Phonetica
Editor: K. Kohler, Kiel

Separatum
Publisher: S. Karger AG, Basel
Printed in Switzerland

Phonetica *41:* 117–124 (1984).

985

Effects of Temporal Stimulus Properties on Perception of the [sl]-[spl] Distinction

Bruno H. Repp
Haskins Laboratories, New Haven, Conn., USA

Abstract. Two studies investigated the influence of the independently varied durations of preceding and following signal portions on the amount of closure silence needed to perceive 'splash' rather than 'slash'. Increases (or decreases) in the durations of the [s] and [l] acoustic segments had opposite effects which cancelled when the silent intervals were short (experiment 1), but yielded a net effect due to [s] duration when the silent intervals were long (experiment 2). These findings, which resolve a conflict between earlier results in the literature, are interpreted as reflecting a perceptual compensation for coarticulatory shortening of [s] before stop consonants, in conjunction with (possibly psychoacoustic) contrastive interactions between the perceived durations of adjacent acoustic segments. The results suggest that local temporal signal properties, as distinct from global perceived speaking rate, are an important factor in phonetic perception.

An important perceptual cue for the distinction between the word-initial clusters [sl] and [spl] is the absence versus presence of a silent interval following the [s] noise [*Bastian* et al., 1961; *Fitch* et al., 1980]. Two fairly recent studies have investigated whether the category boundary on a continuum ranging from 'slit' to 'split', created by varying the duration of the silent closure interval, is affected by reductions in total stimulus duration: *Marcus* [1978] found that temporal compression left the 'slit'-'split' boundary unaffected, whereas *Summerfield* et al. [1981] found that less silence was needed to perceive 'split' in temporally compressed stimuli. Both studies made use of modified natural-speech tokens of 'slit'; *Summerfield* et al. [1981] also used synthetic stimuli, with similar results. In an attempt to explain the difference in outcomes, *Summerfield* et al. pointed out that the category boundaries in the *Marcus* [1978] study were at considerably shorter silences (less than 30 ms) than the boundaries in their own study (around 60 ms). They conjectured [as had *Marcus*, 1978] that a perceptual limit, perhaps related to an articulatory limit, may be encountered at short silences, and that this may be the reason why the boundary refused to shift to even shorter values in the

Marcus study. They interpreted their own findings as reflecting a perceptual adjustment to variations in contextual speech rate.

The principal reason for conducting the present experiments was the author's suspicion that temporal changes in signal portions preceding and following the silence may not be equally relevant. Several earlier perception studies in which closure silence duration was the dependent variable, albeit for different phonetic contrasts, have found that the duration of the preceding acoustic segment has a much stronger effect than that of the following segment [*Port and Dalby*, 1982; *Repp*, 1979]. In addition, there is another reason to expect [s] duration to be important, quite regardless of perceived speaking rate: Fricative noise duration tends to be shorter in [spl] than in [sl] clusters [*Morse* et al., 1982; *Schwartz*, 1969, 1970; *Haggard*, 1973], and listeners may have tacit knowledge of this coarticulatory relationship, as they do of so many others [*Repp*, 1982]. The duration of [l], on the other hand, does not seem to exhibit such coarticulatory variation [*Morse* et al., 1982; *Repp*, unpublished data] and therefore may be perceptually irrelevant. To examine this hypothesis, the durations of the fricative noise and of the lateral resonance were varied independently in the present experiments.

Experiment 1

Experiment 1 used a 'slash'-'splash' continuum [from *Repp*, 1984a: experiment 7] for which the average category boundary happened to be around 25 ms of silence, similar to the short boundary obtained by *Marcus* [1978]. This provided an opportunity to test further the hypothesis of a lower limit for the perception of silence duration in this context. While the reason for the short boundary in *Marcus's* stimuli is not clear, that for the present stimuli was due to inclusion of a labial release burst (from 'splash'), which provided an additional stop manner cue [*Repp*, 1984a].

Unlike the earlier studies, which used only temporal compression, the present experiment introduced both decreases and increases in acoustic segment duration. Although *Marcus* [1978] concluded from his results that the critical silent interval was invariant under changes of speaking rate, he failed to investigate the effects of decreases in simulated rate (i.e., increases in stimulus duration). According to the speaking-rate adjustment hypothesis, the perceptual boundary should shift to longer values of silence in that case, since no perceptual limit is encountered in that direction. The question in experiment 1 was, then, whether either '[s] duration' or '[l] duration', or both, have any effect on a short-silence [sl]-[spl] boundary.

Method

Stimuli. The utterances 'slash' and 'splash' were recorded by a female speaker, low-pass-filtered at 9.6 kHz and digitized at 20 kHz. To avoid strong stop manner cues in the [s] portion, the fricative noise of 'slash' was used in all experimental stimuli. The remainder was taken from 'splash'. This portion included an initial 10-ms release burst, which preceded the first glottal pulse of the [l] segment. The end of the [l] resonance was defined visually by a change in waveform shape coupled with an amplitude increase, and was confirmed by listening. The durations of the [s] noise and of the [l] resonance were varied independently by either removing or duplicating a piece of the waveform. An appropriate piece was selected from the interior of each acoustic segment on the basis of overall and local envelope considerations, and all cuts were made at zero crossings. In the [s] noise (original duration: 142 ms), the piece removed or duplicated was 51 ms long and ended 36 ms before noise offset. In the [l] portion (original duration: 57 ms, or 14 pitch periods), it was 21 ms (5 pitch periods) long and began 28 ms (7 pitch periods) after [l] onset. Thus, the signal portions immediately adjacent to the closure interval were left undisturbed, so as to avoid changing spectral and amplitude envelope cues to stop manner [*Summerfield* et al., 1981: experiment 1]. The ultimate durations were 91, 142, and 193 ms for the [s] noise, and 36, 57, and 78 ms for the [l] resonance. (Note that the changes are proportional and correspond to increases or decreases of about 36%). The orthogonal combination of all [s] and [l] durations resulted in nine stimuli, for each of which silent closure duration was varied from 0 to 50 ms in 10-ms steps. The resulting 54 stimuli were recorded in five different randomizations with interstimulus intervals of 2 s.

Subjects and Procedure. 7 paid volunteers and the author identified the stimuli as 'slash' or 'splash', with 'stlash' as an additional option. The tape was repeated once, so that 10 responses per subject were obtained for each stimulus. Presentation was over TDH-39 earphones at a comfortable intensity in a quiet room.

Results and Discussion

The results are shown in table I in terms of category boundary locations, determined from the average labeling functions by linear interpolation. (Only 3 subjects gave any 'stlash' responses, which were included with 'splash' responses.) Repeated-measures analysis of variance was conducted on individual subjects' response percentages, averaging over silence durations. Increasing silence duration, of course, had the expected effect of increasing the percentage of 'splash' responses; the labeling functions, which are not presented here for the sake of conciseness, were comparable in steepness to those obtained by *Marcus* [1978]. As can be seen in table I, the amount

Table I. Results of experiment 1: average category boundary values (in ms of silence) as a function of [s] and [l] durations

[l] duration, ms	[s] duration, ms			
	91	142	193	mean
36	*24.0*	25.5	27.6	25.7
57	18.8	*23.7*	24.2	22.2
78	17.2	17.9	*23.8*	19.6
mean	20.0	22.4	25.2	22.5

of silence needed to hear a 'p' (or 't') increased as the duration of the [s] noise increased [$F(2,14) = 12.5$, $p < 0.001$], but decreased as the duration of the [l] resonance increased [$F(2,14) = 15.8$, $p < 0.001$]. Both effects were highly consistent across subjects, approximately linear, and of similar magnitude. Their interaction was not significant [$F(4,28) = 1.1$].

Since increases and decreases in acoustic segment duration effected boundary shifts of nearly equal magnitude, it appears that the [sl]-[spl] boundary was not close to a lower limit. In fact, the boundary shifted to as little as 17 ms of silence in the 'short [s], long [l]' condition, which is considerably shorter than any of *Marcus's* [1978] values. This suggests that *Marcus's* failure to find any boundary shifts was not due to the relatively short category boundary for his stimuli. Indeed, closer inspection of table I reveals that *Marcus's* results are replicated by the present study: Due to the opposite and equally sized effects of changes in [s] and [l] duration, simultaneous proportional compression or expansion of both acoustic segments had no effect on the [sl]-[spl] boundary (compare values in italics along the major diagonal in table I; $F(2,14) = 0.1$). Thus, to the extent that the combined [s]+[l] duration conveyed anything about speaking rate, there was no effect of this variable in the present study.

The observed effect of [s] duration on stop manner perception may be attributed to the 'rate' of the speech preceding the silence, which really amounts to merely redescribing the results. An alternative explanation is in terms of a perceptual compensation reflecting listeners' tacit knowledge of the coarticulatory shortening of [s] frication preceding a stop closure. An independent effect of fricative noise duration was also found by *Summerfield* et al. [1981: experiment 1]; however, they tentatively attributed it to a psychoacoustic effect of this variable on the perceived silence duration. This hypothesis cannot be ruled out on the basis of the present data. However, the 'coarticulation-compensation' hypothesis proposed should perhaps be favored in view of many related findings [*Repp,* 1982].

The reversed effect of [l] duration was totally unexpected. Since [l] duration in natural speech does not seem to covary with the presence versus absence of a preceding [p], it is unlikely that [l] duration has any direct cue value for stop manner perception, in the way that [s] duration has. Rather than affecting stop manner perception directly, [l] duration may have its effect by altering the perceived relative duration of the [s] noise [see *Repp* et al., 1978, for a rather similar argument relating to the fricative-affricate contrast]. In other words, the [s] noise may 'sound longer' before a short [l], and shorter before a long [l]. This explanation assumes that the intervening silence does not engage in such contrastive interactions with the surrounding signal portions; this assumption is supported by the absence of any effect of increases or decreases in *both* [s] and [l] duration.

Experiment 2

What is not yet clear is why *Summerfield* et al. [1981] did find an effect of overall stimulus compression. One possibility is that their compression technique affected the amplitude envelopes of the signal surrounding the silence, thus introducing additional stop manner cues that shortened the amount of silence required to hear a 'p'. However, since their technique was similar to *Marcus's* [1978], and in fact left about 10 ms of waveform on either side of the silence undisturbed, this possibility seems unlikely. Another possibility is suggested by the results of experiment 1, however: The hypothesis just proposed to explain the effect of [l] duration predicts that the relational dependence of perceived [s] duration on the context following the silence should decrease with increasing temporal separation. Thus, at the longer silent intervals that characterized the *Summerfield* et al. [1981] stimuli, the effect of [s] duration may have been larger than the (presumably) opposite effect of the signal duration following the silence, thus leading to a net effect in the same direction as that of [s] duration alone.

It is also true, of course, that *Summerfield* et al. [1981] varied the duration of the whole stimulus, and not just of [s] and [l]. It was decided, therefore, to replicate their study using stimuli that had the category boundary at a comparably long silent interval (which was achieved by removing the labial release burst from the stimuli of experiment 1 and by shifting the range of silent intervals employed). The main difference was that, in experiment 2, the durations of [s], [l], and of the final [æʃ] portion were varied independently, so as to determine their separate effects on the 'slash'-'splash' boundary.

Method

Stimuli. The 10-ms release burst was removed from the stimuli of experiment 1. Two [s] and two [l] durations were employed, corresponding to the original and shortened versions of experiment 1. In addition, the final [æʃ] portion was used both in its original version (477 ms) and shortened by 36% (304 ms). Shortening was achieved by deleting two separate pieces of waveform from the interior of the [æ] vowel and one piece from the interior of the [ʃ] noise, thereby reducing each of these two acoustic segments by the same proportional amount. Careful listening indicated no obvious disruptions of spectral continuity caused by the splices. The two [s] durations, two [l] durations, and two [æʃ] durations were combined to yield eight stimuli that were presented with six different silent intervals ranging from 50 to 100 ms in 10-ms steps. The resulting 48 stimuli were recorded in five randomizations with interstimulus intervals of 2 s.

Subjects and Procedure. 10 paid volunteers listened to the tape twice, labeling each stimulus as 'slash' or 'splash'. None of the subjects had participated in experiment 1. The 'stlash' response category was not included, since these responses generally occur only at short closure durations [*Repp,* 1984b]. Otherwise, the procedure was identical to that in experiment 1.

Subjects' comments after a brief preview of the tape revealed that most stimuli were initially perceived as 'splash'. All subjects were consequently encouraged to try to hear more instances of 'slash', and to classify ambiguous stimuli as belonging to this category. No subject had any difficulty carrying out these instructions, which were probably unnecessary, since phonetic boundaries based on closure silence duration are rather sensitive to the range of closure durations employed in a test [*Repp,* 1980]. Most likely, the listeners in experiment 2 rapidly adjusted their perceptual criteria to the prevailing stimulus range.

Results and Discussion

The results are displayed in table II. The average labeling functions from which the boundaries were derived were less steep than in experiment 1 but comparable to those obtained by *Summerfield* et al. [1981]. The boundaries were located at somewhat longer silences than in the *Summerfield* et al. study, probably owing to procedural differences. It can be seen in table II that the basic findings of experiment 1 were replicated: The amount of silence needed to perceive 'splash' increased as [s] duration increased [$F(1,9) = 42.3$, $p < 0.001$], and decreased as [l] duration increased [$F(1,9) = 20.5$, $p < 0.002$]. As in experiment 1, these effects were highly consistent and independent of each other ($F = 0.0$ for their interaction). In contrast to experiment 1, however, and in agreement with the predictions for experiment 2, the effect of [s] duration was larger than the opposite effect of [l] duration, which supports the hypothesis that the latter effect is indirect and decreases with increasing temporal separation between [s] and [l], relative to the effect of [s] duration. In a separate comparison of the results for the two stimuli that differed by a uniform compression of 36% (values in italics in table II), a

Table II. Results of experiment 2: average category boundary values (in ms of silence) as a function of [s], [l], and [æʃ] durations

[l] duration ms	[æʃ] duration ms	[s] duration, ms		
		91	142	mean
36	304	*66.7*	75.4	
	477	69.6	81.7	
	mean	68.2	78.6	73.4
57	304	62.6	75.3	
	477	64.4	*73.2*	
	mean	63.5	74.3	68.9
mean	/	65.9	76.5	71.2

significant 6.5-ms boundary shift was observed [$F(1,9) = 15.2$, $p < 0.004$], which is comparable to the shifts found by *Summerfield* et al. [1981].

The effect of the duration of the [æʃ] portion was less consistent. There was a small but significant main effect [$F(1,9) = 6.2$, $p < 0.04$], as well as a significant interaction with [l] duration [$F(1,9) = 10.4$, $p < 0.02$]. As can be seen in table II, the effect of [æʃ] duration was reversed with respect to that of [l] duration, longer [æʃ] durations leading to longer category boundaries, except in the condition where both [s] and [l] were long. While the reason for this interaction is not clear, the direction of the main effect suggests that, rather than influencing perceived [s] duration, the [æʃ] portion may have modified the perceived [l] duration, which then in turn influenced the perceived [s] duration. In other words, there may be a general contrastive interaction between adjacent energy-carrying acoustic segments of the speech signal with respect to their effective temporal features in phonetic perception. The effect of [æʃ] duration (but not that of [l] duration) is also consistent with a 'contextual speech rate' explanation, but is too small to be of any significance. Clearly, the dominant effect is that of [s] duration.

General Discussion

The present results eliminate the apparent contradiction between the earlier results of *Marcus* [1978] and *Summerfield* et al. [1981], and they also rule out some of the interpretations advanced by these authors. They suggest that *Marcus's* [1978] failure to find a shift of the [sl]-[spl] boundary as a function of stimulus compression was due neither to a perceptual limit nor to any insensitivity of the boundary to contextual influences. Rather, as experiment 1 has shown, even boundaries at very short silences are highly sensitive to context and shift freely to both longer and shorter silences [see also *Repp,* 1983: experiment 4, for a shift to very short silences induced by a restricted stimulus range]. The absence of a net effect of stimulus compression or expansion when the silence duration is short seems to be due to the presence of two opposite effects, of [s] duration and [l] duration, respectively, which are equally strong and thus cancel each other out. Another way of expressing this result is that the [s]/[l] duration ratio remains constant at the phonetic boundary. On the other hand, experiment 2 has shown that, when the silence durations are longer [as in the study by *Summerfield* et al., 1981] the [s] duration effect is larger than the [l] duration effect, so an effect of overall compression is obtained. This overall effect does not seem to reflect an adjustment to perceived global speaking rate but may be due to [s] duration alone, assuming that [s] duration is perceived relative to the context following the silence. As the temporal separation increases, the influence of this context on perceived [s] duration decreases in importance.

The effect of [s] duration is interpreted here as a perceptual compensation for the known reduction in fricative noise duration when it precedes a stop consonant closure. Thus it is considered a purely phonetic effect, deriving from listeners' tacit knowledge of speech patterns [*Repp,* 1982]. This hypothesis predicts that no such effect should be obtained in analogous nonspeech stimuli – a prediction that obviously should be tested. In a more speculative vein, the reversed effect of [l] duration is attributed to some form of perceptual contrast among temporal stimulus properties. It is not clear at what level in perception this contrast might arise, but experiments with non-speech stimuli should also prove revealing in that regard.

While the present results disconfirm the hypothesis that the [sl]-[spl] boundary shifts as a function of global contextual speech rate, the data are compatible with the assumption that listeners compute a variable running estimate of speaking rate on the basis of local temporal properties of the speech signal. In fact, this alternative hypothesis allows for contrastive interactions among adjacent segments whose relative durations deviate from the ratios commonly encountered in natural speech. Accordingly, the effect of [s] duration may be attributed to the listener's estimate of the *local* speaking rate at that time, based on [s] duration relative to the following context. While this account provides an alternative to the hypothesis of perceptual compensation for [s]-stop coarticulation, the latter hypothesis is to be preferred because speaking rate is not a quantity that varies from segment to segment in speech production; hence, to postulate a corresponding, continuously varying perceptual dimension is of questionable explanatory value. Moreover, as *Miller* et al. [1984] have recently shown, perceptual effects of local temporal stimulus properties are independent of subjects' estimates of (global) rate of articulation. The only serious alternative to the coarticulation-compensation account, therefore, is a purely psychoacoustic explanation based on auditory temporal contrast, which needs to be tested directly in future experiments.

Acknowledgments

This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. A short version of this paper was presented at the 107th meeting of the Acoustical Society of America in Norfolk, Virginia, May 1984. I am grateful to *Peter Bailey, Joanne Miller, Richard Pastore*, and especially *Douglas Whalen* for helpful comments on an earlier draft.

References

Bastian, J.; Eimas, P. D.; Liberman, A. M.: Identification and discrimination of a phonemic contrast induced by silent interval (Abstract). J. acoust. Soc. Am. *33:* 842 (1961).

Fitch, H. L.; Halwes, T.; Erickson, D. M; Liberman, A. M.: Perceptual equivalence of two acoustic cues for stop-consonant manner. Perception Psychophysics *27:* 343–350 (1980).

Haggard, M.: Abbreviation of consonants in English pre- and postvocalic clusters. J. Phonet. *1:* 9–24 (1973).

Marcus, S. M.: Distinguishing 'slit' and 'split' – an invariant timing cue in speech perception. Perception Psychophysics *23:* 58–60 (1978).

Miller, J. L.; Aibel, I. L.; Green, K.: On the nature of rate-dependent processing during phonetic perception. Perception Psychophysics *35:* 5–15 (1984).

Morse, P. A.; Eilers, R. E.; Gavin, W. J.: The perception of the sound of silence in early infancy. Child Dev. *53:* 189–195 (1982).

Port, R. F.; Dalby, J.: Consonant/vowel ratio as a cue for voicing in English. Perception Psychophysics *32:* 141–152 (1982).

Repp, B. H.: Influence of vocalic environment on perception of silence in speech. Haskins Lab. Status Rep. Speech Res., SR-57, pp. 267–290 (Haskins Laboratories, New Haven 1979).

Repp, B. H.: A range-frequency effect on perception of silence in speech. Haskins Lab. Status Rep. Speech Res., SR-61, pp. 151–165 (Haskins Laboratories, New Haven 1980).

Repp, B. H.: Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. Psychol. Bull. *92:* 81–110 (1982).

Repp, B. H.: Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. Speech Commun. *2:* 341–362 (1983).

Repp, B. H.: The role of release bursts in the perception of stops after [s]. J. acoust. Soc. Am. *75:* 1219–1230 (1984a).

Repp, B. H.: Closure duration and release burst amplitude cues to stop consonant manner and place of articulation. Lang. Speech (in press, 1984b).

Repp, B. H.; Liberman, A. M.; Eccardt, T.; Pesetsky, D.: Perceptual integration of acoustic cues for stop, fricative, and affricate manner. J. exp. Psychol. human Perception Performance *4:* 621–637 (1978).

Schwartz, M. F.: Influence of vowel environment upon the duration of /s/ and /š/. J. acoust. Soc. Am. *46:* 480 (1969).

Schwartz, M. F.: Duration of /s/ in /s/-plosive blends. J. acoust. Soc. Am. *47:* 1143–1144 (1970).

Summerfield, Q.; Bailey, P. J.; Seton, J.; Dorman, M. F.: Fricative envelope parameters and silent intervals in distinguishing 'slit' and 'split'. Phonetica *38:* 181–192 (1981).

Received: January 25, 1984
Accepted: April 27, 1984

Bruno H. Repp,
Haskins Laboratories,
270 Crown Street,
New Haven, CT 06511-6695 (USA)