The Journal of the Acoustical Society of America

MARCH 01 1977

Consonant duration in American English

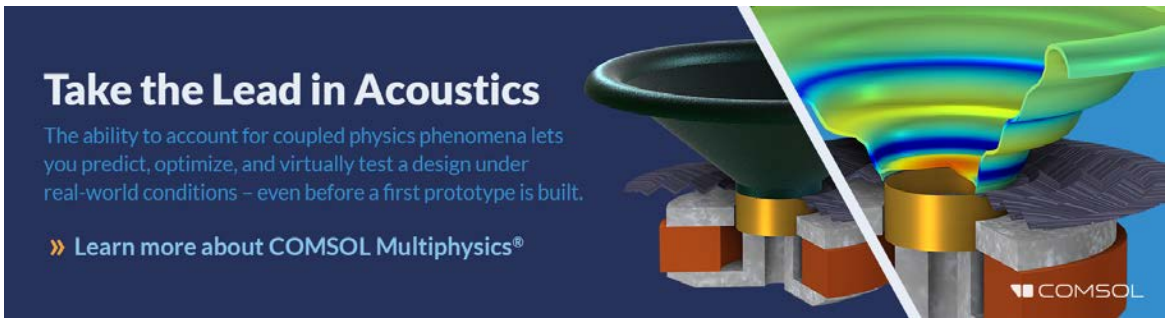
Noriko Umeda

Check for updates

J. Acoust. Soc. Am. 61, 846–858 (1977)
https://doi.org/10.1121/1.381374

View Online
Export Citation
CrossMark

10 December 2023 20:55:07



Consonant duration in American English

Noriko Umeda

Bell Laboratories, Murray Hill, New Jersey 07974
(Received 24 April 1976; revised 17 November 1976)

The paper discusses the temporal behavior of all measurable consonants, detailed in all possible conditions, in an extensive reading by one speaker. The data indicate a strong parallelism in duration distributions among similar kinds of consonants, and interesting similarities and differences between different kinds of consonants in terms of phoneme-sequential constraints and higher-level linguistic factors.

PACS numbers: 43.70.Gr, 43.70.Ve

INTRODUCTION

"Consonant" is a term given to many different kinds of speech sounds whose manners and places of articulation differ and whose phonological distributions (phoneme-sequential constraints) vary widely. Some consonants are occlusive, and some are very vowel-like. Probably for this reason, studies of consonant duration have been limited to a certain group of sounds under particular conditions (Barnwell, 1971; Klatt, 1975a; Lindblom and Rapp, 1973; Lisker and Abrahmson, 1967; Malecot and Lloyd, 1968; Scharf, 1962).

The present paper summarizes the temporal behavior of consonants which have been studied over several years. Data reported in this paper were taken from the same material with which we have studied various properties of speech (Umeda and Coker, 1974; Coker and Umeda, 1975; Umeda, 1975). This material, which is a 20-min reading of an essay by a male speaker (referred to as SP in our previous studies), produced about 650 spectrograms. His data were checked with those of other speakers for any peculiarity in pronunciation. On the spectrogram, a phoneme boundary was determined at a discontinuity in excitation, formant structure, or both. Three people were engaged in the measurements, checking their results with one another.

Some consonants such as /h/, /r/, /w/, /y/, and word-final /l/ were totally impossible to measure, and for many other consonants we had many unmeasurable tokens. Many word-initial /ð/'s were discarded because of unclear boundary with their preceding phoneme. Dental consonants (/s/, /t/, /d/, and /n/) have a far greater number of occurrences in our material than consonants with other places of articulation. These dental consonants provide about 600 to more than 800 tokens (/d/ and /z/ have slightly less tokens), whereas other consonants (/f/, /p/, /m/, /k/, etc.) usually provide 300 or fewer tokens. The number of occurrences of measurable tokens for all consonants reported in this paper is shown in Table I.

Several factors which are claimed to be important by other investigators in the determination of phoneme duration did not demonstrate their importance in our consonant data. They are: (1) number of syllables in the word, (2) phrase-final position without a silence following, and (3) identity of the vowel preceding or following the consonant.

(1) *Number of syllables in the word.* This factor is a mild constraint of rhythm compensation effects. When the material is uniform such as a list of words or a list of carrier phrases, the rhythm in reading becomes quite regular. When the main purpose of reading is to convey constantly occurring new ideas in a discourse to the listener, the importance of the rhythm of the phrase diminishes considerably. The difference between the two situations in the effect of this factor using vowel durations has been shown elsewhere (Harris and Umeda, 1974). In Fig. 1, using the single, stressed, word-initial /s/ as an example, the effect of this factor with the data studied in this paper is shown together with the effect of the frequency of occurrence of the word in English (Kučera and Francis, 1970). (Data for the most frequent category are for /f/ in "for," because there is no function word in the rank order up to 22 that starts with /s/ in English.) This figure indicates that the frequency of occurrence, which is a rough measure of the information load of the word, is a more important factor in discourse than the number of syllables in the word.

TABLE I. Number of measurable tokens for each consonant. Two numbers separated by a slash indicate the number of tokens for closure and the number for aspiration, respectively.

	Word initial	Word medial	Word final	Total
p	124/148	128/128	23/16	275/292
t	145/178	342/342	196/185	682/705
k	120/135	183/179	67/68	370/382
b	118	88	5	211
d	93	68	144	305
g	25	47	1	73
tʃ	6/6	26/26	27/27	59/59
dʒ	18/20	35/35	10/10	63/65
tr	21/23	23/23		44/46
θ	20	15	13	48
f	100	69	41	210
s	162	222	242	626
ʃ	15	121	5	141
ð	152	25	13	190
v	26	90	124	240
z		78	370	448
ʒ		10		10
m	125	188	98	411
n	213	309	318	840
ŋ		13	95	108
l	80	34		114

Copyright © 1977 by the Acoustical Society of America

10 December 2023 20:55:07

IPR2026-00275
Krisp Technologies EX1035 Page 2

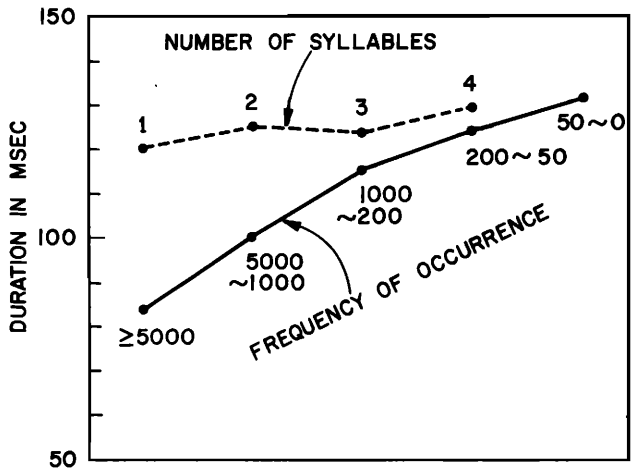


FIG. 1. Duration of word-initial, single, stressed /s/ as a function of number of syllables in the word and the frequency of occurrence of the word (Umeda, 1976).

(2) *Phrase-final effect when no silence follows (pseudopause).* The well-known effect of elongation when the unit (word, syllable, or syllable nucleus) is followed by a pseudopause has been studied in connection with grammatical structure by several investigators (Goldhor, 1975; Klatt, 1975b; Cooper, 1976; Lindblom, Lyborg, and Holmgren, 1976). In dealing with an extensive reading, however, we often find that the speaker does not make a pause at an expected grammatical boundary, and makes one at places where there is no such boundary. This is probably because the material contains a complexity not only in syntactic structure, but in semantic relations between words or phrases. Because of this fact, we classified the data according to perceived boundaries that three listeners have detected (Umeda, Harris, and Forest, 1975). The results show that, unlike vowels, consonants are significantly lengthened before a physical pause, but not before a pseudopause.

(3) *Identity of the vowel adjacent to the consonant.* In isolated, nonsense utterences, a sequential effect of vowels on consonants is seen (Schwartz, 1970), but our data show that this effect is not seen in data from an extensive reading. This is probably because the effect is overshadowed by other higher-level factors when the speech material increases in complexity.

Two kinds of conditions were considered: (i) phoneme-sequential conditions both within and across word boundaries; and (ii) suprasegmental factors such as word and phrase boundaries, stress situations, and the position of the consonant in a word (and in a syllable, if necessary). In vowel environments (i.e., no consonant adjacent to the one being measured) only suprasegmental factors are shown in this paper, for vowels have no severe constriction along the vocal tract, and no strong conflict in articulation with adjacent consonants, and, consequently, no consistent effect of vowel identity is seen in the data distribution.

In consonantal environments, the first conditions, phoneme-sequential ones, have to be taken into account as well as boundary and stress conditions. Phoneme-sequential conditions are manners or places of articulation of the two consonants, their excitation modes, and the relative consonantalness[1] of the two. These conditions influence the duration of a consonant in various ways. Unlike those in vowel environments, consonants in consonantal environments have to be discussed separately according to the kinds of consonants they are adjacent to and their positions relative to word boundary, stress and syllable inside the word.

In short, the consonant durations are a function of the following factors: (1) position of the consonant in the word, (2) its relation to lexical stress and morpheme boundary (if any) within the word, (3) whether it is in the postpausal position, (4) whether it is in the prepausal position, (5) content-function difference of the word, and (6) effect of adjacent consonant both inside the word and across the word boundary.

I. CONSONANTS IN VOWEL ENVIRONMENTS

A. General trend

Major factors that determine the duration of a consonant in a vowel environment (i.e., located between two vowels) are (1) the position of the word-boundary—whether the consonant is immediately after a word boundary, before the boundary, or inside a word; and (2) the position of stress—whether or not the consonant is at the head of a stressed syllable.[2] The possible combinations of the boundary and stress factors are five (a word-final consonant can not occur at the head of a stressed syllable). For some consonants a further subclassification of the word-medial condition is necessary. One is the case where the consonant possesses the feature of an initial allophone (Umeda and Coker, 1974), as in initial position of the second half of a compoundlike word (e.g., "peapod") or at the head of such endings as "-ness" or "-less" (e.g., "happiness"). Another is the case where the consonant falls at the end of a prefix followed by a vowel (e.g., "unaware"). These two cases show consonant durations different from those in the unbounded intervocalic situation.

Postpausal and prepausal situations are also considered. The pause is defined as the presence of a silence (regardless of its length); the so-called pseudopause (for its definition see Umeda, 1975) is excluded from the pause condition. The reason for excluding it is that the presence of a pseudopause does not influence the duration of a preceding consonant so greatly as it does the duration of a preceding vowel. For example, average durations of /f/, /s/, and /n/ at the place where three listeners agreed upon the presence of a pseudopause (Umeda *et al.*, 1975) in vowel environments are 88, 100, and 45 msec, respectively. These values indicate that, although vowels before a pseudopause are greatly lengthened, consonant durations in this position is very close to that in word-final, nonprepausal positions.[3]

In Fig. 2, we see a parallel pattern of durational distributions in three groups of consonants that have closure along the vocal tract: voiceless stops, voiced stops, and nasals. In the figure, /x/ stands for the

10 December 2023 20:55:07

IPR2026-00275
Krisp Technologies EX1035 Page 3

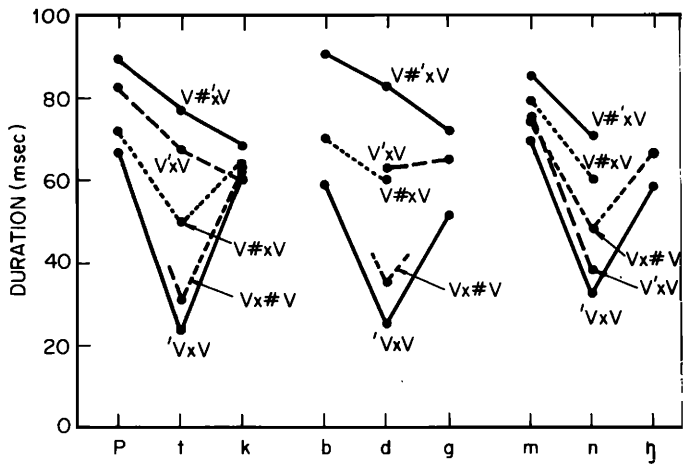


FIG. 2. Measured duration of voiceless stops, voiced stops and nasals in vowel environments under various suprasegmental conditions. For explanation of notation see text.

consonant in question. The symbol ϕ stands for pause, # for word boundary, ' for primary stress, and *V* for vowel. (These notations, as well as *C* for a consonant, *N* for a nasal, and + for an obvious boundary inside a word such as in "bright+ly" and "bath+tub," are used throughout the paper.) Durations of voiceless stops do not include the aspiration period. (For the relationship between closure time and aspiration period, see below.)

The overall patterns of these three groups of consonants are very similar. In the word-initial stressed condition, labial phonemes are the longest, dentals are shorter than labials, and velar consonants are the shortest. Velar consonants show the smallest range of duration difference in terms of the phonological conditions shown in the figure. Dental consonants yield the greatest range of difference, for they become flap consonants in intervocalic position. Other conditions line up neatly between the word-initial stressed and the intervocalic poststressed ones.

Table II summarizes this consistency in a simple additive form with the word-initial stressed condition

TABLE II. Consonant duration difference in vowel environments from the value of the word-initial, stressed condition.

	V#'xV	V'xV	V#xV	Vx#V	'VxV	ϕ'xV	Vxϕ
p	89	−3	−18	0?	−22		+21?
t	77	−10	−27?	−46	−52		−31
k	69	−9	−15	−6	−8		+9
b	90		−20?		−33		
d	83	−20	−23	−45	−57		−7
g	69	−2			−16		
tʃ	58?	+2?		−10?	−19		+9?
dʒ	58	+5?		−9?	−20		
θ	119			−21?	−36?		
f	122	−13		−39	−36		+30
s	129	−9	−23	−34	−39	−10	+20
ʃ	118	−18?			−20	−40?	+72?
v	78	−13		−22	−10		−8
z	85*	−12		−23	−17		+6
ʒ	85*				−13		
m	86	−12	−6	−13	−16	−35	+4
n	71	−33	−11?	−23	−37	−38?	+10
ŋ	70*			−3	−12		+13
l	66	−19			−26	−29	

as the base duration (for the discussion of fricatives, see below). Numbers shown with an asterisk are values hypothesized from the general tendancy only for the sake of a rule making. Consonants in these conditions did not occur in our material. Numbers with a question mark are based on less than ten tokens.

Voiceless stops and voiced stops in the initial position of a word-medial syllable have slightly shorter durations than those in word-initial stressed position. On the other hand, nasals (and probably /l/, too) in the initial position of a word-medial stressed syllable become far shorter than those in a word-initial stressed syllable (see Table II).

When a pause precedes the consonant, it becomes considerably shorter than it does in word-initial but phrase-medial positions. In this condition, voiceless stops are unmeasurable; most voiced stops become unvoiced and so are unmeasurable except for a few cases of /g/. Nasals and /l/ in this situation are very short with very low intensity (Umeda and Coker, 1974). In the prepausal situation, the duration of /p/, /k/, and all nasals becomes greater than that in any other condition. /t/ and /d/ in this situation are shorter than those in the word-initial position, but longer than those in word-medial or word-final situations.

At first inspection of Fig. 3, durational distribution patterns for fricatives look different from those for stops. However, by examining the pattern for voiceless fricatives more closely, one will find that the be-

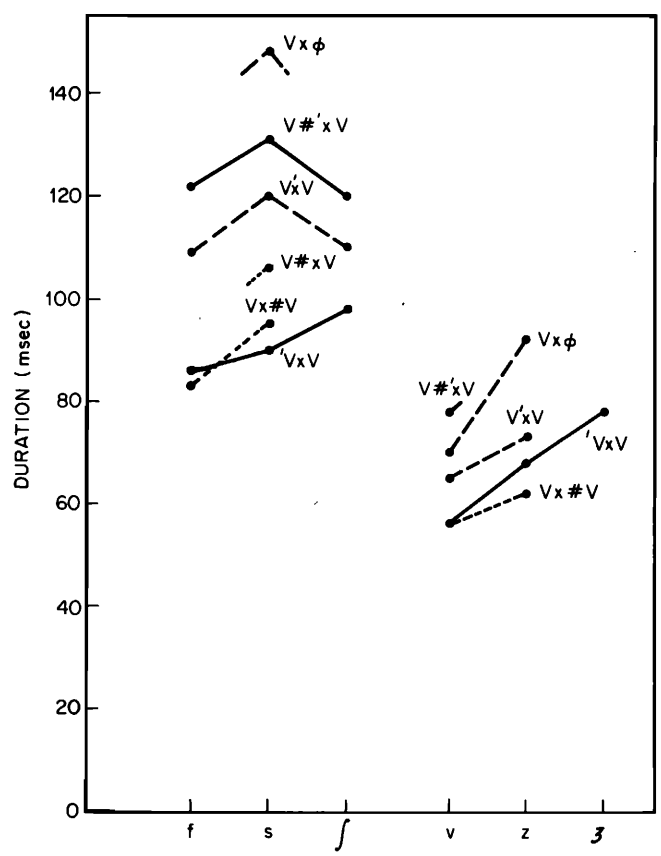


FIG. 3. Measured duration of voiced and voiceless fricatives in vowel environments.

10 December 2023 20:55:07

IPR2026-00275
Krisp Technologies EX1035 Page 4

havior of /f/ (labial) and /ʃ/ (dorsal) is similar to that for stops in the following sense; /f/ is longer than /ʃ/ in word-initial stressed position, and /ʃ/ does not markedly change its duration according to different conditions. Only /s/ acts differently from dental stops; word-initial stressed /s/ is the longest of the voiceless fricatives, and its range of durational distribution is no greater than that of /f/. Although the evidence for voiced fricatives is incomplete (there is not even a single occurrence of /z/ in word-initial positions in our material, and /ʒ/ occurs only in the intervocalic position), the pattern of their durational behavior appears to be similar to that of their voiceless counterparts, with considerably shorter individual durations. The durational tendency of fricatives, including /θ/ and the closure period of /tʃ/ and /dʒ/, is also shown in Table II in an additive form.

Thus, summarizing Table II, one sees, in nonpausal conditions, monotonous decreases in duration along the parameter shown at the top. Exceptions are: voiceless stops /p/ and /k/ become fairly short in the word-initial unstressed position; in nasals, the stressed syllable initial condition (which is next to the longest condition for other consonants) is almost as short as the intervocalic one; and voiced fricatives show one reverse order between word-final and intervocalic situations. In postpausal situations consonants become short. The durational pattern of the prepausal situation is not very consistent among consonants.

/ð/ is a peculiar consonant; it does not appear in the initial position of full content words in English. Its phonetic character is rather glidelike. (This means that the consonant does not hold a constant articulatory manner like /s/ or /ʃ/, but the constriction made by tongue tip often turns into a light closure.) The duration of this consonant, therefore, can not be treated within the general pattern of fricatives. As for word-initial /ð/, the average duration is 52 msec when it is preceded by a vowel. A slight difference is seen between durational values in "those–these" (62 msec) and in other function words such as "that," "this," "the," and the "they" class (48 msec). The consonant is longer when a voiced consonant /v/ or /z/ precedes it (41 msec) than when a voiceless consonant /t/, /k/, /f/, or /tʃ/ precedes it (27 msec). /s/ is an exception among voiceless consonants; when it precedes /ð/, the latter becomes long (53 msec). The duration of the intervocalic, poststressed /ð/ such as in "farther" is 46 msec. The identity of the consonant in "with" is sometimes /θ/ and sometimes /ð/, it is said (see Kenyon and Knott, 1953). However, the duration of the consonant in this word is far shorter (64 msec) than the consonant occurring in the final position of a content word such as "truth," in which the duration is 100 msec.

B. Influence of stress value of the vowel on unstressed consonants

In most cases the stress value of the preceding vowel does not influence the duration of an unstressed consonant, but there are some cases in which it does. One such case is intervocalic /t/. The poststressed /t/ (e.g., in "sitting" and "writer") is always voiced and the shortest consonant in English. However, /t/ after an unstressed vowel (e.g., in "visitor" and "falsity") often becomes devoiced and has a clear aspiration after the closure. Thus the closure duration of /t/ in this situation tends to be longer than that in the poststressed position (39 versus 25 msec; for aspiration see below). The same tendency is seen in the other dental consonants /d/ and /n/ to a lesser degree (/d/ in /VdV/ 31 msec versus 26 msec in /'VdV/; and /n/ in /VnV/ 39 msec versus 33 msec in /'VnV/).

It is necessary to mention the nature of the intervocalic flap /t/. A number of linguists and phoneticians contend that the intervocalic /t/ does not retain the identity of /t/ but rather becomes a /d/. As far as our duration data are concerned, /t/ and /d/ behave very similarly, showing almost identical continua from the condition that requires the longest duration to that which requires the shortest (that is, the intervocalic flap). One reason that the intervocalic flap /t/ is regarded as /d/ is that the consonant is voiced in this position. This is true with our data. However, the consonant is sometimes voiced and sometimes not voiced in situations that are slightly longer than the shortest as in between two unstressed vowels (as in "visitor") or in word-final situations; in much longer situations, the consonant is never voiced. This means that for the excitation manner, too, the voiced intervocalic /t/ is one end of the continuum of /t/ identity. Therefore to regard the flap /t/ as a separate phoneme from /t/ is not an appropriate interpretation of our data.

In other consonants, either there is no difference between poststressed intervocalic consonants and those between two unstressed vowels (/k/, /g/, /m/, /ŋ/, /s/, /f/, /v/, and /z/); or the direction is opposite to that of dental consonants, that is, the consonant after an unstressed vowel is shorter than that in poststressed position (/p/, /b/, and /ʃ/). One probable reason that dentals are different from other consonants is that dentals are very lax in intervocalic positions (as flaps) and are weakened with the increase of the effort made by their preceding vowel.

Another interesting case is /n/ in the /V+nV/ situation, that is, /n/ in the suffix "-ness" preceded by a vowel (e.g., "happiness). /n/ in this situation clearly has the quality of an initial allophone in our intensity data, that is about 3 dB lower than that in the intervocalic situation without such a boundary (Umeda and Coker, 1974). The duration data, too, show that /n/ in "-ness" is far longer (55 msec) than it is in an intervocalic situation without a boundary (33 msec) or it is in word-medial stressed position (38 msec). In fact, the duration of /n/ in "-ness" is very close to the figure for word-initial stressed position. This seems to mean that the speaker makes some special effort to distinguish /n/ in this situation from that in the unbounded intervocalic position.

10 December 2023 20:55:07

IPR2026-00275
Krisp Technologies EX1035 Page 5

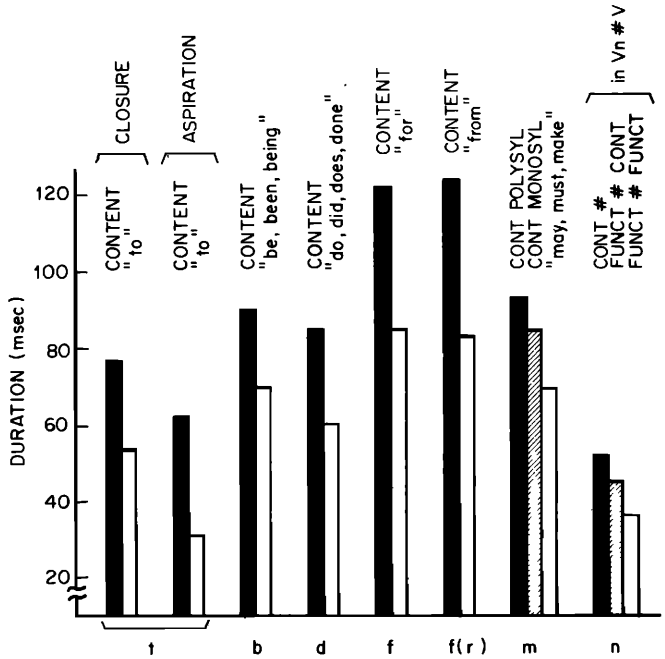


FIG. 4. Comparison between the duration of word-initial consonants in content words and that in function words. The comparison is also made for word-final /n/.

C. Content words versus function words

Content words usually carry important information concerning the content of the message, and so they are pronounced with considerable care, whereas function words are easy to guess from the context and pronounced with minimum effort. Therefore it is not surprising to find that this difference between content words and function words is reflected in the duration of the consonant at the edge of the word as well as in vowel duration and other acoustic parameters. Our data show that there is always a significant difference between the two situations in terms of the duration of the word-initial consonant, but the word-final consonant (except /n/) does not show a consistent difference.

Figure 4 shows the durational difference between word-initial stressed consonants in content words and those in function words, both in vowel environments. Comparisons are made in the figure between closure duration of /t/ in content words and that in "to"; aspiration time in these two cases; duration of /f/ in content words and in the function words "for" and "from"; duration of /b/ in content words and in "be" verbs (be, been, and being); duration of /d/ in content words and in "do" verbs (do, did, does, and done); and duration of /m/ in content words and in "may," "might," and "must." There is a 20- to more than 40-msec difference in the mean duration of initial consonants in terms of content-function distinction. The data distributions of these two situations—content and function—are clearly distinguished.

The rightmost group of bars in Fig. 4 shows the difference in mean durations of three final /n/ situations: (1) the final position of a content word, (2) the end of a function word which is followed by a content word, and (3) the end of a function word followed by another function word (such as "in a"). There is a difference of about 16 msec in mean duration between word-final /n/ in content words and in function words. This difference is smaller than that in any word-initial distinction. Moreover, unlike the word-initial situation, the data distributions of the two situations overlap substantially. Word-final /n/ in a content word can be as brief as in the minimum function-word situation (20 msec) or as much as twice the corresponding maximum (50 msec ×2). There is no consistency in this broad distribution of content-word /n/ in terms of tenseness or laxness in the preceding vowel, of its stressed–unstressed distinction, the number of syllables constituting the word, or whether a content word or a function word is following.

D. Variance of data in vowel environments

The standard deviation for each consonant in various vowel environments is shown in Table III. There is no clear tendency toward correlation between a greater standard deviation and a longer consonant duration; but rather the amount of variance depends on the kind of consonants and their phonological conditions. This implies that the accuracy of control in timing in phoneme production is kept constant regardless of the absolute duration of the consonant.

In general, both voiced and voiceless stops (/p/, /t/, /k/, /b/, /d/, /g/, /tʃ/, and /dʒ/) show a smaller variance than other consonants. /f/ in intraword situations scatters over a greater range than any other consonant in these situations. In cross-condition comparison, the intervocalic situation shows the smallest standard deviations. Before and after a pause, the consonant shows the greatest range. Word-initial, stressed consonants show larger variance than do those in word-medial situations. Word-initial, unstressed consonants, though their occurrences are not great and

TABLE III. Standard deviation of durations for each consonant in vowel environments.

	ϕ'xV	V#'xV	V'xV	V#xV	'VxV	Vx#V	Vxϕ
p		15.1	16.3	11.2	15.1	?	?
t		13.3	12.1	?	8.2	12.0	20.9
k		17.1	11.3	13.8	13.0	15.8	12.3
b		14.1	?	?	7.0		
d		15.0	14.2	11.3	6.6	8.0	18.7
g		12.5	8.9		7.4		
tʃ		?		?	8.3	?	?
dʒ		10.9	?	?	6.6	?	
θ		23.1			?	?	?
f	?	14.1	21.9		20.6	10.5	27.6
s	22.1	20.0	14.8	15.9	12.7	11.2	26.7
ʃ	?	13.4	?	?	9.7		?
v	?	11.9	15.3		12.9	12.5	25.6
z			8.3		13.8	11.4	23.7
ʒ					3.7		
m	27.1	17.1	10.7	17.1	13.1	16.0	22.6
n	?	18.6	14.2	?	11.4	16.2	21.4
ŋ					14.7	12.0	18.3
l	19.1	21.0	13.1		7.8		

IPR2026-00275
Krisp Technologies EX1035 Page 6

10 December 2023 20:55:07

TABLE IV. Measured duration of constants in natural clusters in word-initial, stressed situations.

	V#'xV (msec)	Followed by r	l	p	t	k	tr	s	Preceded by s,f	b,g	p,k
f	122	+2	+15								
s	129			−43	−42	−39	−31				
p	89	−19	−14					−13			
t	77	−30						−36			
k	69		−4					−21			
b	90										
d	83	−18									
g	67		+6								
l	66								−19	−8	+3

are limited to a certain consonants, show smaller variances than their stressed counterparts. Word-final consonants show a similar tendancy to intervocalic ones, with a slightly larger standard deviation.

The interpretation of this result is as follows: The great variance in phrase initial consonants is due to the instability of phonation after silence; and that of phrase-final consonants is due more to the option given to the speaker according to the degree of grammatical and semantic distance from (i.e., continuation to or separation from) the following phrase. The durations of word-initial, stressed consonants scatter more widely than do those inside the word, because there is a wide range in the promotion–demotion value of the word (conventionally called stress) and in the grouping–separation function between words (conventionally, boundary). The reason for the small variance in word-medial and word-final consonants is that they are affected more by lower level factors, that is, phonological environments, than by these higher-level prosodic functions.

II. CONSONANTS IN CLUSTERS

Phoneme-sequential factors act upon the duration of a consonant in various ways: The duration of a consonant is very different when it has a conflicting gesture with its adjacent consonant from when it has an overlapping one. Even the same sequence of consonants is affected differently depending on their position relative to the juncture (i.e., word boundary, stress, or syllable boundary, when definable). There is no general duration rule to govern all consonants in cluster situations as there is for those in vowel environments. Discussion has to be undertaken individually, in other words, consonant by consonant and condition by condition.

A. Clusters in the initial position of word or of stressed syllable

In this section, we will discuss the word-initial or stressed-syllable initial consonant cluster when it is preceded by a vowel. Only stops or voiceless fricatives occur as the first member of such clusters. Voiced-stop clusters with /r/ or /l/ occur very infrequently in our text; only word-initial stressed /dr/ and /gl/ have a reasonable number of occurrences. The /kr/ cluster preceded by a vowel occurs only three times in the entire text and so is excluded from the discussion. The /sn/ or /sm/ clusters never occur in the text; /sl/ occurs only a few times.

1. *Voiceless fricatives in clusters*

/s/ clustered with a voiceless stop is far shorter (about 40 msec less) than its single counterpart. These results are shown in Table IV. This tendancy holds true for fricatives in the initial position of a stressed syllable inside a word, though the number of their occurrences is considerably smaller: the duration of /s/ in /V'sV/ is 120 msec, whereas that in /V'sp/ 80 msec and in /V'st/ 83 msec.

It has been reported that /s/ in a word-initial cluster with /p/ is significantly shorter than that with /t/ or /k/ (Schwartz, 1970). However, our data taken from natural reading of an extensive text indicate that there is only a negligible difference between /s/ in these three situations (in the order of 4 msec, not 20 msec as in Schwartz's data). Only when /tr/ follows does /s/ become significantly longer than in these three situations. Here again, we see the difference between factors controlling timing in an extensive reading and those in carrier phrases (Harris and Umeda, 1974). There is, however, a noticeable difference between the variance of /s/ duration when followed by /p/ and that when followed by /t/; the former scatters from 70 to 110 msec, the latter only from 80 to 90 msec (and over a greater number of tokens). This means that using the same articulator (in this case, tongue tip), the accuracy of timing control for the /st/ sequence is very high.

When these clusters occur at the head of a stressed syllable inside a word, the situation is very similar to that in word-initial ones: /s/ is 40 msec shorter than in the /V'xV/ situation when /p/ follows the fricative, and 37 msec shorter when /t/ follows it. When a consonant precedes these clusters, a great difference is seen between the /s/ duration in word-initial position and that in word-medial stressed position: word-medial, stress-initial /s/ when preceded by one of /p/, /t/, and /k/ or /n/ is 56 msec shorter than that in /V'xV/, whereas when preceded by a consonant across the word boundary it is only 13 msec shorter. The word-initial /s/ in these clusters when preceded by /n/ or

IPR2026-00275
Krisp Technologies EX1035 Page 7

10 December 2023 20:55:07

/l/ is 25 msec shorter than that preceded by a vowel.

/s/ followed by a nasal has not a single occurrence in the text that speaker *SP* read, but according to the data from another speaker, JH, with a different text, /s/ followed by a nasal or a liquid is shortened about 30 msec (in his speech /s/ in /V#'sV/ is 153 msec, whereas in /V#'sn/ it is 124 msec). His reading is considerably slower than SP's. In the /#'sl/ situation /s/ seems to be slightly longer than in /#'SN/ situations, but shorter than in the vowel environment.

Other fricatives are accompanied only by liquids. Among them, there was no single occurrence of /#θr-/, and only one occurrence of /#θr-/. /#'fr-/ and /#'fl-/ were more popular in our material; /f/ in /V#'fr-/ is almost as long as single /f/ in the equivalent position; and /f/ in /V#'fl-/ is about 15 msec longer than its single counterpart.

2. *Stops in clusters*

The voiceless stops /p/, /t/, and /k/ can stand as the first member of a cluster with /r/ or /l/ (no /tl/) as well as the second member with /s/. These consonants in both situations show a very similar durational value (aspiration time in these two cases is very different; see Sec. III). That is, /t/ becomes considerably shorter (30–36 msec) than its single counterpart (see Table IV), but /p/ and /k/ only about 15–20 msec shorter. The cluster /tr/ is recognized as an affricate in phonetics; its closure time is far shorter than that of single /t/; and its aspiration time is far greater (as discussed later).

There are not enough tokens of voiced stops to indicate a trend. Judging from only two cases (/dr/ and /gl/), one may guess that the effect of clustering on the durations of voiced stops is similar to that of voiceless stops but less pronounced.

Stops followed by /w/ occur too infrequently to show data in the paper; /tw/ occurs once, and /kw/ three times.

3. *Liquids and nasals in clusters*

As the second member of consonant clusters, /l/ is the only liquid of which at least half of the occurrences in the text were measurable. The lower part of Table IV indicates that the influence of the preceding consonant in the cluster on the duration of /l/ is greatest when that consonant is a voiceless fricative, and that a voiceless stop has the least influence. /l/ in this situation is almost always partly devoiced or aspirated. The numbers given in the table include the devoiced or aspirated portions, following the closure of the preceding stop or the friction of the preceding fricative. The duration of the devoicing period is shortest with voiceless fricatives (13 msec), medium with voiced stops (22 msec) and longest with voiceless stops (53 msec).

Not a single occurrence of nasals as the second member of the cluster with /s/ was obtained in the material that speaker SP read. For the other speaker, JH, we have a very limited number of /n/'s and /m/'s following /s/. Including the devoiced portion, nasals in this situation show almost no difference in duration from those in the vowel environment /V#'nV/.

TABLE V. Measured duration of word-initial single consonants when preceded by a consonant across the word boundary.

	V#'xV (msec)	Preceded by		
		l	Nasals	Other consonants
#'p	89	−20	−15	−13(k, s, v)
#'t	77	−13?	−33	−37
#'k	69	+6?	−1(m, n), −22(ŋ)	−13 (z = −1),
#'b	90		−27	−2
#'d	83	0?		−20
#'g	67		−9 (no /ŋ/)	−6
#'f	122	+26	+9	−13(t, d, z)
#'s	129		−26	−31(t), −24(θ), −5(d), +14(f, v)
#'ʃ	118		+21(m, n), +62(ŋ)	+35
#'θ	119		+6	−13(t, k), −46(v, z)
#'v	78		−12	−14(s), +7(z)
#'ð	52	−8	no ð portion	−25(t, k, f, tʃ), +1(s), −11(z, v)
#'m	86			−42(t, v), −14(s), +13(k, z)
#'n	71			−21(unvoice, d), −11(v, z)

B. Word-initial consonant when preceded by another consonant

The duration of the word-initial consonant is often significantly different depending on whether a consonant or a vowel terminates the preceding word. An example familiar to some is a word-initial voiceless stop preceded by a nasal, in which the stop has been said to be considerably shortened (Monsen, 1971). This difference, however, is not uniform among all voiceless stops. /p/ and /t/ are shortened equally when preceded by any of the three nasals, but /k/ is shortened only by /ŋ/ and not by /m/ or /n/.

For some consonants, any preceding consonant shortens their duration, as with /p/ and /t/; for some others, any preceding consonant lengthens their duration, as with /ʃ/. Members of a third group of consonants are shortened by certain preceding consonants and lengthened by others (e.g., /s/ and /m/). Taking the duration of the word-initial consonant in the vowel environment /V#'xV/ as the base duration, Table V summarizes the durational deviations for each word-initial consonant with its mean values (in msec) in various preceding conditions. Word-initial consonants except /ð/ in function words are excluded from the table. They are considerably shorter than those in content words.

As shown in Table V, most word-initial consonants tend to shorten when they are preceded by any consonant. There seem to be two situations in which only a small amount of shortening, or even a considerable amount of lengthening, takes place. One is the situation where one of the two consonants (such as tongue-body consonants /ʃ/, /k/, etc., or voiced fricatives) across the word boundary requires more time or effort than the other; and the other is the situation where the two consonants are produced by conflicting tongue gestures. An extra shortening occurs when the two consonants share a similar gesture, as in /ŋ#'k/ or /t#'s/.

10 December 2023 20:55:07

IPR2026-00275
Krisp Technologies EX1035 Page 8

TABLE VI. Measured duration of word-final single consonants when followed by a nasal, a liquid or a semivowel across the word-boundary.

	Vx V (msec)	Followed by l	r	Nasal	Semivowel	h
p#	90?				+7(w)	
t#	31	+43		+39		−2
k#	63			+36(m, n)	+36(w)	
d#	38			+33	+22	
f#	83			+13		−43
s#	95			−18		
v#	56	+16				
z#	62		−5	−17(m), −11(n)	−10(w)	+7
m#	73	+25			+25	
n#	48				+27	+4
ŋ#	67				+43(w)	+10

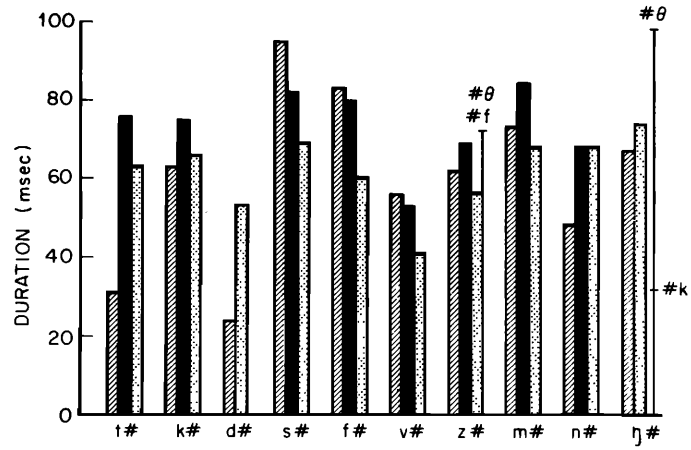


FIG. 5. The comparison between the duration of word-final single consonants followed by a vowel (hatched bars), that followed by a voiced consonant (solid bars) and that followed by a voiceless consonant (dotted bars).

C. Word-final consonants in clusters

1. *Word-final consonant followed by a consonant*

A word-final consonant is shorter when it is followed by a word beginning with a voiceless consonant (stop or fricative) than when it is followed by a voiced one. The difference is greatest in /f/ (20 msec) and smallest in /n/ (0 msec). Figure 5 shows the mean duration in both situations, a following voiced consonant and a following voiceless one, compared with the duration of the same consonant in the vowel environment /Vx# V/. /d/ has no tokens followed by a voiced consonant. Word-final /ŋ/ followed by /θ/ is 24 msec longer than when followed by other voiceless consonants. Word-final /z/ is 16 msec longer when it is followed by /f/ or /θ/ than when it is followed by other voiceless consonants. Word-initial /θ/ in the following word is an exceptional influence on the preceding word-final consonant, in that it shortens itself and lengthens the preceding word-final consonant (see Table V).

Table VI shows the duration of word-final consonants when followed by a softer consonant such as a nasal, a liquid or a semivowel. /l/ and semivowels seem to lengthen the preceding consonant; word-final /s/ before a word-initial nasal is quite short; and /h/ does not have a systematic effect on the various consonants that precede it.

2. *Word-final clusters*

Consonant durations in word-final clusters are summarized in Table VII as deviations from those in the vowel environment /Vx#V/. Tongue-tip consonants such as /t/, /d/ and /n/ are very short, like an intervocalic flap in word-final vowel environments; but in clusters with another consonant they lose the flap character and become longer (except /nd/). A word-final consonant preceded by /l/ is not shortened in /ls#/ and /lf#/; the consonant becomes long with /l/ when it is one of the tongue-tip consonants (see /lt#/ in Table VII). /θ/ and /z/ tend to lengthen their adjacent consonant. The combination of /ŋ/ and /k/ shortens the duration of both consonants.

The voice–voiceless distinction in word-final clusters with /n/ is reported to be made in isolated words by the duration of /n/ and the preceding vowel (Raphael *et al.*, 1975). In the continuous reading of a text, there is no significant difference in vowel duration between the two situations, but /n/ is shorter with a voiceless consonant than with a voiced one (the duration of /n/ is about 15 msec shorter with /t/ than with /d/ and about 30 msec shorter with /s/ than with /z/).

TABLE VII. Measured duration of consonants in word-final natural clusters.

	Vx#V (msec)	Preceded by l	n	Consonant	Followed by Consonant
t#	31	+44		+12(p, t, k, s)	
k#	63		−16(ŋ)	−8(s)	
s#	95	−5	−37	−28(p, t, k), +15(θ)	
s					−30(p, t, k)
f#	83	−1			
f					−38(t, s)
θ#	100	+20	−30		
θ					−80(s)
d#	38		−10		
z#	62	+1	−7(m, n, ŋ)	−16(d), +8(b, v)	
n	48				+15(θ), −2(s, t, tʃ), +14(d), +27(z)
ŋ	67				−18(k), +3(θ), +51(z)

IPR2026-00275
Krisp Technologies EX1035 Page 9

10 December 2023 20:55:07

TABLE VIII. Measured duration of consonants when they are the first member of the natural cluster in the word-medial position.

	'VxV (msec)	Followed by r	l	l̩	n	n̩	Consonant
p	67	+6	+28	−3			+10(t, ʃ)
t	25			+1		+59	
k	61	+3	+3				−15(t, s, tʃ, ʃ)
b	57	+15	+23	+luns +15str			−20(m), +15(s)
m	70						+28(p, f, θ), −θ(bl)
n	33						+11(d, s), −1(t), −5(tr)
f	86	+14					−51(t)
s	90		+7		−10		−19(t, k, b, m)
z	68				0		0(m)

D. Word-medial consonant clusters

The duration of each consonant in word-medial clusters is summarized in Tables VIII and IX, shown as the deviation from its duration in the vowel environment /'VxV/. Table VIII shows the duration of the consonant when another consonant follows, and Table IX when another consonant precedes. /n̩/ and /l̩/ in Table VIII indicate that these consonants are syllabic. As in word-final clusters, /l/ tends to lengthen its preceding or following consonant; and nasals in sequence with their oral cognates such as /mb/ and /ŋk/ shorten themselves as well as their followers. /n/ tends to shorten its following consonant. /s/ (or a fricative) tends to lengthen any following consonant except /k/. Stops shorten the following fricatives.

/s/ is the only consonant that falls between two hard consonants (e.g., /-ksp-/ in "expiration"). In this situation, /s/ receives a double shortening; with one consonant adjacent to it, the reduction of the /s/ duration is about 25 msec, whereas a reduction of 50 msec is observed when /s/ falls between two consonants.

E. Duration of combined consonants

Two consecutive consonants often form one segment in spectrograms when their articulatory gestures are similar. Plenty of such examples are found at word boundaries. A sequence of two identical consonants always runs together; and very frequently, the boundary between two stops, even if one is voiced and the other is voiceless, is not detectable. Two adjoining nasals make a gradual transition on the spectrogram; a voiced fricative becomes voiceless when followed by its voiceless cognate (e.g., /z#s/); and /ð/ seems to disappear completely in sequence with /n/.

Durations of these combined consonant clusters follow, quite closely, the cluster rules shown in Fig. 5 and Tables V and VII. The only exceptions are the tongue-body consonants such as /k/, /g/, or /ŋ/, which tend to shorten their total duration in combined situations more than the rule predicts; and /z/ with /s/ in the /z#s/ situation, which is actually shorter than the rule predicts. The data imply that no abbreviation of duration takes place in combined consonants except in some special situations. Some examples are shown below.

1. Example 1. t#'t

Word-initial stressed single /t/ preceded by C: 40 (Table V).

Word-final /t/ followed by unvoiced C: 63 (Fig. 5).

$D_{t\#'t} = 63 + 40 = 103$.

Measured duration: 100.

2. Example 2. Cs#'s

Word-initial stressed single /s/ preceded by /t/: 98 (Table V).

Word-final single /s/ followed by unvoiced C: 69 (Fig. 5).

Subtract 28 for cluster: 41 (Table VII).

$D_{Cs\#'s} = 98 + 41 = 139$.

Measured duration: 130.

3. Example 3. Nasal combinations

$m\#(V)$: 73 . $(V)\#'m$: 86.

$n\#(V)$: 48 . $(V)\#'n$: 71 (Table II).

$D_{m\#'n} = 144$. Measured duration: 150 .

$D_{n\#'m} = 134$. Measured duration: 130.

$D_{n\#'m} = 119$. Measured duration: 100.

Table X summarizes the durations of the various consonant combinations both as measured and as predicted by rule.

F. Summary of consonant clusters

Several studies have reported that the duration of a consonant is reduced when it forms a cluster with another consonant (Lindblom and Rapp, 1973; Schwartz, 1970; Haggard, 1973; Klatt, 1974). Explanations of this shortening have been attempted in terms of a compensation process to preserve a constant length of syllable or word; of the aerodynamics of the vocal tract; and of coarticulation. Speech materials used in these studies are either nonsense syllables or words in lists or in carrier phrases. Most of these studies discussed

TABLE IX. Measured duration of consonants when they are the second member of the natural cluster in the word-medial position.

	'VxV (msec)	Preceded by l	m	n		Consonant
p	67		−17			+10(s)
t	25	+28		+5		+21(f, s)
k	61			−4		−33(s)
b	57	+16	−20			+16(s, d, r)
d	26			−15		
g	51	0			−16	
m	70					−2(s, ʃ, dʒ)
s	90	+24		−25		−28(p, t, k)
ʃ	98			−20		−12(p, k)
z	68		0	0		

IPR2026-00275
Krisp Technologies EX1035 Page 10

10 December 2023 20:55:07

TABLE X. Measured and predicted duration of combined consonants at the word boundary.

Phoneme sequence	Measured duration	Duration by rule	Remarks
t#'tv	100	103	word-initial /t/ in content words
	80		word-initial /t/ in function words
nt#"to"	45		
t#'pV	137	139	/p/ stressed
t#pV	108		/p/ unstressed
t#bV	106		/b/ in function words or unstressed
t#'dV	125	126	/d/ in content words stressed
	106		/d/ in function words
k#'kV	108	122	
d#bV	99		/b/ in function words
d#tV	71		/t/ in function words
f#'fV	160	169	
Cs#'sV	130	139	
v#'fV	143	150	
v#'pV	125	117	
v#ð	99	97	103 when two are measurable separately
z#'sV	136	160	
z#'sC	113	140	
ð#ð	87		all occurrences are "with th.."
n#'nV	100	119	
n#'mV	130	134	
m#'nV	150	144	
ŋ#'nV	127	153	
ŋ#'mV	116	138	

only the duration of /s/.

Does this shortening take place in a similar way in an extensive reading, and with other consonants? It does generally in word-initial clusters; but most of the word-initial single consonants also become shorter when they are preceded by a consonant than when they are preceded by a vowel across the word boundary. In word-medial and word-final positions, shortening occurs as well as lengthening, depending on the kinds of consonants that are adjacent to each other. In our study, the comparison is made between a consonant surrounded by vowels or adjacent to a pause, and one adjacent to another consonant in the same stress, pausal and positional situation.

The previous investigators have found that as the size of the cluster increases, the duration of the consonant decreases. This is true with the word-medial /s/ (stressed or not) in our data. For word-final /s/ a significant difference is obtained between single /s/ and /s/ in two-consonant clusters (except when /θ/ precedes /s/), but no further reduction is obtained for /s/ in three-consonant clusters where /s/ is between two consonants. Furthermore, /t/ has a different situation. The word-final /t/ in the vowel environment (/Vt#V/) is the shortest among all final situations. The durations of /t/ in /Ct#V/, /VtC#/ or /CtC#/ are all similar and about 15 msec longer than the vowel-surrounded /t/.

When the rhythm is regulated by an external framework of utterances such as in the case of reading a list or reciting a poem, the compensation process seems to be one of the dominating factors for temporal control. In real speech communication, however, the elasticity of timing in speech units according to rules agreed upon by people who use the language is more important than preserving a constant rhythm; the timing variability among speech units conveys information, such as syntactic distance and semantic importance between the units, in the acoustic speech wave.

Our data suggest that timing adjustments result from articulatory constraints on the two adjacent consonants. The general tendency is that a consonant becomes shorter next to a consonant than next to a vowel, independently of whether there is a word boundary between them. There are several cases where a consonant is prevented from shortening; (1) when it is very short in a vowel environment, such as a poststressed intervocalic dental consonant (/t/, /d/, or /n/); (2) when its tongue gesture conflicts with that of the other consonant, as in /ʃ/ with /θ/; or (3) when the consonant is for some reason difficult to sustain. In this situation, the consonant shortens itself and lengthens the other; this occurs with /θ/ and voiced fricatives. An extreme shortening occurs when a stop is preceded by its nasal cognate (e.g., /ŋg/ and /mb/).

III. CLOSURE VERSUS ASPIRATION

In our study of phoneme duration, the closure portion of stop consonants has been considered the entire duration of the consonants, and their aspiration portion [we use this term as an equivalent to VOT used by other investigators (Klatt, 1975a; Lisker and Abramson, 1967)] has been regarded as a part of the following vowel. In this way, we have obtained consistency among vowel durations in various consonantal environments (Umeda, 1975) and parallel patterns of duration among voiceless stops, voiced stops and nasals (see Fig. 2).

It is true, however, that the aspiration is also a property of stop consonants; it carries the information about higher-level prosodic functions such as word-boundary, stress, the content-function distinction, and pause. In fact, the relationship between closure time and aspiration period shows a consistent trend in terms of phoneme-sequential constraints and of higher-level prosody.

Figures 6–8 represent the relationship between closure time on the abscissa and the aspiration period on the ordinate for /p/, /t/, and /k/, respectively. [4] For each consonant, its occurrences in detailed phonological situations form three groups, represented by three broken lines in each figure. One group, which has a linear relationship between closure time and aspiration period and lies in the center of the three groups in the figures, consists of word-initial and -medial allophones, both single and clustered, and allophones clustered with other consonants in word-final position. The second group, which has a minimum aspiration period regardless of closure time, consists of word-final and morpheme-final allophones. These two groups meet at the point with minimum closure time; this point is occupied by the intervocalic poststressed allophone. The third group, consisting of tokens preceded by nasals, occupied a position in the figures above the first group. The presence of this third group indicates that a stop, when preceded by a nasal, maintains the same aspiration period as its oral counterpart with much shorter closure time.

10 December 2023 20:55:07

IPR2026-00275
Krisp Technologies EX1035 Page 11

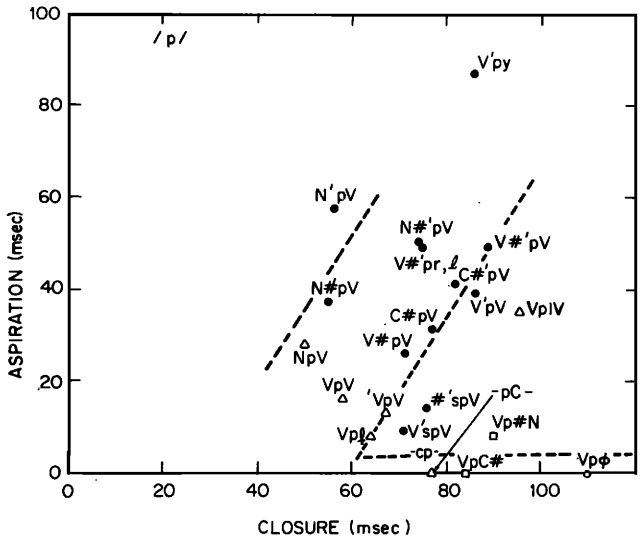


FIG. 6. The relationship between closure time and aspiration period of /p/ under various phonological conditions. A solid circle stands for an initial allophone, a triangle for medial, a square for final, and a blank circle for prepausal.

Among the three stops /t/ produces the greatest number of tokens (Table I), and so the most reliable data.

Data for affricates are shown in Fig. 9. It is said in phonetics that /t/ followed by /r/ is an affricate. And in fact, the closure-aspiration relationship is exactly parallel to that of /tʃ/. In voiceless affricates, the closure time does not vary very much but the aspiration period varies according to conditions. In voiced affricates, the situation is opposite; that is, the aspiration period is almost constant regardless of the closure period. Affricates do not show a separate trend for word-final members.

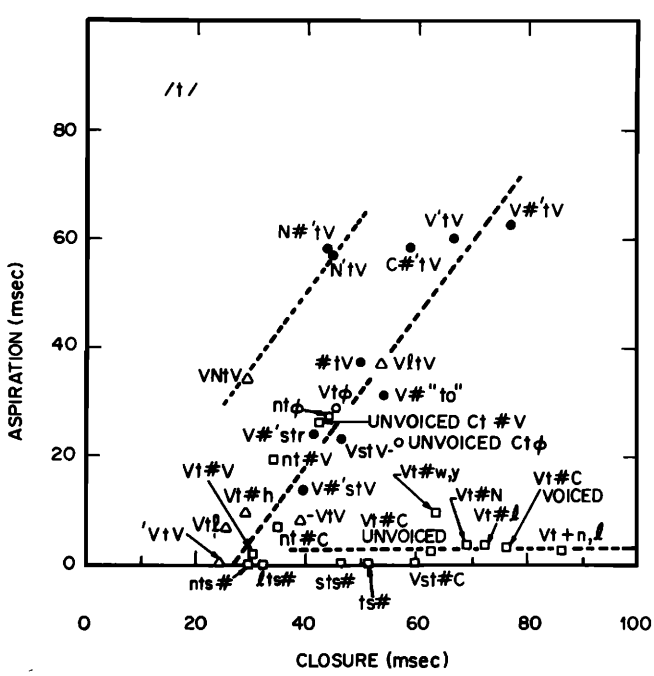


FIG. 7. The relationship between closure time and aspiration period of /t/.

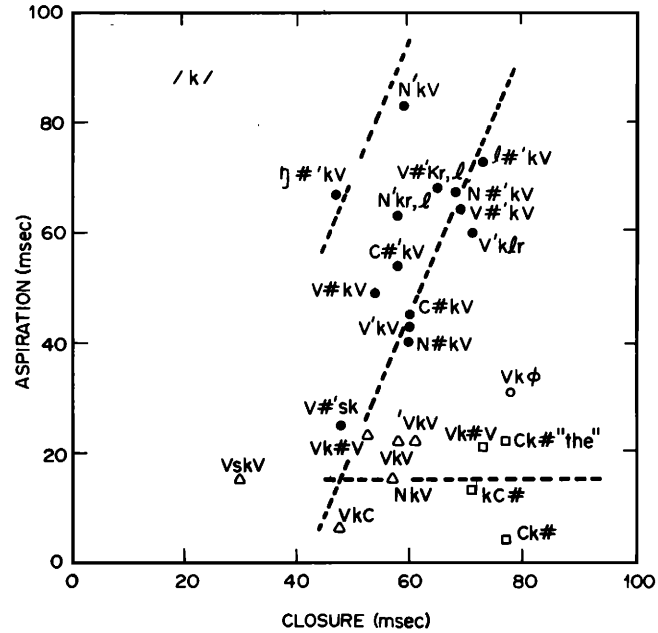


FIG. 8. The relationship between closure time and aspiration period of /k/.

IV. SUMMARY

We have discussed the durations of all consonants in various phoneme-sequential and prosodic situations except those that were unmeasurable. An additive model described here seems to work well with consonants under a wide variety of conditions. However, this model is complex because different constants have to be chosen for different consonants. Because this is the case, it appears that a model where a fixed set of constants is used for all consonants, such as proposed by Klatt (1976), could not predict the complexities of consonant duration.

Consonants show quite regular durational behavior when they are surrounded by vowels, but when adjacent to another consonant their durational patterns become complex. If it occurs 600 or more times in the text, one phoneme can appear in almost all possible situations with each condition having a reasonable number of tokens. But only the most frequent consonants—/s/, /t/, and /n/—have this large number of occurrences in our material of 20 minutes' reading.

The data show interesting consistencies and parallel distribution patterns among members of similar kinds of consonants in their equivalent phonological situations, as we have seen in Figs. 2 and 3, and 6–9. These data points distribute themselves neatly along a continuous scale of duration in terms of phoneme-sequential and prosodic conditions. As with vowel duration controls, we see that the speaker produces each consonant with a high degree of accuracy in timing. Unlike vowel duration, however, the timing control of consonant production results from an intricate harmony between the realization of prosodic functions and the facility of the sequential movement of articulators. This is the reason why consonant duration in consonantal environments yields such complex patterns.

10 December 2023 20:55:07

IPR2026-00275
Krisp Technologies EX1035 Page 12

The discussion of consonant duration in this paper is based almost entirely on one speaker's reading. As one sees throughout this paper, the number of factors by which the temporal structure of consonants in the language is determined is so immense that unless we collect a large amount of data it is impossible to cover many possible constituents of the entire structure. The collection and sorting of this amount of data is certainly very time consuming. Whether one chooses to study a small fragment of speech from many speakers' data, or to use extended material with a very limited number of speakers, depends entirely on one's interest and discipline. The major disadvantage of the former method is that the fragment, when isolated from its mother structure, forms a structure by itself, and never remains as a constituent of the larger structure. Because we wish to grasp the entire nature of temporal structure, we chose to study phoneme duration using an extensive amount of connected text material with one speaker.

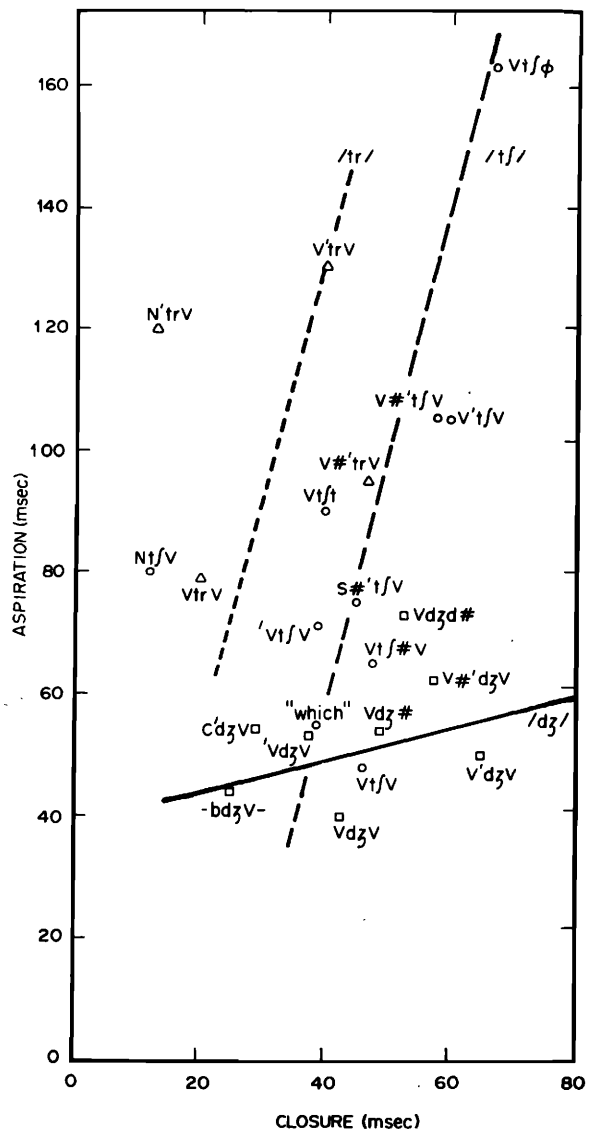


FIG. 9. The relationship between closure time and aspiration period of affricates /tr/, /tʃ/, and /dʒ/. A circle stands for /tr/, a triangle for /tʃ/, and a square for /dʒ/.

TABLE XI. Cross-speaker variability of the duration of closure and aspiration of /t/ in certain situations.

	#'stV		#'strV		(')VstV	
	Closure	aspiration	Closure	aspiration	Closure	aspiration
CC	51	9	61	51	40	30
JH	56	17	58	60	46	41
MM	55	12	68	38	57	24
SP	40	14	42	21	46	24

In order to get an idea of the variability among speakers, the closure time and the aspiration period of /t/ when it is preceded by /s/ is shown in Table XI for four speakers. This may not be a fair comparison, for among them, JH and SP read extensive materials, but the other two speakers read only a list of unrelated sentences. Nevertheless, we see that the closure time is fairy well regulated while the aspiration period seems to have a good deal of room for options. Among three situations, the /#'str/ combination offers the widest range of options in its aspiration period; one may pronounce /tr/ in this situation like an affricate or just like an ordinary stop. On the contrary, the aspiration period in /#'stV/ has the least range of options.

ACKNOWLEDGMENTS

My deepest gratitude goes to Marion Harris for obtaining data at various stages. I also thank Cecil Coker for our collaboration in speech synthesis, through which I could draw a coarse sketch of this complex structure of consonant duration. I appreciate Randall Monsen's help in making sound spectrograms and measuring the duration of segments.

[1]An objective measure of the consonantalness is seen in the order in which initial clusters are formed. The outer consonant is more consonantal than the inner one (e.g., #sp, #spl, #pl, etc.).

[2]Tokens for secondary-stressed conditions are scarce, and so data for this condition are not shown in this paper.

[3]One or two cases where /s/ is lengthened to about 150 msec are obtained. In these cases, the word is extraordinarily emphasized, being surrounded by quotation marks in the text.

[4]A slight difference in the /t/ figure from the one published earlier results from a more detailed phonological classification made recently.

Barnwell, T. P. (1971). "An algorithm for segment durations in a reading machine context," Technical Report 479, MIT Research Laboratory of Electronics, Cambridge, MA (unpublished).

Coker, C. H., and Umeda, N. (1975). "The importance of spectral detail in initial-final contrasts of voiced stops," J. Phonetics 3, 63–68.

Cooper, W. E. (1976). "Syntactic control of timing in speech production: a study of complement clauses," J. Phonetics 4, 151–171.

Goldhor, R. S. (1975). "Sentential determination of duration in speech," Master thesis (MIT) (unpublished).

Haggard, M. (1973). "Abbreviation of consonants in English pre- and post-vocalic clusters," J. Phonetics 1, 9–24.

Harris, M. S., and Umeda, N. (1974). "Effect of speaking mode on temporal factors in speech: Vowel duration," J. Acoust. Soc. Am. 56, 1016–1018.

Kenyon, J. S. and Knott, T. A. (1953). *Pronouncing Dictionary of American English* (Merriam, Springfield, MA).

10 December 2023 20:55:07

IPR2026-00275
Krisp Technologies EX1035 Page 13

Klatt, D. H. (1974). "Duration of [s] in English words." J. Speech Hear. Res. 17, 51–63.
Klatt, D. H. (1975a). "Voice-onset time, friction and aspiration in word-initial consonant clusters," J. Speech Hear. Res. 18, 389–406.
Klatt, D. H. (1975b). "Vowel lengthening is syntactically determined in a connected discourse," J. Phonetics 3, 129–140.
Klatt, D. H. (1976). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. 59, 1208–1221.
Kučera, H., and Francis, W. N. (1970). *Computational analysis of present-day American English* (Brown University, Providence, RI).
Lindblom, B., and Rapp, K. (1973). "Some temporal regularities of spoken Swedish," Papers from the Institute of Linguistics, University of Stockholm, Publ. 21.
Lindblom, B., Lyberg, B., and Holmgren, K. (1976). "Durational patterns of Swedish phonology: Do they reflect short-term motor memory process," Papers from the Institute of Linguistics, University of Stockholm.
Lisker, L., and Abrahmson, A. S. (1967). "Some effects of context on voice onset time in English stops," Language Speech 10, 1–28.
Malecot, A., and Lloyd, P. (1968). "The /t/-/d/ distinction in American alveolar flaps," Lingua 19, 264–272.
Monsen, R. B. (1971). "Durational properties of the American English consonants /p – t – k/," Bell Laboratories Technical Memorandom (unpublished).
Raphael, L. J., Dorman, M. F., Freeman, F., and Tobin, C. (1975). "Vowel and nasal duration as cues of voicing in word-final stop consonants: Spectrographic and perceptual studies," J. Speech Hear. Res. 18, 389–400.
Scharf, D. J. (1962). "Duration of post-stress intervocalic stops and preceding vowels," Language Speech 5, 26–30.
Schwartz, M. F. (1970). "Duration of /s/ in /s/-plosive blends," J. Acoust. Soc. Am. 47, 1143–1144.
Umeda, N., and C. H. Coker, (1974). "Allophonic variation in American English," J. Phonetics 2, 1–5.
Umeda, N., Harris, M. O., and Forrest, K. (1975). "The placement of auditory boundary in fluent speech," J. Phonetics, 3, 191–196.
Umeda, N. (1975). "Vowel duration in American English," J. Acoust. Soc. Am. 58, 434–455.
Umeda, N. (1976). "Linguistic rules for text-to-speech synthesis," Proceedings IEEE 64, 443–445.

10 December 2023 20:55:07

IPR2026-00275
Krisp Technologies EX1035 Page 14