup of three subjects was assigned a within-experiment presentation order according to the latin square and received this same order in all four experiments.

A session consisted of eight blocks of 64 trials (two blocks of trials for each cell of the paradigm shown in Table 1). The first four blocks of trials in a given session consisted of the order specified by the latin square, and this order was then reversed for the remaining four trials in that session. For the orthogonal condition the two blocks of trials for each cell were identical as described above, since all four possible stimuli occurred in random order.

For the control condition the nontarget dimension was held constant at level 1 on one block of trials and at level 2 on another block of trials, as shown in Table 1. For half the subjects the nontarget dimension was held constant at level 1 during the first block of trials and at level 2 during the second block of trials, while the other half of the subjects received the levels of the nontarget dimension in the reverse order. Thus, each subject received two blocks of 64 trials in each cell of each experiment, with order of presentation determined both within and between experiments by a latin square.

At the beginning of each session, the four stimuli for that experiment were presented in a fixed sequence and repeated until the subject reported that they could be easily distinguished. This rarely required more than two or three repetitions of the sequence for any subject. Response buttons were then assigned to levels on each dimension in the following manner. One dimension was held constant and the two levels on the other dimension were presented in alternating order, beginning with the level for button 1. When the subject reported that he knew the correct button for each level, the stimuli were presented in random order until a criterion of eight consecutive correct responses was attained. This procedure was then repeated to assign responses to the levels on the second dimension.

After mastering the button assignments, subjects received one practice block of 64 trials for each cell, in the same order as they would be received in the actual experiment. In addition, each block of trials in the actual experiment was preceded by at least eight practice trials to allow subjects to adapt to the target dimension and stimulus set they would receive on that block. Such extensive practice was designed to maintain performance at a stable optimal level during the experiment, while minimizing the possibility of artifact in the electrical recordings. Between each block of trials subjects received a four to five minute rest period, with a longer rest interval following the fourth block of trials in each session. An entire session, including electrode application, lasted three to four hours.

Data Analysis

As described above, the RT on each trial was recorded by the counter-timer, translated by the LINC, and punched on paper tape together with the code for the stimulus presented on that trial. These tapes were read into a PDP-12 computer which sorted the RTs into appropriate categories as a function of subject, dimension, and condition. These data were then transferred to digital magnetic tape for permanent storage. For statistical analysis a complete four-way factorial analysis of variance was computed on the RTs from each experiment (Subjects x Conditions x Dimensions x Within). The RTs were untransformed except that RTs greater than 1 sec were set equal to 1 sec. This procedure eliminated the few very long RTs (less than 3 percent) resulting from failure to make electrical contact with the response button, etc. Subsequent individual comparisons between main effect and interaction means were made using the Scheffe procedure (Scheffe, 1960; Winer, 1962). Unless otherwise noted, all statements of statistical significance are the $P < .001$ level.

The evoked potentials collected on-line by the LINC during the control condition in each experiment were averaged separately for each dimension. This procedure resulted in an evoked potential for each dimension based on 128 trials at each electrode location for each subject. The within-subject data were averaged across subjects to yield a single average for each dimension consisting of 1,536 trials over the 12 subjects. These across-subject averages correspond directly to the data presented by Wood et al. (1971) and constitute the principal neurophysiological data for each experiment. The statistical significance of differences between evoked potentials in each experiment was evaluated by computing Wilcoxon matched-pairs signed-ranks tests (Siegel, 1956) at each of the 256 time points in a pair of evoked potentials at a given electrode location. This procedure was used by Wood et al. (1971) not only to determine the statistical reliability of differences between evoked potentials for the two dimensions, but also to determine the precise distribution of significant differences in time relative to stimulus onset and subjects' identification responses.

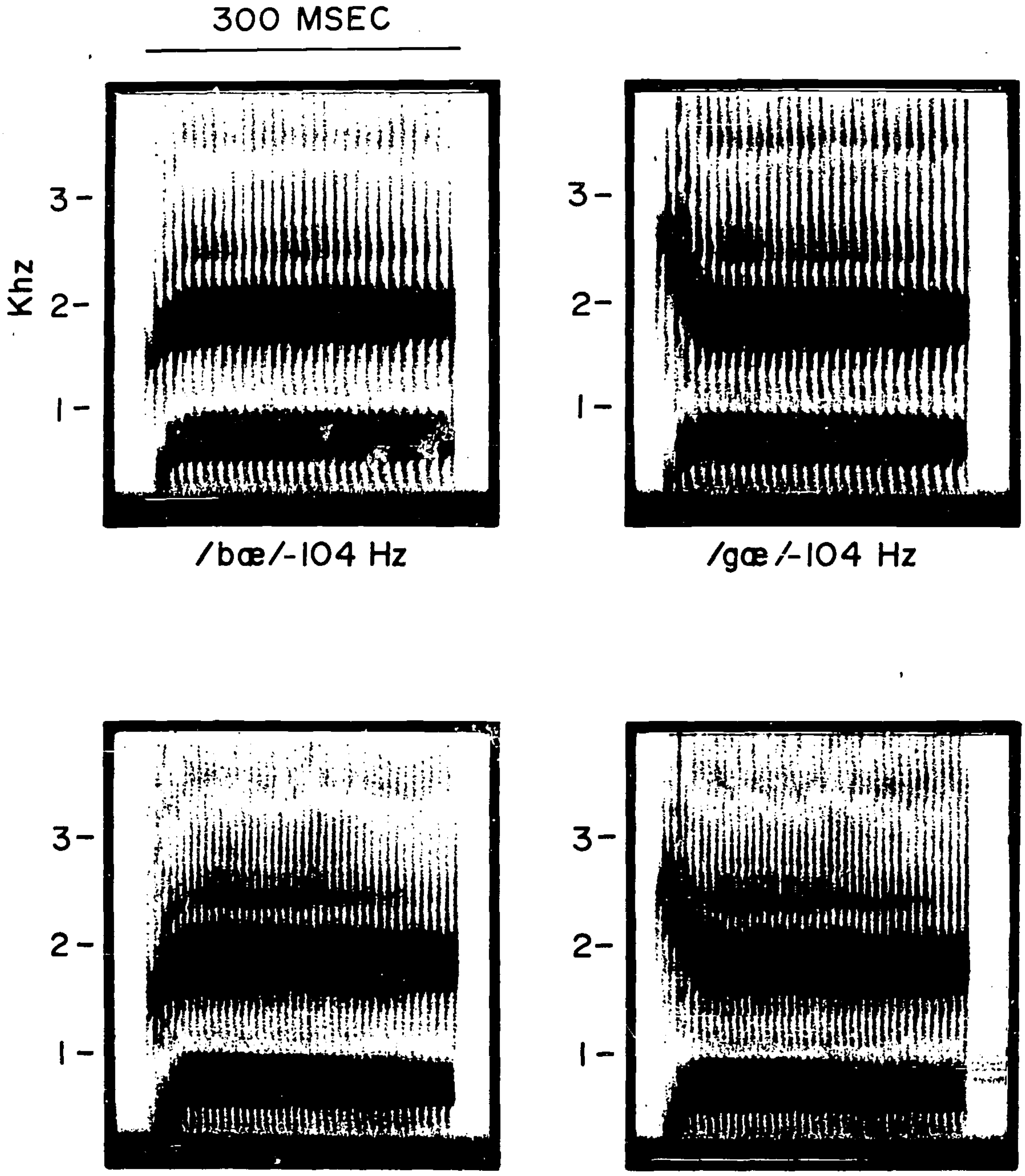
EXPERIMENT 1: PLACE OF ARTICULATION

Experiment 1 was designed to replicate the initial RT and evoked potential experiments of Day and Wood (1972a) and Wood et al. (1971) in a single experiment. The same phonetic and auditory dimensions used in the initial experiments were again used in Experiment 1: place of articulation of voiced stop consonants (Place) and fundamental frequency (Pitch). Place was selected as the phonetic dimension since this cue is an excellent example of the "encoding" of phonetic information in the speech signal (Liberman et al., 1967). In addition, in categorical perception and dichotic listening experiments, stop consonants have consistently produced results which have been interpreted as characteristic of phonetic perception (Liberman et al., 1967; Studdert-Kennedy et al., 1970a; Studdert-Kennedy and Shankweiler, 1970; Pisoni, 1971). Pitch was selected as the auditory dimension since the absolute fundamental frequency of a syllable conveys no phonetic information in English. In addition to replicating the initial experiments, Experiment 1 assessed their generality by using two-formant instead of three-formant synthetic syllables, by using acoustic cues for a different value on the Place dimension (/b/ and /g/ instead of /b/ and /d/), and by using formant frequencies for a different vowel (/ae/ instead of /a/) than were employed in the initial experiments.

Stimuli

The stimuli for Experiment 1 consisted of the four possible combinations of two levels on the Place (/bae/ and /gae/) and Pitch dimensions (104 Hz and 140 Hz). Wide- and narrow-band spectrograms of these four stimuli are shown in Figures 1 and 2, respectively. The spectrograms are three-dimensional displays in which intensity (relative darkness of the display) is plotted as a function of frequency (vertical axis) and time (horizontal axis). The two horizontal bands of highest energy in each stimulus are called formants, and are numbered from low to high frequency (i.e., F1, F2, etc.). The rapid frequency changes at the beginning of each formant are called formant transitions.

Pairs of stimuli which differed on the Place dimension differed only in the direction and extent of the F2 transition, as shown by the two spectrograms on the left versus the two on the right of Figures 1 and 2. All acoustic parameters of these pairs of stimuli were identical except for the initial 45 msec of F2. The direction and extent of the F2 transition is the acoustic cue important for distinguishing among voiced stop consonants (Liberman, Delattre, Cooper, and



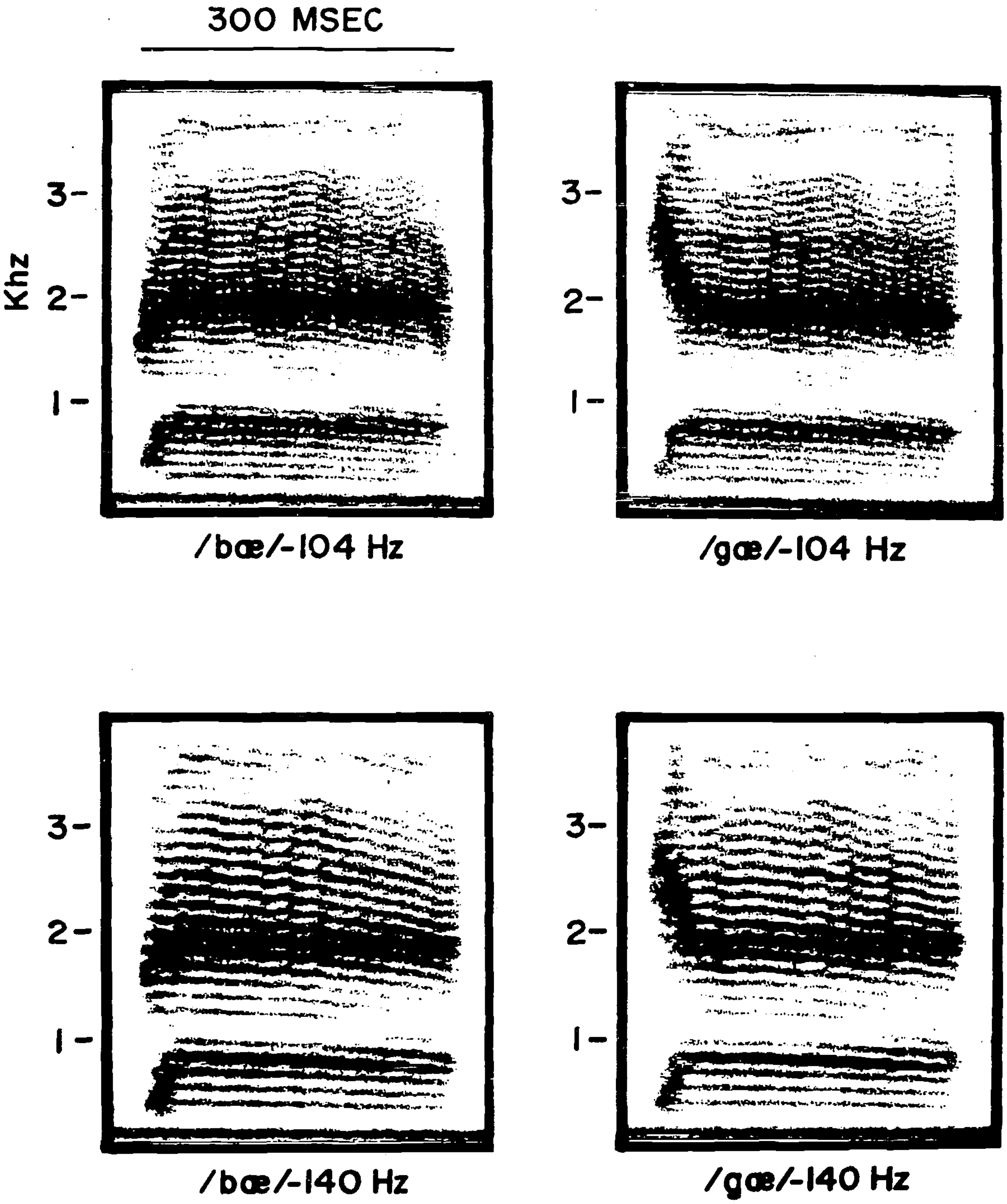
/bæ/-104 Hz

/gæ/-104 Hz

/bæ/-140 Hz

/gæ/-140 Hz

Figure 1: Wide-band spectrograms of the four synthetic stimuli for Experiment 1.



/bœ/-104 Hz

/gœ/-104 Hz

/bœ/-140 Hz

/gœ/-140 Hz

Figure 2: Narrow-band spectrograms of the same stimuli shown in Figure 1.

Gerstman, 1954; Delattre, Liberman, and Cooper, 1955). In the context of the vowel /ae/, a rising F2 transition is the cue for /b/ and a sharply falling F2 transition is the cue for /g/.

Pairs of stimuli which differed on the Pitch dimension (top versus bottom halves of Figures 1 and 2) had identical formant patterns and differed only in fundamental frequency (F_0). The differences in F_0 are shown indirectly in both Figures 1 and 2, but may be seen more clearly in the narrow-band spectrograms of Figures 2. In these spectrograms the bandwidth of the analysis filters was sufficiently narrow to resolve the individual harmonics of the F_0 of each stimulus. The harmonics of the 104 Hz stimuli (upper two spectrograms in Figure 2) are more closely spaced, reflecting the closer spacing of integral multiples of 104 Hz than 140 Hz.

The assignment of levels on each dimension to response buttons was 104 Hz-button 1, 140 Hz-button 2, /bae/-button 1, and /gae/-button 2. This assignment was the same for the control and orthogonal conditions for both dimensions.

Results and Discussion

Reaction time. Mean RTs for the Place and Pitch dimensions are shown in Figure 3, for the control and orthogonal conditions. Each point in the display is the mean of 1,536 observations over the 12 subjects, with a given subject contributing 128 observations to each mean. A 2 x 2 display of this kind will comprise the principal RT results in each experiment.

As described in the introduction, the question of interest in the RT data is whether each dimension could be processed without interference from irrelevant variation in the other dimension. For Place, Figure 3 shows that there was a substantial increase in RT of 50.1 msec from the control to the orthogonal condition, while for Pitch the corresponding difference between conditions was 0.6 msec. The statistical reliability of these results may be determined from the analysis of variance for this experiment presented in Table 2. The term in the analysis corresponding to the 2 x 2 partition of the data in Figure 3 is the Condition x Dimension interaction (B x C in Table 2), which was highly significant. The main effects of Conditions and Dimensions are also significant, but these effects can be completely accounted for by the Condition x Dimension interaction. Individual differences among subjects will be considered in a later section.

The Scheffe procedure for individaul comparisons was applied to the Condition x Dimension interaction means and showed that a difference of 14.6 msec was necessary for significance at the $P < .001$ level. Thus, the orthogonal condition for Place differed significantly from the other three conditions, while the differences among the latter were not significant. The RT results of Experiment 1 therefore constitute a clear replication of those obtained by Day and Wood (1971a). In both experiments there was a significant Condition x Dimension interaction, indicating that irrelevant variation in Pitch produced significantly more interference with the identification of Place than the reverse.

Evoked potentials. The evoked potential data recorded during the control conditions for Place and Pitch in Experiment 1 are shown in Figure 4. Evoked potentials from the two tasks are superimposed at each electrode location to facilitate visual comparison. Each trace is the average of 1,536 trials (128 trials for each of the 12 subjects) and corresponds directly to the RT data from

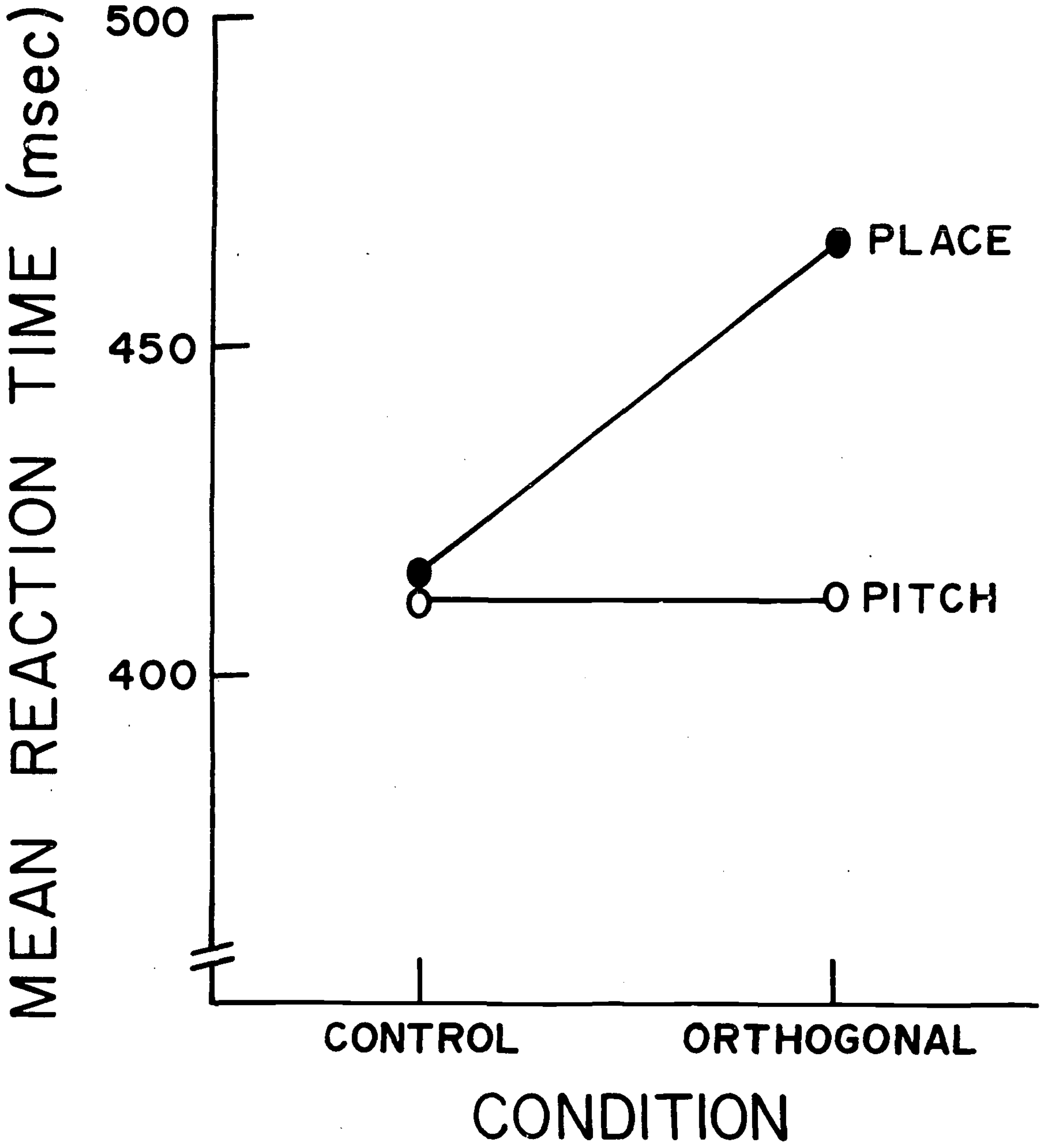


Figure 3: Reaction time data for Experiment 1.

TABLE 2: Summary of analysis of variance for Experiment 1.

SOURCE	df	MS	F
Subjects (A)	11	2005004.86	233.77*
A x D	1397	8576.78	
Conditions (B)	1	985365.38	49.72*
B x D	127	19821.49	
Dimensions (C)	1	1304742.38	128.44*
C x D	127	10158.10	
Within (D)	127	19751.72	
A x B	11	16319.56	2.05
A x B x D	1397	7970.90	
A x C	11	184764.15	25.22*
A x C x D	1397	7326.90	
B x C	1	941242.50	99.65*
B x C x D	127	9445.22	
A x B x C	11	17660.27	1.91
A x B x C x D	1397	9244.43	
TOTAL	6143		

*p < .001

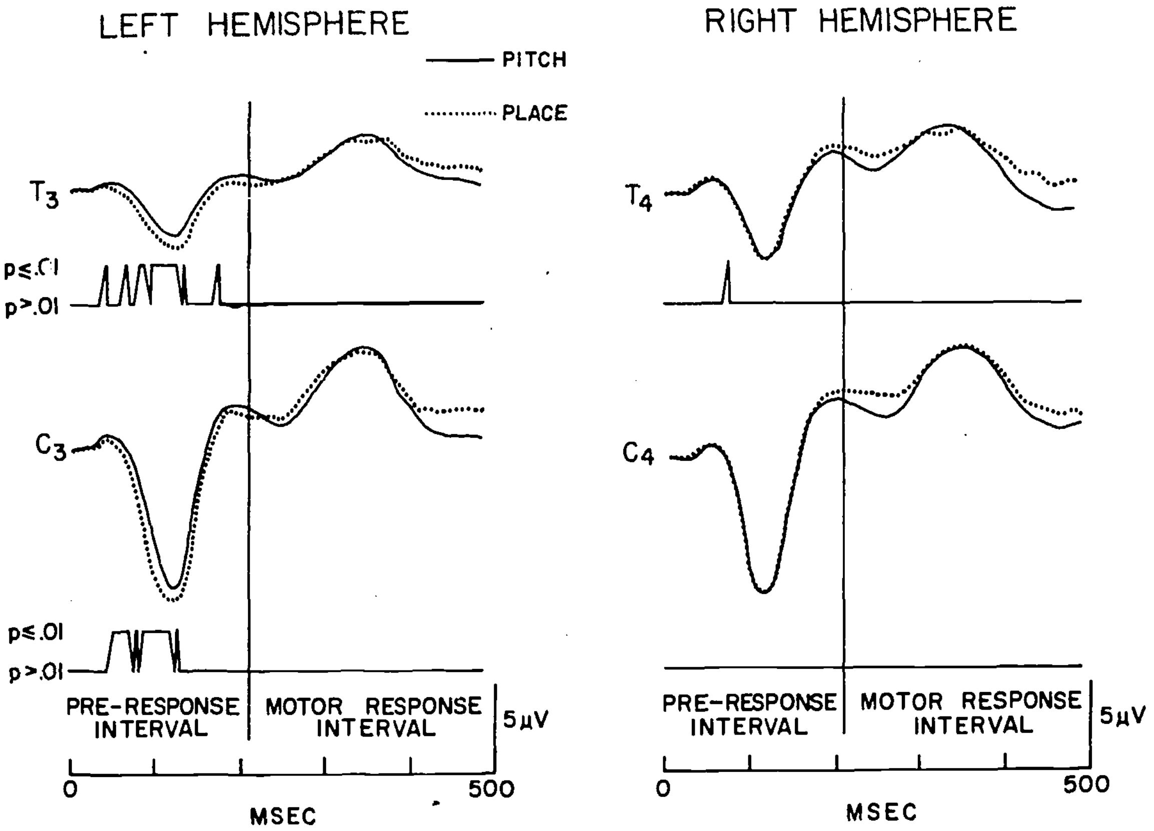


Figure 4: Average evoked potentials during identification of Place and Pitch in Experiment 1.

the control conditions for Place and Pitch shown in Figure 3. The statistical reliability of differences between the evoked potentials for the two dimensions was analyzed using the Wilcoxon procedure described in the Method section above. The results of this analysis are shown in histogram form directly beneath the pair of evoked potentials at each electrode location. Upward deflections from baseline in the statistical traces indicate that the obtained difference in evoked potentials at that individual time point was significant at the $P < .01$ level.

In order to analyze neural activity that occurred during the processing of each dimension and eliminate from consideration possible differences in neural activity associated with the button-press response, the 490 msec averaging epoch was empirically divided into pre-response and motor response intervals. On a single trial, the time interval during which perceptual processing must have occurred begins at stimulus onset and ends with the subject's button-press response. However, in the large number of trials required for evoked potential averaging, the precise termination of this "processing interval" is less clear. The criterion employed by Wood et al. (1971) to estimate the end of the "processing interval" was also used here, namely, the time point after which 99 percent of the button-press responses occurred. This point is shown for the data in Figure 4 by the vertical lines at 207 msec. It should be noted that this criterion is particularly conservative, since it eliminates from consideration activity which occurs in time after the first 1 percent of subjects' motor responses. It is conceivable, of course, that differences in neural activity related to perceptual processing might not occur until the middle of the RT distribution. However, in such a case it would be impossible to determine whether these differences were related to perceptual processing or to the motor response.

As described in the introduction, factors such as the acoustic stimuli, stimulus presentation probability, subjects' motor response and RT, electrode location, and all aspects of the recording apparatus were equated in the evoked potentials from the Place and Pitch tasks. Therefore, if no true difference in neural activity were produced by the Place and Pitch tasks, then the two evoked potentials at each location in Figure 4 would merely be random samples from the same population. Under these conditions, the evoked potentials from the Place and Pitch tasks should differ only to the extent expected by chance alone. For the right hemisphere locations (T4 and C4) shown on the right side of Figure 4, this was indeed the case. No more significant differences than would be expected by chance occurred at either location. In contrast, at corresponding locations over the left hemisphere (T3 and C3), there were significant differences in evoked potentials during the pre-response interval. While 2.56 points significant at the $P < .01$ level would be expected to occur by chance, 48 and 54 significant points were obtained at T3 and C3, respectively. These results provide a clear replication of the evoked potential results of Wood et al. (1971). In both experiments, significant differences between evoked potentials for Place and Pitch were obtained only at left-hemisphere locations.

EXPERIMENT 2: INTENSITY

The conclusions of Experiment 1 and the initial experiments would be strengthened by showing that the patterns of RT and evoked potential results for the Place and Pitch dimensions do not occur when neither dimension requires phonetic processing. That is, if the results of these experiments were actually due to different levels of processing required for the auditory and phonetic dimensions, then similar results should not occur when two auditory dimensions are compared. Experiment 2 provided such a comparison for the auditory dimensions Pitch and Intensity. All conditions of this experiment were identical to Experiment 1, except that Intensity was substituted for Place as the second target dimension.

Stimuli

Two of the stimuli for this experiment were the syllables /bae/-104 Hz and /bae/-140 Hz shown on the left side of Figures 1 and 2. Thus, the Pitch dimension in this experiment (104 Hz versus 140 Hz) was identical to that of Experiment 1. However, instead of varying in Place, variations in Intensity were produced by attenuating the stimuli /bae/-104 Hz and /bae/-140 Hz by 20 db SL. Thus, the four stimuli used in this experiment were 104 Hz-loud, 104 Hz-soft, 140 Hz-loud, and 140 Hz-soft, each with formant patterns identical to those of the syllable /bae/ used in Experiment 1. The 20 db SL attenuation was produced relative to each subject's individual 65 db SL signal level by interposing a Hewlett Packard Model 350B decade attenuator in the stimulus circuit under LINC relay control. The assignment of response buttons to levels on each dimension was similar to Experiment 1: 104 Hz-button 1, 140 Hz-button 2, loud-button 1, and soft-button 2.

Results and Discussion

Reaction time.

Mean RTs for Pitch and Intensity are shown in Figure 5. There was a large increase in RT from the control to the orthogonal condition for both dimensions. For Pitch the increase was 42.7 msec and for Intensity the increase was 36.1 msec. The results of the analysis of variance shown in Table 3 indicate that the difference between control and orthogonal conditions was highly significant, while there was no difference between the Pitch and Intensity dimensions and no Condition x Dimension interaction. The Scheffe procedure for individual comparisons showed that a difference of 15.3 msec was required for significance at the $P < .001$ level. These results demonstrate that irrelevant variation in Pitch and Intensity each produced substantial interference with the identification of each other. This pattern of results is identical to that obtained for "integral" pairs of stimulus dimensions by Garner and Felfoldy (1970). Such mutual interference suggests that both dimensions are automatically processed on each trial, regardless of which dimension subjects are required to identify by the processing tasks.

Evoked potentials.

The evoked potentials recorded during the control conditions of Experiment 2 are shown in Figure 6. In this experiment, the time point which divides the pre-response and motor response intervals was 209 msec, as shown by the vertical lines in Figure 6. The evoked potentials for Pitch and Intensity were not significantly different at any electrode location. These data are therefore consistent with the RT results presented above in that neither response measure provides evidence that different levels of processing are required for the two auditory dimensions. In isolation, these "negative" results could have been produced by measurement insensitivity. However, in the context of the significant differences in both response measures obtained in Experiment 1, one may be confident that the results of this experiment were not due to imprecise measurement.

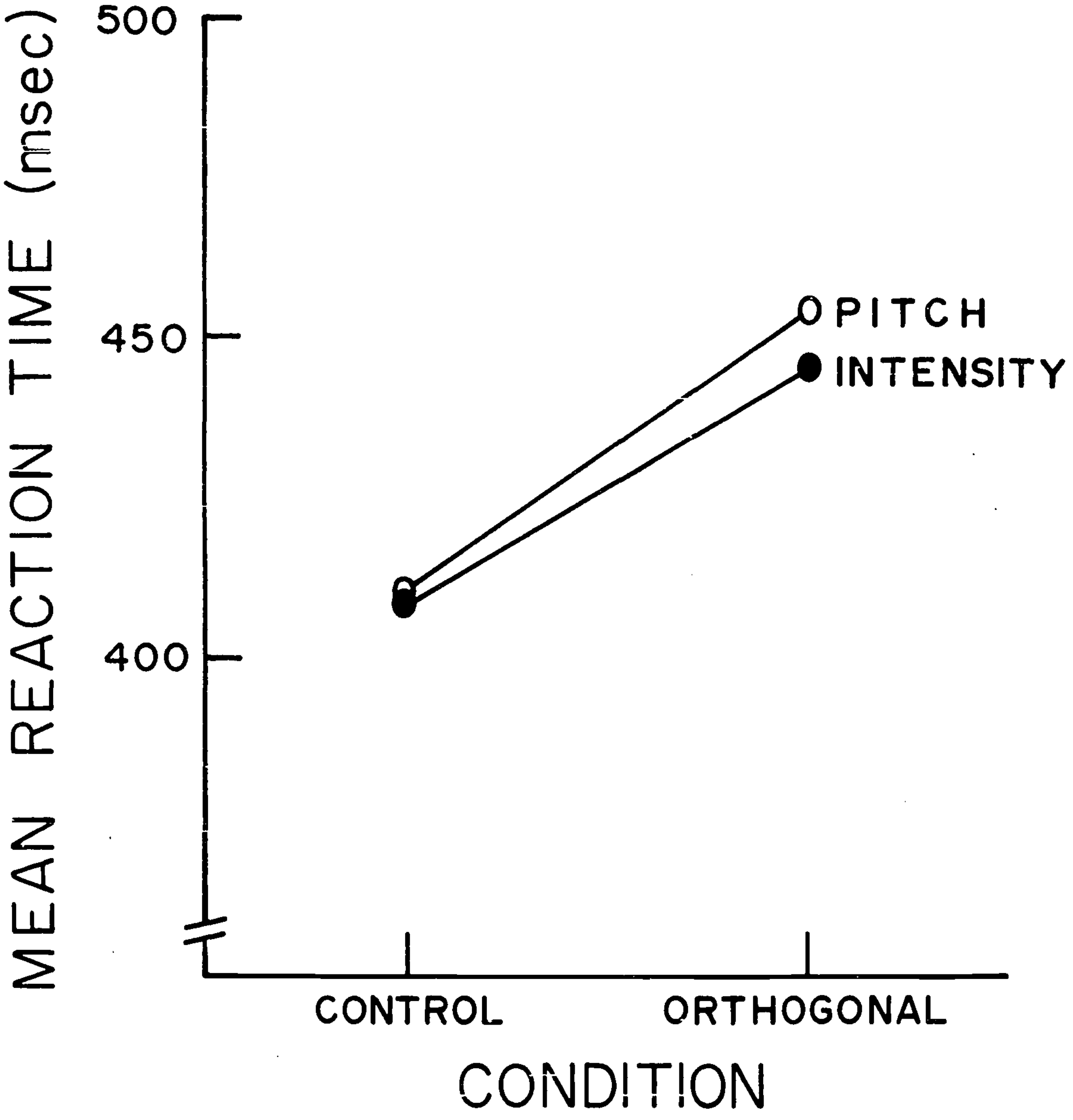


Figure 5: Reaction time data for Experiment 2.

TABLE 3: Summary of analysis of variance for Experiment 2.

SOURCE	df	MS	F
Subjects (A)	11	962771.08	106.99*
A x D	1397	8998.16	
Conditions (B)	1	2366852.75	130.49*
B x D	127	18137.03	
Dimensions (C)	1	39310.13	2.67
C x D	127	14750.24	
Within (D)	127	30277.46	
A x B	11	14748.20	1.78
A x B x D	1397	8278.54	
A x C	11	303186.76	36.23*
A x C x D	1397	8367.31	
B x C	1	15741.59	1.49
B x C x D	127	10542.85	
A x B x C	11	93978.44	8.15*
A x B x C x D	1397	11537.36	
TOTAL	6143		

*p < .001

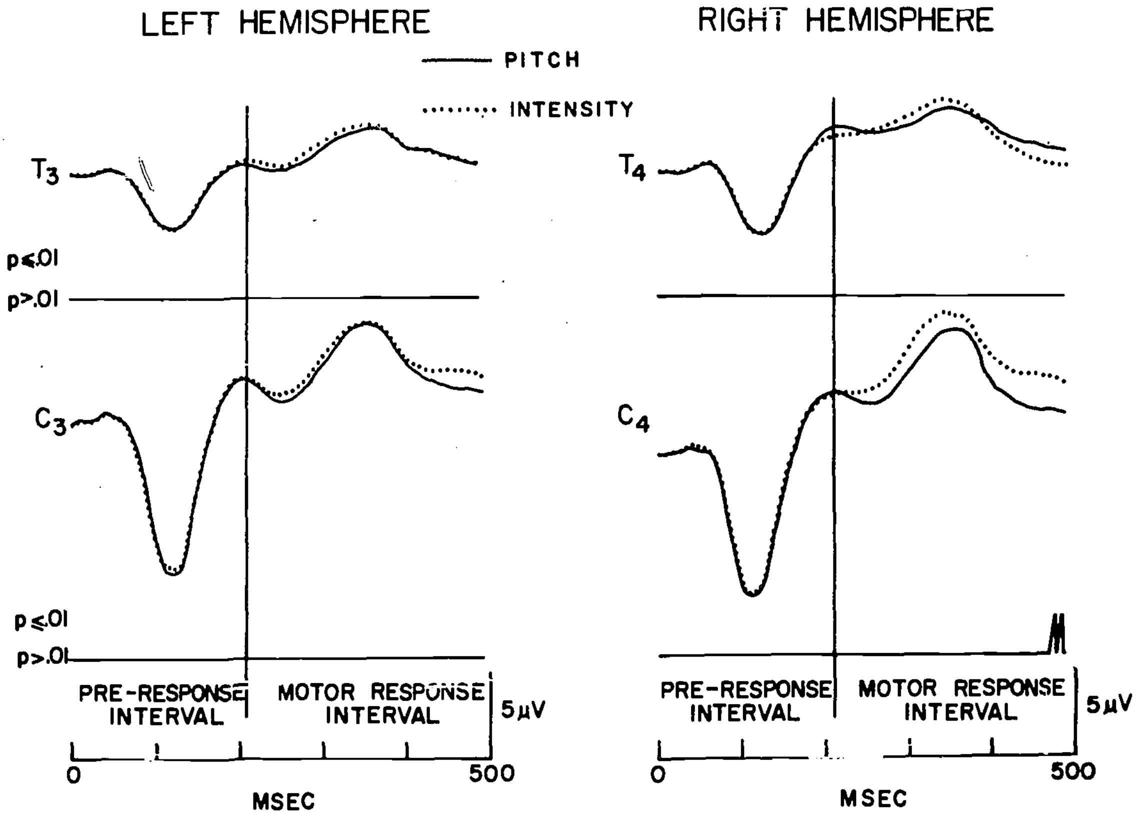


Figure 6: Average evoked potentials during identification of Intensity and Pitch in Experiment 2.

EXPERIMENT 3: SECOND FORMANT TRANSITIONS

As shown in the spectrograms of the stimuli in Experiment 1 (Figures 1 and 2), the F2 transition was the only acoustic difference between the syllables /bae/ and /gae/ at each level of the Pitch dimension. Since these pairs of stimuli were acoustically identical in all other respects, the F2 transitions (or some portion thereof) must have been the acoustic basis for the discrimination between /bae/ and /gae/ in the Place task. Therefore, the phonetic level of processing must also be directly related to the F2 transitions.

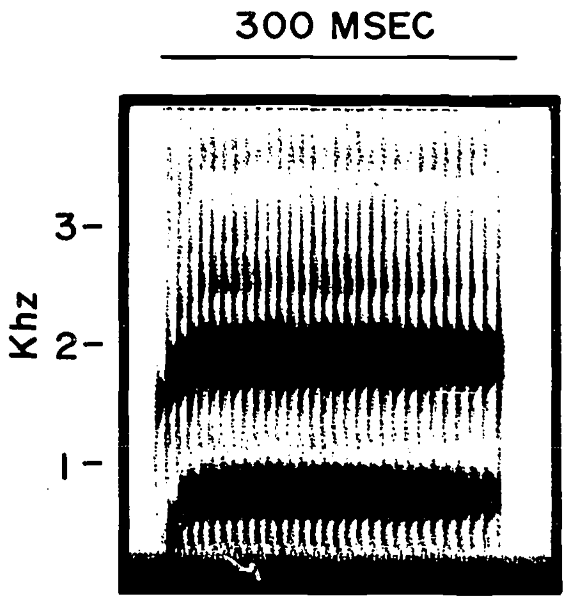
Two alternative modes of processing could characterize the additional level of processing for Place demonstrated for Experiment 1: an *auditory* mode and a *phonetic* mode. According to the auditory mode, what has been termed the "phonetic" level would actually be an auditory process specialized to detect particular *acoustic* cues in the speech signal. The F2 transition is one acoustic cue that such an auditory process would undoubtedly be specialized to detect, since this cue "is probably the single most important carrier of linguistic information in the speech signal" (Liberman et al., 1967). From this point of view, the differences between Place and Pitch in Experiment 1 would be attributed to different processing requirements for the F2 transitions: to be correct in the Place task subjects had to process the F2 transitions, while they could have performed the Pitch task correctly without processing the F2 transitions.

Alternatively, the phonetic mode would suggest that the additional level of processing would be specialized for the extraction of *phonetic features* rather than the detection of particular acoustic events in the speech signal. Instead of attributing the results of Experiment 1 to the requirement for processing of the F2 transitions per se, the phonetic mode would suggest that the important difference between Place and Pitch in Experiment 1 was that the F2 transitions occurred in phonetic context and cued a phonetic distinction.

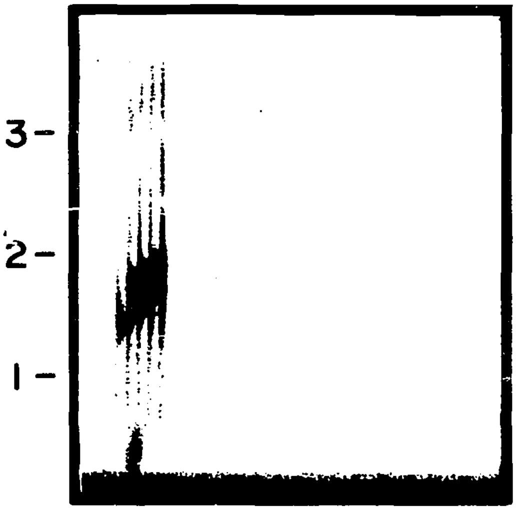
Experiment 3 sought to distinguish empirically between the auditory and phonetic modes described in the previous paragraphs. By isolating the F2 transitions from the syllable context of the stimuli in Experiment 1, it was possible to require identification of the same F2 transitions and the same levels on the Pitch dimension, while eliminating the phonetic distinction normally cued by the F2 transitions. Thus, this experiment investigated whether the differences between Place and Pitch in Experiment 1 were due to the processing of the F2 transitions per se, or rather to the fact that they cued a phonetic distinction. If the results of Experiment 1 were produced by the F2 transitions per se, then identical results should be obtained in the present experiment.

Stimuli

The four stimuli for Experiment 3 were the isolated F2 transitions of the CV syllables in Experiment 1, shown in Figures 1 and 2. These stimuli were constructed by eliminating from the CV syllables all portions of the stimuli which were identical for /bae/ and /gae/. Figure 7 presents an example of this process. The top of Figure 7 shows a wide-band spectrogram of the syllable /bae/-104 Hz, identical to that shown in the upper left of Figure 1. The corresponding F2 transition for this syllable used in Experiment 3 is shown in the bottom of Figure 7. Note that all of F1 and the steady-state vowel portion of F2 have been eliminated, leaving only the F2 transition. Thus, precisely the same acoustic



/bæ/-104 Hz



F2 TRANSITION

Figure 7: An example of the relationship between the CV syllables used in Experiment 1 and the F2 transitions used in Experiment 3.

information was available for distinguishing between /b/ and /g/ in this experiment and Experiment 1, the difference being the presence or absence of redundant phonetic context. Phenomenologically, isolated F2 transitions are perceived as nonspeech "chirps" (Liberman et al., 1967; Mattingly et al., 1971). The resulting set of four stimuli was composed of two levels on the Pitch dimension (104 Hz and 140 Hz) and two F2 transitions (rising, extracted from the syllable /bae/; and falling, extracted from the syllable /gae/).

The assignment of response buttons to stimuli in this experiment was: 104 Hz-button 1, 140 Hz-button 2, rising F2 transition-button 1, and falling F2 transition-button 2. Note that for the Pitch dimension this assignment is identical to that in Experiment 1, and that each F2 transition was assigned the same response as the corresponding syllable in Experiment 1. To avoid biasing the way subjects perceived the isolated F2 transitions, they were not told how the stimuli related to those of Experiment 1, and the neutral label "quality" was used to refer to the variation in the F2 transitions. Subjects were therefore free to distinguish between rising and falling F2 transitions in any manner they wished.

Results and Discussion

Reaction time. Mean RTs for the identification of Pitch and the F2 transitions are shown in Figure 8. For Pitch, there was an increase of 65.7 msec from the control to the orthogonal condition, while for the F2 transitions there was an increase of 53.8 msec. The analysis of variance (Table 4) and the Scheffe procedure showed that both differences were significant. The main effect of Conditions was highly significant, while the main effect of Dimensions and the Condition x Dimension interaction did not reach significance. The results of the Scheffe procedure showed that a difference of 19.8 msec between any pair of means in Figure 8 was necessary for significance at the $P < .001$ level. Thus, there was no difference between Pitch and the F2 transitions in either condition, and all of the variance in the 2 x 2 partition of the data shown in Figure 8 was associated with the difference between control and orthogonal conditions.

Evoked potentials. The evoked potentials recorded during the identification of the F2 transitions and Pitch are shown in Figure 9. In this experiment the pre-response and motor response intervals were divided at 199 msec. As shown by the Wilcoxon traces, there were no more significant differences than would be expected by chance alone at any electrode location.

Thus, neither the RT nor the evoked potential results for the F2 transitions in this experiment duplicated those of the Place dimension in Experiment 1. These results strongly suggest that the same F2 transitions were perceived differently in isolation as opposed to when they occurred in syllable context and cued a phonetic distinction. Moreover, both the RT and evoked potential results in the present experiment correspond exactly to those in Experiment 2 for Pitch and for another auditory dimension, Intensity. Therefore, the perception of the F2 transitions in isolation corresponded much more closely to the perception of Intensity than to the perception of the same F2 transitions when they cued a phonetic distinction.

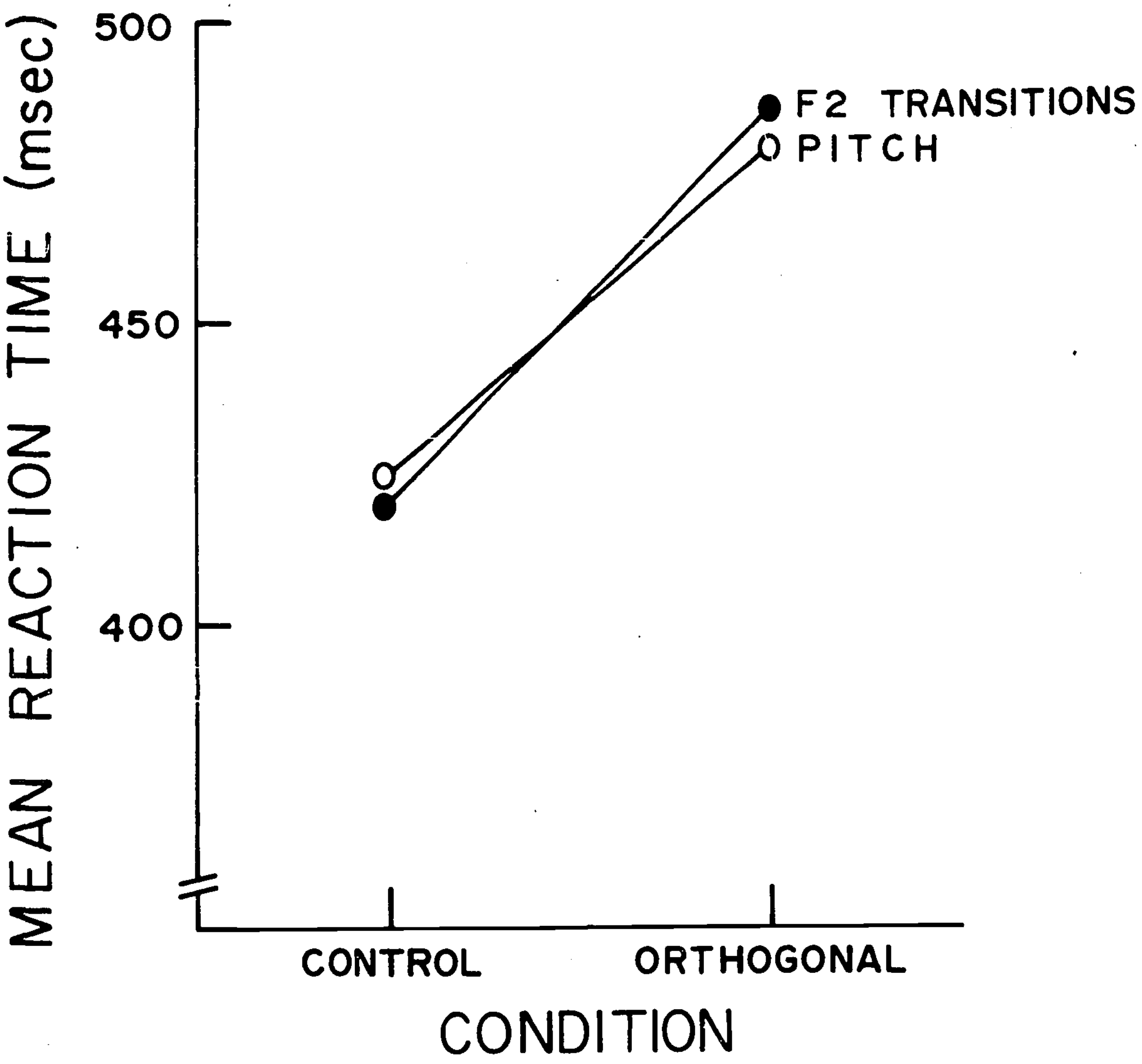


Figure 8: Reaction time data for Experiment 3.

TABLE 4: Summary of analysis of variance for Experiment 3.

SOURCE	df	MS	F
Subjects (A)	11	2013964.78	172.42*
A x D	1397	11680.80	
Conditions (B)	1	5478265.56	191.87*
B x D	127	28551.31	
Dimensions (C)	1	1260.59	.08
C x D	127	17116.80	
Within (D)	127	37839.50	
A x B	11	169639.54	17.98*
A x B x D	1397	9434.80	
A x C	11	217904.69	24.47
A x C x D	1397	8905.44	
B x C	1	54582.47	3.14
B x C x D	127	17391.20	
A x B x C	11	130270.87	10.17*
A x B x C x D	1397	12805.22	
TOTAL	6143		

*p < .001

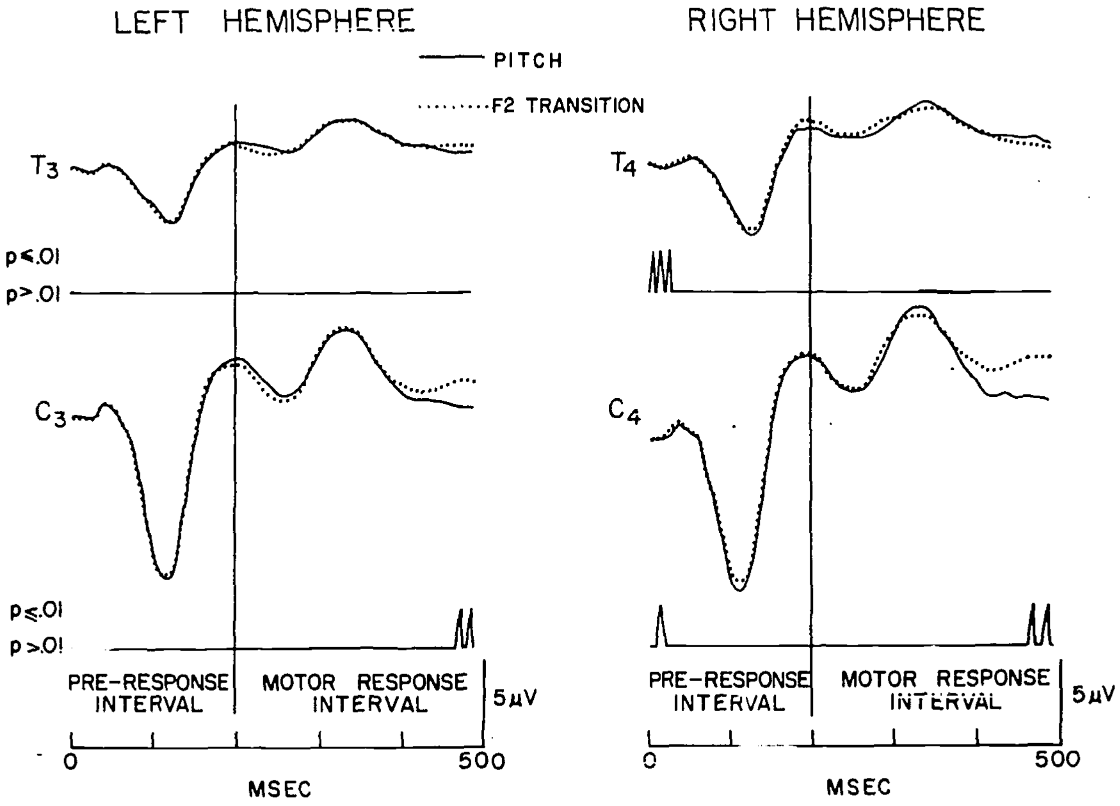


Figure 9: Averaged evoked potentials during identification of F2 transitions and Pitch in Experiment 3.

EXPERIMENT 4: PITCH CONTOUR

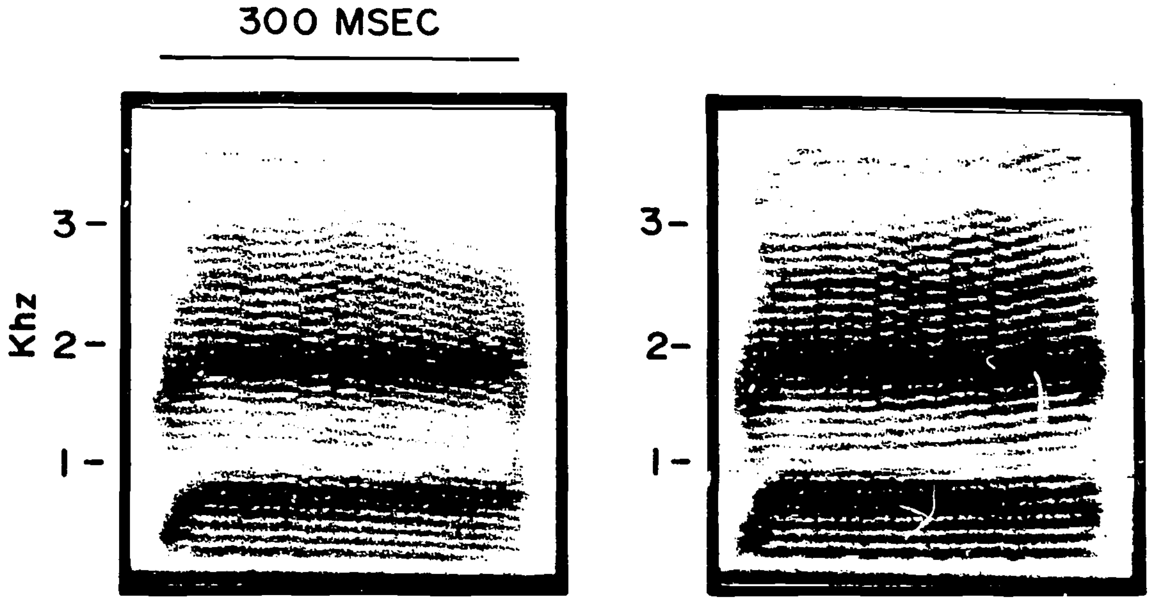
The F2 transitions which cued the phonetic distinction between /bae/ and /gae/ in Experiment 1 are examples of cues which are said to be highly "encoded" or "restructured." These terms refer to acoustic segments in the speech signal which transmit information in parallel about multiple phonetic segments, and which therefore undergo considerable context-conditioned variation as a function of their phonetic environment (cf. Liberman et al., 1967; Studdert-Kennedy and Shankweiler, 1970; Haggard, 1971; Darwin, 1971a). In contrast to such highly "encoded" cues, there are other acoustic parameters which carry linguistic information but which undergo much less context-conditioned variation. For these cues there is more or less a one-to-one correspondence between a given acoustic parameter and the linguistic distinction cued by that parameter. At the phonetic level, examples of such "unencoded" cues are the frequency positions of formants as cues for isolated vowels (Delattre, Liberman, Cooper, and Gerstman, 1952; Peterson and Barney, 1952), and the frequency position of friction noises as cues for certain fricatives (Hughes and Halle, 1956; Harris, 1958).

Another relatively "unencoded" linguistic cue is the direction of change in fundamental frequency of terminal portions of an utterance. Terminal change in fundamental frequency (here called Pitch Contour) is the most important single cue for judging whether an utterance was a question or statement (Lieberman, 1967, in press; Fry, 1968; Lehiste, 1970; Studdert-Kennedy and Hadding, 1971). In contrast to the "unencoded" cues at the phonetic level described above, Pitch Contour can occur over longer durations than a single syllable and the perceived direction of terminal Pitch Contour can be influenced by the Pitch level of earlier syllables in the utterance (Hadding-Koch and Studdert-Kennedy, 1964; Studdert-Kennedy and Hadding, 1971).

Experiment 4 investigated whether the additional level of processing demonstrated by the RT and evoked potential data of Experiment 1 was required for the identification of rising and falling Pitch Contour. Since a given change in Pitch Contour is judged in a similar manner both when it is carried by a speech signal and a pure tone (Studdert-Kennedy and Hadding, 1971), this experiment investigated an acoustic parameter which conveys linguistic information in an "unencoded" auditory form rather than the highly "encoded" form represented by the cues for Place in Experiment 1. In addition, Pitch Contour was also selected for this experiment in order to evaluate the use of the Pitch dimension as the baseline auditory dimension in Experiments 1-3 and the initial experiments. It is possible that the use of Pitch as a nonlinguistic dimension may be inappropriate, since changes in Pitch (i.e., Pitch Contour) can convey linguistic information. This experiment therefore provided a basis for comparing the results for Pitch and Pitch Contour with dimensions that are clearly linguistic (Place in Experiment 1) and nonlinguistic (Intensity in Experiment 2).

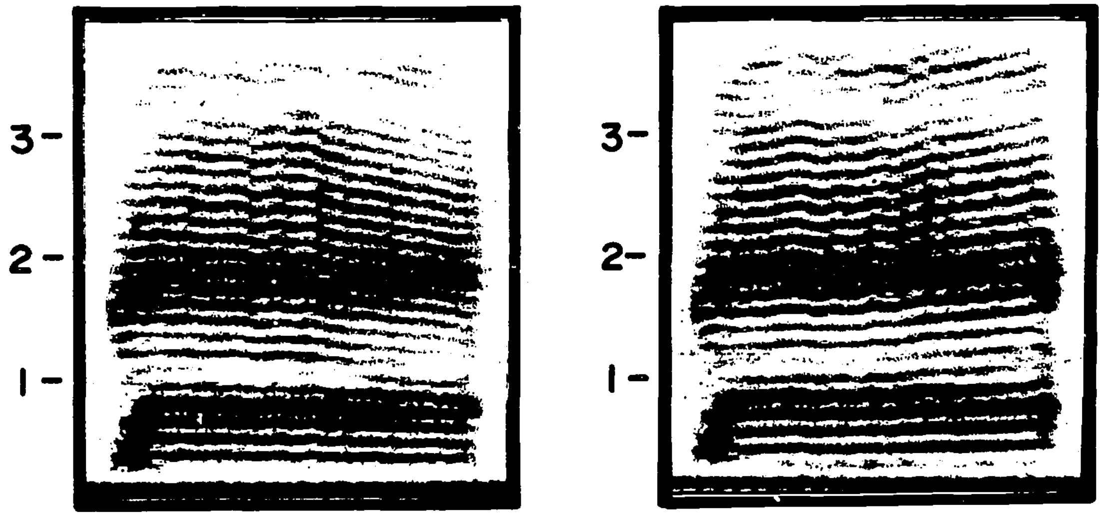
Stimuli

Narrow-band spectrograms of the stimuli for Experiment 4 are shown in Figure 10. The two stimuli on the left side of this figure are identical to the syllables /bae/-104 Hz and /bae/-140 Hz used in Experiment 1 and shown on the left sides of Figures 1 and 2. As indicated by the falling harmonics in the spectrograms, these two stimili had a Pitch Contour which gradually fell over the 300 msec duration of the stimuli. The right half of Figure 10 shows the remaining



104 Hz - FALL

104 Hz - RISE



140 Hz - FALL

140 Hz - RISE

Figure 10: Narrow-band spectrograms of the four stimuli used in Experiment 4.

two stimuli for this experiment. These stimuli are identical to those on the left in all respects, except for their Pitch Contour. In the stimuli on the right the Pitch Contour is gradually rising instead of falling. Thus, the stimuli for Experiment 4 varied in Pitch (104 Hz versus 140 Hz) and Pitch Contour (falling contour, the cue for statement; and rising contour, the cue for question). Assignments of levels on the two dimensions to response buttons in this experiment were: 104 Hz-button 1, 140 Hz-button 2, statement-button 1, and question-button 2.

Results and Discussion

Reaction time. Mean RTs for the identification of Pitch and Pitch Contour are shown in Figure 11. There was a large increase in RT from the control to orthogonal conditions for both dimensions: 101.5 msec for Pitch and 101.1 msec for Pitch Contour. In the analysis of variance shown in Table 5, the main effect of conditions was highly significant, while there was no significant Condition x Dimension interaction. These results suggest that Pitch and Pitch Contour interfere mutually, in a manner similar to Pitch and Intensity in Experiment 2.

However, in contrast to the previous experiments, Figure 11 shows that there was a large difference in RT between Pitch and Pitch Contour in the control condition as well. The statistical reliability of this difference is indicated by the significant main effect of Dimensions in absence of a Condition x Dimension interaction, and by the value of 17.8 msec required for significance at the $P < .001$ level according to the Scheffe procedure. This main effect difference between Pitch and Pitch Contour demonstrates the necessity for obtaining data from both the orthogonal and control conditions for unambiguous interpretation of differences in RT between a given pair of dimensions. The results of the orthogonal condition in this experiment are equivalent to those of Experiment 1 for Place and Pitch, with Pitch significantly faster than the other dimension in both experiments. However, a comparable difference between Pitch and Pitch Contour in the control condition indicates that the results of the orthogonal condition do not reflect differential interference between dimensions, but rather that it took approximately 80 msec longer in both conditions to identify the direction of Pitch change than the absolute level of the Pitch itself.

Evoked potentials. The evoked potential data recorded during the identification of Pitch and Pitch Contour are shown in Figure 12. In this experiment the pre-response and the motor response intervals were divided at 204 msec. In the pre-response interval there were no more significant differences between the evoked potentials for Pitch and Pitch Contour than would be expected by chance at any location. Thus the evoked potential differences characteristic of Place and Pitch in Experiment 1 did not occur in this experiment.

However, in the motor response interval there were significant differences between dimensions at all four electrode locations. This pattern of results provides an excellent illustration of the value in distinguishing between pre-response and motor response intervals in the evoked potential analysis. In absence of such a distinction, differences such as those in Figure 12 might be attributed to differences in the perceptual processing of the two dimensions, in a manner similar to that in Experiment 1 and the experiment of Wood et al. (1971). Clearly, however, there is an alternative explanation for differences during the motor response interval. Since the evoked potential differences between Pitch and Pitch Contour did not occur until after subjects began to make

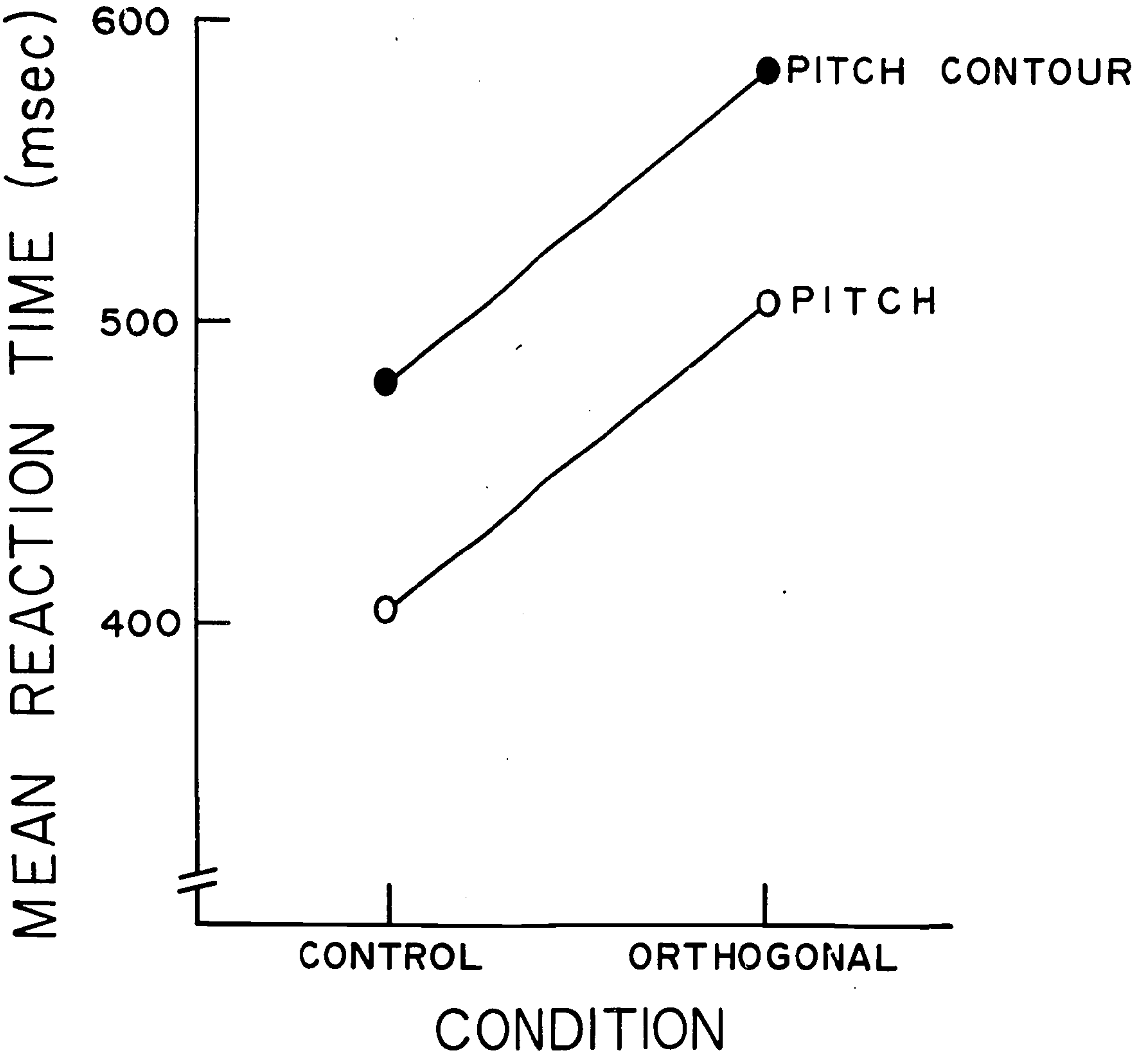


Figure 11: Reaction time data for Experiment 4.

TABLE 5: Summary of analysis of variance for Experiment 4.

SOURCE	df	MS	F
Subjects (A)	11	1657009.45	182.06*
A x D	1397	9101.24	
Conditions (B)	1	15752017.25	587.78*
B x D	127	26799.38	
Dimensions (C)	1	9080791.63	571.40*
C x D	127	15892.06	
Within (D)	127	33872.06	
A x B	11	115329.08	13.59*
A x B x D	1397	8481.84	
A x C	11	155983.69	22.82*
A x C x D	1397	6834.86	
B x C	1	270.13	.02
B x C x D	127	14038.41	
A x B x C	11	53567.78	5.95*
A x B x C x D	1397	9007.67	
TOTAL	6143		

*p < .001

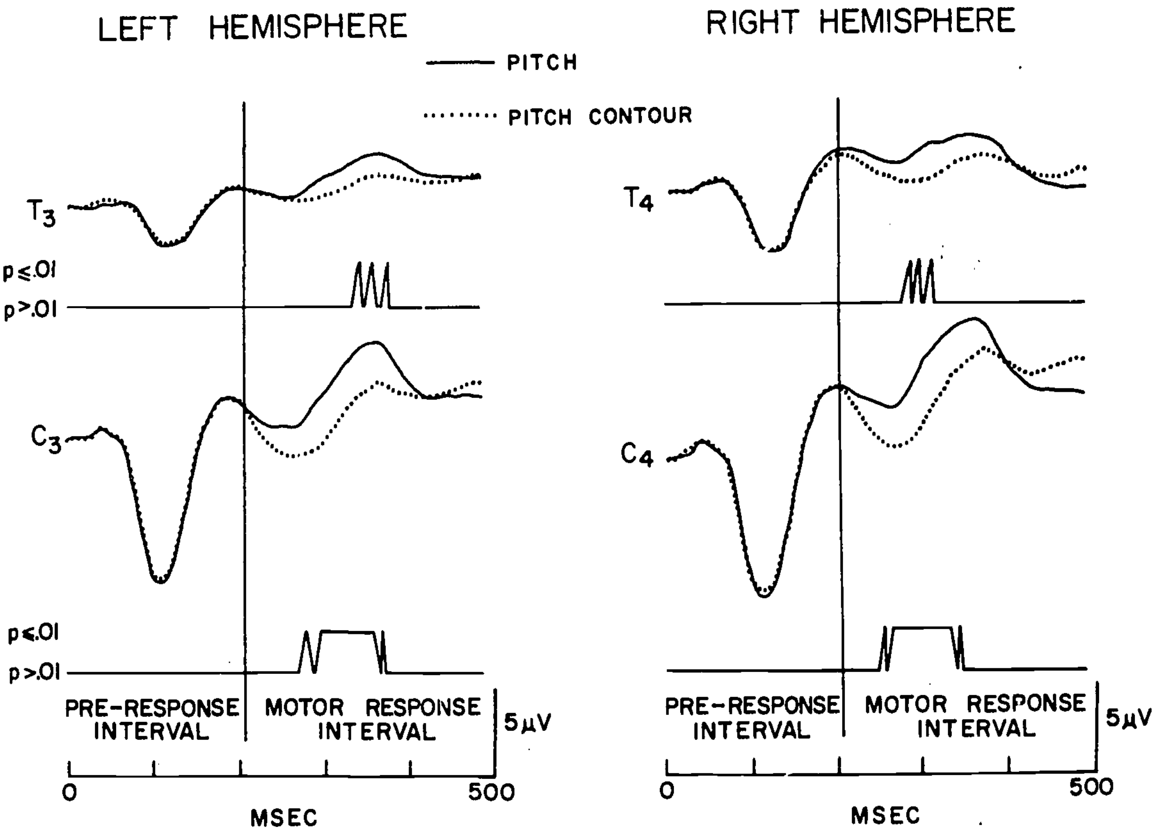


Figure 12: Average evoked potentials during identification of Pitch Contour and Pitch in Experiment 4.

their button-press responses, these could reflect differences between dimensions either in perceptual processes or in RT, or both. This ambiguity is precisely the reason that Wood et al. (1971) noted that evoked potential differences could not be attributed to perceptual variables in cases where RT differences were also obtained.

Evidence from other experiments suggests that the differences between Pitch and Pitch Contour in Figure 12 may be related to differences in RT. Using a simple RT task, Bostock and Jarvis (1970) obtained averaged evoked potentials separately for the fastest, middle, and slowest thirds of the RT distribution. The resulting evoked potentials showed significant differences as a function of RT, and the form of the differences closely paralleled those shown in Figure 12. Evoked potentials from the shortest RT trials were positive in polarity relative to those from longer RT trials, in the same 200-400 msec latency range as the differences in Figure 12. Thus, differences very similar to those between Pitch and Pitch Contour in Figure 12 can be obtained as a function of RT differences alone, without differences in the perceptual task.

In summary, neither perceptual or motor response variables can be eliminated as possible sources of the differences between Pitch and Pitch Contour in Figure 12. However, it is likely that motor response and RT factors were involved. In any case, regardless of the source of the evoked potential differences in this experiment, they do not parallel those in the pre-response interval between Place and Pitch in Experiment 1. Therefore, the RT and evoked potential data from Experiment 4 are consistent in suggesting that the identification of Pitch Contour does not require the additional level of processing required for the Place dimension in Experiment 1.

ADDITIONAL ANALYSES OF NEURAL ACTIVITY DURING THE PROCESSING OF AUDITORY AND PHONETIC DIMENSIONS

The technique of signal averaging used to obtain the evoked potential data presented above is only one of a number of methods for investigating possible differences in neural activity between two processing tasks. This particular technique was selected for the initial evoked potential experiment (Wood et al., 1971) because it is designed to increase the signal-to-noise ratio of activity synchronized to the onset of the acoustic stimuli relative to the "noise" of the background EEG. For general discussions of signal averaging in relation to evoked potentials, see Geisler (1960), Ruchkin (1965), Vaughan (1966), Perry and Childers (1969), and Regan (1972).

However, it is possible that the averaging procedure is neither the most sensitive nor the most appropriate method for the analysis of differences in neural activity between tasks. This possibility may be related to any or all of the following consequences of the signal averaging procedure as used in the present investigation: 1) it automatically eliminates consideration of the background EEG as a measure of possible differences in neural activity between tasks; 2) it precludes analysis of intervals in time other than the 490 msec sampling epoch immediately following stimulus onset; and 3) by synchronizing on the onset of the acoustic stimuli, this procedure implicitly assumes that the neural events of interest are in fact synchronized to stimulus onset. The fact that the averaging procedure actually resolved differences between the Place and Pitch tasks in Experiment 1 and the experiment of Wood et al. (1971) is evidence of its sensitivity and appropriateness under these conditions. However, the success of the averaging procedure does not eliminate the possibility that other methods might be equally or more successful.

In addition to the possibility that differences in neural activity between the Place and Pitch tasks might be reflected in measures other than the averaged evoked potential, a more serious possibility should be considered. The obtained differences in evoked potentials could have been indirect effects of differences in other forms of neural acitvity between the Place and Pitch tasks. A brief example will serve to illustrate this possibility.

A number of experiments have reported that evoked potential amplitude is enhanced under conditions related to psychological variables such as "attention," "stimulus significance," "task relevance," etc. This enhancement effect has been predominantly observed in a component of the evoked potential with positive polarity and a latency of approximately 300 msec, and has therefore been referred to as the "P300" effect (see discussions by Hillyard, Squires, Bauer, and Lindsay, 1971; Ritter, Simson, and Vaughan, 1972; Squires, Hillyard, and Lindsay, 1973). Donchin and Smith (1970) and Karlin (1970) have pointed out that the "P300" effect could in principle be produced by the offset of direct-current potentials (the "contingent negative variation" or CNV; Tecce, 1972) which are known to occur in time before stimulus onset in these tasks; that is, by activity outside the evoked potential averaging epoch. If this suggestion were correct, then the "P300" enhancement effect would be an electrical artifact instead of an enhancement of neural activity related to the perceptual processing of stimulus information.

In an explicit analysis of this question, Donald and Goff (1971) attempted to determine whether the "P300" effect could be completely accounted for by prestimulus CNV. Their results showed that although evoked potential amplitude changes covaried with CNV amplitude as suggested by Donchin and Smith (1970) and Karlin (1970), the "P300" enhancement effect was still obtained when amplitude differences in the CNV were statistically eliminated. Thus, although part of the evoked potential variance was associated with the CNV, a direct evoked potential effect was obtained as well.

The analyses reported in the present section were designed to investigate the two possible relations between the evoked potential results of Experiment 1 and other measures of neural activity: 1) that the evoked potential differences between Place and Pitch might be indirect effects of differences in other measures of neural activity; or 2) that these differences might be direct effects of the Place-Pitch manipulation but might also be accompanied by parallel differences in other measures of neural activity as well. It should be emphasized that in either case the conclusion that different neural activity occurred during the identification of Place and Pitch would still be completely valid. The presence of the evoked potential differences in Experiment 1 for Place and their absence in Experiment 2 for Intensity rule out the possibility that they were associated with logical or experimental artifacts. Thus, the present analyses attempted to determine whether the differences in neural activity during phonetic and auditory processing tasks are evoked potential differences per se, or are also present in other measures of neural activity.

The four analyses in this section were based on the tape recorded EEG for the same blocks of trials in the control condition used to obtain the evoked potentials for Place and Pitch of Experiment 1 (Figure 4). This procedure provides a strong test of the possibility that other changes in neural activity were associated with the obtained evoked potential differences, since these additional analyses were computed from precisely the same raw data as were the evoked potentials. Therefore, if differences in the background EEG, for example, were associated with the obtained differences in evoked potentials, then such changes should be clearly evident in an explicit analysis of the EEG from the same block of trials.

The four analyses described below certainly do not exhaust all possible ways of comparing neural activity between the Place and Pitch tasks. However, they do represent measures of neural activity that have previously been shown to be sensitive to various perceptual variables, and to be possible sources of indirect changes in evoked potentials. The first two analyses considered possible differences in the background EEG from which the evoked potentials were extracted by signal averaging: 1) an analysis of the component frequencies of the EEG using spectral analysis techniques, and 2) an analysis of the amplitude distribution of the EEG using goodness-of-fit tests of EEG amplitude histograms. The third and fourth analyses, like the evoked potential data presented above, were based on signal averaging techniques: 3) averaging over the entire 5-sec intertrial interval between successive stimuli to detect possible baseline differences related to the CNV, and 4) synchronizing the averaging process on the subjects' button-press response instead of stimulus onset. Because of the magnitude of the computations required for most of these analyses, only one of the two electrode locations over each hemisphere was selected for analysis (C3 and C4).

Spectral Analysis of Background EEG

The first method of investigating possible differences in background EEG between the Place and Pitch tasks was to decompose the EEG into its frequency components by the use of spectral analysis techniques. General discussions of spectral analysis techniques are given by Blackman and Tukey (1958), Bendat and Piersol (1966), and Jenkins and Watts (1968), while specific applications to EEG data are presented by Walter (1963); Walter and Adey (1963); Walter, Rhodes, Brown, and Adey (1966); and Dumermuth, Walz, Scollo-Lavizzari, and Kleiner (1972). The resulting frequency spectra plot relative energy in the EEG signal as a function of frequency. The EEG from the Place and Pitch tasks in Experiment 1 was submitted to spectral analysis and the resulting spectra were compared statistically to determine whether there were significant differences between tasks associated with the significant evoked potential differences presented above.

Method. For each subject an average frequency spectrum was computed for the Place and Pitch tasks at electrode locations over the left (C3) and right hemisphere (C4). The EEG recorded on magnetic tape during each session was played into the analog-to-digital converter of a PDP-12 computer, which simultaneously digitized the analog EEG signals from both locations at a rate of 1,024 samples per second (9.73 msec per point) and stored the values on digital magnetic tape. Each block of 64 trials lasted approximately 5.33 minutes, resulting in 128 256-sample magnetic tape blocks for each block of 64 trials. To avoid aliasing (Blackman and Tukey, 1958; Bendat and Piersol, 1966; Jenkins and Watts, 1968), each channel of EEG was low-pass filtered at 40 Hz before input to the analog-to-digital converters by a Krohn-Hite Model 3322 two-channel variable filter (attenuation of 24 db per octave). Following analog-to-digital conversion of the EEG for each block of trials, the frequency spectrum for that block was computed using a Fast Fourier Transform (Cooley and Tukey, 1965), implemented with a modified version of programs developed by J. S. Bryan (DECUS No. L-25). The spectra computed in this manner covered the frequency range from DC to 40 Hz in 0.2 Hz intervals. Each resulting spectrum was stored on digital magnetic tape for later statistical analysis.

Results. Examples of frequency spectra from a single subject during the Place task are shown in Figure 13 for the left hemisphere (C3) and Figure 14 for the right hemisphere (C4). The shaded spectrum at the bottom of each figure is the average for the Place task for that subject which was entered into the across-subject statistical analysis. Above these averages in each figure are three-dimensional plots of individual spectra computed over successive 20-sec epochs during the two blocks of trials in the Place task. For purposes of visual presentation, the individual spectra were smoothed prior to plotting by a three-point smoothing algorithm. However, the average spectra shown at the bottom of each figure were not smoothed prior to plotting. Therefore, the spectra for each subject entered into the statistical analysis were the actual computed spectra without transformations of any kind. In Figures 13 and 14, spectra from the first block of Place trials (the initial 5.33 min in each figure) are followed immediately by those from the second block, even though these blocks did not follow each other directly during the experiment. The data are presented in this way to illustrate the overall stability of the frequency composition of the EEG signal during the two blocks of trials for each task, and to show that the average spectrum over the two blocks does no injustice to any individual spectral estimate within either block.

The across-subject averages for the Place and Pitch tasks are shown in Figure 15, for the left (C3) and right (C4) hemisphere locations. These data are directly analogous to the evoked potential data for C3 and C4 in the Place and Pitch tasks presented in Figure 4 above. In contrast to the significant differences in evoked potentials which occurred on these same blocks of trials, there were no significant differences in the EEG spectra at either location. Visual inspection indicates an almost complete overlap between the spectra for the Place and Pitch tasks throughout the frequency range. This conclusion was substantiated by the results of Wilcoxon tests, identical to those described above for the evoked potential data, which were computed at each 0.2 Hz frequency interval in the spectra. These results indicate that there were no significant differences in the background EEG between the Place and Pitch tasks, either in overall energy level, or in the energy within any specific frequency band. Thus, these data eliminate the possibility that the evoked potential differences between Place and Pitch in Experiment 1 could have been indirect effects of generalized differences in background EEG between tasks (cf. discussions by Broughton, 1969; and Regan, 1972).

Gaussian Characteristics of Background EEG: Analysis of EEG Amplitude Histograms

The results of the spectral analyses described in the previous section do not rule out the possibility that the evoked potential differences might have been accompanied by differences in other measures of background EEG. The present section considers an additional measure, the proportion of time during the Place and Pitch tasks that the EEG was distributed according to a Gaussian (normal) distribution.

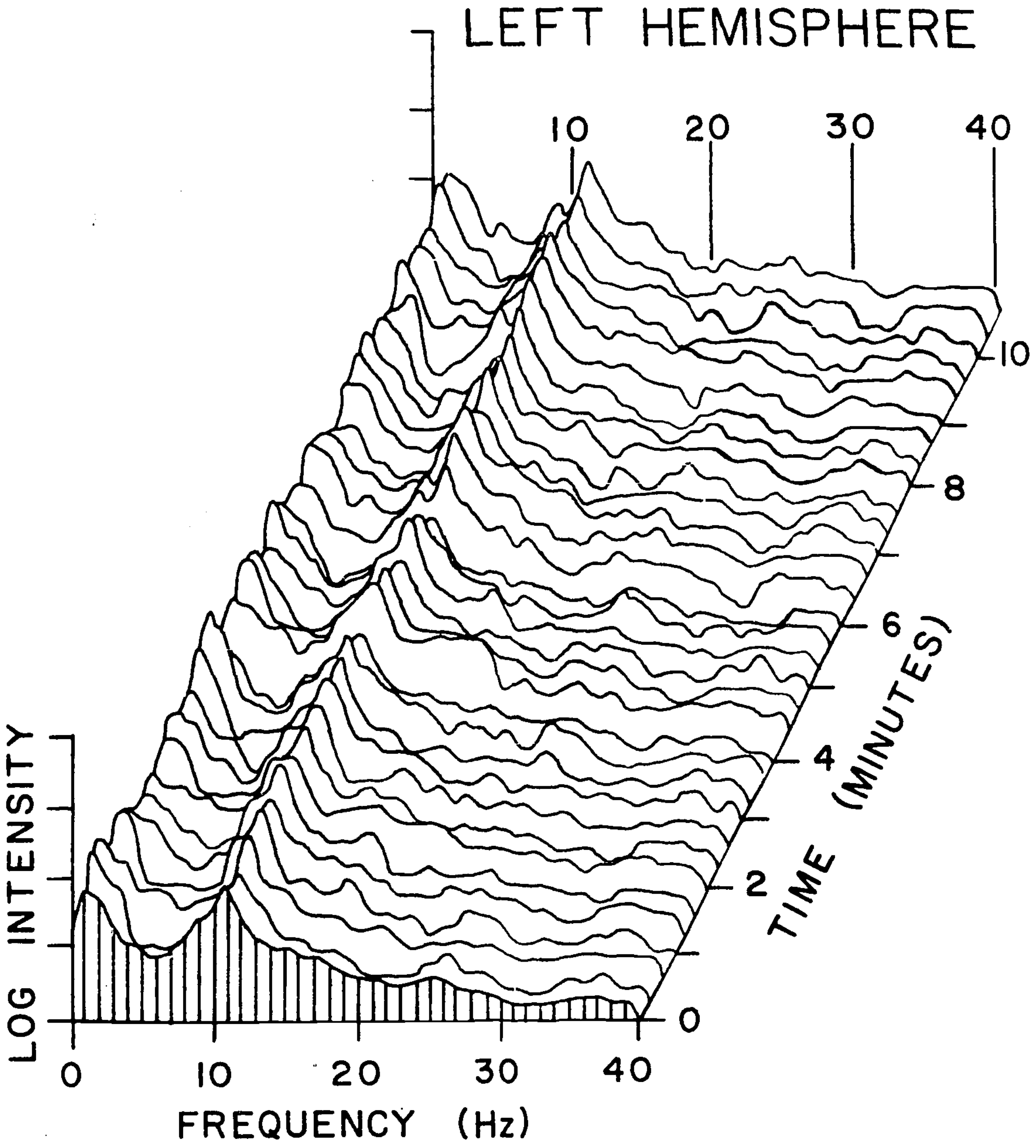


Figure 13: Frequency spectra from the left hemisphere (C3) during identification of Place.

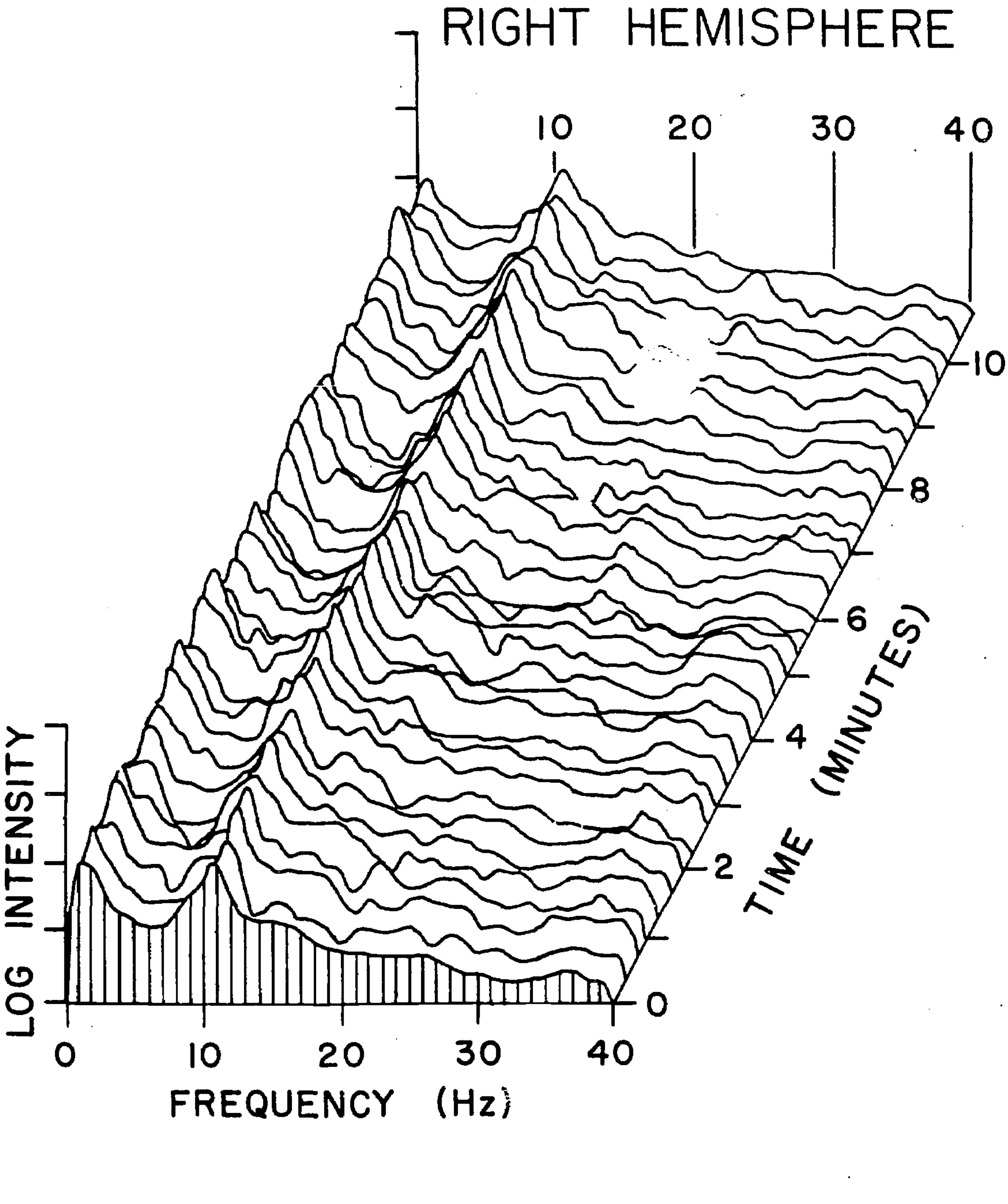


Figure 14: Frequency spectra from the right hemisphere (C4) which occurred simultaneously with those shown in Figure 13.

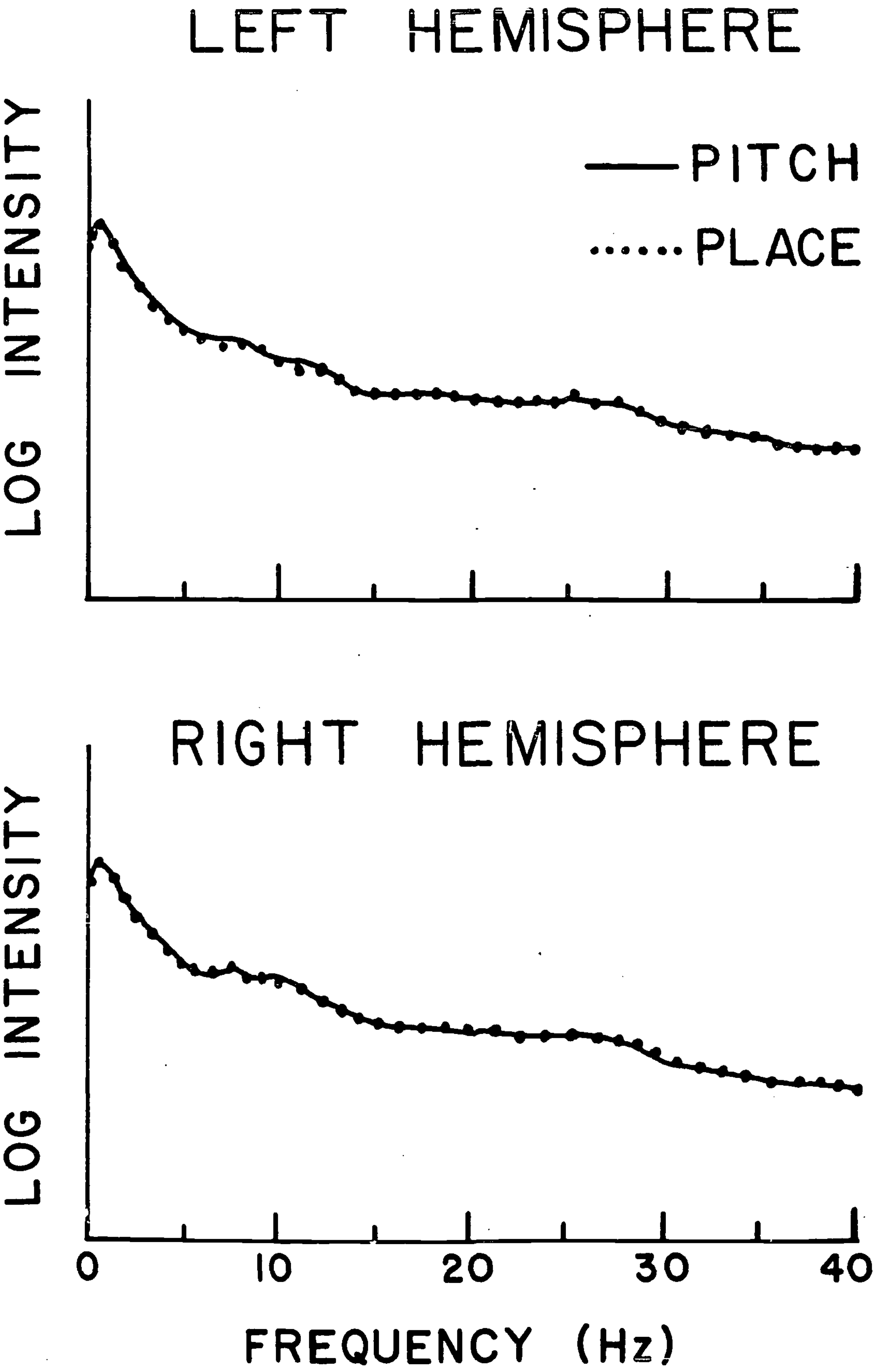


Figure 15: Average frequency spectra across subjects during identification of Place and Pitch.

This analysis was based on a series of investigations by Elul (1967, 1968, 1969), which suggested that inferences about cooperative interactions among individual neurons could be made from the amplitude distribution of the gross EEG. The logic of this inference rests upon the statistical relationship between the gross EEG recorded from the cortical surface or the scalp, and slow-wave activity recorded intra- or extracellularly from single cortical neurons (Elul, 1967, 1968, 1969). In both the cat (Elul, 1967, 1968) and the human (Saunders, 1963; Elul, 1969), the gross EEG is normally distributed, provided suitable epoch lengths are used to compute the distributions. In contrast, the distributions of slow-wave activity recorded from individual neurons are clearly not normal (Elul, 1967, 1968). Elul suggested that this relationship could be accounted for if the individual potential fields summated according to that statistical central limit theorem (cf. Cramer, 1955): "the sum of a large number of probability distributions always tends to assume a normal distribution, regardless of the nature of the component distributions, provided only that these original distributions are independent, or at least nonlinearly related, possess a mean, and a finite standard deviation...(Therefore) the EEG may be accounted for as the normally distributed output ensuing from combination of the activity of many independent (or nonlinearly related) neuronal generators" (Elul, 1968).

If the above interpretation for Gaussian characteristics of the EEG were correct, then situations in which the EEG were *not* normally distributed would indicate changes in the nature of the interaction between the individual generators. For example, Elul (1969) computed amplitude distributions of the EEG from the same electrode location on the human scalp under two conditions: 1) when the subject was resting quietly, and 2) when performing a mental arithmetic task. The EEG in the resting condition was normally distributed approximately 66 percent of the time, while during the mental arithmetic task the percentage decreased to 32 percent. Based on the interpretation of EEG distributions outlined above, Elul (1969) argued that the differences between these proportions reflected an "increase in the cooperative activity of cortical neuronal elements during performance of a mental task."

The analysis of EEG amplitude histograms to be described below was a straightforward adaptation of the experiment of Elul (1969) in order to investigate possible hemisphere differences in the background EEG during the Place and Pitch tasks. It should be noted that while there is strong evidence the postsynaptic slow-wave activity is the major neuronal source of the surface EEG (cf. Jasper and Stefanis, 1965; Creutzfeldt, Watanabe, and Lux, 1966a, 1966b; Creutzfeldt, 1970; Pollen, 1970), the inference from these data that Gaussian and non-Gaussian EEG distributions indicate the degree of "cooperative activity of cortical neuronal elements" remains largely unverified by empirical data. However, since empirical differences in EEG distributions for Place and Pitch would be important independent of Elul's interpretation, questions about the validity of the interpretation do not decrease the validity of the empirical comparison.

Method. The same digitized EEG from the Place and Pitch tasks entered into the spectral analyses in the previous section was used to compute the EEG amplitude histograms. For each subject, goodness-of-fit tests to a Gaussian distribution were computed on EEG segments of 2.5-sec duration throughout each block of trials in the Place and Pitch tasks. Measures of skewness and kurtosis were computed on each 2.5-sec segment of EEG, resulting in 256 tests for skewness and 256 for kurtosis on each electrode location in each task. The null hypothesis of a Gaussian distribution was rejected whenever $P < .05$ on either measure. Skewness was evaluated using the Pearson B_1 statistic, where: $B_1 = \mu_3^2/\mu_2^3$, and

$\mu_k = \Sigma(x_i - \bar{x})^k/N$. Similarly, kurtosis was evaluated using the Pearson B_2 statistic, where: $B_2 = \mu_4/\mu_2^2$, and μ_k is defined as above.

Results. The results of the goodness-of-fit tests are shown in Table 6, which presents for each subject the percentage of the 156 2.5-sec EEG segments in each task in which the null hypothesis of a Gaussian distribution was rejected. As shown in the means at bottom of the table, the Gaussian assumption was rejected in approximately half the segments in both the Place and Pitch tasks at both left and right hemisphere locations. The difference between Place and Pitch tasks failed to reach statistical significance at either the left or right hemisphere locations ($P > .10$ Wilcoxon tests). Thus, like the results of the spectral analyses presented above, these data suggest that the evoked potential differences between Place and Pitch in Experiment 1 were not associated with differences in the background EEG.

Signal Averaging of Activity During the Intertrial Interval Between Successive Stimuli

The possibility was raised above that the evoked potential differences between Place and Pitch in Experiment 1 could have been associated with prestimulus CNV differences between the two tasks. The CNV ("contingent negative variation," Walker, Cooper, Aldridge, McCallum, and Winter, 1964) is a prolonged surface-negative baseline shift in the EEG which, under certain conditions, preceeds a stimulus that requires a response by the subject (see reviews by Cohen, 1969, and Tecce, 1972). Typically, CNV experiments have employed one stimulus as a "warning stimulus" which is followed by a "task stimulus" after an interval of usually 2-4 sec. The CNV baseline shift develops during the interval between the "warning" and "task" stimuli, and has been related to concepts such as "expectancy," "anticipation," and "preparation" (see Walter et al., 1964; Cohen, 1969; McAdam, Knott, and Rebert, 1969; Donald, 1968, 1970; for a discussion of CNV in paradigms more familiar to cognitive psychologists, see Posner and Boies, 1971; Posner, Klein, Summers, and Baggie, 1973).

However, the narrowly defined "warning stimulus--task stimulus" paradigm commonly used to study CNV is not the only situation in which baseline shifts similar to the CNV may be obtained. A number of authors have noted that the most important prerequisites for obtaining CNV are: 1) that the "task stimulus" require some processing or response by the subject; and 2) that the "task stimulus" be predictable in time from some preceding event (cf. Naatanen, 1967, 1969; Cohen, 1969; Karlin, 1970; Donchin and Smith, 1970; Donald and Goff, 1971). Therefore, tasks in which successive stimuli are presented at fixed intertrial intervals are, in principle, sufficient for the development of CNV during the intertrial intervals. In this situation, each stimulus in the sequence would serve both as the "task stimulus" requiring a response by the subject, and also as the "warning stimulus" for the next stimulus in the sequence (cf. Naatanen, 1967, 1969; Donchin and Smith, 1970).

Thus, the identification tasks used in the present investigation could have conceivably produced prestimulus baseline shifts in the EEG similar to the CNV. In addition, as described in the introduction of this section above, it is possible for apparent evoked potential differences to be associated with prestimulus differences in CNV (Naatanen, 1967, 1969; Karlin, 1970; Donchin and Smith, 1970; Donald and Goff, 1971). Therefore, the present analysis investigated the possibility that the evoked potential differences between Place and Pitch in Experiment 1 were associated with differences in CNV between the two tasks.

TABLE 6: Percentage of EEG segments in which the Gaussian hypothesis was rejected.

	Left Hemisphere (C)		Right Hemisphere (C4)	
Subject	Place	Pitch	Place	Pitch
1	52.0	52.3	52.0	55.7
2	51.2	39.1	49.2	36.5
3	69.9	69.9	67.9	67.2
4	46.5	42.4	65.3	55.8
5	36.3	35.9	35.2	37.7
6	41.1	42.8	38.5	46.1
7	45.7	44.9	48.0	50.0
8	48.9	50.2	49.8	41.4
9	46.5	67.1	44.5	66.8
10	67.6	39.2	38.4	66.0
11	40.4	43.4	44.7	45.3
12	45.9	40.4	38.1	36.1
Mean	49.3	47.3	47.6	50.4

Method. The EEG from the control condition for Place and Pitch was averaged in a manner similar to that used to obtain the evoked potential data presented above. However, instead of averaging only 490 msec following each stimulus, the entire interstimulus interval between successive stimuli was averaged in order to observe possible prestimulus baseline shifts. Since the intertrial interval between stimuli was 5 sec, the sampling epoch for this analysis was made to be 6,144 msec, therefore including 2 successive stimulus presentations plus the entire intertrial interval. Thus, for a given subject trials 1 and 2 were included in the first epoch, trials 3 and 4 in the second epoch, and so on. Since each subject received 2 blocks of 64 trials in the control condition for each dimension, this procedure resulted in an average of 64 2-stimulus trials in the Place and Pitch tasks for each subject.

The four channels of EEG (T3, C3, C4, and T4) were played from the Honeywell tape recorder into the analog-to-digital converter of a PDP-12 computer. The pulses on a separate tape channel synchronized to stimulus onset were used to trigger the averaging process. When a trigger occurred on tape the computer began to sample the four EEG channels simultaneously at a rate of 24 msec per point for 256 successive points, resulting in epochs of 6,144 msec. The averages were stored on magnetic tape and were later analyzed using the Wilcoxon statistical procedure used previously for the evoked potential data.

Results. Figure 16 presents the averaged activity during the intertrial intervals of the Place and Pitch tasks at the four electrode locations. Note that two stimulus presentations are included in each trace. One presentation occurred at the beginning of the traces where sampling was initiated, and the second presentation occurred near the end of the trace following the 5-sec intertrial interval. The results of the Wilcoxon analyses for these data indicated that no more significant points than would be expected by chance alone occurred at any electrode location, despite the apparent differences between Place and Pitch at left hemisphere locations during the "evoked potential" portions of each trace. While 2.56 significant points would be expected to occur in each trace by chance alone, 4 and 5 significants were obtained at T3 and C3, respectively, and 2 and 1 significant points were obtained at C4 and T4.

It is interesting to note the time intervals in which the significant points occurred, even though they were not sufficient in number for rejection of the null hypothesis. At the two left hemisphere locations (T3 and C3), eight of the nine significant points were clustered in each trace at approximately 100-200 msec following the onset of the acoustic stimuli; that is, in the same time interval in which significant differences were obtained in the 490 msec sampling epochs (Figure 4). In contrast, the significant points at right hemisphere locations (T4 and C4) were apparently random with respect to stimuli in each trace.

These data indicate that the evoked potential differences between Place and Pitch in Experiment 1 were not accompanied by differences between tasks during the intertrial interval. Therefore, the evoked potential differences could not have been produced by an indirect influence of the CNV (Naatanen, 1967, 1969; Karlin, 1970; Donchin and Smith, 1970; Donald and Goff, 1971).

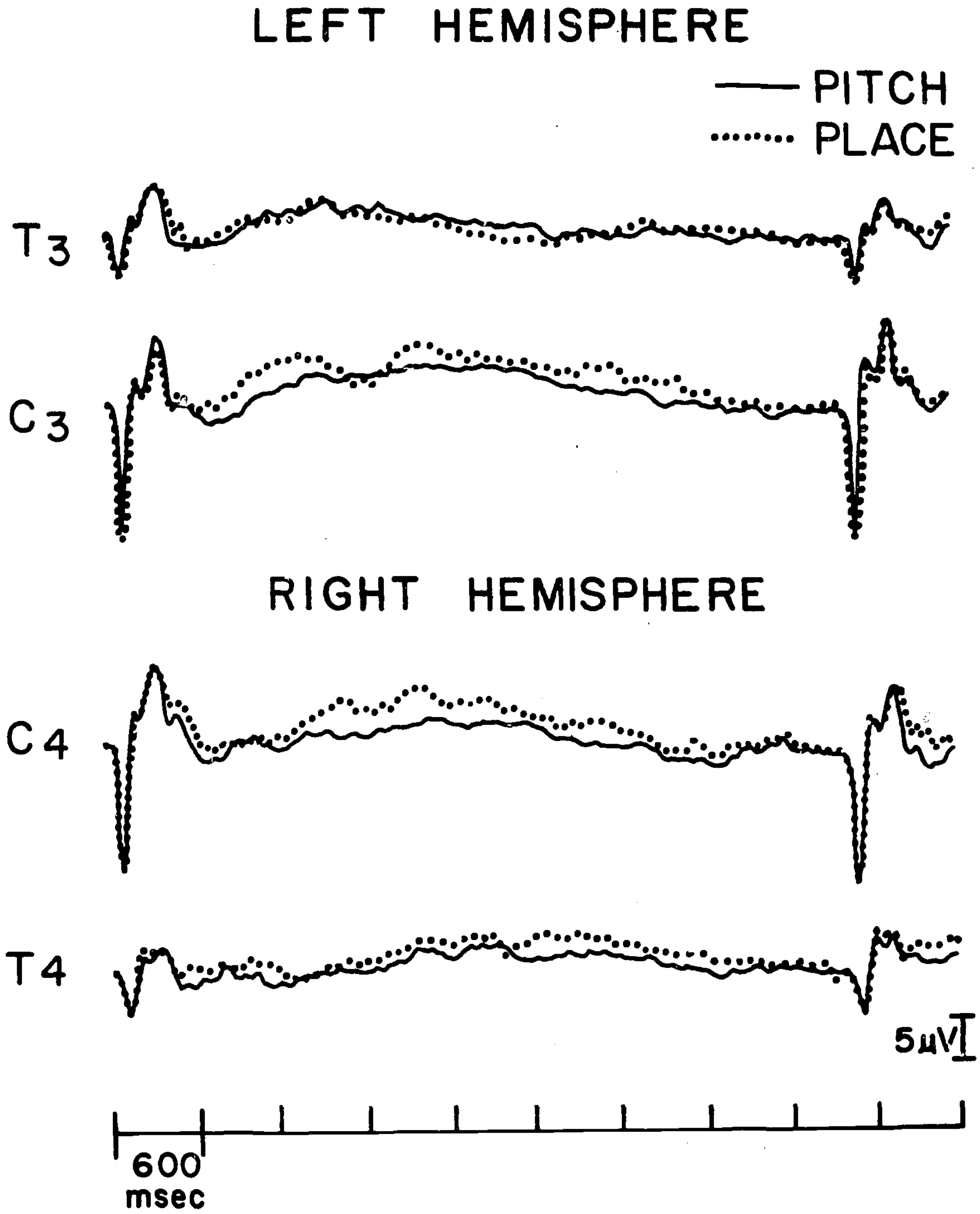


Figure 16: Averaged activity during the intertrial intervals between successive stimuli.

Signal Averaging Synchronized to Subjects' Button-Press Responses Instead of Stimulus Onset

The signal averaging analysis in both the 490 msec and 6,144 msec sampling epochs was based on the implicit assumption that differences in neural activity between Place and Pitch would be synchronized to the onset of the acoustic stimuli. The fact that significant differences between dimensions were obtained by synchronizing upon stimulus onset indicates that this assumption was valid at least to some degree. However, it is possible that the differences between Place and Pitch might be better synchronized to the "end" of perceptual processing rather than the beginning as implicitly assumed by the previous analyses. The present analysis investigated this possibility by synchronizing the signal averaging process to subjects' button-press responses.

A second reason for this analysis was to investigate more directly the possibility that the evoked potential differences between Place and Pitch might have been indirect effects of differences in the motor response between tasks. As pointed out by Wood et al. (1971), both the actual button-press responses and RT were nominally the same in the Place and Pitch tasks: the same buttons were equally distributed across stimuli in both tasks, and there were no significant differences in RT between tasks. In addition, Wood et al. (1971) showed that partitioning the evoked potential data into blocks for fast versus slow RT did not produce evoked potential differences in the preresponse interval similar to those obtained when the data were partitioned into blocks for Place and Pitch.

Despite such indirect evidence against an explanation of the evoked potential results in motor response terms, it is still possible for such an explanation to be correct. For example, slight differences in the degree of pressure exerted for "identical" button-press responses in the two tasks might have been sufficient to produce the obtained differences in evoked potentials. Therefore, the present analysis investigated this possibility more directly, by determining whether neural activity synchronized to the button-press response was significantly different in the two tasks.

Method. The same raw EEG from the control conditions for Place and Pitch used in previous analyses was played from the Honeywell tape recorder into the analog-to-digital converter of a LINC computer. Two electrode locations were analyzed, one over the left hemisphere (C3) and one over the right hemisphere (C4). The LINC was programmed to sample and average 1.28 sec before and after each trigger pulse at a sampling rate of 10 msec per point. The pulses used to trigger the LINC were those which had been generated simultaneously with subjects' button-press responses and recorded on a separate tape channel during each task. For each subject, this procedure resulted in an average of 128 trials in the control conditions for Place and Pitch at each electrode location, corresponding to the evoked potential data from Experiment 1 shown in Figure 4. These data were then entered into the Wilcoxon statistical analysis to determine whether there were significant differences between the Place and Pitch tasks.

Results. The across-subject averages for these data are shown in Figure 17. Note that in contrast to previous figures, the point of synchronization in Figure 17 is the middle of the time scale (shown by the arrow), so that activity both before and after the button-press response is shown. Visual inspection of these data suggests that there were no differences between the Place and Pitch tasks. This conclusion was substantiated by the results of the Wilcoxon analysis

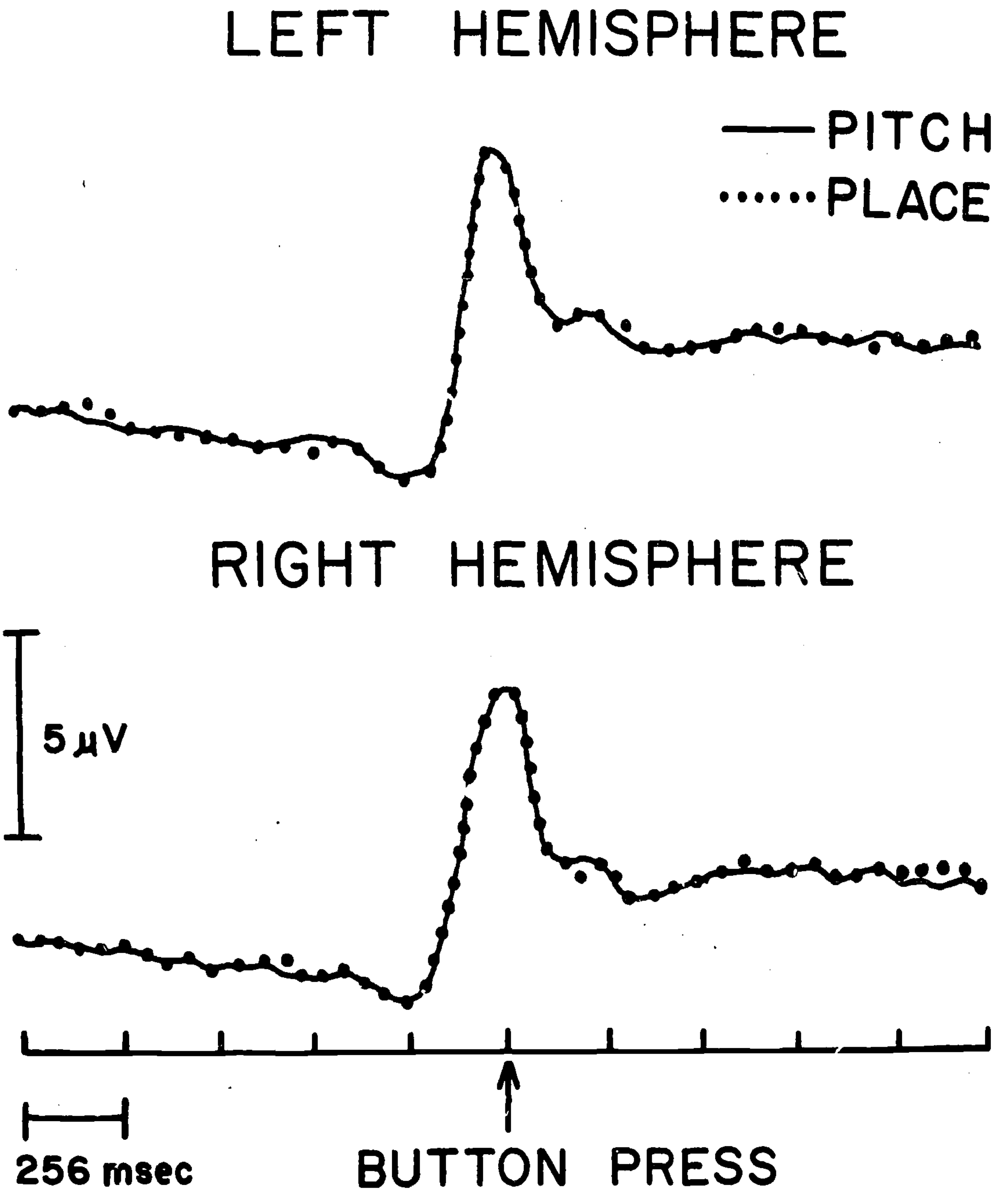


Figure 17: Averaged activity synchronized to subjects' button-press responses.

which showed that no more significant points than would be expected by chance alone occurred at either location. Two significant points were obtained at C3 and zero significant points were obtained at C4. These results indicate that possible differences in activity produced by the button-press response could not have produced the significant differences between Place and Pitch in the pre-response interval in Experiment 1. Moreover, they indicate that the neural activity associated with phonetic processing was better synchronized to stimulus onset than the button-press response.

GENERAL DISCUSSION

Summary of Experiments

The present investigation had three major goals: 1) to specify in greater detail the nature of the acoustic stimuli and processing tasks responsible for the RT and evoked potential differences between auditory and phonetic dimensions obtained by Day and Wood (1971a) and Wood et al. (1971); 2) to make a stronger test of the convergence of the RT and evoked potential measures upon the distinction between auditory and phonetic levels of processing; and 3) to obtain additional information about the neurophysiological characteristics of the neural activity associated with phonetic processing. The results of the four main experiments and the additional neurophysiological analyses will be summarized briefly in the context of these goals.

Experiment 1 was a replication of the separate initial experiments for Place and Pitch in a single experiment. Both the RT and evoked potential results of Experiment 1 confirmed those obtained in the initial experiments. These results provide strong support for the conclusion that differences in both response measures reflect different levels of processing required for the Place and Pitch dimensions. In addition, the results of Experiment 1 lend further generality to those of the initial experiments since it employed two-formant instead of three-formant synthetic syllables, a different level on the Place dimension, and a different vowel context.

Experiment 2 was a control experiment designed to insure that the differences between Place and Pitch in Experiment 1 were due to different levels of processing for the auditory and phonetic dimensions. If this were the case, then two auditory dimensions should not have produced the RT and evoked potential differences attributed to the phonetic level of processing. All conditions of Experiment 2 were identical to those of Experiment 1, except that a second auditory dimension, Intensity, was substituted for the phonetic dimension Place. In contrast to the RT and evoked potential differences between Place and Pitch, there were no differences between Intensity and Pitch in either response measure. These results indicate that the differences between Place and Pitch were not artifacts of some aspect of the experimental design or measurement techniques, since these were identical in both experiments.

Thus, Experiments 1 and 2 provide a firm empirical basis for the conclusion that the differences in RT and evoked potentials between Place and Pitch reflect different levels of processing involved in the identification of the two dimensions. Characteristics of the Place dimension which might have been responsible for the differential results of Experiments 1 and 2 were investigated in Experiments 3 and 4.

Experiment 3 investigated the acoustic cue for the Place distinction, the F2 transition, in isolation rather than in phonetic context as in Experiment 1. The stimuli for Experiment 3 were those used in Experiment 1, but with all identical portions of the stimuli /bae/ and /gae/ eliminated. Therefore, all acoustic information that could have been used to distinguish between /bae/ and /gae/ in Experiment 1 was also present in the stimuli for Experiment 3, with only the redundant phonetic context removed. The RT and evoked potential data for Experiment 3 were identical to those for Pitch and Intensity in Experiment 2: there were no differences between Pitch and the F2 transitions in either response measure. These results suggest that it was not the processing of the F2 transitions per se that produced the differences between Place and Pitch, since the same F2 transitions had to be processed in Experiment 3. Rather, the important difference between Experiments 1 and 3 was that in Experiment 1 the F2 transitions occurred in phonetic context and cued a phonetic distinction, while in Experiment 3 they occurred in isolation and did not cue a phonetic distinction. Therefore, the phonetic level of processing appears to be specialized for the extraction of phonetic information and not for the processing of particular acoustic events in the speech signal.

Experiment 4 compared Pitch with Pitch Contour, a dimension which is basically auditory like Pitch and Intensity, but which is the cue for a linguistic distinction under some conditions. The RT and evoked potential results of this experiment were more comparable to those for auditory dimensions in Experiments 2 and 3 than for Place in Experiment 1. Therefore, the mere fact that an acoustic parameter is known to be the cue for a linguistic distinction under certain conditions does not necessarily imply that the perception of this parameter requires processing in addition to the auditory level.

The EEG from which the evoked potentials for Place and Pitch in Experiment 1 were averaged was subjected to additional analyses in order to determine: 1) whether the evoked potential differences were indirect effects of changes in other measures of neural activity; or 2) whether these were valid differences but were accompanied by parallel differences in other measures of neural activity. The results of all four additional analyses were consistent. There were no differences between the Place and Pitch tasks in background EEG (assessed by two independent techniques), no differences in averaged activity during the intertrial intervals, and no differences in activity synchronized to the subjects' button-press response. These results suggest that the differences in neural activity between auditory and phonetic dimensions are limited to the actual interval during which the phonetic processing occurs, and that this activity is better synchronized to the onset of a speech stimulus than to a motor response about that stimulus.

Relative Discriminability of the Stimulus Dimensions in Each Experiment

The principal difference between the RT data of Experiment 1 and Experiments 2-4 was that in the latter there were mutual interference effects between Pitch and the second dimension in each experiment. In contrast, the interference between Pitch and Place in Experiment 1 was unidirectional in that irrelevant variation in Pitch significantly interfered with the identification of Place but not the reverse. This unidirectional interference effect has been attributed to the hypothesis that identification of Place involved an additional level of processing over that required for the identification of Pitch. However, an alternative factor should be considered: namely, the relative discriminability of the stimulus dimensions in each experiment.

Imai and Garner (1965) showed that speed of card-sorting was highly dependent upon the discriminability of the stimulus dimension used to define the sorting categories. Discriminability was varied by varying the physical distance between the two levels on a given dimension, and the effects were measured in a two-category card-sorting task. Sorting times were found to decrease significantly with increases in discriminability, until an asymptotic minimum was reached. Further increases in discriminability beyond the value necessary to reach this asymptote would therefore have little effect on sorting speed. The same general relation between speed and discriminability obtained by Imai and Garner (1965) would be expected to occur in discrete RT tasks such as those used in the present investigation (cf. Biederman and Checkosky, 1970; Biederman and Kaplan, 1970; Well, 1971; Biederman, 1972). Therefore, it is important to ask whether the failure of Place to interfere with Pitch in Experiment 1 could have been due to differences in discriminability rather than subjects' ability to process Pitch selectively without processing Place as suggested above.

First, it should be pointed out that the particular levels on the dimensions in all four experiments were selected a priori to be as equal in discriminability as possible. The absence of significant differences between the control RTs within the first three experiments, and the marked similarity of control RTs across the four experiments suggest that this attempt was successful. However, this evidence is not conclusive, since it would be possible for two dimensions to have equal control RTs and yet differ in discriminability. This could have happened in Experiment 1, for example, if the Pitch dimension were so highly discriminable that it fell on the asymptotic minimum of the hypothetical RT-discriminability function (Imai and Garner, 1965). In this case irrelevant variations in Place could have interfered with the processing of Pitch, but not enough to alter the RT for Pitch identification.

Two kinds of evidence suggest that RT for Pitch was not at an asymptotic minimum in Experiment 1. 1) The same levels on the Pitch dimension used in Experiment 1 were also used in Experiments 2-4, in which there were large differences in RT for Pitch between the control and orthogonal conditions. Since Pitch discriminability was identical in all four experiments, differences in their results cannot be attributed to differences in Pitch discriminability across experiments. 2) Direct evidence that Pitch was not an asymptotic minimum was obtained in an experiment by Wood (1973), using the same acoustic stimuli and the same control and orthogonal conditions used in Experiment 1 above. The results from this part of the experiment were identical to those for Place and Pitch in Experiment 1, despite the fact that control RTs averaged 25 msec faster. In addition, a third condition was employed in which the levels on the Place and Pitch dimensions were completely correlated. In this correlated condition, RTs for both Place and Pitch were significantly <u>faster</u> than the control conditions by more than 40 msec. These results clearly indicate that RT for Pitch was not at an asymptotic minimum in Experiment 1, and therefore could not have produced the unidirectional interference between Place and Pitch.

A second way that relative discriminability of Place and Pitch could have affected the results of Experiment 1 would be if Pitch were in the middle of the RT-discriminability function (as the data above indicate), but Place were far <u>less</u> discriminable. Again, the equal RTs for Place and Pitch provide suggestive, but not conclusive, evidence that Place and Pitch did not differ in discriminability. In addition, the experiment of Day and Wood (1972b) showed that irrelevant variations in Place produced substantial interference with identification of vowels, a dimension which was far more discriminable than Pitch in the present experiments. Therefore, neither relatively high discriminability of Pitch nor relatively low discriminability of Place could have been responsible for the unidirectional interference between these dimensions in Experiment 1.

Individual Differences

Previous speech perception experiments have obtained systematic individual differences in the degree to which subjects are perceptually "bound" by the constraints of the language (Day, 1970). That is, some subjects appear to be able to "disengage" the linguistic processing of speech sounds and process the acoustic aspects of the stimuli. Other subjects, however, have difficulty with such a task and their perception of speech stimuli appears to be governed to a much greater extent by linguistic constraints. In light of these data, it is important to consider the possible role of individual differences in the present experiments.

The occurrence of individual differences may be evaluated directly in the analysis of variance for each experiment (Tables 2-5). In all four experiments, the main effect of subjects was highly significant, indicating that individual subjects differed considerably in their baseline RTs. In Experiment 1, for example, the main effect means for subjects ranged from 337 msec to 531 msec. In addition, the interactions between subjects and the other factors in each experiment were also significant. The Subject x Condition interaction was marginally significant at the $P < .05$ level in Experiments 1 and 2 and was clearly significant in Experiments 3 and 4, while the Subject x Dimension interaction was significant in all four experiments. Finally, the Subject x Condition x Dimension interaction was marginally significant at the $P < .05$ level in Experiment 1 and significant in Experiments 2-4.

These significant interactions between subjects and the other factors in each experiment indicate that individual subjects differed both in the particular stimulus dimension identified faster in each experiment, and also in the absolute magnitude of the interference produced by each dimension. A similar pattern of individual differences was obtained by Garner and Felfoldy (1970) in card-sorting experiments with conditions analogous to the control and orthogonal conditions of the present experiments. In addition, Garner and Felfoldy (1970) found that amount of interference was not significantly correlated with the sorting speed for the single dimensions. A similar analysis was performed separately for each dimension in the four present experiments with identical results. There were no significant correlations between the RT for a given dimension in the control condition and the interference produced by that dimension when it was irrelevant in the orthogonal condition. Garner and Felfoldy (1970) noted that significant correlations between interference and base speed might have been expected if there were an identical relation between discriminability and processing speed for all subjects. The results of the present experiments support their conclusion that more complex speed-discriminability relations between subjects appear to be involved.

Finally, the degree to which subjects differed in their ability to "disengage linguistic processing (Day, 1970) in the present experiments may be evaluated in the data for Place and Pitch in Experiment 1. It should be pointed out that the overall pattern of results shown in Figure 3 clearly indicates that over the entire experiment the subjects as a group were able to identify Pitch with minimal interference from irrelevant variation in Place. However, this pattern of results in the group means does not preclude the possibility of individual differences in this pattern among subjects. The appropriate term for this question in the analysis of variance (Table 2) is the Subject x Condition x Dimension interaction, which indicates whether the 2 x 2 partition of the data shown in Figure 3 differed significantly across subjects. This interaction was barely significant at the $P < .05$ level, indicating a slight tendency toward different amounts of interference for different subjects. However, like the significant Subject x Condition x Dimension interactions in the other three experiments (Tables 3-5), this interaction mainly reflects the sensitivity of the within-subject design to differences in the magnitude of interference for different subjects. Since there was much more interference in the orthogonal condition for Place than Pitch for all 12 subjects, the pattern of results shown in Figure 3 is a valid summary of the data for any individual. The subjects in the present experiment were extremely well practiced (see Method section), and it is possible that more systematic individual differences might have been observable in earlier stages of practice. Preliminary experiments are under way to investigate this question further.

Relation of the Present Experiments to the Distinction Between Integral and Nonintegral Stimulus Dimensions

Following previous distinctions by Torgerson (1958), Attneave (1962), Shepard (1964), Lockhead (1966), and Hyman and Well (1968), Garner and Felfoldy (1970) distinguished between "integral" and "nonintegral" stimulus dimensions: "...the distinction phenomenologically being between (nonintegral) dimensions which can be pulled apart, seen as unrelated, or analyzable, and those (integral dimensions) which cannot be analyzed but are somehow perceived as single dimensions" (p. 325). The integral-nonintegral distinction resolves a number of apparently conflicting findings obtained when the particular stimulus dimensions employed in a given experiment are not taken into account (Garner, 1970).

In operational terms, integral dimensions "...produce a redundancy gain when the dimensions are correlated and some measure of speed or accuracy of discrimination is used, and produce interference in speed of classification when selective attention is required with orthogonal stimulus dimensions" (Garner and Felfoldy, 1970:328). In contrast, nonintegral dimensions produce no redundancy gain when correlated and no interference when orthogonal. The control and orthogonal conditions of the present experiments were directly analogous to two of the three conditions used by Garner and Felfoldy (1970) to distinguish empirically between integral and nonintegral dimensions. Therefore, it is of interest to compare the results for the dimensions in the present experiments with those typical of integral and nonintegral dimensions in the experiment of Garner and Felfoldy (1970).

In the discrete RT tasks of the present experiments, integral dimensions would be expected to produce increased RT in the orthogonal conditions relative to the control conditions for both dimensions. This pattern of results is exactly what was obtained in Experiments 2-4. The correspondence between the results of Experiments 2-4 in the auditory modality and the integral visual dimensions studied by Garner and Felfoldy (1970) suggests that the concept of stimulus integrality is not limited to the visual modality but is a more general characteristic of human information processing. The fact that two physical dimensions can be manipulated independently in the environment does not guarantee that they are perceptually independent (cf. Garner and Morton, 1969).

In contrast to the similarity between Experiments 2-4 and results typical of integral dimensions, the results for Place and Pitch in Experiment 1 were clearly not consistent with either the integral or nonintegral dimensions of Garner and Felfoldy (1970). Day and Wood (1972a) referred to the pattern of results for Place and Pitch as unidirectional, as opposed to mutual, interference between dimensions. These results appear to be the only case to date of unidirectional interference in speeded classification tasks of this type. However, previous experiments using different perceptual tasks have obtained results which may be related.

For example, Egeth and Pachella (1969, Experiment 4) required subjects to make absolute judgments of three dimensions (color, size, and eccentricity) of ellipses. Their results showed clear differences in the interference produced by color and the other two dimensions, just as Place and Pitch differed in this respect in the present investigation. In a condition in which all three dimensions varied orthogonally, judgment of color was unimpaired relative to a condition in which color varied alone. In contrast, judgments of both size and eccentricity were significantly impaired in the orthogonal condition relative to the corresponding single dimension conditions. By implying a directional dependence between dimensions, this lack of mutual interference between color and the other dimensions in an absolute judgment task is formally equivalent to the lack of mutual interference between dimensions in a speeded classification task.

Another result which may be related to the unidirectional interference between Place and Pitch is based on the "physical-match" versus "name-match" paradigm of Posner and Mitchell (1967; Posner, 1969). In these experiments, subjects were presented pairs of letters and had to indicate whether they were "same" or "different" as rapidly as possible on each trial. On some trials the same-different decision was to be based on the physical identity of the letters (e.g., AA), while on other trials it was to be based on their name identity (e.g., Aa), whether they were the same or different physically. Subjects took significantly longer to decide that the same stimulus pair differed in name than physical identity, leading Posner and Mitchell (1967) to conclude that an additional level of processing was required for the name match over that required for the physical match.

The data of Egeth and Pachella (1969) and Posner and Mitchell (1967), and the unidirectional interference between Place and Pitch (Day and Wood, 1972a, Experiment 1) indicate that the integral-nonintegral distinction alone is not sufficient to account for situations in which different levels of processing may be involved in the perception of different stimulus attributes. In addition to consideration of the particular dimensions in a given experiment as advocated by the integral-nonintegral distinction, consideration must also be given to the particular role in the experiment played by each dimension. As long as the results for both members of a pair of dimensions are equivalent, as in the experiments of Garner and Felfoldy (1970), there is no need to distinguish between individual members of the pair. However, when different dimensions of the same physical stimuli produce reliably different results, such a distinction is clearly required.

Converging Operations and the Concept of Specialized Neural Mechanisms for Speech Perception

Evidence from a number of sources converges upon the distinction between auditory and phonetic levels of processing, and upon the idea that the phonetic level involves specialized neural mechanisms which are lateralized in one cerebral hemisphere:

1) the right-ear advantage in dichotic listening (Studdert-Kennedy and Shankweiler, 1970; Darwin, 1971a; Haggard, 1971; and references cited therein);

2) the tendency for phoneme discrimination to be limited by linguistic categories rather than physical differences in the stimuli (i.e., "categorical perception," Liberman et al., 1967; Studdert-Kennedy et al., 1970a; Pisoni, 1971);

3) the difference in temporal-order judgment accuracy for dichotically presented linguistic and nonlinguistic dimensions (Day and Bartlett, 1971; and references cited therein);

4) the unidirectional interference between auditory and phonetic dimensions of the same speech stimuli in speeded-classification tasks (Day and Wood, 1972a; RT data of Experiment 1); and

5) the differences in neural activity over the left hemisphere during the processing of auditory and phonetic dimensions of the same speech stimuli (Wood et al., 1971; evoked potential data of Experiment 1).

The "lag effect" for the identification of temporally offset dichotic syllables (Studdert-Kennedy et al., 1970b) has been proposed by some authors as another result reflecting the operation of the phonetic level (cf. Liberman et al., 1971; Kirstein, 1971). However, while the "lag effect" is a consistent finding for speech stimuli (Berlin et al., 1970; Lowe, Cullen, Thompson, Berlin, Kirkpatrick, and Ryan, 1970; Studdert-Kennedy et al., 1970b; Darwin, 1971b), it has been obtained with nonspeech stimuli as well (Darwin, 1971b).

Recently, different investigators have used the strategy of varying the stimuli and tasks in these paradigms to determine the particular characteristics which distinguish the phonetic level of processing from the general auditory system. Although precisely the same experiments have not been conducted in all four paradigms, the pattern of results across paradigms forms a consistent, if incomplete, characterization of the phonetic level of processing.

The first important observation about the nature of the phonetic level is that not all linguistic dimensions require specialized phonetic processing for their perception. Using both natural and synthetic speech, a number of experiments have obtained results with certain linguistic dimensions which are more typical of those obtained with nonspeech. For example, while stop consonants have consistently produced evidence of phonetic level processing, steady-state vowels have not. Vowels have not produced a right-ear advantage in dichotic listening (Studdert-Kennedy and Shankweiler, 1970; Darwin, 1969, 1971a); they have not resulted in the phoneme boundary effect of categorical perception (Liberman, Harris, Hoffman, and Griffith, 1957; Fry, Abramson, Eimas, and Liberman, 1962; Stevens, Liberman, Studdert-Kennedy, and Ohman, 1969; Pisoni, 1971); and they have not produced the unidirectional interference with the auditory dimension Pitch in speeded-classification tasks (Day and Wood, in preparation). Evoked potential data for vowels corresponding to the experiments of Wood et al. (1971) and Experiment 1 for stop consonants are not available at present. It is interesting to note in this context that consonants and vowels also differ in characteristics of short-term memory (Crowder, 1971; Pisoni, 1971; Cole, 1973).

The differences between stop consonants and vowels have led to the suggestion that the phonetic level may be related to the nature of the acoustic cues which underlie the perception of these stimuli (Liberman et al., 1967). The cues for the stop consonants are highly "encoded" in the sense that information about both the consonant and adjacent vowel is transmitted simultaneously in the speech signal. Therefore, large variations in the acoustic cues for stop consonants occur commonly as a function of phonetic environment (Liberman et al., 1967). Thus, the mapping from sound to phoneme for stop consonants is both one-to-many and many-to-one. In contrast, the acoustic cues for vowels are much less dependent upon their phonetic context, and as a result the phonetic categories for vowels are more directly related to the acoustic cues. Based on these observations, Liberman et al. (1967) and Studdert-Kennedy and Shankweiler (1970), among others, have suggested that a crucial determinant for the involvement of the phonetic level may be the "encodedness" of the phonemes in a given experiment. Support for this hypothesis comes from a number of different sources in addition to the basic difference between the highly "encoded" stop consonants and the "unencoded" vowels.

Right-ear advantages can be obtained for vowels under conditions which presumably increase their "encodedness" by decreasing the one-to-one relationship between the vowel categories and their acoustic cues. Darwin (1971a) obtained a significant right-ear advantage for vowels when the stimuli varied in vocal tract size, but no ear advantage when the stimuli were from a single vocal tract. Similar results were obtained by Haggard (1971) with vowels that varied in fundamental frequency as well as vocal tract size. Spellacy and Blumstein (1970) found a significant right-ear advantage for dichotic vowels when they were embedded in a sequence which included dichotically contrasting stop consonants, but a left-ear advantage for the same vowel stimuli when embedded in a series of dichotic nonspeech stimuli. Thus, vowels can yield a right-ear advantage similar to that obtained for stop consonants, but only under conditions which appear to require more complex "decoding" than is normally required for isolated steady-state vowels.

Day and Vigorito (1973) and Cutting (1973) assumed an "encodedness continuum" in which "stop consonants appear to be the most highly encoded speech sounds, vowels the least encoded, with other sounds in the middle" (Day and Vigorito, 1973). In both temporal-order judgment (Day and Vigorito, 1973) and ear-monitoring tasks (Cutting, 1973), the magnitude of the right-ear advantage paralleled the degree of "encodedness" for stops, liquids, and vowels: stops produced large right-ear advantages, liquids produced reduced right-ear advantages, and vowels produced either no ear advantage or one in favor of the left. Darwin (1971a) found that identification of dichotic fricatives produced a right-ear advantage when cued by friction noise plus formant transitions, but did not produce a right-ear advantage when cued by the friction noise alone. Friction noises are relatively "unencoded" since they show little variation as a function of phonetic context, while the formant transitions for fricatives are "encoded" just as they are for stop consonants. Therefore, the experiment of Darwin (1971a) demonstrates that the use of the same phonemic response categories can produce different ear advantages, depending upon the "encodedness" of the acoustic cues which underlie the phonemic distinction.

All of the evidence supporting the "encodedness" hypothesis described above comes from the dichotic listening paradigm. Except for the basic difference between stop consonants and vowels, variations in "encodedness" have not been widely investigated in other paradigms. However, Experiment 4 of the present investigation provides suggestive evidence in favor of the "encodedness" hypothesis from the speeded-classification and evoked potential paradigms. The dimension compared with Pitch in this experiment was Pitch Contour, the major cue for judging whether an utterance is a statement or question (Lieberman, 1967, in press; Fry, 1968; Lehiste, 1970; Studdert-Kennedy and Hadding, 1971). In contrast to the highly "encoded" phonetic cues, the linguistic categories of question and statement correspond much more directly to particular values of the acoustic cue. Both the RT and evoked potential data for Pitch Contour indicated that the perception of Pitch Contour did not require processing in addition to the auditory level: the results were much more similar to those for the nonlinguistic dimension Intensity, than for the phonetic dimension Place. The data of Studdert-Kennedy and Hadding (1971; Hadding-Koch and Studdert-Kennedy, 1964) are also consistent with the suggestion that Pitch Contour does not require processing in addition to the auditory level. With only minor exceptions, judgments of Pitch Contour were the same when subjects were instructed to judge them on auditory grounds (falling versus rising) and on linguistic grounds (statement versus question). Moreover, Studdert-Kennedy and Hadding (1971) found that Pitch Contour was judged similarly when the Pitch changes were carried by a speech stimulus and by a pure tone.

The experiments described in the previous paragraphs suggest that the phonetic level of processing may be related to the specialized "decoding" process required for speech perception. This conclusion has been reached by a number of other investigators, and is most clearly summarized by Studdert-Kennedy and Shankweiler (1970): "...specialization of the dominant hemisphere in speech perception is due to its possession of a linguistic device, not to specialized capacities for auditory analysis... [W]hile the general auditory system common to both hemispheres is equipped to extract the auditory parameters of a speech signal, the dominant hemisphere may be specialized for the extraction of linguistic features from those parameters" (p. 579).

However, there is an alternative interpretation for most of the experiments cited above, which is equally consistent with the data. With few exceptions, the experiments which required processing of highly "encoded" cues also involved the processing of formant transitions. Therefore, it is possible that the results which have been attributed to a phonetic or linguistic decoding process were actually due to the processing of particular *acoustic* features in the stimuli like the formant transitions. The possibility that the phonetic level may be specialized for processes which are basically *auditory* rather than *linguistic* has been noted in passing by other authors (cf. Liberman, 1970; Studdert-Kennedy and Shankweiler, 1970), and has been explicitly suggested as a possible basis for speech perception by Abbs and Sussman (1971). Only recently has direct evidence regarding this possibility become available.

In a categorical perception experiment, Mattingly et al. (1971) compared the perception of the same F2 transitions under two conditions: 1) when they occurred in phonetic context and cued the phonetic distinction among /bae/, /dae/, and /gae/; and 2) when they occurred in isolation and sounded like nonspeech "chirps." If the phonetic level were specialized for the processing of formant transitions, identical results would be expected in both conditions. When the F2 transitions occurred in phonetic context, results typical of categorical perception were obtained: high discrimination peaks near the phoneme boundaries with performance near chance within the phonemic categories. In contrast, the perception of the same F2 transitions was clearly not categorical when they occurred in isolation.

A similar pattern of results for isolated F2 transitions was obtained in the speeded-classification and evoked potential paradigms in Experiment 3 of the present investigation. In both sets of data the perception of the isolated F2 transitions more closely resembled the nonspeech dimension Intensity than the same F2 transitions when they cued a phonetic distinction (Experiment 1). Taken together, the results of Mattingly et al. (1971) and Experiment 3 provide strong evidence that the phonetic level of processing is specialized for the extraction of abstract phonetic features, not for the detection of particular acoustic features which occur in speech. While a process formally resembling "feature detection" may be involved in speech perception (cf. Eimas and Corbit, 1973), the evidence presented above clearly indicates that the features must be specified in linguistic rather than auditory or acoustic terms (Mattingly et al., 1971; Studdert-Kennedy et al., 1972; Eimas and Corbit, 1973; Experiment 3).

Serial Versus Parallel Organization of Auditory and Phonetic Levels [Abridged]

The experiments described in the previous section firmly establish the distinction between auditory and phonetic levels of processing in speech perception, and provide additional information about the specialized decoding process performed by the phonetic level. However, these experiments provide relatively little information concerning the general organization of these two levels or the nature of the interaction between them. Two process models appear to be consistent with existing data: 1) a serial or sequential model, in which the auditory level would occur first in sequence followed by the phonetic level; and 2) a parallel model in which at least some portion of auditory and phonetic processing could proceed simultaneously.

Direct evidence regarding serial versus parallel organization of auditory and phonetic levels was obtained in a recent experiment by Wood (1973).* This experiment employed control and orthogonal conditions identical to those of the

---

*Editor's Note: In the original body of the thesis, the remainder of this section presented a detailed discussion of the rationale, results, and interpretation of the Wood (1973) experiment. This section has been abridged here since a complete version of the Wood (1973) experiment appears elsewhere in this issue of the Haskins Laboratories Status Report on Speech Research (Parallel Processing of Auditory and Phonetic Information in Speech Perception).

present experiments, and used the same CV syllables varying in Place and Pitch shown in Figures 1 and 2 (Experiment 1). The possibility of parallel processing of Place and Pitch was evaluated in a correlated condition in which the two dimentions varied in a completely redundant manner. The RTs in the correlated condition for both Place and Pitch were significantly faster than the corresponding single dimension control conditions. This "redundancy gain" was not attributable to speed-accuracy trades, to selective serial processing (Garner, 1969; Morton, 1969; Felfoldy and Garner, 1971), or to differential transfer between conditions. These results are consistent only with a model in which auditory and phonetic information can be processed in parallel (cf. Biederman and Checkosky, 1970; Lockhead, 1972).

However, the conclusion that auditory and phonetic processing can occur in parallel contradicts the widely held idea that linguistic processing is dependent upon preliminary processes performed by the general auditory system (cf. Stevens and House, 1970; Studdert-Kennedy and Shankweiler, 1970). As noted by Wood (1973), this idea must be correct at least to some degree, since to be perceived all acoustic signals must first be transduced by the receptor apparatus. To account for these observations, Wood (1973) presented a hypothetical organization of auditory and phonetic levels consisting of three components: 1) a common peripheral component for the transduction and preliminary analysis of all acoustic signals; 2) a "central" auditory component for the additional processing of non-linguistic auditory information; and 3) a "central" phonetic component for the extraction of abstract phonetic features from the results of the preliminary auditory analysis. The second two components would be capable of functioning in parallel, but both would be dependent upon the output of the prior peripheral processing. According to this hypothesis, the term "auditory level" as used by Studdert-Kennedy and as used in the initial portions of the present paper would actually consist of two parts, the first occurring before any phonetic processing is begun, and the second occurring simultaneously with phonetic processing. As a working hypothesis, this "hybrid" organization appears to be consistent with the evidence that distinguishes between auditory and phonetic levels of processing, and with the demonstration that processing of auditory and phonetic information can occur in parallel.

SUMMARY AND CONCLUSIONS

The RT and evoked potential data of the present experiments provide two new sources of evidence for a distinction between auditory and phonetic levels of processing in speech perception, and provide additional insight into the nature of the specialized processes performed by the phonetic level. Experiments 1 and 2 verified the conclusions of Day and Wood (1972a) and Wood et al. (1971) that the differences in RT and evoked potentials between Place and Pitch were the result of different levels of processing required for auditory and phonetic dimensions. Both sets of data suggest that identification of the phonetic dimension involved additional processing mechanisms which were not required for identification of an auditory dimension of the same physical stimuli. Experiment 3 showed that the phonetic level is specialized for the extraction of abstract phonetic features, not for the detection of particular acoustic features in the speech signal. However, while the processes performed by the phonetic level are basically linguistic in nature, Experiment 4 showed that the phonetic level is not required for the processing of all acoustic dimensions that can carry linguistic information. The additional neurophysiological analyses demonstrated that the differences in neural activity between auditory and phonetic dimensions occurred only during the actual processing of the two dimensions, and were not accompanied by more generalized differences in neural activity. Taken together, these experiments provide a strong set of converging operations upon the distinction between auditory and phonetic levels of processing in speech perception, and upon the idea that the phonetic level involves specialized linguistic mechanisms which are lateralized in one cerebral hemisphere.

REFERENCES

Abbs, J. H. and H. M. Sussman. (1971) Neurophysiological feature detectors and speech perception: A discussion of theoretical implications. J. Speech Hearing Res. 14, 23-36.

Attneave, F. (1962) Perception and related areas. In Psychology: A Study of a Science, Vol. 4, ed. by S. Koch. (New York: McGraw-Hill) 619-659.

Bendat, J. S. and A. G. Piersol. (1966) Measurement and Analysis of Random Data. (New York: Wiley).

Berlin, C. I., M. E. Willett, C. Thompson, J. K. Cullen, and S. S. Lowe. (1970) Voiceless versus voiced CV perception in dichotic and monotic listening. J. Acoust. Soc. Amer. 47, 75(A).

Biederman, I. (1972) Human performance in contingent information-processing tasks. J. Exp. Psychol. 93, 219-238.

Biederman, I. and S. F. Checkosky. (1970) Processing redundant information. J. Exp. Psychol. 83, 486-490.

Biederman, I. and R. Kaplan. (1970) Stimulus discriminability and S-R compatibility: Evidence for independent effects in choice reaction time. J. Exp. Psychol. 86, 434-439.

Blackman, R. B. and J. W. Tukey. (1958) The Measurement of Power Spectra. (New York: Dover).

Bostock, H. and M. J. Jarvis. (1970) Changes in the form of the cerebral evoked response related to the speed of simple reaction time. EEG Clin. Neurophysiol. 29, 137-145.

Broughton, R. (1969) Discussion. In Average Evoked Potentials, ed. by E. Donchin and D. B. Lindsley, NASA SP-191. (Washington, D.C.) 143-198.

Cohen, J. (1969) Very slow brain potentials relating to expectancy: The CNV. In Average Evoked Potentials, ed. by E. Donchin and D. B. Lindsley, NASA SP-191. (Washington, D.C.) 143-198.

Cole, R. A. (1973) Different memory functions for consonants and vowels. Cog. Psychol. 4, 39-54.

Cooley, J. W. and J. W. Tukey. (1965) An algorithm for the machine calculation of complex Fourier series. Mathematics of Computation 19, 297-301.

Cooper, F. S. (1971) How is language conveyed by speech? Haskins Laboratories Status Report on Speech Research SR-27, 3-21.

Cooper, F. S. and I. G. Mattingly. (1969) Computer-controlled PCM system for investigation of dichotic speech perception. Haskins Laboratories Status Report on Speech Research SR-17/18, 17-21.

Cramer, E. G. H. (1955) The Elements of Probability Theory. (New York: Wiley).

Creutzfeldt, O. D. (1970) Neuronal mechanisms underlying the EEG. In Basic Mechanisms of the Epilepsies, ed. by H. H. Jasper, A. A. Ward, Jr., and A. Pope. (Boston, Mass.: Little, Brown) 397-410.

Creutzfeldt, O. D., S. Watanabe, and H. D. Lux. (1966a) Relations between EEG phenomena and potentials of single cortical cells. I. Evoked responses after thalamic and epicortical stimulation. EEG Clin. Neurophysiol. 20, 1-18.

Creutzfeldt, O. D., S. Watanabe, and H. D. Lux. (1966b) Relations between EEG phenomena and potentials of single cortical cells. II. Spontaneous and convulsoid activity. EEG Clin. Neurophysiol. 20, 19-37.

Crowder, R. G. (1971) The sound of vowels and consonants in immediate memory. J. Verbal Learn. Verbal Behav. 10, 587-596.

Cutting, J. E. (1973) A parallel between degree of encodedness and the ear advantage: Evidence from an ear-monitoring task. J. Acoust. Soc. Amer. 53, 368(A).

Darwin, C. J. (1969) Auditory perception and cerebral dominance. Unpublished Doctoral dissertation, University of Cambridge.

Darwin, C. J. (1971a) Ear differences in the recall of fricatives and vowels. Quart. J. Exp. Psychol. 23, 46-62.

Darwin, C. J. (1971b) Dichotic backward masking of complex sounds. Quart. J. Exp. Psychol. 23, 386-392.

Day, R. S. (1970) Temporal-order judgments in speech: Are individuals language-bound or stimulus bound? Haskins Laboratories Status Report on Speech Research SR-21/22, 71-87.

Day, R. S. and J. C. Bartlett. (1972) Separate speech and nonspeech processing in dichotic listening? J. Acoust. Soc. Amer. 51, 79(A).

Day, R. S. and J. M. Vigorito. (1973) A parallel between degree of encodedness and the ear advantage: Evidence from a temporal-order judgment task. J. Acoust. Soc. Amer. 53, 368(A).

Day, R. S. and C. C. Wood. (1972a) Interactions between linguistic and nonlinguistic processing. J. Acoust. Soc. Amer. 51, 79(A).

Day, R. S. and C. C. Wood. (1972b) Mutual interference between two linguistic dimensions of the same stimuli. J. Acoust. Soc. Amer. 52, 175(A).

Day, R. S. and C. C. Wood. (in preparation) Interactions between the processing of vowels and pitch.

Delattre, P. C., A. M. Liberman, and F. S. Cooper. (1955) Acoustic loci and transitional cues for consonants. J. Acoust. Soc. Amer. 27, 769-773.

Delattre, P. C., A. M. Liberman, F. S. Cooper, and L. J. Gerstman. (1952) An experimental study of the acoustic determinants of vowel color: Observations on one- and two-formant vowels synthesized from spectrographic patterns. Word 8, 195-210.

Donald, M. W. (1968) Electrocortical correlates of fixed foreperiod decision tasks. Unpublished Doctoral dissertation, McGill University.

Donald, M. W. (1970) Direct-current potentials in the human brain evoked during timed cognitive performance. Nature 227, 1057-1058.

Donald, M. W. and W. R. Goff. (1971) Attention-related increases in cortical responsitivity dissociated from the contingent negative variation. Science 172, 1163-1166.

Donchin, E. and D. B. D. Smith. (1970) The CNV and late positive wave of the evoked potential. EEG Clin. Neurophysiol. 29, 201-203.

Dumermuth, G., W. Walz, G. Scollo-Lavizzari, and B. Kleiner. (1972) Spectral analysis of EEG activity in different sleep stages in normal adults. Europ. Neurol. 7, 265-296.

Egeth, H. and R. Pachella. (1969) Multidimensional stimulus identification. Percept. Psychophys. 5, 341-346.

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

Elul, R. (1967) Statistical mechanisms in generation of the EEG. In Progress in Biomedical Engineering, Vol. 1. (Washington, D.C.: Spartan) 131-150.

Elul, R. (1968) Brain waves: Intracellular recording and statistical analysis help clarify their physiological significance. In Data Acquisition and Processing in Biology and Medicine, Vol. 5. (New York: Pergamon) 93-115.

Elul, R. (1969) Gaussian behavior of the electroencephalogram: Changes during performance of a mental task. Science 164, 328-331.

Fant, G. (1967) Auditory patterns of speech. In Models for the Perception of Speech and Visual Form, ed. by W. Wathen-Dunn. (Cambridge, Mass.: MIT Press) 111-125.

Felfoldy, G. L. and W. R. Garner. (1971) The effects on speeded classification of implicit and explicit instructions regarding redundant dimensions. Percept. Psychophys. 9, 289-292.

Fry, D. B. (1956) Perception and recognition in speech. In For Roman Jakobson, ed. by M. Halle, H. G. Lunt, and C. H. van Schooneveld. (The Hague: Mouton) 169-173.

Fry, D. B. (1968) Prosodic phenomena. In Manual of Phonetics, ed. by B. Malmberg. (Amsterdam: North-Holland) 365-411.

Fry, D. B., A. S. Abramson, P. D. Eimas, and A. M. Liberman. (1962) The identification and discrimination of synthetic vowels. Lang. Speech 5, 171-189.

Garner, W. R. (1969) Speed of discrimination with redundant stimulus attributes. Percept. Psychophys. 6, 221-224.

Garner, W. R. (1970) The stimulus in information processing. Amer. Psychol. 25, 350-358.

Garner, W. R. and G. L. Felfoldy. (1970) Integrality of stimulus dimensions in various types of information processing. Cog. Psychol. 1, 225-241.

Garner, W. R., H. W. Hake, and C. W. Ericksen. (1956) Operationism and the concept of perception. Psychol. Rev. 63, 149-159.

Garner, W. R. and J. Morton. (1969) Perceptual independence: Definitions, models, and experimental paradigms. Psychol. Bull. 72, 233-259.

Geisler, G. D. (1960) Average responses to clicks in man recorded by scalp electrodes. MIT Research Laboratory of Electronics Tech. Rep. 380.

Geschwind, N. (1970) The organization of language and the brain. Science 170, 940-944.

Hadding-Koch, K. and M. Studdert-Kennedy. (1964) An experimental study of some intonation contours. Phonetica 11, 175-185.

Haggard, M. P. (1971) Encoding and the REA for speech signals. Quart. J. Exp. Psychol. 23, 34-45.

Harris, K. S. (1958) Cues for the discrimination of American English fricatives in spoken syllables. Lang. Speech 1, 1-7.

Hillyard, S. A., K. C. Squires, J. W. Bauer, and P. H. Lindsay. (1971) Evoked potential correlates of auditory signal detection. Science 172, 1357-1360.

Hughes, G. W. and M. Halle. (1956) Spectral properties of fricative consonants. J. Acoust. Soc. Amer. 28, 303-310.

Hyman, R. and A. Well. (1968) Perceptual separability and spatial models. Percept. Psychophys. 3, 161-165.

Imai, S. and W. R. Garner. (1965) Discriminability and preference for attributes in free and constrained classification. J. Exp. Psychol. 69, 596-608.

Jasper, H. (1958) The ten twenty electrode system of the International Federation. EEG Clin. Neurophysiol. 10, 371-375.

Jasper, H. and C. Stefanis. (1965) Intracellular oscillatory rhythms in pyramidal tract neurones in the cat. EEG Clin. Neurophysiol. 18, 541-553.

Jenkins, G. M. and D. G. Watts. (1968) Spectral Analysis and its Applications. (San Francisco, Calif.: Holden-Day).

Karlin, L. (1970) Cognition, preparation, and sensory-evoked potentials. Psychol. Bull. 73, 122-136.

Kimura, D. (1961a) Some effects of temporal lobe damage on auditory perception. Canad. J. Psychol. 15, 156-165.

Kimura, D. (1961b) Cerebral dominance and the perception of verbal stimuli. Canad. J. Psychol. 15, 166-171.

Kirstein, E. (1971) Temporal factors in perception of dichotically presented stop consonants and vowels. Doctoral dissertation, University of Connecticut. [Also in Haskins Laboratories Status Report on Speech Research SR-24 (1970).]

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

Liberman, A. M. (1970) The grammars of speech and language. Cog. Psychol. 1, 301-323.

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

Liberman, A. M., P. C. Delattre, F. S. Cooper, and L. J. Gerstman. (1954) The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychol. Monogr., Whole No. 68.

Liberman, A. M., K. S. Harris, H. S. Hoffman, and B. C. Griffith. (1957) The discrimination of speech sounds within and across phoneme boundaries. J. Exp. Psychol. 53, 358-368.

Liberman, A. M., I. G. Mattingly, and M. T. Turvey. (1971) Language codes and memory codes. Haskins Laboratories Status Report on Speech Research SR-27, 59-88.

Lieberman, P. (1967) Intonation, Perception, and Language. (Cambridge, Mass.: MIT Press).

Lieberman, P. (in press) A study of prosodic features. In Current Trends in Linguistics, Vol XII, ed. by T. A. Sebeok. (The Hague: Mouton). [Also in Haskins Laboratories Status Report on Speech Research SR-23 (1970) 15-48.]

Lockhead, G. R. (1966) Effects of dimensional redundancy on visual discrimination. J. Exp. Psychol. 72, 95-104.

Lockhead, G. R. (1972) Processing dimensional stimuli: A note. Psychol. Rev. 79, 410-419.

Lowe, S. S., J. K. Cullen, C. Thompson, C. I. Berlin, L. L. Kirkpatrick, and J. T. Ryan. (1970) Dichotic and monotic simultaneous and time-staggered speech. J. Acoust. Soc. Amer. 47, 76(A).

Mattingly, I. G. (1968) Synthesis by rule of General American English. Supplement to Haskins Laboratories Status Report on Speech Research.

Mattingly, I. G. and A. M. Liberman. (1969) The speech code and the physiology of language. In Information Processing in the Nervous System, ed. by K. N. Leibovic. (New York: Springer-Verlag) 97-114.

Mattingly, I. G., A. M. Liberman, A. K. Syrdal, and T. Halwes. (1971) Discrimination in speech and nonspeech modes. Cog. Psychol. 2, 131-157.

McAdam, D. W., J. R. Knott, and C. S. Rebert. (1969) Cortical slow potential changes in man related to interstimulus interval and to pre-trial prediction of interstimulus interval. Psychophysiol. 5, 349-358.

Milner, B. (1967) Brain mechanisms suggested by studies of temporal lobes. In Brain Mechanisms Underlying Speech and Language, ed. by C. H. Millikan and F. L. Darley. (New York: Grune and Stratton) 122-131.

Morton, J. (1969) The use of correlated stimulus information in card sorting. Percept. Psychophys. 5, 374-376.

Naatanen, R. (1967) Selective attention and evoked potentials. Annales Academiae Scientiarum Fennicae 151, 1-226.

Naatanen, R. (1969) Anticipation of relevant stimuli and evoked potentials: A comment on Donchin's and Cohen's "Averaged evoked potentials and intramodality selective attention." Percept. Mot. Skills 28, 639-646.

Perry, N. W. and D. G. Childers. (1969) The Human Visual Evoked Response. (Springfield, Mass.: Thomas).

Peterson, G. E. and H. L. Barney. (1952) Control methods used in a study of the vowels. J. Acoust. Soc. Amer. 24, 175-184.

Pisoni, D. B. (1971) On the nature of categorical perception of speech sounds. Doctoral dissertation, University of Michigan. (Issued as Supplement to Haskins Laboratories Status Report on Speech Research.)

Pollen, D. A. (1970) On the generation of neocortical potentials. In Basic Mechanisms of the Epilepsies, ed. by H. H. Jasper, A. A. Ward, Jr., and A. Pope. (Boston, Mass.: Little, Brown) 411-420.

Posner, M. I. (1969) Abstraction and the process of recognition. In Psychology of Learning and Motivation, Vol. 3, ed. by G. Bower and J. T. Spence. (New York: Academic Press).
Posner, M. I. and S. J. Boies. (1971) Components of attention. Psychol. Rev. 78, 391-408.
Posner, M. I., R. Klein, J. Summers, and S. Buggie. (1973) On the selection of signals. Mem. Cog. 1, 2-12.
Posner, M. I. and R. F. Mitchell. (1967) Chronometric analysis of classification. Psychol. Rev. 74, 392-409.
Regan, D. (1972) Evoked Potentials in Psychology, Sensory Physiology, and Clinical Medicine. (London: Chapman and Hall).
Ritter, W., R. Simson, and H. G. Vaughan. (1972) Association cortex potentials and reaction time in auditory discrimination. EEG Clin. Neurophysiol. 33, 547-556.
Ruchkin, D. S. (1965) An analysis of average response computations based upon aperiodic stimuli. IEEE Trans. Biomed. Engin. 12, 87-94.
Saunders, M. G. (1963) Amplitude probability density studies on alpha and alpha-like patterns. EEG Clin. Neurophysiol. 15, 761-767.
Scheffe, H. A. (1960) The Analysis of Variance. (New York: Wiley).
Shepard, R. N. (1964) Attention and the metric structure of the stimulus space. J. Math. Psychol. 1, 54-87.
Siegel, S. (1956) Nonparametric Statistics. (New York: McGraw-Hill).
Spellacy, F. and S. Blumstein. (1970) The influence of language set on ear preference in phoneme recognition. Cortex 6, 430-439.
Squires, K. C., S. A. Hillyard, and P. H. Lindsay. (1973) Cortical potentials evoked by confirming and disconfirming feedback following an auditory discrimination. Percept. Psychophys. 13, 25-31.
Stevens, K. N. and M. Halle. (1967) Remarks on analysis by synthesis and distinctive features. In Models for the Perception of Speech and Visual Form, ed. by W. Wathen-Dunn. (Cambridge, Mass.: MIT Press) 88-102.
Stevens, K. N. and A. S. House. (1972) Speech perception. In Foundations of Modern Auditory Theory, ed. by J. V. Tobias. (New York: Academic Press) 1-62.
Stevens, K. N., A. M. Liberman, M. Studdert-Kennedy, and S. E. G. Ohman. (1969) Cross-language study of vowel perception. Lang. Speech 12, 1-23.
Studdert-Kennedy, M. (in press) The perception of speech. In Current Trends in Linguistics, Vol. XII, ed. by T. A. Sebeok. (The Hague: Mouton). [Also in Haskins Laboratories Status Report on Speech Research SR-23 (1970) 15-48.]
Studdert-Kennedy, M. and K. Hadding. (1971) Auditory and linguistic processes in the perception of intonation contours. Haskins Laboratories Status Report on Speech Research SR-27, 153-174.
Studdert-Kennedy, M., A. M. Liberman, K. Harris, and F. S. Cooper. (1970a) The motor theory of speech perception: A reply to Lane's critical review. Psychol. Rev. 77, 234-249.
Studdert-Kennedy, M. and D. Shankweiler. (1970) Hemispheric specialization for speech perception. J. Acoust. Soc. Amer. 48, 579-594.
Studdert-Kennedy, M., D. Shankweiler, and D. Pisoni. (1972) Auditory and phonetic processes in speech perception: Evidence from a dichotic study. Cog. Psychol. 3, 455-466.
Studdert-Kennedy, M., D. Shankweiler, and S. Schulman. (1970b) Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. J. Acoust. Soc. Amer. 48, 599-602.
Tecce, J. J. (1972) Contingent negative variation (CNV) and psychological processes in man. Psychol. Bull. 77, 73-108.

Torgerson, W. S. (1958) *Theory and Methods of Scaling*. (New York: Wiley).

Vaughan, H. G. (1966) The perceptual and physiologic significance of visual evoked responses recorded from the scalp in man. *Clinical Electroretinography*, Supplement to Vision Research, 203-223.

Walter, D. O. (1963) Spectral analysis for electroencephalograms: Mathematical determination of neurophysiological relationships from records of limited duration. Exp. Neurol. *8*, 155-181.

Walter, D. O. and W. R. Adey. (1963) Spectral analysis of electroencephalograms recorded during learning in the cat, before and after subthalamic lesions. Exp. Neurol. *7*, 481-501.

Walter, D. O., J. M. Rhodes, B. S. Brown, and W. R. Adey. (1966) Comprehensive spectral analysis of human EEG generators in posterior cerebral regions. EEG Clin. Neurophysiol. *20*, 224-237.

Walter, W. G., R. Cooper, V. J. Aldridge, W. C. McCallum, and A. L. Winter. (1964) Contingent negative variation: An electrical sign of sensorimotor association and expectancy in the human brain. Nature *203*, 380-384.

Well, A. D. (1971) The influence of irrelevant information on speeded classification tasks. Percept. Psychophys. *10*, 79-84.

Winer, B. J. (1962) *Statistical Principles in Experimental Design*. (New York: McGraw-Hill).

Wood, C. C. (1973) Parallel processing of auditory and phonetic information in speech perception. Haskins Laboratories Status Report on Speech Research *SR-35/36* (this issue).

Wood, C. C., W. R. Goff, and R. S. Day. (1971) Auditory evoked potentials during speech perception. Science *173*, 1248-1251.

III. PUBLICATIONS AND REPORTS

IV. APPENDIX

PUBLICATIONS AND REPORTS

Publications and Manuscripts

Effect of Speaking Rate on Labial Consonant Production: A Combined Electromyographic/High-Speed Motion Picture Study. T. Gay and H. Hirose. Phonetica (1973) 27, 44-56.

Laryngeal Control in Vocal Attack: An Electromyographic Study. H. Hirose and T. Gay. Folia Phoniatrica (1973) 24, 203-213.

Hemiretinae and Nonmonotonic Masking Functions with Overlapping Stimuli. Claire Farley Michaels and M. T. Turvey. Bulletin of the Psychonomic Society (1973) 2, 163-164.

The following papers were reprinted in Speech Synthesis (Benchmark Papers in Acoustics), ed. by James L. Flanagan and Lawrence R. Rabiner. (Stroudsburg, Pa.: Dowden, Hutchinson & Ross, 1973):

The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech. F. S. Cooper, A. M. Liberman, and J. M. Borst. [From Proceedings of the National Academy of Sciences (1951) 37, 318-325.] 59

Some Experiments on the Perception of Synthetic Speech Sounds. F. S. Cooper, P. C. Delattre, A. M. Liberman, J. M. Borst, and L. J. Gerstman. [From Journal of the Acoustical Society of America (1952) 24, 597-606.] 67

Minimal Rules for Synthesizing Speech. A. M. Liberman, F. Ingemann, L. Lisker, P. Delattre, and F. S. Cooper. [From Journal of the Acoustical Society of America (1959) 31, 1490-1499.] 330

Are You Asking Me, Telling Me, or Talking to Yourself? Kerstin Hadding and Michael Studdert-Kennedy. Working Papers, Phonetic Laboratory, Lund University, Sweden (1973) 7, 111-125. (Also in SR-33.)

The Physiological Substrate of Speaking. Katherine S. Harris. In Articulation and Hearing: New Dimensions in Research, Diagnostics and Therapy, ed. by Wolfe. (Springfield, Ill.: Charles C Thomas, 1973).

The Relationships Between Speech and Reading. Ignatius G. Mattingly and James F. Kavanagh. [Adapted from an article in The Linguistic Reporter, October 1972.] (Bethesda, Md.: National Institutes of Health, 1973) DHEW Publication No. (NIH) 73-475.

Speaker Identification by Speech Spectrograms: Some Further Observations. (Letter to the Editor). Richard H. Bolt, Franklin S. Cooper, Edward E. David, Peter B. Denes, James M. Pickett, and Kenneth N. Stevens. Journal of the Acoustical Society of America (1973) 54, 531-534.

The Specialization of the Language Hemisphere. A. M. Liberman. In The Neurosciences: Third Study Program, ed. by F. O. Schmitt and F. G. Worden. (Cambridge, Mass.: MIT Press) 43-56.

Research on Audible Outputs of Reading Machines for the Blind. F. S. Cooper, J. H. Gaitenby, I. G. Mattingly, P. W. Nye, and G. N. Sholes. Bulletin of Prosthetics Research (Fall 1972) BPR 10-18, 272-276; and (Spring 1973) BPR 10-19, 225-230.

On the Evolution of Language: A Unified View. Philip Lieberman. Cognition (1973) 1, 59-94.

*State-of-the-Art Report on Language Processing. A. M. Liberman. In Sensory Capabilities of Hearing-Impaired Children (Proceedings of a conference held in Baltimore, Md., October 26 and 27, 1973), ed. by Rachel E. Stark. (Baltimore, Md.: University Park Press, in press).

*Audible Outputs of Reading Machines for the Blind. Franklin S. Cooper, Jane H. Gaitenby, Ignatius G. Mattingly, Patrick W. Nye, and George N. Sholes. Bulletin of Prosthetics Research (Fall 1973) BPR 10-20, in press.

An Experimental Approach to the Problem of Articulation in Aphasia. D. Shankweiler and K. S. Harris. In Psycholinguistics and Aphasia, ed. by Goodglass and Blumstein. (Baltimore, Md.: Johns Hopkins Press, in press). [Reprint, with comment, of Cortex (1966) 2, 277-292.]

Hemispheric Specialization for Speech Perception in Six-Year-Old Black and White Children from Low and Middle Socioeconomic Classes. M. F. Dorman and Donna S. Geffner. Cortex (in press). (Also in SR-34, 1973.)

The following five papers were presented at a National Institute of Child Health and Human Development Conference, "The Role of Speech in Language," held at Columbia, Md., October 1973; and will be published in the conference proceedings, ed. by J. F. Kavanagh and J. E. Cutting. (Cambridge, Mass.: MIT Press):

Structure of the Conference. Alvin M. Liberman and James F. Kavanagh

The Human Aspect of Speech. Ignatius G. Mattingly

The Evolution of Speech and Language: A Convergent Process. Philip Lieberman

From Continuous Signal to Discrete Message: The Distinctive Properties of Phonetic Perception. Michael Studdert-Kennedy

Phonetic Feature Analyzers and the Processing of Speech of Infants. Peter D. Eimas and J. E. Cutting

---

*Appears in this report, SR-35/36.

The Two Brains and the Partition of Intelligence. D. Shankweiler. Invited paper presented to the Minnesota Conference on Cognition, Knowledge, and Adaptation, University of Minnesota, August 1973; to be published in the conference proceedings, ed. by Robert Shaw and John Bransford.

Auditory Evoked Potential Correlates of Speech Sound Discrimination. Michael F. Dorman. Perception and Psychophysics (in press). (Also in SR-29/30, 1972.)

Short-Term Habituation of the Infant Auditory Evoked Response. Michael F. Dorman and Robert Hoffmann. Journal of Speech and Hearing Research (in press). (Also in SR-29/30, 1972.)

The Development of Auditory Feedback Monitoring: DAF Studies on Vocalizations of Children Between 6 and 19 Months. N. F. Belmore, N. Fargo, D. Kewley-Port, R. L. Mobley, and V. E. Goodman. Journal of Speech and Hearing Research (in press).

Explicit Syllable and Phoneme Segmentation in the Young Child. I. Y. Liberman, D. Shankweiler, F. W. Fischer, and B. Carter. Journal of Experimental Child Psychology (in press).

*The Lag Effect in Dichotic Speech Perception. Emily F. Kirstein

*On the Identification of Place and Voicing Features in Synthetic Stop Consonants. David B. Pisoni and James R. Sawusch

*Perception of Speech and Nonspeech, with Speech-Relevant and Speech-Irrelevant Transitions. James E. Cutting

Reports and Oral Presentations

Speculations on the Linguistic Competence Required for Learning to Read. D. Shankweiler. Invited talk, Center for Research in Human Learning, University of Minnesota, Minneapolis, April 1973.

Speech and Handedness. D. Shankweiler. Colloquium presented at University of Iowa, College of Medicine, Iowa City, May 1973.

Acting and Perceiving; and Perception in a Single Glance. Michael T. Turvey. Two invited addresses at the Conference on Cognition and Knowledge, University of Minnesota, Minneapolis, July-August 1973.

An Analysis of Natural versus Synthetic Speech Intelligibility: A Preliminary Appraisal of a Reading Machine for the Blind. P. W. Nye, J. H. Gaitenby, and J. D. Hankins. Presented at the annual conference of the Association for Computing Machinery, Atlanta, Ga., 27-29 August 1973.

Linguistic and Paralinguistic Interchange; and On the Evolution of Language: A Unified View. Philip Lieberman. Two papers presented at the 9th International Congress of Ethnological and Anthropological Sciences, Chicago, Ill., 28 August - 2 September 1973. (Both papers appeared in SR-33.) Also, discussant on a three-day panel "On the Antecendents of Man."

Electrophysiology and the Distinctive Properties of Speech Perception. Michael Studdert-Kennedy. Invited paper read before the 8th International Congress of Electroencephalography and Clinical Neurophysiology, Marseille, France, 5 September 1973.

The Physiological Concomitants of Stress. Katherine S. Harris. Colloquium, University of Pennsylvania, Departments of Linguistics and Psychology, 20 September 1973.

Computer Simulation of the Human Speech Production. Paul Mermelstein. Computer Science Colloquium, University of Connecticut, Storrs, 21 September 1973.

*Stress and Syllable Duration Change. Katherine S. Harris. Presented to the American Association of Phonetic Sciences, Detroit, Mich., 11 October 1973.

Informal comment on symposium, "Oral Structure, Articulation, and Intelligibility," at the meeting of the American Speech and Hearing Association. Katherine S. Harris. Detroit, Mich., 13 October 1973.

Identification of Synthetic Vowel Sounds in Two Dialect Regions. E. Fenton, F. Freeman, E. Aschkenasy, and K. S. Harris. American Speech and Hearing Association meeting, Detroit, Mich., 14 October 1973.

*Parallel Processing of Auditory and Phonetic Information in Speech Perception. Charles C. Wood. Presented at the 86th meeting of the Acoustical Society of America, Los Angeles, Calif., 30 October 1973.

Differences Between Language-Bound and Stimulus-Bound Subjects in Solving Word-Search Puzzles. Ruth S. Day. Presented at the 86th meeting of the Acoustical Society of America, Los Angeles, Calif., October 1973.

*The Motor Organization of Some Speech Gestures. Fredericka Bell-Berti and Katherine Safford Harris. Presented at the 86th meeting of the Acoustical Society of America, Los Angeles, Calif., November 1973.

*Effect of Speaking Rate on Labial Consonant-Vowel Articulation. Thomas Gay, Tatsujiro Ushijima, Hajime Hirose, and Franklin S. Cooper. Presented at the 86th meeting of the Acoustical Society of America, Los Angeles, Calif., November 1973.

Backward and Forward Masking of Brief Vowels. M. F. Dorman, D. Kewley-Port, S. Brady, and M. T. Turvey. Presented at the 14th annual meeting of the Psychonomic Society, St. Louis, Mo., November 1973.

Individual Differences in Cognition. Ruth S. Day. Paper presented at the Psychonomic Society meeting, St. Louis, Mo., November 1973.

Dissertations

The Velopharyngeal Mechanism: An Electromyographic Study. Fredericka Bell-Berti. Ph.D. dissertation, City University of New York, 1973. (Issued as a Supplement to Haskins Laboratories Status Report on Speech Research, September 1973.)

*Levels of Processing in Speech Perception: Neurophysiological and Information-Processing Analyses. Charles C. Wood. Ph.D. dissertation, Yale University, 1973. (Issued as a Supplement bound with SR-35/36, this issue.)

An Examination of Hemispheric Asymmetry in the Visual Processing of Linguistic Items. Claire Farley Michaels. Ph.D. dissertation, University of Connecticut, 1973. (Abstract in SR-35/36, this issue.)

Audio-Visual Materials

A Prototype Automatic Reader for the Blind. (Color film with sound track.) P. W. Nye, writer, producer, and director. Presented at the National Convention of Workers for the Blind, Cleveland, Ohio, 22-25 July 1973.

Some Reminiscences on Speech Research. Franklin S. Cooper. Soundsheet recording of remarks and speech synthesis examples given upon his receipt of the 1972 Pioneer in Speech Award, presented by the IEEE Audio and Electroacoustics Group. In IEEE Transactions on Audio and Electroacoustics (June 1973) AU-21, insert between pages 144 and 145.

APPENDIX

DDC (Defense Documentation Center) and ERIC (Educational Resources Information Center) numbers:

SR-21/22 to SR-31/32

Status Report		DDC	ERIC
SR-21/22	January - June 1970	AD 719382	ED-044-679
SR-23	July - September 1970	AD 723586	ED-052-654
SR-24	October - December 1970	AD 727616	ED-052-653
SR-25/26	January - June 1971	AD 730013	ED-056-560
SR-27	July - September 1971	AD 749339	ED-071-533
SR-28	October - December 1971	AD 742140	ED-061-837
SR-29/30	January - June 1972	AD 750001	ED-071-484
SR-31/32	July - December 1972	AD 757954	
SR-33	January - March 1973		
SR-34	April - June 1973	AD 766178	

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
Leasco Information Products, Inc.
P. O. Drawer O
Bethesda, Maryland 20014

STATUS REPORT VOLUME NUMBERS

1965: SR-1, SR-2, SR-3, SR-4

1966: SR-5/6, SR-7/8

1967: SR-9, SR-10, SR-11, SR-12

1968: SR-13/14, SR-15/16

1969: SR-17/18, SR-19/20

1970: SR-21/22, SR-23, SR-24

1971: SR-25/26; SR-27; SR-28

1972: SR-29/30, SR-31/32

1973: SR-33, SR-34, SR-35/36

SUPPLEMENTS

1966: Singer (October)

1967: Gay (June)

1968: Mattingly (April)

1969: Halwes (September)

1970: Kirstein (Bound with SR-24)

1971: Raphael (August), Pisoni (November)

1973: Bell-Berti (September), Wood (Bound with SR-35/36)

CUMULATIVE INDEX

1965-1971 (SR-1 - SR-28): SR-28

1972-1973 (SR-29/30 - SR-35/36): SR-35/36

AUTHOR INDEX

Haskins Laboratories Status Report on Speech Research

SR-29/30 (January-June 1972) - SR-35/36 (July-December 1973)

Abramson, Arthur S. and Leigh Lisker. Voice-Timing Perception in Spanish Word-Initial Stops. SR-29/30 (1972), 15.

__________ (See Hirose, SR-31/32, 183)

__________ (See Erickson, SR-31/32, 231)

Atkinson, James E. Aspects of Intonation in Speech: Implications from an Experimental Study of Fundamental Frequency. (Abstract) SR-34 (1973), 211.

Bartlett, James C. (See Day, SR-31/32, 49)

Bell-Berti, Fredericka and H. Hirose. Stop Consonant Voicing and Pharyngeal Cavity Size. SR-31/32 (1972), 207.

__________ and H. Hirose. Velar Activity in Voicing Distinctions: A Simultaneous Fiberoptic and Electromyographic Study. SR-31/32 (1972), 223.

__________ (See Raphael, SR-33, 203)

__________ and H. Hirose. Patterns of Palatoglossus Activity and Their Implications for Speech Organization. SR-34 (1973), 203.

__________ The Velopharyngeal Mechanism: An Electromyographic Study. (Ph.D. Dissertation) Supplement to the Status Report, September 1973.

__________ and K. S. Harris. The Motor Organization of Some Speech Gestures. SR-35/36 (1973), 1.

Borden, Gloria J. Some Effects of Oral Anesthesia Upon Speech: An Electromyographic Investigation. SR-29/30 (1972), 27.

__________, K. S. Harris, and L. Catena. Electromyographic Study of Speech Musculature During Lingual Nerve Block. SR-31/32 (1972), 213.

__________, K. S. Harris, and W. Oliver. Oral Feedback, Part I: Variability of the Effect of Nerve-Block Anesthesia Upon Speech. SR-34 (1973), 159.

__________, K. S. Harris, and L. Catena. Oral Feedback, Part II: An Electromyographic Study of Speech Under Nerve-Block Anesthesia. SR-34 (1973), 167.

Brady-Wood, Susan. (See Dorman, SR-33, 93)

__________ and D. Shankweiler. Effects of Amplitude Variation on an Auditory Rivalry Task: Implications Concerning the Mechanism of Perceptual Asymmetries. SR-34 (1973), 119.

Carter, Bonnie. (See Liberman, I. Y., SR-31/32, 145)

Catena, Lorne. (See Borden, SR-31/32, 213)

__________ (See Borden, SR-34, 167)

Cooper, Franklin S., Jane H. Gaitenby, Ignatius G. Mattingly, Patrick W. Nye, and George N. Sholes. Audible Outputs of Reading Machines for the Blind. SR-29/30 (1972), 91; and SR-35/36 (1973), 117.

__________ (See Nye, SR-29/30, 97)

__________ Machines and Speech. SR-31/32 (1972), 159.

__________ (See Gaitenby, SR-31/32, 177)

__________ (See Gay, SR-33, 221; and SR-35/36, 7)

Croog, Joanne F. (See Turvey, SR-31/32, 67)

Cutting, James E. A Parallel Between Encodedness and the Magnitude of the Right-Ear Effect. SR-29/30 (1972), 61.

__________ (See Day, SR-31/32, 49)

__________ Ear Advantage for Stops and Liquids in Initial and Final Position. SR-31/32 (1972), 57.

__________ A Preliminary Report on Six Fusions in Auditory Research. SR-31/32 (1972), 93.

__________ Phonological Fusion in Synthetic and Natural Speech. SR-33 (1973), 19.

__________ Perception of Speech and Nonspeech, with and without Transitions. SR-33 (1973), 37.

__________ Speech Misperception: Inferences About a Cue for Cluster Perception from a Phonological Fusion Task. SR-33 (1973), 57.

__________ Levels of Processing in Phonological Fusion. (Ph.D. Dissertation) SR-34 (1973), 1.

__________ Phonological Fusion of Synthetic Stimuli in Dichotic and Binaural Presentation Modes. SR-34 (1973), 55.

__________ Phonological Fusion of Stimuli Produced by Different Vocal Tracts. SR-34 (1973), 61.

__________ Perception of Speech and Nonspeech, with Speech-Relevant and Speech-Irrelevant Transitions. SR-35/36 (1973), 55.

Day, Ruth S. and C. C. Wood. Mutual Interference Between Two Linguistic Dimensions of the Same Stimuli. SR-29/30 (1972), 69.

__________ and J. M. Vigorito. A Parallel Between Degree of Encodedness and the Ear Advantage: Evidence from a Temporal Order Judgment Task. SR-31/32 (1972), 41.

__________, J. C. Bartlett, and J. E. Cutting. Memory for Dichotic Pairs: Disruption of Ear Report Performance by the Speech-Nonspeech Distinction. SR-31/32 (1972), 49.

__________ Digit-Span Memory in Language-Bound and Stimulus-Bound Subjects. SR-34 (1973), 127.

__________ On Learning "Secret Languages." SR-34 (1973), 141.

Dorman, Michael F. Auditory Evoked Potential Correlates of Speech Sound Discrimination. SR-29/30 (1972), 111.

__________ and R. Hoffman. Short-Term Habituation of the Infant Auditory Evoked Response. SR-29/30 (1972), 121.

__________ Discrimination of Intensity Differences on Formant Transitions In and Out of Syllable Context. SR-33 (1973), 13.

__________, D. Kewley-Port, S. Brady-Wood, and M. T. Turvey. Forward and Backward Masking of Brief Vowels. SR-33 (1973), 93.

__________ and D. S. Geffner. Hemispheric Specialization for Speech Perception in Six-Year-Old Black and White Children from Low and Middle Socioeconomic Classes. SR-34 (1973), 151.

Erickson, Donna and A. S. Abramson. Electromyographic Study of the Tones of Thai. SR-31/32 (1972), 231.

__________, I. G. Mattingly, and M. T. Turvey. Phonetic Coding of Kanji. (Abstract) SR-29/30 (1972), 155.

__________, I. G. Mattingly, and M. T. Turvey. Phonetic Activity in Reading: An Experiment with Kanji. SR-33 (1973), 137.

Fischer, F. William. (See Liberman, I. Y., SR-31/32, 145)

Fujimura, O. (See Miyawaki, SR-33, 67)

Gaitenby, Jane H. (See Cooper, SR-29/30, 91; and SR-35/36, 117)

__________, G. N. Sholes, and G. M. Kuhn. Word and Phrase Stress by Rule for a Reading Machine. SR-29/30 (1972), 105.

__________, G. Sholes, T. Rand, G. Kuhn, P. Nye, and F. Cooper. An Automated Reading Service for the Blind. SR-31/32 (1972), 177.

__________ (See Nye, SR-33, 77)

Gay, Thomas. (See Hirose, SR-29/30, 49)

__________ and H. Hirose. Effect of Speaking Rate on Labial Consonant Production: A Combined Electromyographic-High Speed Motion Picture Study. SR-31/32 (1972), 193.

__________, T. Ushijima, H. Hirose, and F. S. Cooper. Effect of Speaking Rate on Labial Consonant-Vowel Articulation. SR-33 (1973), 221; and SR-35/36 (1973), 7.

Geffner, Donna S. (See Dorman, SR-34, 151)

Hadding, Kerstin and M. Studdert-Kennedy. Are You Asking Me, Telling Me, or Talking to Yourself? SR-33 (1973), 1.

Hankins, J. D. (See Nye, SR-29/30, 97)

Harris, Katherine S. (See Borden, SR-31/32, 213)

__________ (See Borden, SR-34, 159)

__________ (See Borden, SR-34, 167)

__________ (See Bell-Berti, SR-35/36, 1)

__________ Stress and Syllable Duration Change. SR-35/36 (1973), 31.

Hirose, Hajime and T. Gay. Laryngeal Control in Vocal Attack: An Electromyographic Study. SR-29/30 (1972), 49.

__________, L. Lisker, and A. S. Abramson. Physiological Aspects of Certain Laryngeal Features in Stop Production. SR-31/32 (1972), 183.

__________ (See Gay, SR-31/32, 193)

__________ (See Bell-Berti, SR-31/32, 207)

__________ (See Bell-Berti, SR-31/32, 223)

__________ (See Gay, SR-33, 221; and SR-35/36, 7)

__________, Charles Y. Lee, and Tatsujiro Ushijima. Laryngeal Control in Korean Stop Production. SR-34 (1973), 191.

__________ (See Bell-Berti, SR-34, 203)

Hoffmann, Robert. (See Dorman, SR-29/30, 121)

Jenkins, J. J. (See Miyawaki, SR-33, 67)

Kavanagh, James F. (See Mattingly, SR-29/30, 83)

Kewley-Port, Diane and Malcolm S. Preston. Early Apical Stop Production: A Voice Onset Time Analysis. SR-29/30 (1972), 125.

__________ (See Turvey, SR-31/32, 131)

__________ (See Dorman, SR-33, 93)

__________ Computer Processing of EMG Signals at Haskins Laboratories. SR-33 (1973), 173.

Kirstein, Emily. The Lag Effect in Dichotic Speech Perception. SR-35/36 (1973), 81.

Kuhn, Gary M. The Phi Coefficient as an Index of Ear Differences in Dichotic Listening. SR-29/30 (1972), 75.

__________ (See Gaitenby, SR-29/30, 105)

__________ (See Gaitenby, SR-31/32, 177)

Lee, Charles Y. (See Hirose, SR-34, 191)

Leidner, David R. An Electromyographic Study of the American English Liquids. SR-33 (1973), 195.

Liberman, Alvin, M. The Specialization of the Language Hemisphere. SR-31/32 (1972), 1.

__________ (See Miyawaki, SR-33, 67)

__________ State-of-the-Art Report on Language Processing. SR-35/36 (1973), 107.

Liberman, Isabelle Y., Donald Shankweiler, Bonnie Carter, and F. William Fischer. Reading and the Awareness of Linguistic Segments. SR-31/32 (1972), 145.

__________ Segmentation of the Spoken Word and Reading Acquisition. SR-33 (1973), 157.

Lieberman, Philip. Linguistic and Paralinguistic Interchange. SR-33 (1973), 167.

__________ On the Evolution of Language: A Unified View. SR-33 (1973), 229.

Lisker, Leigh. (See Abramson, SR-29/30, 15)

__________ (See Hirose, SR-31/32, 183)

Lukatela, Georgije. Pitch Determination by Adaptive Autocorrelation Method. SR-33 (1973), 185.

Mattingly, Ignatius G. and James F. Kavanagh. The Relationships Between Speech and Reading. SR-29/30 (1972), 83.

__________ (See Cooper, SR-29/30, 91; and SR-35/36, 117)

__________ (See Nye, SR-29/30, 97)

__________ (See Erickson, SR-29/30, 155)

__________ (See Erickson, SR-33, 137)

__________ Phonetic Prerequisites for First-Language Acquisition. SR-34 (1973), 65.

Michaels, Claire Farley and M. T. Turvey. Hemiretinae and Nonmonotonic Masking Functions with Overlapping Stimuli. SR-31/32 (1972), 127.

__________ An Examination of Hemispheric Asymmetry in the Visual Processing of Linguistic Terms. (Ph.D. Dissertation) Abstract. SR-35/36 (1973), 121.

Miyawaki, Kuniko, A. M. Liberman, O. Fujimura, W. Strange, and J. J. Jenkins. Cross-Language Study of the Perception of the F3 Cue for [r] versus [l] in Speech- and Nonspeech-Like Patterns. SR-33 (1973), 67.

Morse, Philip Allen. The Discrimination of Speech and Nonspeech Stimuli in Early Infancy. (Ph.D. Dissertation) Abstract. SR-29/30 (1972), 151.

Nye, Patrick W. (See Cooper, SR-29/30, 91; and SR-35/36, 117)

__________ J. D. Hankins, T. Rand, I. G. Mattingly, and F. S. Cooper. Field Evaluation of an Automated Reading System for the Blind. SR-29/30 (1972), 97.

__________ (See Gaitenby, SR-31/32, 177)

__________ and J. H. Gaitenby. Consonant Intelligibility in Synthetic Speech and in a Natural Speech Control (Modified Rhyme Test Results). SR-33 (1973), 77.

Oliver, William. (See Borden, SR-34, 159)

Pisoni, David B. (See Turvey, SR-31/32, 67)

__________ Perceptual Processing Time for Consonants and Vowels. SR-31/32 (1972), 83.

__________ and Jeffrey Tash. Reaction Times to Comparisons Within and Across Phonetic Categories: Evidence for Auditory and Phonetic Levels of Processing. SR-34 (1973), 77.

__________ The Role of Auditory Short-Term Memory in Vowel Perception. SR-34 (1973), 89.

__________ and J. R. Sawusch. On the Identification of Place and Voicing Features in Synthetic Stop Consonants. SR-35/36 (1973), 65.

Port, Diane Kewley. (See Kewley-Port)

Porter, Robert John. The Effect of Delayed Channel on the Perception of Dichotically Presented Speech and Nonspeech Sounds. (Ph.D. Dissertation) Abstract. SR-29/30 (1972), 153.

Preston, Malcolm S. (See Kewley-Port, SR-29/30, 125)

Rand, Timothy C. (See Nye, SR-29/30, 97)

__________ (See Gaitenby, SR-31/32, 177)

__________ Dichotic Release from Masking for Speech. SR-33 (1973), 47.

Raphael, Lawrence J. and F. Bell-Berti. The Role of the Extrinsic and Intrinsic Tongue Muscles in Differentiating the English Tense-Lax Vowel Pairs. SR-33 (1973), 203.

Sawusch, James R. (See Pisoni, SR-35/36, 65)

Shankweiler, Donald. (See Studdert-Kennedy, SR-31/32, 23)

__________ (See Liberman, I. Y., SR-31/32, 145)

__________ (See Brady-Wood, SR-34, 119)

Sholes, George N. (See Cooper, SR-29/30, 91; and SR-35/36, 117)

__________ (See Gaitenby, SR-29/30, 105)

__________ (See Gaitenby, SR-31/32, 177)

Strange, Winifred. (See Miyawaki, SR-33, 67)

Studdert-Kennedy, Michael and D. Shankweiler. A Continuum of Cerebral Dominance for Speech Perception? SR-31/32 (1972), 23.

__________ A Right-Ear Advantage in Choice Reaction Time to Monaurally Presented Vowels: A Pilot Study. SR-31/32 (1972), 75.

__________ (See Hadding, SR-33, 1)

Sullivan, Edie V. and M. T. Turvey. On the Short-Term Retention of Serial, Tactile Stimuli. SR-33 (1973), 123.

Tash, Jeffrey. (See Pisoni, SR-34, 77)

Turvey, Michael T. Some Aspects of Selective Readout from Iconic Storage. SR-29/30 (1972), 1.

__________ (See Erickson, SR-29/30, 155)

__________, David Pisoni, and Joanne F. Croog. A Right-Ear Advantage in the Retention of Words Presented Monaurally. SR-31/32 (1972), 67.

__________ Constructive Theory, Perceptual Systems, and Tacit Knowledge. SR-31/32 (1972), 109.

__________ (See Michaels, SR-31/32, 127)

__________, C. F. Michaels, and D. Kewley-Port. Visual Storage or Visual Masking?: An Analysis of the "Retroactive Contour Enhancement" Effect. SR-31/32 (1972), 131.

__________ (See Dorman, SR-33, 93)

__________ and R. A. Weeks. Effects of Proactive Interference and Rehearsal on the Primary and Secondary Components of Short-Term Retention. SR-33 (1973), 101.

__________ (See Sullivan, SR-33, 123)

__________ (See Erickson, SR-33, 137)

__________ A Note on the Relation Between Action and Perception. SR-34 (1973), 71.

Ushijima, Tatsujiro. (See Gay, SR-33, 221; and SR-35/36, 7)

__________ (See Hirose, SR-34, 191)

Vigorito, James M. (See Day, SR-31/32, 41)

Weeks, Robert A. A Speech Perception Paradox?: The Right-Ear Advantage and the Lag Effect. SR-33 (1973), 29.

__________ (See Turvey, SR-33, 101)

Wood, Charles C. (See Day, SR-29/30, 69)

__________ Levels of Processing in Speech Perception: Neurophysiological and Information-Processing Analyses. (Ph.D. Dissertation) Abstract--SR-34 (1973), 215; Text--Supplement bound with SR-35/36 (1973), 123.

__________ Parallel Processing of Auditory and Phonetic Information in Speech Perception. SR-35/36 (1973), 39.

TITLE INDEX

Haskins Laboratories Status Report on Speech Research
SR-29/30 (January-June 1972) - SR-35/36 (July-December 1973)

Are You Asking Me, Telling Me, or Talking to Yourself? K. Hadding and M. Studdert-Kennedy. SR-33 (1973), 1.

Aspects of Intonation in Speech: Implications from an Experimental Study of Fundamental Frequency. J. E. Atkinson. (Ph.D. Dissertation) Abstract. SR-34 (1973), 211.

Audible Outputs of Reading Machines for the Blind. F. S. Cooper, J. H. Gaitenby, I. G. Mattingly, P. W. Nye, and G. N. Sholes. SR-29/30 (1972), 91; and SR-35/36 (1973), 117.

Auditory Evoked Potential Correlates of Speech Sound Discrimination. M. F. Dorman. SR-29/30 (1972), 111.

An Automated Reading Service for the Blind. J. Gaitenby, G. Sholes, T. Rand, G. Kuhn, P. Nye, and F. Cooper. SR-31/32 (1972), 177.

Computer Processing of EMG Signals at Haskins Laboratories. D. Kewley-Port. SR-33 (1973), 173.

Consonant Intelligibility in Synthetic Speech and in a Natural Speech Control (Modified Rhyme Test Results). P. W. Nye and J. H. Gaitenby. SR-33 (1973), 77.

Constructive Theory, Perceptual Systems, and Tacit Knowledge. M. T. Turvey. SR-31/32 (1972), 109.

A Continuum of Cerebral Dominance for Speech Perception? M. Studdert-Kennedy and D. Shankweiler. SR-31/32 (1972), 23.

Cross-Language Study of the Perception of the F3 Cue for [r] versus [l] in Speech- and Nonspeech-Like Patterns. K. Miyawaki, A. M. Liberman, O. Fujimura, W. Strange, and J. J. Jenkins. SR-33 (1973), 67.

Dichotic Release from Masking for Speech. T. C. Rand. SR-33 (1973), 47.

Digit-Span Memory in Language-Bound and Stimulus-Bound Subjects. R. S. Day. SR-34 (1973), 127.

Discrimination of Intensity Differences on Formant Transitions In and Out of Syllable Context. M. F. Dorman. SR-33 (1973), 13.

The Discrimination of Speech and Nonspeech Stimuli in Early Infancy. P. A. Morse. (Ph.D. Dissertation) Abstract. SR-29/30 (1972), 151.

Ear Advantage for Stops and Liquids in Initial and Final Position. J. E. Cutting. SR-31/32 (1972), 57.

Early Apical Stop Production: A Voice Onset Time Analysis. D. K. Port and M. S. Preston. SR-29/30 (1972), 125.

The Effect of Delayed Channel on the Perception of Dichotically Presented Speech and Nonspeech Sounds. R. J. Porter, Jr. (Ph.D. Dissertation) Abstract. SR-29/30 (1972), 153.

Effect of Speaking Rate on Labial Consonant Production: A Combined Electromyographic-High Speed Motion Picture Study. T. Gay and H. Hirose. SR-31/32 (1972), 193.

Effect of Speaking Rate on Labial Consonant-Vowel Articulation. T. Gay, T. Ushijima, H. Hirose, and F. S. Cooper. SR-33 (1973), 221; and SR-35/36 (1973), 7.

Effects of Amplitude Variation on an Auditory Rivalry Task: Implications Concerning the Mechanism of Perceptual Asymmetries. S. Brady-Wood and D. Shankweiler. SR-34 (1973), 119.

Effects of Proactive Interference and Rehearsal on the Primary and Secondary Components of Short-Term Retention. M. T. Turvey and R. A. Weeks. SR-33 (1973), 101.

Electromyographic Study of Speech Musculature During Lingual Nerve Block. G. J. Borden, K. S. Harris, and L. Catena. SR-31/32 (1972), 213.

An Electromyographic Study of the American English Liquids. D. R. Leidner. SR-33 (1973), 195.

Electromyographic Study of the Tones of Thai. D. Erickson and A. S. Abramson. SR-31/32 (1972), 231.

An Examination of Hemispheric Asymmetry in the Visual Processing of Linguistic Items. C. F. Michaels. (Ph.D. Dissertation) Abstract. SR-35/36 (1973), 121.

Field Evaluation of an Automated Reading System for the Blind. P. W. Nye, J. D. Hankins, T. Rand, I. G. Mattingly, and F. S. Cooper. SR-29/30 (1972), 97.

Forward and Backward Masking of Brief Vowels. M. Dorman, D. Kewley-Port, S. Brady-Wood, and M. T. Turvey. SR-33 (1973), 93.

Hemiretinae and Nonmonotonic Masking Functions with Overlapping Stimuli. C. F. Michaels and M. T. Turvey. SR-31/32 (1972), 127.

Hemispheric Specialization for Speech Perception in Six-Year-Old Black and White Children from Low and Middle Socioeconomic Classes. M. F. Dorman and D. S. Geffner. SR-34 (1973), 151.

The Lag Effect in Dichotic Speech Perception. E. F. Kirstein. SR-35/36 (1973), 81.

Laryngeal Control in Korean Stop Production. H. Hirose, C. Y. Lee, and T. Ushijima. SR-34 (1973), 191.

Laryngeal Control in Vocal Attack: An Electromyographic Study. H. Hirose and T. Gay. SR-29/30 (1972), 49.

Levels of Processing in Phonological Fusion. J. E. Cutting. (Ph.D. Dissertation) SR-34 (1973), 1.

Levels of Processing in Speech Perception: Neurophysiological and Information-Processing Analyses. C. C. Wood. (Ph.D. Dissertation) Abstract--SR-34 (1973), 215; Text---Supplement bound with SR-35/36 (1973), 123.

Linguistic and Paralinguistic Interchange. P. Lieberman. SR-33 (1973), 167.

Machines and Speech. F. S. Cooper. SR-31/32 (1972), 159.

Memory for Dichotic Pairs: Disruption of Ear Report Performance by the Speech-Nonspeech Distinction. R. S. Day, J. C. Bartlett, and J. E. Cutting. SR-31/32 (1972), 49.

The Motor Organization of Some Speech Gestures. F. Bell-Berti and K. S. Harris. SR-35/36 (1973), 1.

Mutual Interference Between Two Linguistic Dimensions of the Same Stimuli. R. S. Day and C. C. Wood. SR-29/30 (1972), 69.

A Note on the Relation Between Action and Perception. M. T. Turvey. SR-34 (1973), 71.

On Learning "Secret Languages." R. S. Day. SR-34 (1973), 141.

On the Evolution of Language: A Unified View. P. Lieberman. SR-33 (1973), 229.

On the Identification of Place and Voicing Features in Synthetic Stop Consonants. D. B. Pisoni and J. R. Sawusch. SR-35/36 (1973), 65.

On the Short-Term Retention of Serial, Tactile Stimuli. E. V. Sullivan and M. T. Turvey. SR-33 (1973), 123.

Oral Feedback, Part I: Variability of the Effect of Nerve-Block Anesthesia Upon Speech. G. J. Borden, K. S. Harris, and W. Oliver. SR-34 (1973), 159.

Oral Feedback, Part II: An Electromyographic Study of Speech Under Nerve-Block Anesthesia. G. J. Borden, K. S. Harris, and L. Catena. SR-34 (1973), 167.

A Parallel Between Degree of Encodedness and the Ear Advantage: Evidence from a Temporal Order Judgment Task. R. S. Day and J. M. Vigorito. SR-31/32 (1972), 41.

A Parallel Between Encodedness and the Magnitude of the Right-Ear Effect. J. E. Cutting. SR-29/30 (1972), 61.

Parallel Processing of Auditory and Phonetic Information in Speech Perception. C. C. Wood. SR-35/36 (1973), 39.

Patterns of Palatoglossus Activity and Their Implications for Speech Organization. F. Bell-Berti and H. Hirose. SR-34 (1973), 203.

Perception of Speech and Nonspeech, with and without Transitions. J. E. Cutting. SR-33 (1973), 37.

Perception of Speech and Nonspeech, with Speech-Relevant and Speech-Irrelevant Transitions. J. E. Cutting. SR-35/36 (1973), 55.

Perceptual Processing Time for Consonants and Vowels. D. B. Pisoni. SR-31/32 (1972), 83.

The Phi Coefficient as an Index of Ear Differences in Dichotic Listening. G. M. Kuhn. SR-29/30 (1972), 75.

Phonetic Activity in Reading: An Experiment with Kanji. D. Erickson, I. G. Mattingly, and M. T. Turvey. SR-33 (1973), 137.

Phonetic Coding of Kanji. D. Erickson, I. G. Mattingly, and M. Turvey. (Abstract) SR-29/30 (1972), 155.

Phonetic Prerequisites for First-Language Acquisition. I. G. Mattingly. SR-34 (1973), 65.

Phonological Fusion in Synthetic and Natural Speech. J. E. Cutting. SR-33 (1973), 19.

Phonological Fusion of Stimuli Produced by Different Vocal Tracts. J. E. Cutting. SR-34 (1973), 61.

Phonological Fusion of Synthetic Stimuli in Dichotic and Binaural Presentation Modes. J. E. Cutting. SR-34 (1973), 55.

Physiological Aspects of Certain Laryngeal Features in Stop Production. H. Hirose, L. Lisker, and A. S. Abramson. SR-31/32 (1972), 183.

Pitch Determination by Adaptive Autocorrelation Method. G. Lukatela. SR-33 (1973), 185.

A Preliminary Report on Six Fusions in Auditory Research. J. E. Cutting. SR-31/32 (1972), 93.

Reaction Times to Comparisons Within and Across Phonetic Categories: Evidence for Auditory and Phonetic Levels of Processing. D. B. Pisoni and J. Tash. SR-34 (1973), 77.

Reading and the Awareness of Linguistic Segments. I. Y. Liberman, D. Shankweiler, B. Carter, and F. W. Fischer. SR-31/32 (1972), 145.

The Relationships Between Speech and Reading. I. G. Mattingly and J. F. Kavanagh. SR-29/30 (1972), 83.

A Right-Ear Advantage in Choice Reaction Time to Monaurally Presented Vowels: A Pilot Study. M. Studdert-Kennedy. SR-31/32 (1972), 75.

A Right-Ear Advantage in the Retention of Words Presented Monaurally. M. T. Turvey, D. Pisoni, and J. F. Croog. SR-31/32 (1972), 67.

The Role of Auditory Short-Term Memory in Vowel Perception. D. B. Pisoni. SR-34 (1973), 89.

The Role of the Extrinsic and Intrinsic Tongue Muscles in Differentiating the English Tense-Lax Vowel Pairs. L. J. Raphael and F. Bell-Berti. SR-33 (1973), 203.

Segmentation of the Spoken Word and Reading Acquisition. I. Y. Liberman. SR-33 (1973), 157.

Short-Term Habituation of the Infant Auditory Evoked Response. M. F. Dorman and R. Hoffmann. SR-29/30 (1972), 121.

Some Aspects of Selective Readout from Iconic Storage. M. T. Turvey. SR-29/30 (1972), 1.

Some Effects of Oral Anesthesia upon Speech: An Electromyographic Investigation. G. J. Borden. SR-29/30 (1972), 27.

The Specialization of the Language Hemisphere. A. M. Liberman. SR-31/32 (1972), 1.

Speech Misperception: Inferences About a Cue for Cluster Perception from a Phonological Fusion Task. J. E. Cutting. SR-33 (1973), 57.

A Speech Perception Paradox?: The Right-Ear Advantage and the Lag Effect. R. A. Weeks. SR-33 (1973), 29.

State-of-the-Art Report on Language Processing. A. M. Liberman. SR-35/36 (1973), 107.

Stop Consonant Voicing and Pharyngeal Cavity Size. F. Bell-Berti and H. Hirose. SR-31/32 (1972), 207.

Stress and Syllable Duration Change. K. S. Harris. SR-35/36 (1973), 31.

Velar Activity in Voicing Distinctions: A Simultaneous Fiberoptic and Electromyographic Study. F. Bell-Berti and H. Hirose. SR-31/32 (1972), 223.

The Velopharyngeal Mechanism: An Electromyographic Study. F. Bell-Berti. (Ph.D. Dissertation). Supplement to the Status Report, September 1973.

Visual Storage or Visual Mssking?: An Analysis of the "Retroactive Contour Enhancement" Effect. M. T. Turvey, C. F. Michaels, and D. K. Port. SR-31/32 (1972), 131.

Voice-Timing Perception in Spanish Word-Initial Stops. A. S. Abramson and L. Lisker. SR-29/30 (1972), 15.

Word and Phrase Stress by Rule for a Reading Machine. J. H. Gaitenby, G. N. Sholes, and G. M. Kuhn. SR-29/30 (1972), 105.