395

LANGUAGE AND SPEECH, Vol. 25, Part 3, 1982

AN EFFECT OF SENTENCE FINALITY ON THE PHONETIC SIGNIFICANCE OF SILENCE*

BRAD RAKERD,** DAVID R. DECHOVITZ,†
and
ROBERT R. VERBRUGGE
Haskins Laboratories
and
The University of Connecticut

The presence or absence of silence was found to be a relevant cue for the distinction between affricate /tʃ/ and fricative /ʃ/ when it occurred in sentence-medial position, but not when it occurred at a sentence boundary. This was so in all cases in which target words and their precursors were both produced by the same talker and in some cases in which they were produced by different talkers. The effect was not dependent on specific durations of the silent interval, nor upon listeners' perceptions of the number of talkers who had produced the utterances. These results are, along with previous findings, taken to be consistent with the principle that silence can have phonetic significance for a listener only when it is perceived to have occurred in a stretch of speech that was articulated continuously.

INTRODUCTION

A brief interval of silence can be an important cue for consonantal perception. It has, for example, been variously shown to affect perception of the consonantal features of voicing, place, and manner. With respect to voicing, Lisker (1957) reported that in intervocalic position a voiceless stop consonant (such as the /p/ in *rapid*) will be heard as voiced (i.e., as the /b/ in *rabid*) when the silent interval corresponding to closure is shortened. Further shortening of this interval affects the perceived place of production as well: the bilabial /p/ of *rapid* is heard as the American alveolar flap [ɾ] of *ratted* (Port, 1979). And finally, given appropriate values for the formant transitions, silence is an effective cue for stop manner, both in prevocalic (Bastian, Eimas and Liberman, 1961; Dorman, Raphael and Liberman, 1979; Fitch, Halwes, Erickson and Liberman 1980) and postvocalic (Repp, Liberman, Eccardt and Pesetsky, 1978; Dorman *et al.*, 1979) position.

* *This research was supported by National Institute for Child Health and Human Development grant HD01994 and by BRS grant RR05596 to Haskins Laboratories. We wish to thank Alvin Liberman, Bruno Repp, and Patti Jo Price for their many useful comments on earlier versions of this manuscript.*

** *Now at The University of Minnesota.*

† *Now at the Department of Defense.*

A number of investigators have proposed that these effects are primarily phonetic in origin (see, for example, Liberman and Studdert-Kennedy, 1979). That is to say, they reflect the fact that a listener utilizes some uniquely linguistic knowledge in order to interpret the silence cue in an appropriate way. Accepting this view, Dorman, Raphael and Liberman (1979) sought to characterize some aspects of the knowledge underlying perception of silence as a cue for yet another manner contrast, that between fricative and affricate.[1] To begin with, they demonstrated that a natural utterance of *please say shop* (*shop* beginning with the fricative /ʃ/) can be heard as *please say chop* (*chop* beginning with the homorganic affricate /tʃ/) if a brief interval of silence is introduced between the words *say* and *shop*. The complement is also true: removal of the naturally produced silence from an utterance of *please say chop* cues perception of *please say shop* (Raphael and Dorman, 1977). The authors noted that two relevant facts of speech production are these: (1) an affricate is most prominently distinguished from its homorganic fricative by an initial period of vocal tract closure; and (2) closure is acoustically marked by silence. They concluded that in order for silence to cue the fricative-affricate distinction, their listeners must at least have had knowledge of this linguistically-relevant relation between articulation and acoustics.

In a second experiment, they went on to determine whether listeners also had knowledge of a particular condition in which the relation did not hold. Their experimental question was whether silence could still be an effective cue when the identity of the talker was changed across the silent interval. While a single talker cannot produce an utterance like *please say chop* without generating an interval of silence during the affricate closure, two talkers can produce this word string with no such silence (since the second talker can begin speaking before the first has stopped). To test listeners' sensitivity to this, the authors engineered a *please say shop/chop* stimulus from the speech of two talkers, taking *please say* from a man's recording of *please say shop* and the target word from a woman's recording. They found that the presence or absence of silence had no perceptual significance in this condition, and concluded that listeners do take account of talker identity when interpreting the silence cue.

We thought it possible that listeners likewise take account of other conditions that affect the relationship between silence and articulation, and the multiple-talker result suggested a principle that might be applied to predict what certain of those conditions are. That principle is simply that silence should serve as a phonetic cue only when it occurs in a stretch of speech that has been articulated continuously. This certainly fits the multiple-talker case, since with one talker all three words were articulated as a continuous utterance and with two talkers there was an articulatory break immediately before the target word. And, it makes sense in that silence is invariantly related to vocal tract closure only in the context of continuous articulation.

1 *There is some disagreement as to whether or not the affricate is a single phoneme, distinguished from the fricative by the manner of its production. Some would argue that it is, in fact, two phonemes – a stop followed by a fricative (see Chao, 1934, for a discussion). The resolution of this issue is not critical to the current study, however. Our concern is only with the conditions that determine when silence in the signal will cue the fricative-affricate distinction.*

The purpose of the current study was to apply this principle to a case involving the speech of a *single* talker. That is to say, we wanted to manipulate the articulatory continuity of one-talker utterances. In order to do so, it was necessary to find a natural articulatory unit that could either link a talker's target word with its precursor or separate them. The sentence seemed to us appropriate for this purpose. Its role as a unit of articulatory organization has long been recognized. There is considerable evidence, for example, that a talker modulates intonation (Hadding-Koch and Studdert-Kennedy, 1964; Lieberman, 1967) and the relative duration of constituent words (Umeda, 1975, 1977; Klatt, 1975; Lindblom, Lyberg and Holmgren, 1981) so as to mark sentence boundaries acoustically. Also, Lieberman (1967) has pointed out that breath groups generally span whole sentences, and he has proposed that intonation is most prominently regulated in terms of the subglottal air pressure maintained over the course of a breath group (but see Vanderslice, 1967; Ohala, 1974, 1977). And finally, it is relevant that the articulatory organization of successive phonemes has been shown to vary depending on whether they are produced within the same syntactic unit (such as a sentence) or in separate units (Cooper, Lapointe and Paccia, 1977; Cooper, Egido and Paccia, 1978).[2]

With all of this in mind, we designed an experiment to assess the effectiveness of silence as a cue to the fricative-affricate distinction under two different sentence conditions: a *sentence-nonfinal* condition in which the precursor and target word were in the same sentence; and a *sentence-final* condition in which there was a sentence break immediately before the target. Our hypothesis was that silence would effectively cue the *shop/chop* distinction in the first condition but not in the second. So as to be able to directly compare any perceptual effect of sentence finality with that of talker change, we generated both one-talker and two-talker versions of these stimuli.

EXPERIMENT 1

The effects of sentence finality and talker change were tested in Experiment 1 with appropriate pairings of precursors, silent intervals, and target words. After previous investigators (Raphael and Dorman, 1977; Dorman *et al.*, 1979), the target words chosen were *shop* and *chop*. So as to minimize semantic bias, it was thought advisable that the precursor word string be meaningful both as a stand-alone sentence and in combination with these targets. We therefore used the words *let's go*. Two different talkers (a man and a woman) recorded a sentence-nonfinal version of this precursor as part of the statement **Lets go* shop.*[3] and a sentence-final version as the statement **Let's go.**. Also, to

[2] *Many of these observations can be made, as well, about other syntactic junctures such as the phrase. Vowel lengthening occurs at phrase boundaries for example (Klatt, 1975; Umeda, 1977), as does syntactic blocking of coarticulation (Cooper* et al.*, 1977). The sentence juncture was examined in the present experiment because it represents a particularly clear case.*

[3] *The various utterances recorded by our talkers, and the precursors and target words derived from them, are referenced here with "star" (*) notation. Stars bracket an utterance under consideration, and all internal punctuation marks and parenthetical comments refer to it. Hence, for example, *(Male)* Let's go?* *is a two-word utterance of the question "let's go?" spoken by the male talker.*

provide some sense of the generalizability of any demonstrable sentence effect, the male talker recorded an additional sentence-final version in the form of the question *Let's go?*.

Method

Stimuli. Each stimulus paired a precursor with a particular silent interval and target word, the precursors and targets having been derived from natural utterances in the manner described below.

As noted, there were two target words: *shop* and *chop*. The *shop* target was taken from a male talker's recording of the utterance **let's go shop.**. It was excised from that utterance at the first zero-crossing following the voicing for the vowel of *go*. The *chop* target was then generated by modifying *shop:* 20 msec of signal were deleted from the onset of the *shop* target and 30 msec from more medial portions of the frication, thereby creating an abrupt onset and shortened overall duration which are cues for the affricate (Gerstman, 1957). When played in isolation, the targets were unambiguously heard as *shop* and *chop* respectively.

It was necessary to derive the *chop* target from a naturally recorded *shop* because, as noted by Dorman *et al.* (1979), when such other cues to the fricative-affricate distinction as the rate of onset of the noise or the total noise duration take on sufficiently extreme values, they "bias" perception to such a degree that a silence effect may no longer be demonstrable. In the context of the *Let's go shop/chop* utterance, we found this to be particularly the case for naturally produced *chop* tokens. Listeners generally identified such tokens as *chop* even after all of the silence had been removed from an utterance. However, with our derived *chop* target, we were able to select values for the onset and duration cues which specified *chop* when the token was presented in isolation, but which were (at least for some listeners) sufficiently "neutral" to allow the silence cue to have an effect in an appropriate utterance.

For the first experiment, the *shop* target was always paired with a 75-msec silent interval. This duration was chosen because it was well beyond the silence "boundary" (40 msec) previously reported by Dorman *et al.* (1979), and because informal listening indicated that it was sufficient for the *shop* target to be heard as *chop*. The *chop* target was always paired with a 0-msec silent interval; since it is the brevity of silence that cues the fricative, this 0-msec interval was the optimal choice.

As to the precursors, recall that they were versions of the two-word string *let's go*. All were derived from a male and a female talkers' recordings of the statements **Let's go.** and **Let's go shop.**, and from the male talker's additional recording of the question **Let's go?**. The male talker recorded his utterances first, speaking in time with a metronome (150 beats per minute). This metronomic constraint was included to ensure that all utterances were spoken at roughly the same rate.[4] This talker was instructed to speak each word on a beat and to produce the entire utterance on a single breath. The

[4] *It was deemed important to roughly equate the rate at which all precursors were spoken becasue Repp* et al. *(1978) have shown that listeners are sensitive to this variable when interpreting the silence cue.*

TABLE 1

Acoustic parameters of the precursor tokens

Sentence Condition	Talker	Token	F_0 Change (Hz)[a]	Duration (msec)		$F2$ Change (Hz)	
				let's	go	Interval 1[b]	Interval 2[c]
Nonfinal	Male	**Let's go . . .**	0	100	170	–40	+45
	Female	**Let's go . . .**	–25	90	170	–30	+55
Final	Male	**Let's go.**	–75	110	210	–35	–25
		Let's go?	+45	100	230	–25	–15
	Female	**Let's go.**	–65	110	210	–30	–30

[a]F_0 change was measured over the voiced region of the word *go*.

[b,c]Interval 2 was the final 75 msec of voicing for *go*. Interval 1 was the immediately preceding 75 msec.

female talker was given the same instructions; in addition, she was asked to listen to the male's recordings and to mimic his prosody in her own productions.

Acoustic analyses showed that several different cues appropriately distinguished the sentence-final tokens from the nonfinal tokens. These cues are summarized in Table 1. First of all, we report, for each token, the amount of fundamental-frequency (F_0) change (in Hz) measured over the voiced interval of the word *go*. As one would expect, for the nonfinal tokens there was only a negligible F_0 change, while for the sentence-final tokens there was a substantial modulation of F_0. That modulation was, of course, a fall in the case of the statement and a rise for the question. As an index of sentence-final word lengthening, we measured the duration of voicing for each of the two precursor words. Note in the table that the duration of the initial word (*let's*) is about the same for all tokens, while that of the final word (*go*) clearly distinguishes the sentence-nonfinal tokens from the sentence-final tokens, the duration being substantially longer for the latter.

Although it is not strictly a cue to sentence finality, we report one additional parameter in the table – the frequency change (in Hz) of the second formant. This was measured for two intervals: the final 75 msec of the word *go;* and the immediately preceding 75 msec. For all sentence-final tokens, $F2$ fell throughout both of these intervals as would be expected since the vowel /o/ is produced with an offglide to the high back tongue position appropriate for /u/. With the nonfinal tokens, however, $F2$ had a somewhat different trajectory. It fell over the first interval and rose sharply over the second. Apparently, the tongue was initially shifted up and back for the /o/ offglide, but then,

TABLE 2

Set of precursor utterances for Experiment 1

Sentence Conditions	Source Conditions: One Talker	Source Conditions: Two Talkers
Nonfinal	*(Male) *Let's go . . .**	*(Female) *Let's go . . .**
Final	*(Male) *Let's go.**	*(Female) *Let's go.**
	*(Male) *Let's go?**	
	*(Male) *Let's go.* (Modif)*	

owing to anticipatory coarticulation with the following word *shop*, was moved forward toward the alveolar place of articulation appropriate for /ʃ/. This cue to coarticulation was doubtless another source of information about the articulatory continuity of nonfinal tokens.

We created one additional precursor token, a modified version of the male talker's sentence-final statement (**Let's go.**). That statement was distinguished from his nonfinal token (**Let's go . . .**) by a marked amplitude declination that occurred over the final 250 msec of the utterance. Such a declination is to be expected in that subglottal air pressure is generally falling at the end of a sentence (Lieberman, 1967). Nevertheless, it posed a problem in that the reduced acoustic energy in the sentence-final token made it somewhat less effective as a potential forward-masking stimulus than its nonfinal counterpart. So as to avoid confounding our sentence conditions with this psychoacoustic variable, we generated the modified token by amplifying the final 250 msec of the **Let's go.** statement so as to approximate the acoustic energy of the nonfinal precursor. The resulting precursor token (**Let's go.* (Modif)*) sounded unnatural and appreciably different from the **Let's go.** statement from which it was derived. It was nevertheless evident that it was a sentence-final token.

There were, then, six precursors in all. These are listed in Table 2 according to the conditions of sentence finality and source (number of talkers) that they satisfied.[5] Twelve stimuli were generated for Experiment 1. Six of these paired the *shop* target (and 75 msec of silence) with each of the precursors. The other six paired the *chop* target (and 0 msec of silence) with each precursor.

[5] *It can be seen in Table 2 that the stimuli comprised all possible combinations of source (one talker, two talkers) and* sentence *(final, nonfinal). This made it possible to make statistical statements about the independent influences of these two factors (their main effects) and about the nature of their interaction.*

Procedure. Ten repetitions of each of the 12 stimuli were randomized into a single test list. This list was presented over headphones at a comfortable listening level, the stimuli being separated by a three-second interstimulus interval. Listeners were tested in groups of four or five in a sound-attenuated room. They were instructed to listen carefully to each three-word stimulus and to identify the final word as *shop* or *chop*. Responses were indicated by writing *s* or *c* on an answer sheet.

Subjects. A total of 29 listeners participated in this experiment. All of them had normal hearing and were native speakers of English. They were all naive as to the purpose of the experiment. There was wide variability in their sensitivity to the silence cue. As noted above (see the section on Stimuli), silence is not the only cue to the fricative-affricate distinction, and it can be "overridden" by other cues if their values are sufficiently extreme. There appear to be some individual differences in listeners' sensitivity to these other cues.

Since our interest was in assessing the perceptual significance of sentence finality, rather than in the fricative-affricate distinction per se, we needed to restrict our attention to just those subjects for whom the silence manipulation was effective. Therefore, we devised a screening procedure: listeners were screened on the basis of their responses to the standard (one-talker sentence-nonfinal) **Let's go shop/chop.** stimulus. They were included in all analyses of the *shop*-target stimuli if they identified that target as *chop* on at least 50% of the trials in which it appeared in the standard stimulus. Likewise, they were included in all analyses of the *chop*-target stimuli if they identified that target as *shop* on at least 50% of the trials in which it appeared in the standard stimulus. Of 29 listeners tested, 24 met the criterion for the *shop* target and 12 met it for the *chop* target.

Results

The results are summarized in Figures 1 and 2. Figure 1 displays the mean percentage of trials in which the *shop* target was judged to be *chop* for each of the experimental conditions and Figure 2 the percentage in which the *chop* target was judged to be *shop*. Two-way analyses of variance, comparing the effects of sentence finality and source, were performed on these data. The analysis of the *shop* target showed a highly significant sentence effect, $F(1,23) = 348.33$, $p < 0.001$, while the effect of source and the interaction term were not significant, $F < 1.0$. In the analysis of the *chop* target, sentence finality, $F(1,11) = 52.17$, $p < 0.001$, source, $F(1,11) = 11.15$, $p < 0.01$, and the interaction, $F(1,11) = 10.89$, $p < 0.01$, were all found to be significant.

In both Figures 1 and 2, the value displayed for the one-talker sentence-final condition is the mean for three different precursors: the statement, the question, and the modified statement (see Table 2). Analyses of variance showed no significant differences among these precursors, either for the *shop* target, $F(2,46) = 2.65$, or the *chop* target, $F(2,22) < 1.0$. This indicates that the effect of sentence finality was sufficiently general as to be manifest for both the statement (**Let's go.**) and the question (**Let's go?**). Also, the effect was not diminished by modification of the amplitude declination at the end of the statement, which speaks against any attempt to account for these data in terms of the phenomenon of forward masking.

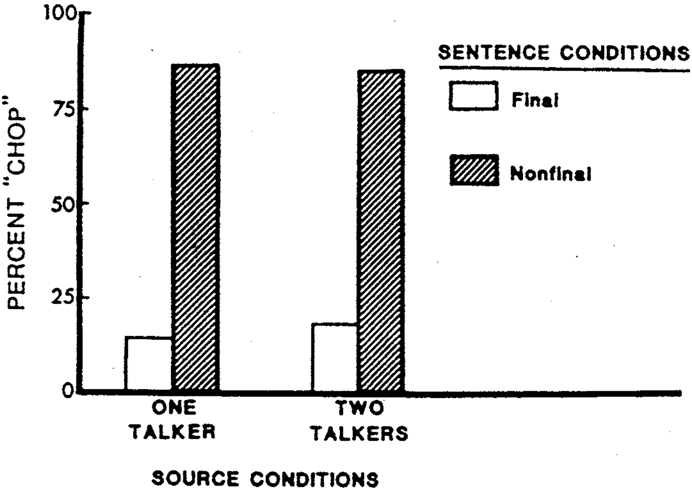

Fig. 1. Mean percentage of *chop* judgments given for the *shop* target in Experiment 1.

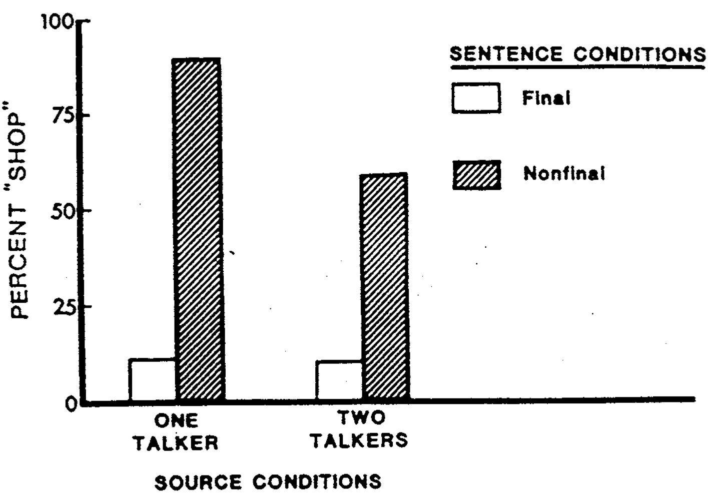

Fig. 2. Mean percentage of *shop* judgments given for the *chop* target in Experiment 1.

Discussion

The primary experimental question was whether a sentence boundary (and presumably a corresponding break in articulation) could be shown to affect the phonetic significance of silence, and the answer to that question is clearly yes: a significant sentence effect was observed for both the *shop* and *chop* targets. It was expected that this result, along with the finding that a change of talkers likewise affects the silence cue, would provide evidence that listeners interpret the silence according to the principle that it can have phonetic significance only when it occurs in a stretch of speech that has been produced as a continuous articulation. The interpretation cannot be quite so straightforward however, given our observations concerning the effect of a talker change.

Those observations are usefully compared with the previously reported findings of Dorman *et al.* (1979). First of all, we confirmed their finding that talker change can affect the silence cue, since a main effect of source was observed for the *chop* target. Somewhat surprisingly, however, we also found that this effect could be "overridden": no main effect of source was observed for the *shop* target. To be more specific, we found that silence could often be a cue in sentence-nonfinal *two-talker* stimuli, while Dorman *et al.* did not. It is our view that this discrepancy can be most reasonably explained in terms of the conflicting cues to articulatory continuity that were present in such stimuli. Cues specific to the source of the utterance presumably signalled an articulatory break, while, at the same time, other cues (such as those listed in Table 1), indicated that the articulation of a sentence was still in progress. Apparently, listeners resolved this conflict by "ignoring" the talker change. Results of this type are common whenever listeners are confronted with conflicting information. For instance, as noted earlier, the fricative-affricate distinction is itself marked by several different cues and if one of them assumes a sufficiently extreme value listeners may attend to it and "ignore" the remaining cues. Likewise, with respect to articulatory continuity, it appears that in the case of our *shop* stimuli, though not in that of the *chop* stimuli, the several cues for continuous articulation overrode those for an articulatory break. Given other tokens and other talkers, however, we would not be surprised to see somewhat different results. Nevertheless, this finding indicates that it is not strictly articulatory continuity, but rather *perceived* articulatory continuity that is critical for determining the phonetic significance of silence.

While we were inclined to this account of our multiple-talker data from the outset, there was a second possibility that needed to be considered. Listeners may simply have failed to hear the source change in the two-talker stimuli and, perceiving only one talker, processed the silence phonetically. This possibility was particularly disturbing from our standpoint since it suggested an alternative account of the sentence effect as well. It is conceivable that listeners correctly heard all sentence-nonfinal stimuli to have been spoken by one talker, but somehow misheard the sentence-final stimuli to have been spoken by two talkers. If so, then sentence finality – the variable of principal interest here – was confounded with perceived source, and the latter variable alone might account for the silence effects observed in this experiment. It became important, therefore, to determine how many talkers were heard in each experimental condition. This was done in Experiment 2.

A second, and rather different, methodological point also limited our ability to draw conclusions from this first experiment. We have asserted that the nature of the sentence effect was such that silence had phonetic significance when it followed nonfinal precursors but had no significance when it followed sentence-final precursors. At least for the *shop*-target stimuli, however, this observation is subject to some qualification. All of these stimuli contained the same silent interval (75 msec), and it is well known that the specific duration of silence needed to cue a contrast can vary with context. For example, Repp *et al.* (1978) reported that in a rapid utterance context (as defined by a rapidly spoken precursor) subjects (paradoxically) required more silence to cue the perception of an affricate than was needed to cue it in a slow context. In Experiment 1, the perceptual effect of the sentence break may similarly have been to shift the silence cue "boundary" (rather than to negate the cue entirely). Thus, in the context of sentence finality, silence might still be a cue that would have phonetic significance for a listener at longer silent intervals.

It should be noted that this issue does not arise for the *chop*-target stimuli. For these, it is clear that the silence cue was negated by the sentence break since it is the brevity of silence that cues the fricative and the silent interval tested here (0 msec) was maximally brief. Nevertheless, for the *shop* target, it was important to test for the sentence effect over a range of silent intervals. This was done in Experiment 3.

EXPERIMENT 2

It was hypothesized earlier that the data of Experiment 1 might be explained, in whole or in part, by listeners' misperceptions of the number of talkers who produced the stimuli. For example, this could explain why, contrary to Dorman *et al.* (1979), we found that under certain conditions silence cued phonetic perception despite a change of talkers. Perhaps in these cases listeners failed to hear the change. A more disturbing possibility is that this *perceived-source* hypothesis could also account for the sentence effect observed in Experiment 1. This could be the case if the two sentence conditions (final and nonfinal) differentially affected the number of talkers heard. In order to test these possibilities, we ran a second group of subjects on a subset of the stimuli used in Experiment 1, and asked them both to judge the number of talkers that they heard and to make phonetic judgments.

Method

The *shop* target and 75 msec of silence were paired with a sentence-final precursor and a sentence-nonfinal precursor, each spoken by the male talker and the female talker. Ten repetitions of each of the four stimuli were then randomized to produce a test list. This was presented to eight new listeners under the same testing conditions as those described in Experiment 1. For each stimulus, they were asked, first, to identify the number of talkers as *one* or *two*, and then to identify the final word as *shop* or *chop*.

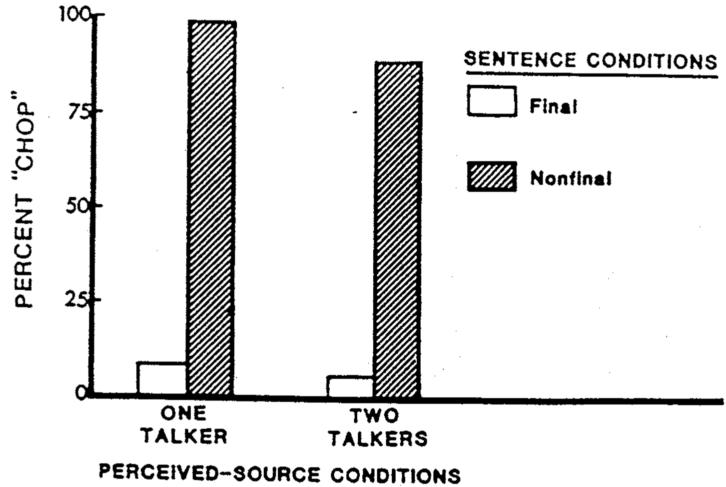

Fig. 3. Mean percentage of *chop* judgments (given for the *shop* target) depending on whether listeners heard one talker or two.

Results and discussion

The results speak against any account of the earlier data that assumes a misperception of source. First of all, listeners made relatively few errors in source identification. Their judgments were correct for 85.7% of all one-talker stimuli and 95.6% of all two-talker stimuli. More importantly, the errors that were made were not significantly associated with the sentence conditions, χ^2 (1) = 1.52, making it impossible for the perceived-source factor to have mediated the sentence effect.[6]

A look at the phonetic judgments data confirms this last observation. In Figure 3, these data have been grouped according to the factors of sentence finality and *perceived* source (i.e., the number of talkers heard). A two-way analysis of variance was computed to compare these factors. The sentence effect was significant, $F(1,7) = 226.73$, $p < 0.001$, but the perceived-source effect, $F(1,7) = 1.25$, and the interaction term, $F(1,7) < 1.0$, were not. Thus, there is no evidence that perceived source influenced listeners' phonetic judgments, either in this experiment or in Experiment 1.

6 *It might be added that in post-experimental interviews the subjects informally reported that they had great confidence in both their source judgments and their phonetic judgments. They said that while many of the stimuli sounded "unusual" to them in that there was a change of talkers in the middle, that change was easily detected and it created no difficulties for identifying the target words as* shop *or* chop.

EXPERIMENT 3

This experiment speaks to two competing hypotheses concerning the effect of sentence finality. One is that it shifts the location of the "boundary" for cueing the affricate with silence. This could be so, for example, if sentence finality alters the perceived rate of articulation which, in turn, affects the "boundary," or if the rapidly modulating fundamental frequency of sentence-final precursors somehow interacts with relevant cues for the fricative-affricate distinction. The second hypothesis concerning the effect of sentence finality is that it negates the silence cue entirely. In Experiment 1 these hypotheses could not be distinguished, since all of the *shop*-target stimuli contained the same 75-msec silent interval. In this experiment, we overcame that limitation by constructing stimuli with a range of silent intervals. If sentence finality does shift the "boundary," then somewhere along this range silence should begin to cue the affricate consistently for sentence-final stimuli (just as it did for sentence-nonfinal stimuli in Experiment 1). Alternatively, if the silence cue is negated by the sentence break, then we should expect relatively few *chop* responses at every silent interval. We also included sentence-nonfinal stimuli in this experiment. For these, the prediction is clear: at silent intervals longer than some "boundary" value silence should consistently cue perception of the affricate.

Method

Two sets of stimuli were generated: one paired a range of silent intervals and the *shop* target with the male talker's sentence-nonfinal precursor (**Let's go . . .**); the other paired them with one of his sentence-final precursors (the statement **Let's go.**). The silent intervals were 0, 15, 30, 45, 60, 75, 90, 105, and 120 msec. Eight repetitions of each of the stimuli were randomized into a single test list. This was presented to nine listeners who had not participated in either of the first two experiments. Again, these listeners had normal hearing, were native speakers of English, and had no knowledge of the purpose of the study. They were tested in groups and performed the same task as that of Experiment 1.

Results and discussion

The results of the listening test are presented in Figure 4. Note first that, as expected, listeners' phonetic judgments varied with the silent interval when the precursor was nonfinal (the condition represented by the solid line). At the shorter intervals, listeners generally reported hearing *shop;* at the longer ones they reported *chop.* This is clear evidence that when the target word and its precursor are produced as parts of the same sentence a fricative-affricate phonetic "boundary" can be found for the silence cue.

The location of the "boundary" is somewhere around 15 msec. It should be noted that the 75-msec silent interval used in Experiment 1 was well beyond this point, which further validates that earlier choice. Also of interest is a comparison of this 15-msec "boundary" with that of 40 msec reported by Dorman *et al.* (1979). The difference may, at least in part, reflect the fact that the respective utterances were spoken at somewhat

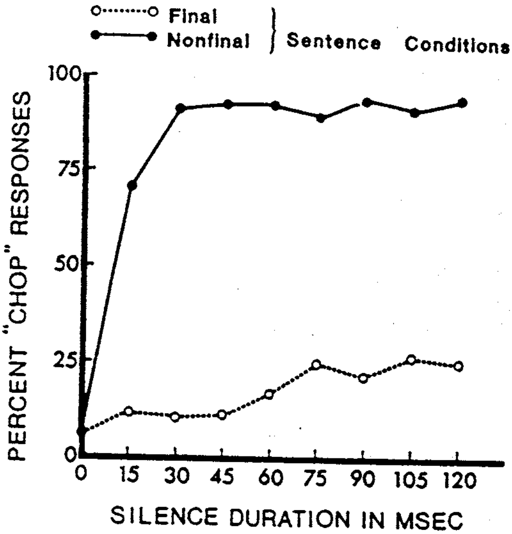

Fig. 4. Mean percentage of *chop* judgments (given for the *shop* target) as a function of the duration of the silent interval.

different rates (see the earlier discussion of rate effects). A more interesting possibility is that it reflects differences in the manner and/or degree of coarticulation of the precursor and target. Dorman *et al.* used the utterance frame *please say shop/chop.* Tongue movement for the vowel /e/ (of *say*) offglides toward the high front position appropriate for /i/, the lips remaining relatively retracted. By contrast, with our utterance frame of *let's go shop/chop,* the tongue must offglide (for /o/) into a high back position roughly appropriate for /u/, with the lips rounding throughout the gesture. It is not clear why these differences should shorten the duration of silence needed to cue the affricate; it is clear, however, that there is a general affricate "bias" in this utterance frame.

Let us return to the question of major interest here: can a boundary also be found when there is a sentence break? If so, we may conclude that when listeners detect such a break the perceptual effect is to shift the boundary for the silence cue. If we find no boundary, we may conclude that the discontinuity negates the silence cue entirely. Clearly the latter conclusion is supported. In the sentence-final condition (represented by the dotted line), listeners generally heard *shop* at everv silent interval tested. Thus, there is no evidence of a boundary. The significant implication of this finding is that manipulation of sentence finality affects the interval over which listeners integrate cues to a phonetic contrast. In the nonfinal case, integration of cues for the fricative-affricate distinction begins with the period of silence that precedes onset of the frication. In the sentence-final case, on the other hand, silence is negated as a cue and integration begins with the frication itself.

General Discussion

In a previous study (Dorman *et al.*, 1979), it was shown that the presence or absence of silence affected perception of the fricative-affricate distinction when it occurred in a three-word string spoken by a single talker, but had no effect in an identically worded string in which talker identity was changed over the silent interval. One way of thinking about that result is in terms of the articulatory continuity of the stimuli as a whole. Articulation was interrupted when there was a change of talkers, but it was continuous for the stimulus spoken by a single talker. Appropriately, listeners interpreted the silence phonetically only in the continuous case.

In the current study, we manipulated articulatory continuity in a way that could affect the speech of one talker as well as two. Target words and their precursors were either linked as parts of a single sentence production, or were separated by a sentence boundary. Silence was shown to have phonetic significance only for the former case. We take this sentence effect, along with the previously demonstrated source effect (which was partially replicated here), to be consistent with the general principle that silence will have phonetic significance for a listener only when it is perceived to have occurred in a stretch of speech that was articulated continuously.

It is noteworthy that articulatory continuity is defined at the utterance level, which makes its influence on the perception of an individual phonetic segment (/ʃ/ vs. /tʃ/) a "top-down" influence. The sentence effect may be particularly interesting in this regard since it is possible to achieve it in cases where there can be no question of response bias. In Experiment 1, for instance, we held the word string constant and, by altering local phonetic cues, created both *shop*-to-*chop* and *chop*-to-*shop* perceptual changes. These complementary results could not both be attributed to bias.

We have mentioned a second aspect of this phenomenon that we would like to highlight here. By manipulating the articulatory continuity of an utterance, it is possible to affect the interval over which a listener integrates cues in phonetic perception. In the case of continuous articulation, the integration of cues for the fricative-affricate distinction begins with the interval of silence preceding the onset of frication, while with a break in articulation it begins with the frication itself. Liberman (1982) and others (Repp, 1978; Oden and Massaro, 1978) have pointed out that since acoustic cues have both proactive and retroactive influences on phonetic perception, a listener must somehow determine the appropriate temporal interval over which to integrate the several cues bearing on any individual contrast. The present data indicate that, for the fricative-affricate contrast at least, the onset of the interval is dictated by the perceived onset of the suprasegmental articulatory unit of which the relevant phonetic segment is a part.

Finally, we would comment that our results are consistent with previous proposals (e.g., Liberman and Studdert-Kennedy, 1979) that the various silence effects are primarily phonetic in origin. On this general view, silence effects are said to arise because listeners use certain linguistically relevant knowledge to interpret silence cues appropriately. The experiments reported here can be taken to demonstrate a specific example of this general principle: listeners use the knowledge that silence is invariantly related to vocal-tract closure only when it occurs in the context of continuous articulation.

The alternative to this phonetic account would be to ascribe our results to some sort of auditory effect, but no currently understood principles of auditory psychophysics can be generalized to explain the several results observed here. Nor do auditory principles offer a basis for relating our single-talker sentence effect to the multiple-talker source effect. The phonetic principle of articulatory continuity, on the other hand, makes this relation straightforward.

REFERENCES

BASTIAN, J., EIMAS, P. and LIBERMAN, A.M. (1961). Identification and discrimination of a phonemic contrast induced by silent interval. *Journal of the Acoustical Society of America*, **33**, 842 (Abstract).

CHAO, Y.R. (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of Historial Philology Academia Sinica*, **IV**, 363-397.

COOPER, W.E., EGIDO, C. and PACCIA, J.M. (1978). Grammatical control of a phonological rule: Palatalization. *Journal of Experimental Psychology: Human Perception and Performance*, **2**, 264-272.

COOPER, W.E., LAPOINTE, S.G. and PACCIA, J.M. (1977). Syntactic blocking of phonological rules in speech production. *Journal of the Acoustical Society of America*, **61**, 1314-1320.

DORMAN, M.F., RAPHAEL, L.J. and LIBERMAN, A.M. (1979). Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, **65**, 1518-1532.

FITCH, H.L., HALWES, T., ERICKSON, D.M. and LIBERMAN, A.M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception and Psychophysics*, **27**, 343-350.

GERSTMAN, L. (1957). Cues for distinguishing among fricatives, affricates, and stop consonants. Ph.D. dissertation. New York University.

HADDING-KOCH, K. and STUDDERT-KENNEDY, M. (1964). An experimental study of some intonation contours. *Phonetica*, **11**, 175-185.

KLATT, D. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, **3**, 129-140.

LIBERMAN, A.M. (1982). On finding that speech is special. *American Psychologist*, **37**, 148-167.

LIBERMAN, A.M. and STUDDERT-KENNEDY, M. (1979). Phonetic perception. In R. Held, H. Leibowitz and H.L. Teuber (eds.), *Handbook of Sensory Physiology*, Volume VIII, "Perception," (Heidelberg), pp. 145-178.

LIEBERMAN, P. (1967). *Intonation, Perception, and Language* (Boston).

LINDBLOM, B., LYBERG, B. and HOLMGREN, K. (1981). Durational patterns of Swedish phonology: Do they reflect short-term memory processes? Indiana University Linguistics Club (Bloomington, IN).

LISKER, L. (1957). Closure duration and the voiced-voiceless distinction in English. *Language*, **33**, 42-49.

ODEN, G.C. and MASSARO, D.W. (1978). Integration of featural information in speech perception. *Psychological Review*, **85**, 172-191.

OHALA, J. (1974). A mathematical model of speech aerodynamics. In G. Fant (ed.), *Proceedings of Speech Communication Seminar*, Volume II (Stockholm), pp. 65-72.

OHALA, J. (1977). The physiology of stress. In L.M. Hyman (ed.), *Studies in Stress and Accent* (Los Angeles), pp. 145-168.

PORT, R.F. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics*, **7**, 45-56.

RAPHAEL, L.J. and DORMAN, M.F. (1977). Perceptual equivalence of cues for the fricative-affricate contrast. *Journal of the Acoustical Society of America*, 61, S45 (Abstract).
REPP, B.H. (1978). Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception and Psychophysics*, 24, 471-485.
REPP, B.H., LIBERMAN, A.M., ECCARDT, T. and PESETSKY, D. (1978). Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 621-637.
UMEDA, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America*, 58, 434-445.
UMEDA, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America*, 61, 846-858.
VANDERSLICE, R. (1967). Larynx vs lungs: Cricothyrometer data refuting some recent claims concerning intonation and archetypality. *UCLA Working Papers in Phonetics*, 7, 67-79.