ON THE AVAILABILITY OF DURATIONAL CUES

Thomas H. Crystal & Arthur S. House

Communications Research Division, Institute for Defense Analyses, Princeton, NJ 08540-3699 U. S. A.

ABSTRACT

Ongoing research to identify phones and measure their durations in recordings of read speech has resulted in the analysis of 10,300 phones produced by six talkers. The texts, the marking technique and some preliminary results were reported previously [2]. This report extends the earlier findings and tests for the presence of well-established durational cues cited in the literature. The analysis found, in general, that most of the cited effects are not clearly evident in continuous (read) speech signals. Some findings to be discussed are (a) completeness of stops; (b) stop variation in context; and (c) vowel lengthening.

INTRODUCTION

This is a progress report in an on-going program dealing with segmental durations in connected speech signals. An earlier report [2] described, in detail, the speech materials, talkers and methods of analysis. The study of these recorded materials has continued with an emphasis on segment durations and modeling of distributions. This report deals with measurements made on two scripts as spoken by six typical talkers—three from the original *slow* group (Nos. 1, 4 & 7; Table II [2]) and three from the *fast* group (Nos. 22, 34 & 43.) The scripts total approximately 600 words in 33 sentences [2, Appendix].

As before, the speech-sound segments of the readings were identified by studying a computer-graphics spectrogram and/or waveform display while simultaneously listening to the signal and by applying most of the standard criteria of acoustic and auditory phonetics. For stops and affricates the *hold* portions were measured (with occasional exceptions), as well as the plosive *release*. For stops with nonplosive releases—nasal, lateral, *etc.*—the released portion generally was included in the following segment. Word and pause boundaries are marked and have been used in the analyses.

SEGMENTAL DURATIONS

On the completeness of stops. A finding in Crystal & House [2] was the low percentage of "complete" stops (hold + plosive release) in the sample. Recently, stop closure duration was studied [6] with "sytematic conditions," using a corpus in which more than 95% of the stops were complete. Such a corpus may be very uncharacteristic of standard speech.

In this corpus the over-all frequency of occurrence of complete stops is 59%. There is a tendency for voiceless stops to be complete a higher percentage of the time than voiced stops (over-all, 65% *vs.* 51%), particularly in word-final position (42% *vs.* 18%.) As expected, word-initial stops are complete more often than word-final ones (85% *vs.* 33%.) There is a tendency, also, for velars to be complete more often than more fronted stops. Stop completeness is examined more closely in Table 1. The table entries display individual stops in various contexts, as indicated. *Caveat lector:* Validity is limited by small sample size and consequent atypical phoneme distributions!

The finding that plosions for /t/ and /k/ were always, essentially, measurable following /s/ is a little unexpected; they are considerably shorter, however, than the plosions of singletons. In the case of stops followed by /l/ or /r/, the plosion releases that occur generally are lallized or rhoticized. It is interesting to notice, also, that a higher percentage of plosions occurs in prepausal word-final stops (col. *F*) than in word-final stops in general.

Table 1. Proportion (*Pr*) of occurrence of complete stops in various contexts. Symbols: *SI* = silence; # = word boundary; — = undefined context; *N* = total tokens in category. Six talkers; two scripts. [*A*] *#stop—*; [*B*] *#stop+l/r*; [*C*] *#s+stop+vowel*; [*D*] *#s+stop+r*; [*E*] *—stop#*; [*F*] *—stop#SI*.

		Position in Word						
		Any	*Initial*				*Final*	
Stop			*A*	*B*	*C*	*D*	*E*	*F*
p	*Pr*	.55	.88	.94	–	.34	.23	–
	N	108	42	18	0	6	43	0
t	*Pr*	.61	.84	.67	1.00	.96	.36	.49
	N	734	234	18	66	24	363	40
k	*Pr*	.77	.98	1.00	1.00	–	.69	.95
	N	310	137	17	6	0	73	21
b	*Pr*	.79	.79	.62	–	–	–	–
	N	208	206	29	–	–	0	0
d	*Pr*	.34	.81	.92	–	–	.18	.31
	N	471	116	12	–	–	320	51
g	*Pr*	.87	.93	.76	–	–	.34	1.00
	N	60	54	17	–	–	6	1
All	*Pr*	.59	.85	.79	*	*	.38	.48
	N	1891	798	111	*	*	838	163

*(voiced cognates do not occur)

Completeness of stops appears to be related to talking rate. The counts in Table 2 show that the *fast* talkers had about 10% lower completion than the *slow* talkers.

Differentiation of stop occlusion. Table 2 displays the duration of stop occlusions (holds) as a function of voicing characteristic and of place of articulation. (The results for *all* stops are highly similar.)

Table 2. Analysis of hold portions of *complete* stops according to voicing characteristic (two left cols.) and place of articulation (three right cols.) Three *slow* and three *fast* talkers. Two scripts. *N* = number of tokens. *Dur* = duration in ms.

		Voicing		*Place*		
Talkers		*Voiced*	*V'less*	*Labial*	*Alveol.*	*Velar*
Slow	*Dur*	54	55	58	50	62
	N	202	388	120	322	148
Fast	*Dur*	54	50	56	48	53
	N	173	356	104	281	144
All	*Dur*	54	53	57	49	58
	N	375	744	224	603	292

The entries indicate that the hold portions of the *slow* talkers tend to be a few ms longer than those of the *fast* talkers. The durations of the holds of voiced and voiceless stops are not substantially different, contradicting experiments using citation forms [1] or words in a frame [10]. This confirms earlier observations [2] questioning the potential usefulness of a putative perceptual cue [1, 5] based on such a difference. (On the other hand, the average *plosions* of voiceless stops are about twice as long as those of voiced stops, as noted earlier by Zue [11].)

The effect of place of articulation is complicated. The average durations of the hold portions for the three (putative) places of stop articulation (right portion of Table 2), while not very different, show a definite tendency for alveolar stops to be shortest. The plosions give a different pattern, however, with duration increasing, on the average, as the point of contact moves from the lips to the velum. This results in total stop duration that, on the average, is about 80 ms for alveolars and labials and about 100 ms for velars.

O'Shaughnessy [9] measured durations of sounds in French words embedded in a sentence frame. In his materials labial stops were about 20 ms longer than lingual stops, with both types being considerably longer than the present results. He also has reported average stop (hold) durations for a read French passage [8] that are more comparable to the values in Table 3, but reports that voiceless stops are 10–15 ms longer, on the average, than voiced stops (63 ms *vs.* 76 ms.) Zue's [11] finding of longer releases for velars compared to labials and alveolars is supported in these materials, but his finding of longer hold portions for /p/ *vs.* /t/ and /k/ is not.

The corpus also contained 705 hold-only stops, *viz.*, without plosion *per se*. The average hold duration for these stops is the same, essentially, as that for complete stops, and the tempo-group differences are comparable.

The over-all conclusion, supported by [6], is that, in continuous speech, the hold portions of stop consonants are not strong indicators of voicing characteristic or place of articulation.

Vocalic variation. A contextual effect that is well-studied in English—lately in [6]—is the change of vocalic duration as a function of the voicing characteristic of the following consonant in the same syllable—the so-called *lengthening-before-voicing* effect. In [2] it was found for long (that is, *tense*) vowels preceding stops, but not for short (*lax*) vowels preceding stops, nor for either type of vowel preceding fricatives. In the present data the effect was investigated when the consonants are word-final and when they are word-final and prepausal, *viz.*, followed by a pause (but see *caveat* above.) Two general facts emerge: (1) the average duration of vowels preceding word-final prepausal consonants is considerably longer than that of vowels preceding word-final consonants in general, and (2) with the prepausal constraint, the data demonstrate the lengthening-before-voicing effect. The only exception noted was for the few cases of short vowels preceding fricatives. With this exception, there is an average 20-ms lengthening associated with vowels preceding prepausal voiced (*vs.* voiceless) obstruents. Without the prepausal constraint, however, the effect is not evident. It can be noted, also, that the progressive lengthening of short vowels before /t/, /s/, /n/, /d/ and /z/, pointed out in Lehiste [4], is not found in the present materials.

O'Shaughnessy [9] described two "strong" preconsonantal effects on vocalic duration in French: *lengthening* before voiced fricatives and *shortening* before voiceless obstruents. Neither effect is obvious in the present data, but there is a tendency for long vowels to lengthen before voiced fricatives. In [9] there also was a "weak" tendency for vocalic duration to vary inversely with vowel height. The present data confirm this for high (long) vowels (*viz.*, /i/ & /u/: $N = 379$) which are, on the average, shorter—108 ms—than other long vowels. However, the relation fails when mid (long) vowels (/e/ & /o/: $N = 318$, $Dur = 141$ ms) are compared to low (long) vowels (/a/ & /æ/: $N = 464$, $Dur = 132$ ms.)

Chen [1] reported that the lengthening usually attributed to the voicing characteristic of a postvocalic consonant functions across intervening sonorants separating a vowel and an obstruent (*sent vs. send.*) In his citation-form data, both sonorant and vowel were lengthened before a voiced, compared to a voiceless, obstruent. A rough test—long and short vowels, separately, before nasals and liquids followed by /p/, /t/, /k/, /č/, /f/, /s/ and their voiced cognates—shows the effect to be quite robust in the present data.

Table 3. Mean durations (*Dur*) and standard deviations (*SD*), in ms, for five "matched" pairs of back and front vowels preceding word-final stops and nasals grouped by place of articulation. N = number of tokens. Types in groups not equated.

	Back Vowels			*Front Vowels*		
Consonant Class	*N*	*Dur*	*SD*	*N*	*Dur*	*SD*
Labial	80	128	42	82	116	56
Dental	262	125	58	411	84	48
Velar	12	67	14	82	92	35

Another potential influence of consonantal context on vocalic duration is a place-of-articulation effect discussed by Fischer-Jørgensen [3] in which, before labials and dentals, back vowels > front vowels, but before velars, back vowels < front vowels. Data for examining this effect are presented in Table 3 (see *caveat* above.) For each consonant class the vowel category that, on the average, is longest, is the one predicted by Fischer-Jørgensen. The Fischer-Jørgensen study followed one by Maack [7], which claimed the relation "vowel+velar > vowel+dental > vowel+labial," but this relation does not hold in the present data. Further tests of vowels preceding word-final labial, dental and velar consonants using (1) 10 vowels (six long; four short) and (2) using all vowels occurring in the context resulted in an ordering by vocalic length that was the reverse of that described by Maack [7]. There are reports also on durational variation according to voicing characteristic for vowels *following* stops [3, 9]. Some of these phenomena are found in the present data.

REFERENCES

1. M. Chen, Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22, 1970, 129-159.

2. T. H. Crystal & A. S. House, Segmental durations in connected speech signals: Preliminary results. *J. acoust. Soc. Amer.* 72(3), 1982, 705–715.

3. E. Fischer-Jørgensen, Sound duration and place of articulation. *Z. Phonetik usw.* 17, 1964, 175–207.

4. I. Lehiste, Segmental features of speech. In *Contemporary issues in experimental phonetics*, N. J. Lass, ed. (Academic Press, 1976), pp. 225–239.

5. L. Lisker, Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, 33, 1957, 42-49.

6. P. A. Luce & J. Charles-Luce, Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *J. acoust. Soc. Amer.* 78(6), 1985, 1949–1957.

7. A. Maack, Die Beeinflussung der Sonantendauer durch die Nachbarkonsonanten. *Z. Phonetik usw.* 7, 1953, 104–128.

8. D. O'Shaughnessy, A multispeaker analysis of durations in read French paragraphs. *J. acoust. Soc. Amer.* 76, 1984, 1664–1672.

9. D. O'Shaughnessy, A study of French vowel and consonant durations. *J. Phonetics*, 9, 1981, 385–406.

10. L. J. Raphael, Durations and contexts as cues to word-final cognate oppositions in English. *Phonetica*, 38, 1981, 126–147.

11. V. W. Zue, *Acoustic characteristics of stop consonants: A controlled study.* Unpublished ScD dissertation, Massachusetts Institute of Technology, 1976.