Manner of Articulation, Parallel Processing, and the Perception of Duration*

Ilse Lehiste

A number of problems are connected with the study of the temporal aspects of speech production and perception. An important one is the problem of segmentation. Any study involving the measurement of duration presupposes the establishment of boundaries. The researchers in the field are by no means unanimous regarding the boundaries of speech sounds. It has to be decided at which level the boundaries are to be located--the articulatory or the acoustic level--and whether these boundaries have any perceptual reality.

Some time ago (Peterson and Lehiste (1960)) I established some practical guidelines for segmenting an utterance on the basis of the acoustic characteristics of the sound wave. Naeser (1969) has recently elaborated these rules, and they have been used by several investigators of speech sound duration. Underlying these rules was a basic assumption which I would now like to make explicit: the production and perception of timing patterns takes place with reference to major changes in manner of articulation.

Before presenting some new evidence that, im my opinion, supports this hypothesis, I should review some of the arguments against the possibility of segmentation.

One of the arguments is connected with the continuous nature of the speech wave, and the fact that there is no one-to-one correspondence between acoustic segments and linguistic segments. Fant has devoted a considerable amount of attention to this problem (Fant (1962)). He has made the observation that although the speech wave is basically continuous, spectrographic pictures of speech often display quite distinct boundaries between successive parts along the time axis. These boundaries are, according to Fant, related to switching events in the speech production mechanism, such as a shift in the primary sound source (e.g. from voice to noise), or the opening and closing off of a passage within the vocal cavities. Boundaries between sound segments are due to the beginning or end of at least one of simultaneously present sound features. But sound segment boundaries are not to be confused with phoneme boundaries. Several adjacent sounds of connected speech may carry information on several adjacent phonemes. A typical example would be the influence exerted by a consonant on a following vowel.

The notion that the same sound segment may carry information on several adjacent phonemes is intimately connected with the hypothesis of parallel processing. In essence, parallel processing means that the same physical signal may carry more than one kind of information, which in the process of speech perception may be extracted simultaneously, as if over separate channels. Parallel processing has been discussed extensively in various recent publications (Neisser (1967); Chistovich and Kozhevnikov (1969-1970); Liberman (1970)). Given the continuous nature of the acoustic signal and the fact that perceptual cues may overlap in time, it is quite understandable that some linguists have claimed that it is not possible to establish the duration of segments in any perceptually meaningful way.

Granted that the acoustic signal is continuous and that speech processing may take in parallel fashion, I still believe that the picture is unduly complicated by not making a distinction between manner of articulation and point of articulation characteristics. It is the point of articulation cues that are continuous; they may be spread out over several adjacent segments, while the manner of articulation cues provide the abrupt changes that are seen in the visual display of an acoustic waveform. While several adjacent segments may carry information on the point of articulation, the manner of articulation can usually be determined from an examination of the pertinent segments themselves. Duration of segments relates to the manner of articulation rather than to the point of articulation, and timing information is extracted primarily from manner of articulation cues.

This is, of course, a hypothesis that should be supported by experimental evidence. I shall try to present some toward the end of this paper. But first let me bring up some further considerations that lead me to believe in the possibility of making rather precise and perceptually meaningful measurements of the duration of various sound segments.

It is an established fact that differences in duration are perceptible. In principle, the ear is capable of distinguishing between durations, be it the duration of a continuous signal (like gated white noise) or the duration of a silent interval embedded in a continuous signal. If it is claimed that listeners can distinguish the durations of non-linguistic stimuli, but cannot perceive differences in the duration of linguistic stimuli, it is assumed that speech sounds have some characteristics that make their boundaries perceptually blurred. Being a native speaker of a quantity language, in which differences in duration carry high linguistic significance, I find this notion intuitively quite unacceptable. If durational differences can serve as part of the linguistic signaling system, listeners must be able to compare the durations of speech sounds (or whatever unit possesses contrastive duration). And if they compare durations, they must know at which moment a given sound begins and ends. In other words, there must exist unambiguous boundaries to which the listener may refer in comparing either two durations heard in succession, or a perceived duration with a stored "durational image". I propose that major changes in manner of articulation constitute the acoustic and perceptual correlates of such boundaries.

It has been suggested that differences in duration are perceived as qualitative rather than quantitative differences. While sounds differing in duration may also differ in quality in a number of cases, it is not necessarily true in other instances. As a limiting case, I would like to quote my own experiments with a set of synthetic Estonian words, in which certain stimuli differed solely in the duration of the intervocalic plosive gap (Lehiste (1970a)). The duration of the rest of the word, including the duration of the transitions to and from the consonant, was held constant. The listeners had no difficulty in assigning different linguistic labels to words falling into different intervocalic plosive duration categories. It should be obvious that there was no qualitative difference in the silence corresponding to the duration of the plosive; thus the judgments must have been based on the perception of differences in the duration of the plosive gap.

I find no contradiction between the claim that listeners can compare the durations of segments and that they perceive speech by a kind of parallel processing of the incoming signal. The timing information is simply another feature which is extracted at the same time as the information concerning the segmental nature of the incoming speech waeve.

There is additional, somewhat circumstantial evidence of the importance of the manner of articulation in speech perception. In a study of the perceptual parameters of consonant sounds, Sharf (1971) established seven-point scales for duration, loudness, frequency, sharpness, and contact. Substantial numbers of significant differences were obtained only for duration comparisons based on manner of articulation. In an earlier study, Denes (1963) showed that manner of articulation carries by far the greatest functional load in the English sound system, and suggested that the acoustic correlates of manner of articulation might be used for segmentation in automatic speech recognition systems.

One way to test the hypothesis that the timing of speech sounds takes place with reference to major changes in manner of articulation would be to find some instances in which the application of a timing rule depends on such differences. The study I want to report about during the rest of this paper deals with a limited attempt to find such a situation.

I investigated the duration of segments in monosyllabic English words beginning and ending in an obstruent consonant and containing syllable nuclei consisting of long and short vowels, preceded and/or followed by resonants. The boundaries between obstruents and both vowels and resonants are clearly manifested and should be easily detectable, whereas the boundaries between vowels and resonants are relatively less well defined and do not correspond to what I would call major changes in manner of articulation. If the hypothesis is true that timing takes place with reference to major changes in manner of articulation, the span between the release of the initial obstruent and the onset of the final obstruent should function as a unit of timing, regardless of the position or even the presence of the resonant; the resonants should fuse with the vowels into syllable nuclei functioning as a whole with regard to some timing rules.

The structure of the test words may be symbolized as C_1 R_1 V R_2 C_2 (consonant-resonant-vowel-resonant-consonant). The first consonant was either a voiced or voiceless plosive. In the latter case, aspiration was present. The duration of initial consonants was not measurable, especially in the case of voiceless plosives; thus only the duration of aspiration will be presented in the following tables. The first resonant could be present or absent. If present, it was either /r/ or /l/. The vowel was always present; it could be either long or short. For purposes of this study, all vowels were considered long with the exception of [I ε ʌ U]. The second resonant could be present or absent; if present, it was either /m/, /n/, /r/, or /l/. The final consonant was a voiced or voiceless obstruent (in most cases, plosive). The tables report the duration of the closure part of the final consonant; release and aspiration (if present) are not included in the tables. For purposes of processing, the sounds were coded in the following manner:

Aspiration	Resonant	Vowel	Resonant	Consonant
1 = +	2 = /r/	1 = long	4 = /m/	1 = voiced
0 = -	1 = /l/	0 = short	3 = /n/	0 = voiceless
	0 = -		2 = /r/	
			1 = /l/	
			0 = -	

For example, the code 00011 refers to a word beginning with a voiced initial consonant, containing no first resonant, a short vowel, /l/ as second resonant, and ending in a voiced final consonant. An example would be the word <u>build</u>. The code 000R1 refers to all words of this type in which a resonant was present in the slot indicated by R.

There were 156 test words of this general structure. Each word was produced five times by three native speakers of English; thus the data consist of 15 x 156 = 2350 productions. A list of the test words, together with their codes, is presented at the end of this paper in Appendix A.

The tapes were processed by means of a Frøkjær-Jensen Pitch Meter and Intensity Meter and displayed on an Elema-Schönander Mingograph, operated at a speed of 10 cm/sec. The boundaries of segments were established mainly on the basis of duplex oscillograms, using principles summarized by Naeser (1969). Durations of all segments were measured, and average durations were computed for all segments in all word types. These average durations are presented at the end of the paper in Appendix B.

The results of the study will be discussed with reference to three summary tables and five figures. Table 1 presents average durations of segments in words with syllable nuclei consisting of vowel + resonant.

TABLE 1

AVERAGE DURATIONS OF SEGMENTS IN WORDS WITH SYLLABLE NUCLEI CONSISTING OF VOWEL + RESONANT

Word type	N	Asp.	R	V	R	C	SN	SN+C
OOOR1	105			229.1	174.4	54.8	403.5	458.3
OOORO	105			163.4	103.6	106.7	267.0	373.7
OO1R1	150			290.0	141.1	95.6	431.1	526.7
OO1RO	150			196.9	89.0	112.9	285.9	398.8
1OOR1	75	78.7		212.1	156.9	55.4	447.7	503.1
1OORO	75	72.2		143.7	92.7	104.6	308.6	413.2
1O1R1	120	86.6		284.4	134.4	61.9	505.4	567.3
1O1RO	135	83.6		192.4	79.7	102.2	355.7	457.9
OR1R1	15		90.7	347.0	100.4	38.7	538.1	576.8
OR1RO	15		75.3	253.5	72.4	78.9	401.2	480.1

The word type is given in Column 1. Five pairs of word types are presented, differing in the voicing of the final obstruent consonant. The first pair consists of words beginning with a voiced plosive, followed by a short vowel and a resonant. The second pair is similar, except that the vowels are long. The third pair consists of words beginning with a voiceless plosive, followed by a short vowel and a resonant. The fourth pair contains a long vowel. The fifth pair finally consists of words in which a voiced initial plosive was followed by a sequence of resonant, long vowel and resonant, followed by a voiced and voiceless plosive. (Only one example of each type was available--bland and blank--and therefore the averages have to be interpreted with caution.)

The second column contains the number of productions used for averaging. Since there were three speakers, each producing the word five times, the number of different words may be obtained by dividing N by 15.

The third column contains the duration of aspiration, which was present in words beginning with a voiceless plosive. (All durations are in milliseconds). The fourth column shows the average duration of the prevocalic resonant, where present. The fifth column contains the average duration of the vowel. The next column shows the average duration of the postvocalic resonant. This is followed in the next column by the average duration of the final consonant. The duration is that of the hold of the consonant and does not include release and/or aspiration. The following column gives the duration of the span from the release of the first obstruent to the onset of the second. The last column gives the sum of the syllable nucleus (consisting of vowel and one or two resonants) and the final consonant. Figure 1 presents the same information graphically.

It is a well known rule in English that vowels are shortened before a voiceless final consonant and lengthened before a voiced final consonant. In an earlier study (Peterson and Lehiste (1960)), we had established the ratio between vowel durations before voiceless and voiced final consonants as 0.66, i.e. approximately 2/3. The present set of data shows that both the vowel and the postvocalic resonant are subject to either shortening or lengthening, depending on the voicing of the final obstruent.

When all parts of the syllable nucleus from the release of the initial plosive to the onset of the postvocalic resonant were combined, the ratio between their average durations before a voiceless and a voiced plosive was 0.73. The duration ratio for postvocalic resonants was 0.62. The ratio of the durations of the whole span from the release of the initial plosive to the closure of the final plosive was 0.69.

In a recent study devoted to vowel length variation as a function of the voicing of the consonant environment, Chen (1970) included 96 word tokens containing vowel + resonant sequences. He obtained comparable ratios: 0.73 for the vowel, 0.60 for the resonant, and 0.66 for the whole vowel + resonant sequence. This, as may be remembered, is identical with the ratio obtained for vowels by Peterson and Lehiste (1960), and very close to the 0.69 ratio obtained in the present study.

I believe it to be obvious that with regard to the timing rule in question, the sequence vowel + resonant functions indeed as a unitary syllable nucleus, albeit a segmentally complex one. The timing of the sequence appears to proceed indeed from the release of the initial obstruent to the formation of the closure of the final obstruent, which constitute major changes in the manner of articulation.

The question whether sequences of resonant + vowel function in the same manner turned out to be somewhat more complicated. Table 2 presents four sets of words in which resonants, if present, preceded and/or followed vocalic syllable nuclei.

TABLE 2
AVERAGE DURATIONS OF SEGMENTS, SYLLABLE NUCLEI AND WORDS
IN VARIOUS WORD TYPES INVOLVING RESONANTS

Word type	N	Asp.	R	V	R	C	SN	SN+C
10100	135	75.1		234.9		135.0	310.0	445.0
101R0	135	83.6		192.4	79.7	102.2	355.7	457.9
1R100	165	95.5	42.9	211.4		125.0	349.8	474.8
1R1R0	30	105.3	20.8	223.5	65.7	88.6	415.3	503.9
00101	45			428.1		76.4	428.1	504.5
001R1	150			290.0	141.1	95.6	431.1	526.7
0R101	60		83.2	394.6		68.1	477.8	545.9
0R1R1	15		90.7	347.0	100.4	38.7	538.1	576.8
10000	45	71.0		180.3		130.2	251.3	381.5
100R0	75	72.2		143.7	92.7	104.6	308.6	413.2
1R000	75	94.0	45.8	172.0		120.2	311.8	432.0
1R0R0	15	89.2	42.5	149.7	79.6	75.4	361.0	436.4
00001	45			292.5		84.9	292.5	377.4
000R1	105			229.1	174.4	54.8	403.5	458.3
0R001	60		83.6	258.4		85.5	342.0	427.5
0R0R1	15		90.6	238.1	144.9	49.7	473.6	523.3

The first set of four consists of words containing long vowels and beginning and ending in a voiceless plosive. The second set is similar, except the words began and ended in voiced plosives. The third set is analogous to the first, except for the vowel being short; the fourth set is in the same way analogous to the second. Sets one and two are shown on Figure 2; Figure 3 presents comparable material for sets three and four.

The differences in the average duration of syllable nuclei (including vowels and resonants) range from 3.0 msec (for 00101 - 001R1) to 181.1 (for 00001 - 0R0R1). In trying to assess the relative significance of the differences, it appears resonable to ask first whether the differences are perceptible. Just noticeable differences (jnd's, or difference limens - DLs) in duration have been studied by several investigators (summarized in Lehiste (1970b)). Table 3 gives the differences between the average durations for all pairs within each set of word types presented in Table 2.

TABLE 3
DIFFERENCE IN THE AVERAGE DURATIONS OF SYLLABLE NUCLEI INVOLVING RESONANTS BEFORE AND AFTER THE VOWEL

Word type	N	Average duration of SN, in msec	Difference in SN, in msec	Nearest absolute DL, in msec
10100	135	310.0		
101R0	135	355.7	45.7	48.0 (Stott, 1935)
10100	135	310.0		
1R100	165	349.8	39.8	48.0 (Stott, 1935)
10100	135	310.0		
1R1R0	30	415.3	105.3	48.0 (Stott, 1935)
101R0	135	355.7		
1R100	165	349.8	5.9	48.0 (Stott, 1935)
00101	45	428.1		
001R1	150	431.1	3.0	48.0 (Stott, 1935)
00101	45	428.1		
0R101	60	477.8	49.7	68.64 (Henry, 1948)
00101	45	428.1		
0R1R1	15	538.1	110.0	69.0 (Stott, 1935)
001R1	150	431.1		
0R101	60	477.8	46.7	68.64 (Henry, 1948)
10000	45	251.3		
100R0	75	308.6	57.3	47.64 (Henry, 1948)
10000	45	251.3		
1R000	75	311.8	60.5	47.64 (Henry, 1948)
10000	45	251.3		
1R0R0	15	361.0	109.7	48.0 (Stott, 1935)
100R0	75	308.6		
1R000	75	311.8	3.2	47.64 (Henry, 1948)
00001	45	292.5		
000R1	105	403.5	111.0	48.0 (Stott, 1935)
00001	45	292.5		
0R001	60	342.0	49.5	47.64 (Henry, 1948)
00001	45	292.5		
0R0R1	15	473.6	181.1	48.0 (Stott, 1935)
000R1	105	403.5		
0R001	60	342.0	61.5	48.0 (Stott, 1935)

The last column contains the absolute DL, in msec, established for reference durations that are closest to the duration of the syllable nuclei under consideration. Table 4 summarizes the pertinent data for durational difference limens. The information is presented graphically in Figures 4 and 5.

TABLE 4

Reference duration (msec)	Δ T/T	Absolute DL (msec)
200	.142	28.4 (Stott, 1935)
277	.172	47.64 (Henry, 1948)
400	.120	48.0 (Stott, 1935)
480	.143	68.64 (Henry, 1948)
600	.115	69.0 (Stott, 1935)

Looking at the first set, we see one difference that is clearly nonperceptible; another that is considerably above threshold and should be perceptible; and two that hover around the difference limen. Similar observations may be made with regard to the other sets.

Some generalizations may be drawn from comparing all four sets. The picture seems a little more systematic with long vowels than with short vowels. Here all differences are below or near the threshold with the exception of that between a vowel occurring alone and a vowel flanked on both sides by a resonant. In words with short vowels and voiceless initial and final plosives, the relative shortness of the vowel raises these differences slightly above threshold in those word pairs in which a vowel occurring alone is compared with vowel preceded and/or followed by a resonant. However, the ordering of the resonant before or after the vowel does not affect the timing in any significant way, as had also been the case with long vowels.

In words with short vowels and voiced plosives, there are two pairs whose differences are clearly above threshold, and two that are close to threshold value.

If the sets are combined according to the voicing or voicelessness of the plosives (ignoring the intrinsic differences in vowel duration), the differences drop below threshold except for the word types containing two resonants. These are longer than the other words by approximately the average duration of one resonant.

Table 2 reveals a number of other interesting facts about the temporal structure of the test words which are, however, not directly relevant to the question under consideration. For example, the duration of final consonants stands in a compensatory relationship to the duration of syllable nuclei, so that the differences between the average durations of words are usually smaller than those between syllable nuclei. For many word types, these differences are likewise below the perceptual threshold.

Consideration of words with resonants preceding and following vowels thus adds some further support to the hypothesis that the timing patterns are related to major changes in the manner of articulation. Roughly speaking, long vowels seem to fuse into a timing unit with either a preceding or a following resonant; with short vowels the evidence is less clear, but at least with voiced initial and final consonants, the vowel and a preceding resonant seem to have the same average duration as the vowel by itself. The ordering of the vowel - resonant sequence is irrelevant for overall duration. Number of segments begins to play a part when more than one resonant is involved; these cases thus provide the limit to which the argument can be carried. It is possible that some of the exceptions to the general pattern are due to the accidents of test word selection; a larger corpus, with a better balanced set of test words, might yield a clearer picture.

Footnote

*The research on which the paper is based was supported in part by the National Science Foundation through Grant GN-534.1 from the Office of Science Information Service to the Computer and Information Science Research Center, The Ohio State University.

References

Chen, Matthew (1970) "Vowel length variation as a function of the voicing of the consonant environment." Phonetica 22. 129-159.

Chistovich, L. A. and V. A. Kozhevnikov (1969) "Perception of speech." in: Voprosy teorii i metodov issledovanija vosprijatija rečevyx signalov, Leningrad; translated as L. A. Chistovich et al. Theory and methods of research on perception of speech signals, JPRS 50423, (1970).

Denes, Peter B. (1963) "On the statistics of spoken English." JASA 53.892-904.

Fant, C. Gunnar M. (1962) "Descriptive analysis of the acoustic aspects of speech." Logos 5.3-17.

Lehiste, Ilse (1970) "Experiments with synthetic speech concerning quantity in Estonian." Proceedings of the 3rd International Congress of Fenno-Ugricists, Tallinn (in press).

Lehiste, Ilse (1970) Suprasegmentals, M.I.T. Press, Cambridge.

Liberman, A. M. (1970) "The grammars of speech and language." Cognitive Psychology 1.301-323.

Naeser, Margaret (1969) "Criteria for the segmentation of vowels on duplex oscillograms." Technical Report No. 124, Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin, Madison.

Neisser, Ulric (1967) Cognitive Psychology, Appleton-Century-Crofts, New York.

Peterson, Gordon E. and I. Lehiste (1960) "Duration of syllable nuclei in English." JASA 32.693-703.

Sharf, Donald J. (1971) "Perceptual parameters of consonant sounds." Language and Speech 14.169-177.

APPENDIX A

List of test words used in the study, arranged according to their structural code.

1.	felt	00010
2.	guilt	00010
3.	built	00010
4.	dug	00001
5.	dead	00001
6.	bed	00001
7.	guild	00011
8.	build	00011
9.	felled	00011
10.	sent	00030
11.	dint	00030
12.	bent	00030
13.	shunt	00030
14.	bent	00030
15.	bend	00031
16.	send	00031
17.	bend	00031
18.	dinned	00031
19.	shunned	00031
20.	beat	00100
21.	doubt	00100
22.	beach	00100
23.	back	00100
24.	goat	00100
25.	gape	00100
26.	died	00101
27.	bayed	00101
28.	bird	00101
29.	goos	00101
30.	molt	00110
31.	bolt	00110
32.	bold	00111
33.	mold	00111
34.	sort	00120
35.	mart	00120
36.	marred	00121
37.	sword	00121
38.	fount	00130
39.	mount	00130
40.	joint	00130
41.	daunt	00130
42.	bank	00130
43.	faint	00130
44.	burnt	00130
45.	feigned	00131
46.	joined	00131
47.	dawned	00131
48.	mound	00131
49.	band	00131
50.	found	00131
51.	burned	00131
52.	bled	01001
53.	blend	01031
54.	gloat	01100
55.	bleat	01100
56.	black	01100
57.	blurt	01100
58.	blade	01101
59.	blurred	01101
60.	glues	01101
61.	blank	01130
62.	bland	01131
63.	bread	02001
64.	dread	02001
65.	drug	02001
66.	grape	02100
67.	drought	02100
68.	breach	02100
69.	dried	02101
70.	brayed	02101
71.	puck	10000
72.	tuck	10000
73.	kick	10000
74.	pug	10001
75.	ted	10001
76.	tilt	10010
77.	cult	10010
78.	tilled	10011
79.	culled	10011
80.	tent	10030

81.	tint	10030
82.	tent	10030
83.	tend	10031
84.	tend	10031
85.	tinned	10031
86.	cap	10100
87.	peach	10100
88.	pot	10100
89.	coke	10100
90.	cape	10100
91.	peat	10100
92.	tout	10100
93.	tight	10100
94.	kirk	10100
95.	tooth	10100
96.	pod	10101
97.	cause	10101
98.	cowed	10101
99.	pan	10101
100.	paid	10101
101.	tied	10101
102.	goos	10101
103.	poise	10101
104.	colt	10110
105.	cold	10111
106.	cart	10120
107.	tart	10120
108.	court	10120
109.	card	10121
110.	tarred	10121
111.	cord	10121
112.	cant	10130
113.	pint	1013
114.	paint	10130
115.	pant	10130
116.	canned	10131
117.	panned	10131
118.	pined	10131
119.	pained	10131
120.	camp	10140
121.	click	11000
122.	pluck	11000
123.	plug	11001
124.	cloak	11100
125.	clap	11100
126.	pleat	11100
127.	plot	11100
128.	clerk	11100
129.	cloud	11101
130.	plod	11101
131.	played	11101
132.	plan	11101
133.	claws	11101
134.	clues	11101
135.	ploys	11101
136.	plant	11130
137.	planned	11131
138.	clamp	11140
139.	truck	12000
140.	crick	12000
141.	tread	12001
142.	trent	12030
143.	trend	12031
144.	preach	12100
145.	crepe	12100
146.	crap	12100
147.	trout	12100
148.	croak	12100
149.	trite	12100
150.	truth	12100
151.	tried	12101
152.	crowd	12101
153.	craws	12101
154.	prayed	12101
155.	crews	12101
156.	cramp	12140

APPENDIX B

Average durations, in milliseconds, of segments occurring in the test words, each produced 5 times by three speakers. N = number of words of a given type.

Word type	N	Asp.	R	V	R	C	SN	SN + C
00001	3			292.5		84.9	292.5	377.4
00010	2			147.1	102.9	125.0	250.0	375.0
00011	2			227.1	181.2	65.5	408.3	473.8
00030	5			179.7	104.2	88.4	283.9	372.3
00031	5			231.1	167.5	44.1	398.6	442.7
00100	6			270.6		152.3		422.9
00101	3			428.1		76.4		504.5
00110	2			190.7	97.5	127.7	288.2	415.9
00111	2			298.6	148.1	62.0	446.7	508.7
00120	1			161.1	85.4	122.5	246.5	369.0
00121	1			208.1	150.1	174.2	358.2	532.4
00130	7			238.8	84.1	88.5	322.9	411.4
00131	7			363.4	125.1	50.6	488.5	539.1
01001	1		89.0	256.1		82.7	345.1	427.8
01031	1		90.6	238.1	144.9	49.7	473.6	523.3
01100	4		89.2	225.4		131.1	314.6	445.7
01101	2		91.1	376.6		70.0	467.7	537.7
01130	1		75.3	253.5	72.4	78.9	401.2	480.1
01131	1		90.7	347.0	100.4	38.7	538.1	576.8
02001	3		78.2	260.6		88.4	338.8	427.2
02100	3		72.8	243.2		158.3	316.0	474.3
02101	2		75.3	412.5		66.3	487.8	554.1
10000	3	71.0		180.3		130.2	251.3	381.5
10001	2	71.6		260.6		95.4	332.2	427.6
10010	2	77.9		114.8	90.8	116.3	283.5	399.8
10011	2	87.6		205.2	153.5	66.3	446.3	512.6
10030	3	66.4		172.6	94.6	92.9	333.6	426.5
10031	3	69.7		219.0	160.3	44.6	449.0	493.6
10100	9	75.1		234.9		135.0	310.0	445.0
10101	7	77.6		373.9		77.2	451.5	528.7
10110	1	83.5		154.2	88.1	125.0	325.8	450.8
10111	1	86.5		252.2	146.6	62.5	485.3	547.8
10120	3	90.9		165.1	73.6	118.9	329.6	348.5
10121	3	97.1		260.9	126.6	70.7	484.6	555.3

Word type	N	Asp.	R	V	R	C	SN	SN+C
10130	4	76.8		218.6	84.8	83.6	380.2	463.8
10131	4	76.2		340.0	130.1	52.6	546.3	598.9
10140	1	83.1		231.9	72.1	81.3	387.1	468.4
11000	2	91.1	47.8	156.1		129.5	295.0	424.5
11001	1	100.7	47.5	244.4		90.3	392.6	482.9
11100	5	94.5	47.3	206.0		116.2	347.8	464.0
11101	5	95.1	51.9	412.0		82.9	559.0	641.9
11130	1	97.9	46.7	248.6	85.3	69.3	478.5	547.8
11131	1	92.9	38.9	322.8	132.3	46.5	586.9	633.4
11140	1	117.1		211.4	64.0	94.4	392.5	486.9
12000	3	96.8	43.9	187.9		110.9	328.6	439.5
12030	1	89.2	42.5	149.7	79.6	75.4	361.0	436.4
12031	1	103.3	50.3	208.9	162.3	47.1	524.8	571.9
12100	6	96.5	38.4	216.8		133.9	351.7	485.6
12101	4	95.6	47.7	375.4		72.2	518.7	590.9
12140	1	93.5	41.6	235.5	67.5	82.8	438.1	520.9

Fig. 1

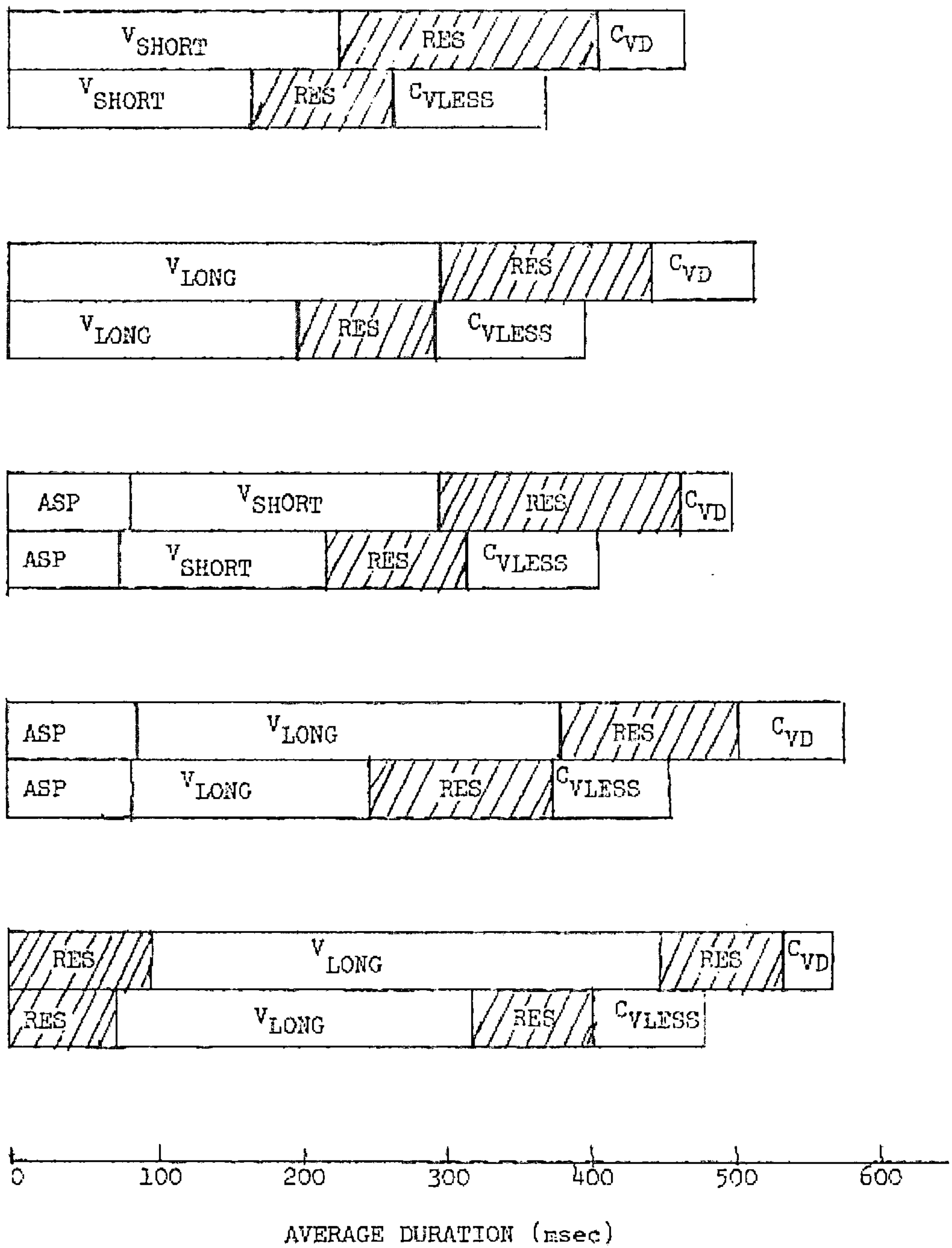

V SHORT
RES
C VD
V SHORT
RES
C VLESS
V LONG
RES
C VD
V LONG
RES
C VLESS
ASP
V SHORT
RES
C VD
ASP
V SHORT
RES
C VLESS
ASP
V LONG
RES
C VD
ASP
V LONG
RES
C VLESS
RES
V LONG
RES
C VD
RES
V LONG
RES
C VLESS
0
100
200
300
400
500
600
AVERAGE DURATION (msec)

Fig. 2

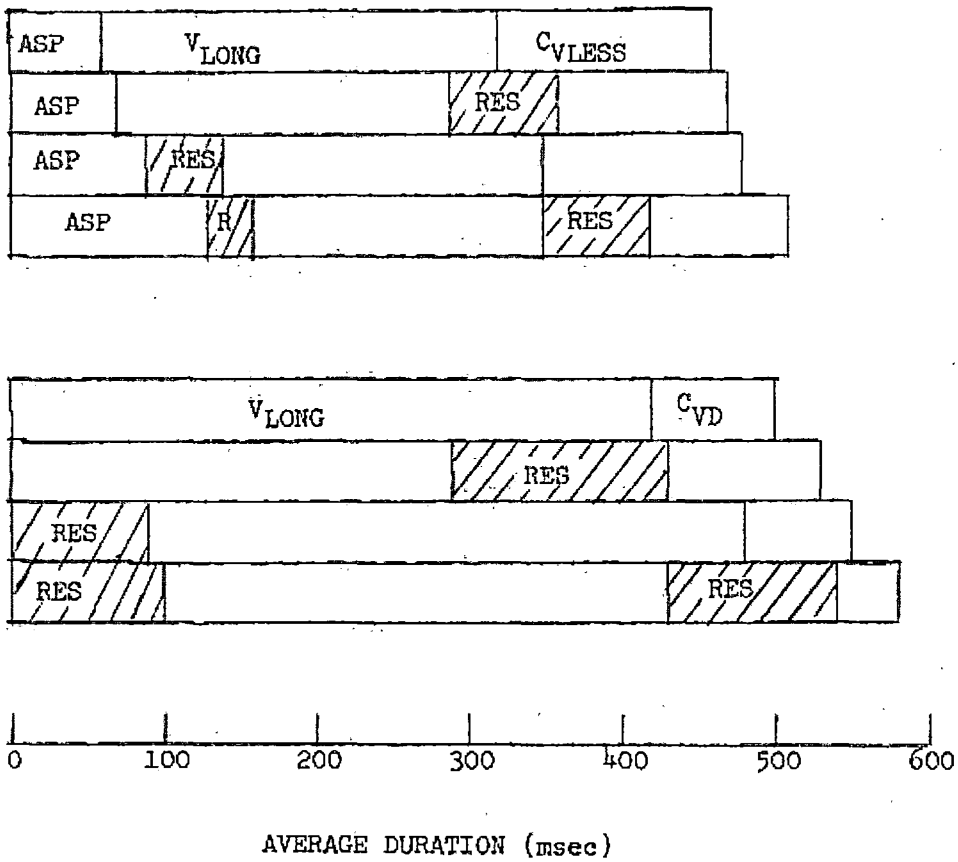
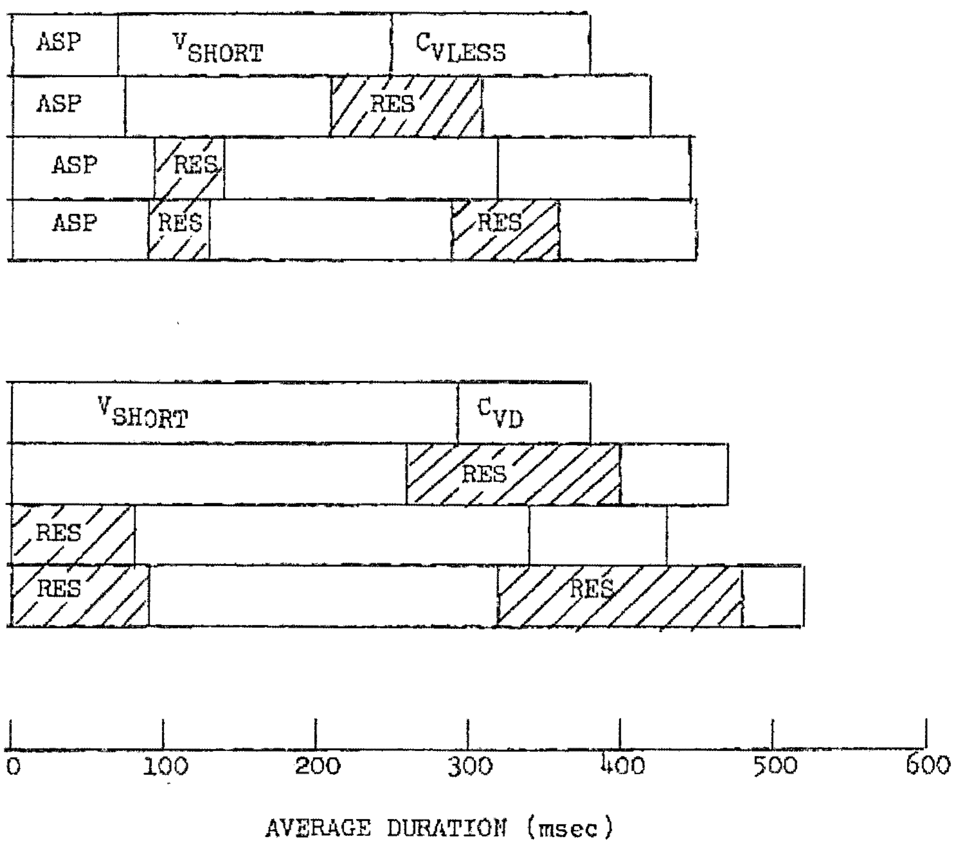

ASP
V LONG
C VLESS
ASP
RES
ASP
RES
ASP
R
RES
V LONG
C VD
RES
RES
RES
RES
0
100
200
300
400
500
600
AVERAGE DURATION (msec)

Fig. 3



ASP
VSHORT
CVLESS
ASP
RES
ASP
RES
ASP
RES
RES
VSHORT
CVD
RES
RES
RES
RES
0
100
200
300
400
500
600
AVERAGE DURATION (msec)

Fig. 4

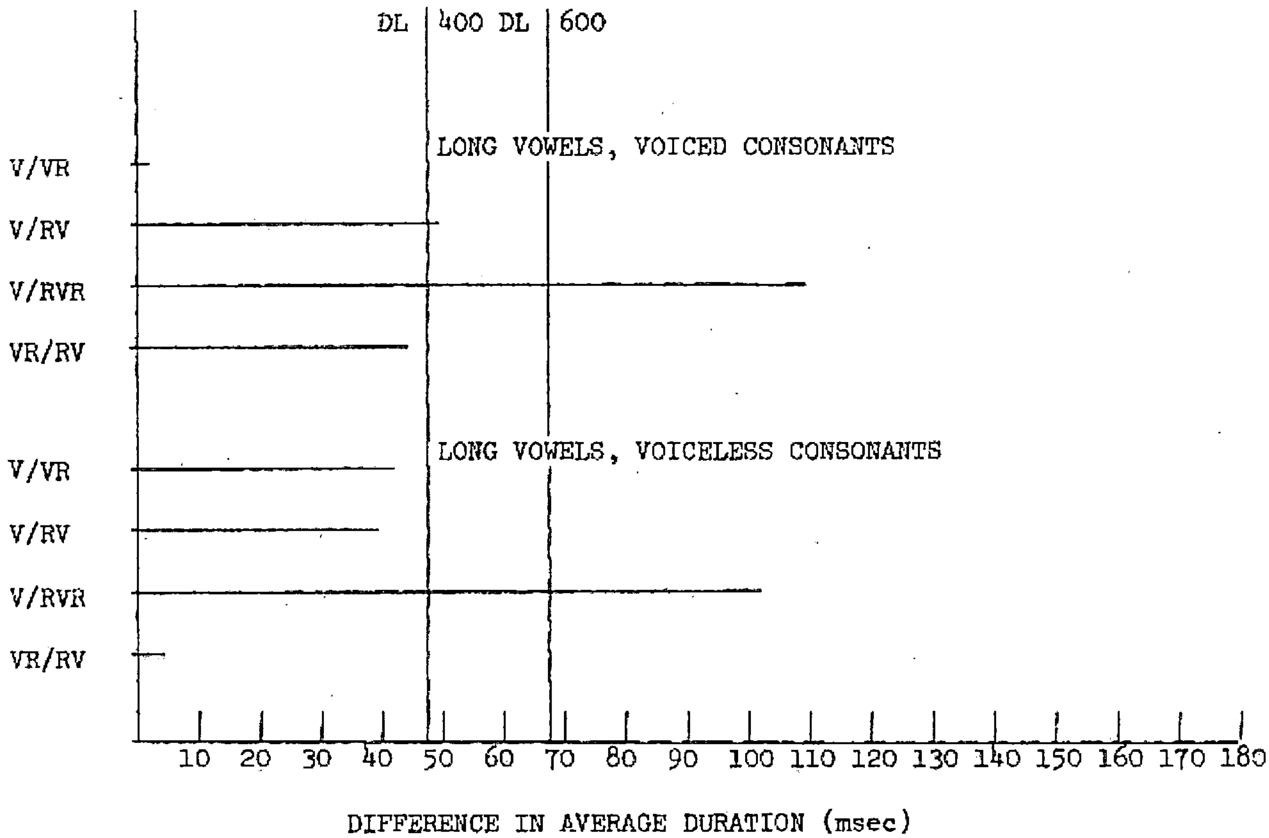
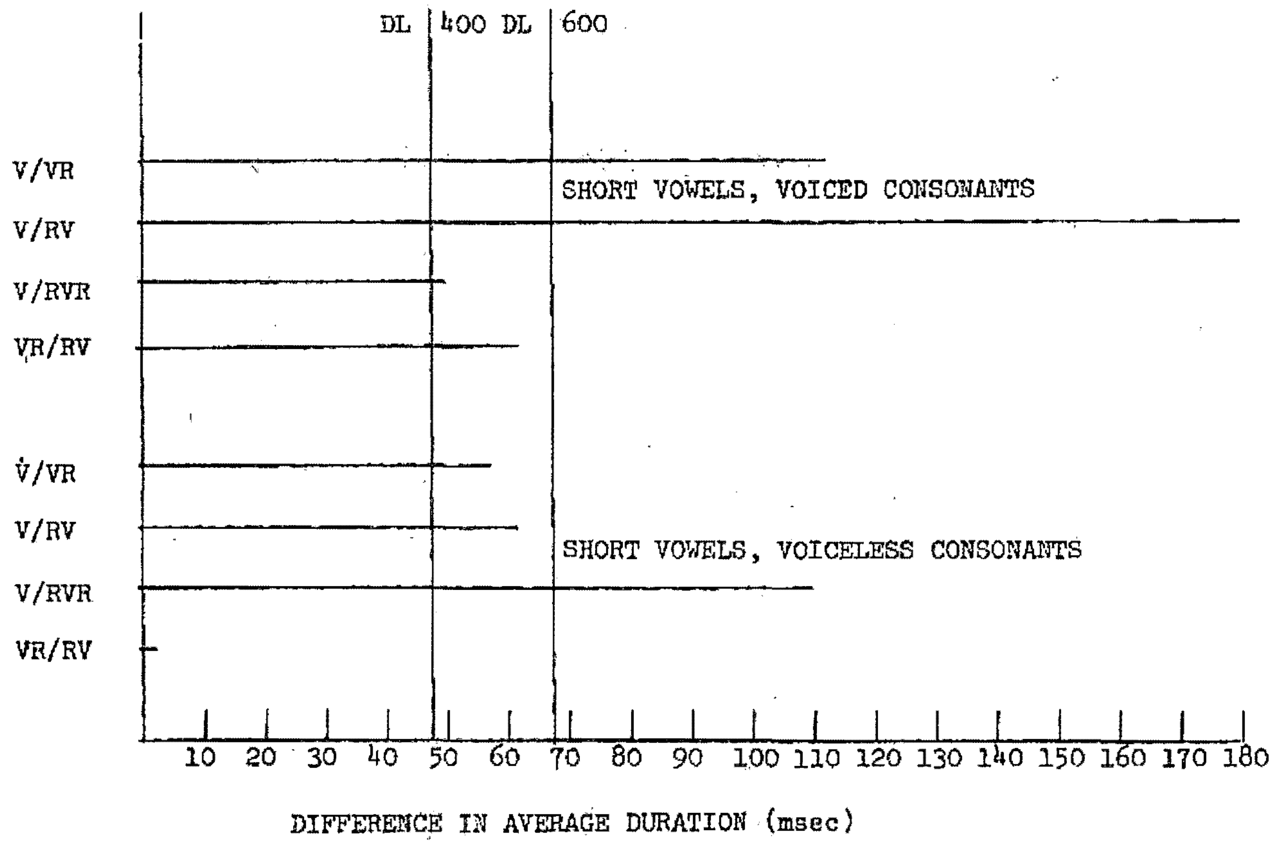

DL
400 DL
600
LONG VOWELS, VOICED CONSONANTS
V/VR
V/RV
V/RVR
VR/RV
LONG VOWELS, VOICELESS CONSONANTS
V/VR
V/RV
V/RVR
VR/RV
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180
DIFFERENCE IN AVERAGE DURATION (msec)

Fig. 5


DL
400 DL
600
V/VR
V/RV
V/RVR
VR/RV
SHORT VOWELS, VOICED CONSONANTS
V̇/VR
V/RV
V/RVR
VR/RV
SHORT VOWELS, VOICELESS CONSONANTS
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180
DIFFERENCE IN AVERAGE DURATION (msec)