Distinguishing "slit" and "split" : an invariant timing cue in speech perception

Perception & Psychophysics
1978, Vol. 23 (1), 58-60

Distinguishing "slit" and "split"—an invariant timing cue in speech perception

STEPHEN M. MARCUS
MRC Applied Psychology Unit, Cambridge, England

The effect of speech rate on the distinction between "slit" and "split" was investigated. This distinction may be cued solely by a silent interval (SI) of sufficient duration between the [s] and the [l], and the boundary SI, at which "slit" and "split" are perceived with equal probability, determined. In this experiment, although subjects showed good identification of the stimulus categories within the range of silent intervals used, no shift in the boundary SI was found for a twofold increase in speech rate. This finding is contrasted with most recent experiments which demonstrate "compensation" for changes in speaker characteristics, such as speech rate. Implications for models of speech perception are discussed.

For some time, experiments in speech perception have been demonstrating the listener's ability to adapt to characteristics of a given speaker. For example, Ladefoged and Broadbent (1957) found that judgments of acoustically identical test words could be shifted from "bit" to "bet" or "bet" to "bat" by modifying the formant frequency range, and thus the perceived voice quality, of a synthetic carrier sentence.

Similarly, the listener can adjust to various temporal aspects of speech: The distinction between a single- and a double-stop consonant ("topic" v. "top pick") can be cued solely by an extended silent interval (SI) during the stop closure; the boundary SI duration, at which a single or double consonant is perceived with equal probability, is dependent on overall speech rate (Pickett & Decker, 1960). More recently, Ainsworth (1974) used simple synthetic CV stimuli consisting of a noise burst followed by a steady-state vowel to investigate interactions between categorical phoneme perception and syllable duration. Subjects categorized the stimuli as [di, ti, *or* si], depending on the noise-burst duration, and phoneme boundaries were determined for two vowel lengths. Boundary durations were shorter for the shorter vowel context. Further, Haggard (1972) has shown that judgments of a synthetic velar stop, V in [Vil], varying in voice onset time (VOT) are influenced both by formant transition rate and steady-state vowel duration. In both cases, the condition which corresponded to "faster" speech resulted in more [k] judgments. Therefore, he concluded that VOT is measured *relative* to average speech rate.

It is well documented that phonemic distinctions may be cued by SIs in the acoustic stimulus (Bastian, Eimas, & Liberman, 1961; Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Lisker, 1957). For example, the insertion of an SI of sufficient duration between [s] and [l] in [slɪt] results in the perception of "split" (Bastian et al., 1961).

The following experiment aimed to investigate the effect of speech rate on the phonemic significance of such SIs. In this experiment, speech rate was manipulated by changing the duration of surrounding segments within the word containing the SI. It was therefore expected that boundary SI would be closely dependent on speech rate, as in Pickett and Decker's (1960) experiment.

The author is now at Instituut voor Perceptie Onderzoek, Eindhoven, The Netherlands.

METHOD

All stimuli were produced by computer modification of a single token of "slit." Subjects judged stimuli as either "split" or "slit." Blocks of stimuli were either uncompressed, or 25% or 50% compressed. Within each block, SIs were randomized, and ranged from 0 to 64 msec in 8-msec steps.

Stimuli

A single instance of the word "slit," spoken in isolation by a male native English speaker, was sampled digitally at a rate of 20,000 Hz and stored on a general-purpose computer. Pitch-synchronous compressed versions of the original stimulus were produced by deleting glottal periods (about 10 msec) during voiced segments, and similar sized sections of unvoiced segments. Care was taken not to introduce abrupt transients when visually selecting start and end points of the sections to be deleted. Two further versions were produced from the original uncompressed (C0) stimulus, the first with one glottal period, or similar durations of unvoiced segments, in every four removed (25% compression, C25) and the second with every other period omitted (50% compression, C50). The three stimuli thus encompassed a twofold increase in speech rate (see Table 1). The fastest was still highly intelligible. From each of these basic stimuli, eight more were generated by inserting an SI of 8 to 64 msec in 8-msec steps at the [s]-[l] juncture, which was easily determined by visual examination of the digitized amplitude waveform. The 27 stimuli were stored on the computer disk, ensuring that all further productions of each stimulus were identical.

Table 1
Segment Durations of Basic Stimuli ("Slit") in Milliseconds

Rate	[s]	[lɪt]	Overall	Percent C0
C0	134	328	462	(100)
C25	97	245	342	74
C50	67	168	235	51

Procedure

For each rate of compression (C0, C25, and C50), three blocks of 84 stimuli were recorded. Recordings were made at 7½ ips on a Ferrograph Series Seven tape recorder. Each block contained nine presentations at each SI duration in a different random order, preceded by three practice items. Stimuli were recorded at 3-sec intervals. The first block at each rate was used as practice and the responses collected discarded from the subsequent data analysis.

The subjects were two groups of 10 Cambridge housewives, who were paid for participating. Stimuli were played at a comfortable listening level over a loudspeaker. They were told that the experiment concerned how fine a discrimination could be made between speech sounds, and were told to judge each stimulus as either "slit" or "split" and mark it accordingly on prepared response sheets. Group 1 was tested on uncompressed stimuli (C0) followed by 25% compression (C25) and then 50% compression (C50). Group 2 was tested with the order reversed.

RESULTS

Percentage "split" responses were determined for each SI by Group by Subject by Rate of Compression summed over Blocks 2 and 3. A minimum normit chi-square solution (Pearson & Hartley, 1972) was used to determine the boundary SI. Mean boundary SIs for each Group and Rate of Compression are shown in Table 2. An analysis of variance revealed no significant difference between conditions [$F(2,36) = 2.51$; $p > .05$] or between groups [$F(1,18) = 1.97$; $p > .05$). No change in boundary SI was measured at any rate of compression, and there was no evidence of effect of order of presentation. Figure 1 shows percentage "split" responses at each SI for each rate of compression, pooled over all subjects. This represents a mean "phoneme boundary" curve for each rate of compression, although it is broadened by individual differences in boundary placement. Despite this, good identification performance can be seen, the changeover from 75% "slit" to 75% "split" occurring in about one-and-a-half steps in SI (12 msec) under all three compression conditions. It can be seen that the range and steps of SI used gives good coverage of the response continuum and would be expected to have given optimal sensitivity to any shift in boundary SI.

DISCUSSION

The lack of compensation between speech rate and boundary SI was somewhat unexpected, especially considering the apparent parallel with Pickett and Decker's (1960) experiment. There are, however, a number of differences. First, their target word was embedded in a full sentence, rather than spoken in isolation. This could have resulted in a reduction in boundary shift, but the blocked design of this experiment gave the subjects ample opportunity to experience the "context" of the different speech rates used in each block. Secondly, Pickett and Decker used natural variations in speech rate, which had differential effects on consonant and vowel durations, vowels becoming shortened proportionately less than consonants with increasing rate (Karlsson & Nord, Note 1). The uniform compression used in this experiment would thus have resulted in an overshortening of vowels and excessively long consonants for the intended speech rate; despite some loss of naturalness, the overall perceptual effect was nonetheless one of considerably increased speech rate. Thirdly, whereas Pickett and Decker were examining the perception of an open juncture cued by changes in segment duration (Lehiste, 1960), this experiment examined the perception of [p] cued by intervals of considerably shorter duration. It has been suggested that physiological constraints place a lower limit, or "time barrier" on stop closure duration (Hudgins & Stetson, 1937; Huggins, 1972; Ohala, Note 2). Closure duration in Pickett and Decker's stimuli would have been well above any physiological limit since these stimuli could be ambiguously interpreted

Table 2
Group Mean Boundary SIs and Standard Error in Milliseconds

	C0		C25		C50	
	Mean	SD	Mean	SD	Mean	SD
Group 1	25.7	5.1	25.9	4.2	26.3	6.2
Group 2	30.4	3.5	26.0	2.7	28.8	5.2

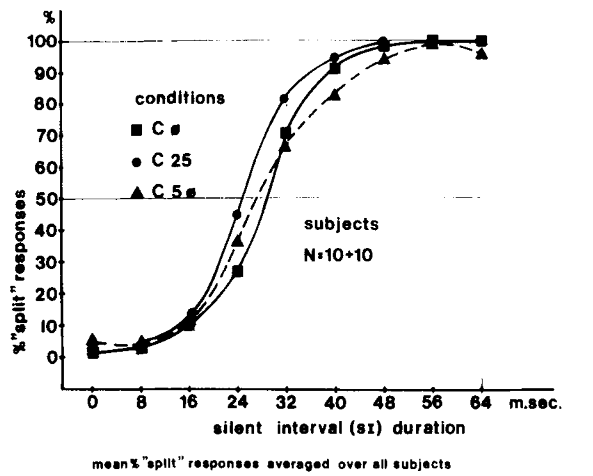

Figure 1. Mean percent "split" responses averaged over all subjects.

as either a single [p] ("topic") or a juncture [p#p] ("top pick"); therefore, SI duration would have been intermediate between that for clear exemplars of phoneme and juncture closures. In this experiment, 75% "split" responses were produced by an SI of 30 msec with the uncompressed stimuli, and almost 100% "split" responses were produced by an SI of 45 msec. These durations are also in accord with Huggins' minimum acceptable [p] closure duration of 40 msec which he predicts from "physiological limits" and the range of 35-45 msec which he observes in his perceptual experiments. Since this duration is the *minimum* that is perceptually acceptable, it does not follow that in production of "split" at different rates the [p] closure duration will have a constant value with increasing rate. Indeed, as speech becomes very fast, individual phones become less well articulated, and in the case of "split," voicing of the [l] may begin during the [p] closure, resulting in a sound with a very short closure, which would more accurately represent the nonword "sblit." In such normal rapid conversational speech, the speaker uses his knowledge of the language to deduce the speaker's intention.

The suggestion that the lack of effect of compression on boundary SI duration may be a perceptual consequence of physiological constraints on production is not held to necessarily imply either analysis-by-synthesis or innate properties of the perception apparatus, rather that experience of the properties of language have become embodied in it. Models which suggest a simple multiplicative relation between speech rate and segment duration, even as a first approximation over a small change in rate (Allen, 1973) require modification, both due to Karlsson and Nord's (Note 1) results, and in the limit due to the "time barrier" which this experiment suggests may have a perceptual as well as a physiological status.

REFERENCE NOTES

1. Karlsson, I., & Nord, L. *Stops and CV segment duration.* International Conference of Speech Communication and Processing, Bedford, Massachusetts, 1972, Paper F5, 210-213, New York, IEEE.

2. Ohala, J. *Aspects of the control and production of speech.* UCLA Working Papers in Phonetics, 15, 1970.

REFERENCES

Ainsworth, W. A. Phoneme boundary shifts. *Proceedings of the Eighth International Congress on Acoustics, London 1974* (Vol. 1). Trowbridge, England: Goldcrest Press, 1974.

Allen, G. D. Segmental timing control in speech production. *Journal of Phonetics*, 1973, **1**, 219-237.

Bastian, J., Eimas, P. D., & Liberman, A. M. Identification and discrimination of a phonemic contrast induced by a silent interval. *Journal of the Acoustical Society of America*, 1961, **33**, 842 (A).

Haggard, M. P. Speech rate effects in the perception of voicing. *Speech Synthesis and Perception*, 1972, **6**, 1-12. (Psychological Laboratory, Cambridge, England.)

Hudgins, C. V., & Stetson, R. H. Relative speed of articulatory movements. *Archives Néerlandaises de Phonétique Expérimentale*, 1937, **13**, 85-94.

Huggins, A. W. F. Just noticeable differences for segment duration in natural speech. *Journal of the Acoustic Society of America*, 1972, **51**, 1270-1278.

Ladefoged, P., & Broadbent, D. E. Information conveyed by vowels. *Journal of the Acoustical Society of America*, 1957, **29**, 98-104.

Lehiste, I. An acoustic-phonetic study of internal open juncture. *Phonetica Supplement*, 1960, **5**, 1-54.

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., & Bastian, J. An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech*, 1961, **4**, 175-195.

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

Pearson, E. S., & Hartley, H. O. *Biometrika tables for statisticians* (Vol. 2). Cambridge, England: Cambridge University Press, 1972. Pp. 91-95.

Picket, J. M., & Decker, L. R. Time factors in the perception of a double consonant. *Language & Speech*, 1960, **3**, 11-17.

(Received for publication May 6, 1977;
revision accepted September 10, 1977.)