Journal of Experimental Psychology:
Human Perception and Performance
1978, Vol. 4, No. 4, 621–637

Perceptual Integration of Acoustic Cues for Stop, Fricative, and Affricate Manner

Bruno H. Repp, Alvin M. Liberman, Thomas Eccardt, and David Pesetsky
Haskins Laboratories, New Haven, Connecticut

Introducing a short interval of silence between the words SAY and SHOP causes listeners to hear SAY CHOP. Another cue for the fricative–affricate distinction is the duration of the fricative noise in SHOP (CHOP). Now, varying both these temporal cues orthogonally in a sentence context, we find that, within limits, they are perceived in relation to each other: The shorter the duration of the noise, the shorter the silence necessary to convert the fricative into an affricate. On the other hand, when the rate of articulation of the sentence frame is increased while holding noise duration constant, a longer silent interval is needed to hear an affricate, as if the noise duration, but not the silence duration, were effectively longer in the faster sentence. In a second experiment, varying noise and silence durations in GRAY SHIP, we find that given sufficient silence, listeners report GRAY CHIP when the noise is short but GREAT SHIP when it is long. Thus, the long noise in the second syllable disposes listeners to displace the stop to the first syllable, so that they hear not a syllable-initial affricate (i.e., stop-initiated fricative) but a syllable-final stop (followed by a syllabe-initial fricative). Repeating the experiment with GREAT SHIP as the original utterance, we obtain the same pattern of results, together with only a moderate increase in GREAT responses. In all such cases, the listeners integrate a numerous, diverse, and temporally distributed set of acoustic cues into a unitary phonetic percept. These several cues have in common only that they are the products of a unitary articulatory act. In effect, then, it is the articulatory act that is perceived.

When a speaker makes an articulatory gesture appropriate for a phonetic segment, the acoustic consequences are typically numerous, diverse, and distributed over a relatively long span of the signal. In the articulation of an intervocalic stop consonant, for example, the characteristically rapid closing and opening of the vocal tract has acoustic consequences that include, among others, the following: various rising and falling transitions of the several formants; a period of significantly reduced sound intensity; and then a second, acoustically different set of formant transitions, plus (in the case of voiceless stops in iambic stress patterns) a transient burst of sound, a delayed onset of the first formant, and, for the duration of that delay, band-limited noise in place of periodic sound in the higher formants.

Despite their obvious diversity and their distribution over periods as long as 300

This research was supported by National Institute for Child Health and Human Development Grant HD01994 and by BRSG Grant RR05596. We would like to thank Georgann Witte for her help in data analysis. A short version of this paper was presented at the 94th meeting of the Acoustical Society of America in Miami Beach, Florida, December 1977.
Additional affiliations of the authors are as follows: Alvin M. Liberman, University of Connecticut and Yale University; Thomas Eccardt, Yale University; David Pesetsky, Massachusetts Institute of Technology.

Request for reprints should be sent to Bruno H. Repp, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

msec (Repp, Note 1), these acoustic features—usually referred to as "cues"—are nevertheless integrated into the unitary perception of a phonetic segment. In such cases of integration we find trading relations among the several cues that take part: Within limits, one cue can be exchanged for another without any change in the phonetic percept; in that sense, the cues are perceptually equivalent, though they may differ greatly in acoustic (and presumably auditory) terms and be quite far removed from each other in time.

To find the basis for the perceptual integration and for the perceptual equivalence it implies, we should first ask what it is that these diverse features have in common. As already implied, we have not far to look: Each is one of the normal products of the same linguistically significant act. Given that commonality, and given the convergence on a unitary phonetic percept, we find it most parsimonious to suppose that the acoustic cues are processed by a system specialized to perceive the phonetically significant act by which they were produced. On that assumption, the boundaries of the integration would be set not by the number, diversity, or temporal distribution of the cues but rather by a decision that they do (or do not) plausibly specify an articulatory act appropriate for the production of a single phonetic segment.

Just how the various cues contribute, separately and in various combinations, to the integrated phonetic percept has been the subject of the many experimental studies of speech perception carried out over the last 30 years. These have established the more or less important roles of the cues and, either directly or by implication, have outlined the trading realtions—hence perceptual equivalences—among them. In one of the most recent of these studies, Summerfield and Haggard (1977) made explicit how a trading relation among the cues for the voicing distinction is to be understood by taking account of the fact that they are the common products of the same articulatory act. A more general discussion of this matter, with examples of the several classes of cues that engage in such trading relations, is to be found in Liberman and Studdert-Kennedy (in press).

In the experiments reported here, we put our attention on simple cues of a temporal sort: duration of silence and duration of fricative noise. We examined their integration in the perception of the distinction between fricative and affricate, and we also investigated the effect on that integration of a still more widely distributed temporal variable, namely, the rate at which the surrounding speech is articulated. In the second experiment, we studied the effects of those same temporal cues, but now in connection with the perception of juncture. That provided us with an opportunity to examine a case in which the integration occurs across syllable boundaries: A syllable-final stop is perceived or not, depending on a cue in the next syllable that simultaneously determines whether the initial segment in that syllable is taken to be a fricative or an affricate.

Experiment 1

In this experiment we selected two cues for study, both temporal in nature and both relevant to the fricative–affricate distinction. One of them is silence. A short period of silence (or near silence) in the acoustic signal tells the listener that the speaker has closed his vocal tract, a gesture characteristic of stop consonants and affricates. That silence is a powerful and often sufficient cue for the perception of stop or affricate manner can be experimentally demonstrated by inserting silence at the appropriate place in an utterance. So, for example, SLIT can be converted into a convincing SPLIT by inserting a sufficient amount of silence between the fricative noise and the vocalic (LIT) portion. That was done originally in tape-splicing experiments (Bastian, Eimas, & Liberman, 1961; Bastian, Notes 2 & 3). For the same phonetic contrast, investigators have more recently explored the range of effective silence durations (Dorman, Raphael, & Liberman, Note 4) and, in another study, revealed a trading relation between silence and a spectral cue (Erickson, Fitch, Halwes, & Liberman, Note 5).

Other contrasts—similar in that they, too, are based on the presence or absence of stop or stoplike manner—have also been found to depend in important ways on the silence cue. Thus, with appropriate insertions of silence, SI can be made to sound like SKI, or SU like SPU (Bailey & Summerfield, Note 6). Silence can also be sufficient to cue the fricative–affricate contrast in, for example, SAY SHOP versus SAY CHOP (Dorman et al., Note 4); it is this contrast that concerns us here.[1]

For the fricative–affricate contrast, there are, as always, other cues besides silence. The one we used in our experiment is duration of (fricative) noise, a cue shown originally by Gerstman (1957) to be important. Thus, we had two temporal cues, duration of silence and duration of noise, and we shall see how they are integrated to produce the perception of affricate manner. Then we shall examine the effect on that integration of a variable that is also temporal in nature: rate of articulation. Our interest in introducing that variable springs from several sources. We might expect, first of all, that the effect of articulatory rate would be especially apparent on cues that are themselves durational in nature. Several studies tend to confirm that expectation (e.g., Ainsworth, 1974; Fujisaki, Nakamura, & Imoto, 1975; Pickett & Decker, 1960; Summerfield, Note 7). Indeed, one of these studies dealt with the same fricative–affricate contrast we studied, and it reported a seemingly paradoxical effect: Having determined that increasing the duration of silence between SAY and SHOP was sufficient to convert the utterance PLEASE SAY SHOP to PLEASE SAY CHOP, Dorman et al. (Note 4) found that when the rate of the precursor PLEASE SAY was increased, more silence was needed to produce the affricate in CHOP. We wished to test for that effect at each of several durations of the fricative noise, and in a larger sentence context. The results may then bear on an interpretation of the paradoxical effect that is consistent with the hypothesis we have advanced to account for the integration of the segmental cues themselves—namely, that perception takes account of production.

To appreciate the point, we should take note of the claim by students of speech production (e.g., Kozhevnikov & Chistovich, 1965) that changes in rate of articulation do not stretch or compress all portions of the speech signal proportionately. In that connection, the data most relevant to our purposes are owing to Gay (1978). He found that durations of silence associated with stop consonants change less with rate than do the durations of the surrounding vocalic portions. It is possible, then, that the somewhat corresponding cue elements of our experiment—duration of silence and duration of fricative noise—are, in like fashion, differentially affected by changes in speaking rate. If so, and if perception is indeed guided, as it were, by tacit knowledge of the consequences of articulation, then we should expect that the perceptual integration of the two cues would reflect such inequalities as the production may have caused.

Method

Subjects. Seven paid volunteers (Yale University undergraduates) participated, as did three of the authors (Repp, Eccardt, & Pesetsky). All except Repp are native speakers of American English (he learned German as his first language). The results of all 10 subjects were combined since there were no substantial differences among them.

Stimuli. A male talker recorded the sentence WHY DON'T WE SAY SHOP AGAIN at two different speaking rates, using a monotone voice and avoiding emphatic stress on any syllable. The fast sentence lasted 1.26 sec, and the slow sentence lasted 2.36 sec—a ratio of .53. The sentences were low-pass filtered at 4.9 kHz and digitized at a sampling rate of 10 kHz. This was done with the Haskins Laboratories Pulse Code Modulation (PCM) system. Monitoring the waveforms on high-resolution oscillograms, we excerpted the SH noise of the slow utterance (110 msec in duration) and substituted it for the SH noise in the fast utterance (originally 92 msec). Thus, the two utterances had identical noise portions.

---

[1] It may be noted that the stop consonants (affricates) in the three examples given have different places of articulation. Perceptual information about place of articulation is provided by spectral cues preceding and following the silence (Bailey & Summerfield, Note 6). In our experiments we are concerned only with cues for stop manner and not with place distinctions. Therefore, we pass over the question why, in the last example, listeners hear SAY CHOP (SAY TSHOP) and not SAY PSHOP or SAY KSHOP.

Knowing that rate of onset of the fricative noise is an important cue for the fricative–affricate distinction (Cutting & Rosner, 1974; Gerstman, 1957), we were concerned that it be not too extreme. Preliminary observations suggested that the noise onset in our stimuli was so gradual as to bias the perception strongly toward fricative and even perhaps to override the effects of the two duration cues we wished to study. To remove, or at least reduce, that bias, we removed the initial 30 msec of the noise, leaving 80 msec. That maneuver had the effect of creating a more abrupt onset.[2]

We used the PCM system to vary the two temporal cues under study: noise duration and silence duration. Three different noise durations were created by either duplicating or removing 20 msec from the center of the 80-msec noise, leaving the onset and offset unchanged. Thus, the noise durations were 60, 80, and 100 msec in both sentence frames. In each of the resulting six sentences, varying amounts of silence were inserted before the fricative noise. Silence duration was varied from 0 to 100 msec in 10-msec steps. Eleven silence durations, three noise durations, and two speaking rates resulted in 66 sentences. These were recorded in five different randomizations, with 2 sec intervening between successive sentences.

To determine how the different noise durations were perceived outside the sentence context, we prepared a separate tape containing isolated SHOP (CHOP) words excerpted from the test sentences. (The stimuli consisted of the portion from the beginning of the fricative noise to the beginning of the P-closure.) Three different noise durations and two speaking rates yielded six stimuli; these were duplicated 10 times and recorded in a random sequence, separated by 3-sec intervals. The different speaking rates were reflected in the durations of the vocalic portions of the test words; they were 140 msec (slow) and 113 msec (fast).

Procedure. The subjects listened in a quiet room over an Ampex Model 620 amplifier-speaker, as the tapes were played back on an Ampex Model AG-500 tape recorder. Intensity was set at a comfortable level. All subjects listened to the isolated words first, except for the three authors, who took this brief test at a later date. The task was to identify each word as either SHOP or CHOP, using the letters S and C for convenience in writing down the responses and guessing when uncertain. The same responses were required in the sentence test. The listeners were informed about the different speaking rates but not about the variations in noise and silence duration (obviously this does not apply to those authors who participated). After a pause, the sentence test was repeated, so that 10 responses per subject were obtained for each sentence.

Results

Consider first the results obtained for isolated words. Although the original utterance had contained SHOP, the isolated words were predominantly perceived as CHOP. Presumably, this was a consequence of our having cut back the original fricative noise and thus creating not only a shorter noise duration but also a more abrupt onset; both changes would be expected to bias perception toward affricate manner (Gerstman, 1957). Despite the bias, there was a clear effect of the variations in noise duration: The percentages of CHOP responses to the three noise durations (60, 80, and 100 msec) were 99, 91, and 81 (slow rate) and 99, 90, and 73 (fast rate), respectively. Thus, as expected, the probability of hearing an affricate decreased as noise duration increased. In addition, there seemed to be a slight effect of vowel duration at the longest noise duration, again in the expected direction: When the vocalic portion was shorter—this being the only manifestation of the faster speaking rate in the isolated words—the probability of hearing CHOP was lower, which indicated that the noise duration was to some extent effectively longer at the fast speaking rate.

We turn now to the results of the main experiment. That silence was an effective cue for the fricative–affricate distinction in sentence context is shown in Figure 1. There we see that the listeners heard SHOP or CHOP, depending on the duration of the silence that separated the fricative noise from the syllable (SAY) immediately preceding it. This replicates earlier findings (Dorman et al., Note 4). If, as is reasonable, we consider an affricate to be a stop-initiated fricative, then our result is also perfectly consistent with those of other investi-

[2] This manipulation merely created a situation favorable for obtaining the desired effect and in no way affected the validity of the experiment. In fact, our cutting back the noise resulted in a moderate bias in the opposite direction, toward hearing an affricate (CHOP). It should be noted in this connection that not only does SAY SHOP turn into SAY CHOP when silence is inserted but that a natural SAY CHOP can also be turned into SAY SHOP by removing the silence that precedes the fricative noise. Both effects have limits, however: A noise with an extremely abrupt onset will not easily be heard as a fricative even in the absence of silence, and a noise with an extremely gradual onset will not easily be heard as an affricate even if sufficient silence is present.

gators who have found silence to be important in the perception of stop-consonant manner.

We see, further, that duration of fricative noise had a systematic effect, as indicated by the horizontal displacement of the three functions in each panel of Figure 1. The proportion of SHOP responses increased significantly with noise duration, $F(2, 18) = 32.36$, $p < .001$. That effect establishes a trading relation between silence and noise durations: As noise duration increases, more silence is needed to convert SHOP into CHOP.[3]

The effect of speaking rate can be seen by comparing the two panels of Figure 1.

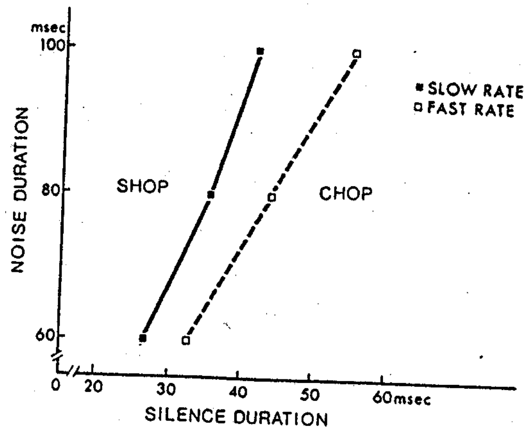

Figure 2. Boundaries between perceived fricative (SHOP) and affricate (CHOP) at each speaking rate as joint functions of the duration of silence and the duration of fricative noise.

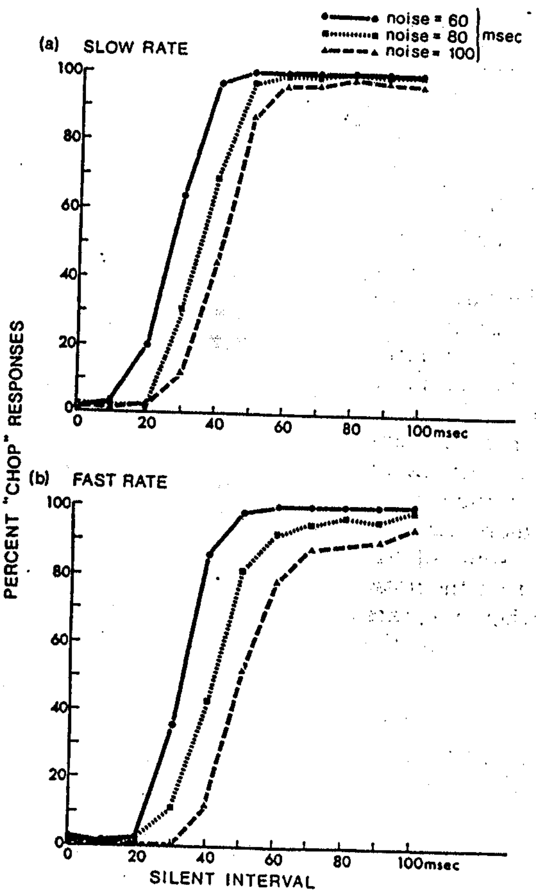

Figure 1. Effect of duration of silence and duration of fricative noise on the perceived distinction between fricative (SHOP) and affricate (CHOP). (This is shown for each of the two rates at which the sentence frame was articulated.)

We see that the paradoxical effect first discovered by Dorman et al. (Note 4) was indeed replicated: For equivalent noise durations, more silence was needed in the fast sentence frame than in the slow sentence frame to convert the fricative into an affricate, $F(1, 9) = 16.35$, $p < .01$.

The foregoing results are represented more concisely in Figure 2 where the data points are the SHOP–CHOP boundaries (i.e., the 50% crossover points of the six labeling functions) as estimated by the method of probits (Finney, 1971). This procedure fits cumulative normal distribution functions to the data; it also yields estimates of standard deviations and standard errors of the

[3] Strictly speaking, the term "trading relation" may not be appropriate for a positve relation between two cues, but for want of a better term, we use it. The positive covariation of the two perceptual cues is a direct consequence of their negative covariation in production: Fricatives have a long noise duration and no silence, whereas affricates have a shorter noise duration preceded by a closure interval. Genuine perceptual trading relations (negative covariation) are observed when two acoustic properties are positively correlated in production, such as, for example, silence and the extent of the first formant transition as cues for stop manner (Bailey & Summerfield, Note 6). In any case, a positive trading relation can be turned into a negative one by simply reversing the directionality of the scale on which one of the cues is measured.

boundaries.[4] To show the trading relation between the temporal cues more clearly, Figure 2 plots the SHOP–CHOP boundaries (abscissa) as a function of noise duration (ordinate) and speaking rate, (the two separate functions). Each function describes a trading relation between noise duration and silence duration by connecting all those combinations of silence and noise durations for which SHOP and CHOP responses are equiprobable. The joint dependence of perceptual judgments on both durational cues is indicated by the fact that the trading functions are neither perfectly vertical nor perfectly horizontal but have intermediate slopes. Both functions are strikingly linear.

Although an increase in speaking rate left the linear form of the trading relation unchanged, it shifted the function toward longer silence durations, simultaneously changing its slope and indicating that rate of articulation had a differential effect on the effective durations of silence and noise. In fact, the trading functions in Figure 2 coincide well with straight lines through the origin of the coordinate system, which means that within each speaking-rate condition, the fricative–affricate boundary is associated with a constant ratio between silence and noise durations—approximately .44 at the slow rate and .55 at the fast rate. A separate analysis of variance of silence/noise ratios showed only a significant effect of speaking rate, $F(2, 18) = 14.60, p < .01$; the effect of noise duration and the interaction term were far from significance. Thus, the consequence of changing the rate of articulation was a change in the ratio of silence to noise required for the same phonetic perception.[5]

Discussion

It is not novel to find that variations in rate of articulation have an effect on the perception of temporal cues in speech. Nor is it entirely novel to find, as we have, that variations in rate have an unequal effect on the several temporal cues—duration of silence and duration of noise—that are effective in the perception of the fricative–affricate distinction; as we pointed out in the introduction, that conclusion was suggested by an experiment that is owing to Dorman et al. (Note 4). What we have done is to extend that finding. Having varied both the duration of silence and the duration of noise, we saw that the inequality is not peculiar to a particular duration of noise, and we saw, moreover, a trading relation between the two duration cues. That trading relation becomes now a component of one interpretation of the seemingly paradoxical rate effect.

To appreciate that interpretation in its broadest form, we should take note once again of the comments by several students of speech production that variations in rate of articulation do not affect all portions of the speech signal equally. To the extent that this is so, a listener cannot adjust for rate variations by applying a simple scale factor but must rather make a more complex correction, one that embodies a tacit knowledge, as it were, of the inequalities in the signal that rate variations generate. Perhaps the results of our experiment are an instance of that correction and that tacit knowledge. Suppose that in the case of utterances like those of our experiment, variations in rate of articulation cause the duration of the fricative noise to change more than the duration of the silence. If the listener's perception reflects an accurate understanding of that inequality, then he or she should expect that given an increase in rate, the noise would shorten more than the silence. But on hearing, as in some of the conditions of our experiment, that the noise duration remains constant when the rate increases, the listener would assign to the

[4] The boundary estimates obtained from the average data of all subjects were virtually identical with the averages of the estimates for individual subjects, so the former have been plotted in Figure 1. The response function for the longest noise seemed to reach asymptote below 100% CHOP responses, especially at the fast speaking rate. This caused the estimated boundaries to fall at somewhat longer silence durations than the 50% intercepts shown in Figure 1.

[5] It must be kept in mind that this description is true only within the limits of the present experiment. Had the noise duration been increased beyond 100 msec, a point would have been reached where no amount of silence would have led to a substantial percentage of CHOP responses (cf. Experiment 2).

noise an effectively greater (relative) length. As we know, a longer noise duration biases the perception toward fricative, though, as shown by the trading relation in our results, that bias can be overcome by an increase in the duration of silence. A consequence of all that would be just the effect of rate we found in our experiment: When the rate was increased as the duration of noise was held constant, listeners required more silence to perceive an affricate.

The foregoing interpretation depends, among other considerations, on a determination that variations in rate do, in fact, produce the particular inequality we are here concerned with. As we pointed out earlier, Gay (1978) found in utterance types somewhat analogous to ours that rate variations produced smaller variations in the silence associated with stop consonants than in the durations of the surrounding vocalic portions. There are no data, unfortunately, on exactly those utterances we used in our experiment. We have made efforts in that direction, but the results so far are inconclusive. Until such time as we know more clearly just what happens in speech production, the interpretation we have offered here is, of course, quite tentative.

The interpretation must be tentative for yet another reason: It does not reckon with the possibility that certain other cues for the fricative–affricate distinction might have been at work in ways that we do not yet thoroughly understand. We have in mind, in particular, the rise time of the fricative noise. From the work of Gerstman (1957) and of Cutting and Rosner (1974), we know that it is a relevant cue. Dorman, Raphael, and Liberman (Note 8) recently varied noise rise time and silence duration to produce the distinction between DISH and DITCH, which is essentially similar to the SAY SHOP—SAY CHOP contrast investigated here, and showed that the two cues engage in an orderly trading relation, as we might have expected. However, we do not know how, or even whether, noise rise time varies with rate of articulation. Information on this matter might conceivably affect our interpretation of the present results.

Returning now to the most important results of our experiment, we should emphasize that there are two. The one has to do with the trading relation between duration of silence and duration of noise as joint cues for the fricative–affricate distinction. It is to us provocative that these cues, diverse and distributed as they seem, are nevertheless integrated into the unitary phonetic percept we call fricative or affricate. In our view, this integration occurs because cues such as these converge through a single decision process that takes account of their common origin: They are the consequences of the same articulatory act. The other result, which we have already discussed at some length, is that the two duration cues were affected unequally by a change in rate of articulation. We would now simply emphasize the inequality, which is a very reliable effect, for it does imply that perceptual correction for variations in rate is not made in this case by applying a simple scale factor but that it may rather require some more sophisticated computation that, like the integration of the duration cues, takes account of particular facts about speech production.

Experiment 2

While exploring the boundaries of the phenomenon reported in Experiment 1, we observed an effect that we have undertaken to investigate more systematically in Experiment 2. We reported in Experiment 1 that with increases in the duration of silence between SAY and SHOP, the fricative in SHOP changed to the affricate in CHOP. However, when the fricative noise was at its longest (100 msec), it occasionally seemed to us that CHOP changed back to SHOP and that the stoplike effect was displaced to the end of the preceding syllable, converting SAY to SAYT. If confirmed, that effect would be interesting from our point of view because it bespeaks an integration of perceptual cues across syllable (word) boundaries. It is also relevant to the problem of "juncture," so long a concern of linguists (see Lehiste, 1960).

Our concern, then, is with the perceptual integration of the cues that affect perception and placement of stop-consonant manner, either as a final segment added to one syllable or as the conversion of the first segment of the next syllable from fricative to affricate. The cues we examined were the same as those of Experiment 1, duration of silence between the syllables and duration of the fricative noise at the beginning of the second syllable. However, there are two changes. To offer maximum opportunity for the stoplike effect to be transferred from the second syllable to the first, we included durations of fricative noise longer than those used in Experiment 1, thus providing a stronger bias against affricate percepts; and to make the alternative responses equally plausible to our subjects, we used a new sentence, DID ANYBODY SEE THE GRAY (GREAT) SHIP (CHIP). The sentence context was employed to make the test as natural as possible. (Rate of articulation was not a variable in this experiment.)

In a second part of the experiment (Experiment 2b), we assessed the effects of those spectral and durational cues that distinguish GRAY and GREAT. For that purpose, we investigated how the results depend on whether, in the original recording, the word was pronounced as GRAY or as GREAT.

Method

Subjects. The subjects were the same as in Experiment 1.

Stimuli: Experiment 2a. The sentence DID ANYBODY SEE THE GRAY SHIP was produced by a male speaker in a monotone voice and recorded in digitized form. Using the editing facilities of the Haskins Laboratories PCM System, we varied the duration of silence inserted before the word SHIP for 0 to 100 msec in steps of 10 msec. The duration of the fricative noise in SHIP was also varied. Starting with the duration of the noise as recorded, which was 122 msec, we excised or duplicated 20-msec portions from its center, thus shortening or lengthening it without changing the characteristics of its onset or offset. In this way we created four durations of noise—62, 102, 142, and 182 msec—for use in the experiment. Four noise durations and 11 silence durations led to 44 test utterances. These were recorded in five different randomizations, with intervals of 2 sec between sentences.

To see how the fricative–affricate distinction is affected by noise duration alone, we excised the word SHIP (CHIP) and varied the duration of the noise as described above, but in steps of 20 rather than 40 msec. These isolated words were recorded in a randomized sequence containing 10 repetitions of each stimulus. The interstimulus interval was 3 sec.

Stimuli: Experiment 2b. A second sentence, DID ANYBODY SEE THE GREAT SHIP, was recorded by the same speaker who had produced the sentence, DID ANYBODY SEE THE GRAY SHIP, of Experiment 2a. He attempted to imitate the intonation and speaking rate of the first-produced sentence. That he succeeded well was suggested by our own listening and by comparison of the waveforms. Using the PCM System, we excerpted the fricative noise from the SHIP of Experiment 2a and substituted it for the noise in the corresponding word of the new sentence. Thus, the two stimulus sentences had exactly the same fricative noise in the final word SHIP. Both sentences were used in Experiment 2b: the original sentence, DID ANYBODY SEE THE GRAY SHIP, and the new sentence, DID ANYBODY SEE THE GREAT SHIP; the important difference was simply in the opposition between the words GRAY and GREAT.

Inspection of waveforms and spectrograms revealed that there was only a slight difference in duration between the two utterances; this difference was almost entirely accounted for by the additional closure period between GREAT and SHIP in the second sentence. The final transitions of the second and third formants were, as expected, somewhat steeper in GREAT than in GRAY. Also, the GREAT syllable had a longer duration (210 msec, not including the following closure period) than GRAY (187 msec).[6] Their offset characteristics were similar.

Only two noise durations, 82 and 142 msec, were used, as against the four (62, 102, 142, and 182 msec) of Experiment 2a. There were more silence durations, on the other hand, covering the (wider) range from 0 to 150 msec in steps of 10 msec. Thus, with 2 noise durations, 16 silence durations, and 2 sentence frames, there were 64 test sentences in all. These were recorded in five randomized sequences.

Procedure. Experiments 2a and 2b were conducted in a single session of about 2-hours duration. The isolated word sequence was presented first (the response alternatives being SHIP and CHIP), followed by the sentences of Experiments 2a and 2b, in that order. Each set of sentences was repeated once, so that each subject gave a total of 10 responses to each sentence. The subjects chose from four response alternatives, using letter codes in writing down their responses: A = GRAY SHIP, B = GREAT SHIP, C = GRAY CHIP, D = GREAT CHIP. No subject had any difficulties in using this system.

Results

Experiment 2a. In Figure 3 are shown the effects of the two cues, duration of si-

[6] Our intuition may tell us that GRAY should have been longer than GREAT. However, these intuitions are based on the pronunciation of these words in isolation, where word-final lengthening extends the vowel in GRAY. When followed by SHIP, on the other hand, the longer duration of GREAT is quite plausible. However, we do not know whether this observation has any generality.

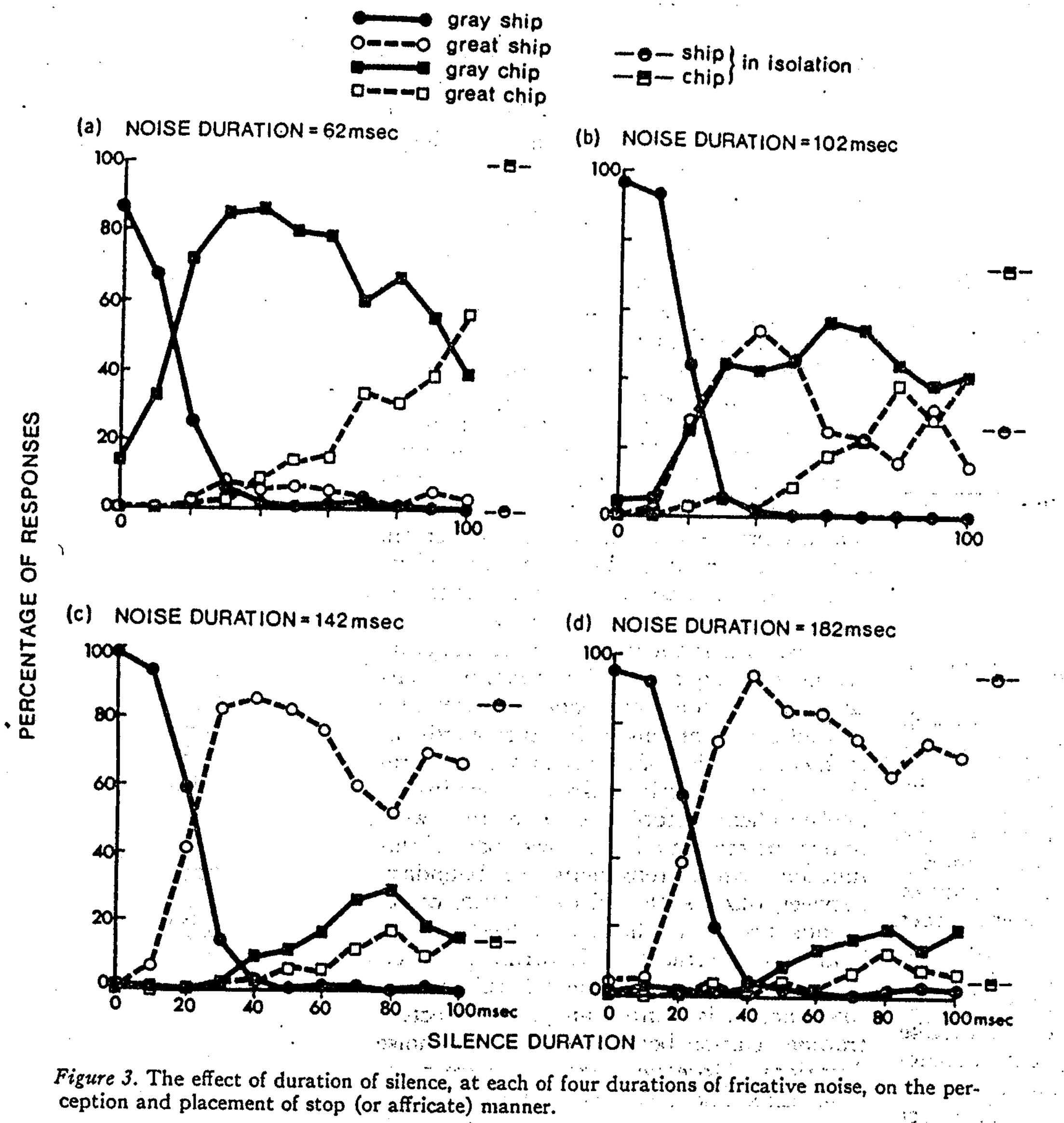

Figure 3. The effect of duration of silence, at each of four durations of fricative noise, on the perception and placement of stop (or affricate) manner.

lence and duration of fricative noise, on the perception of stop or stoplike manner in the utterance DID ANYBODY SEE THE GRAY (GREAT) SHIP (CHIP). Duration of silence is the independent variable; the four panels correspond to the durations of fricative noise. At the right of each panel, we also show the results obtained when the second of the key words, SHIP (CHIP), was presented in isolation.

Let us consider first the responses to the isolated word SHIP (CHIP). At noise durations of 62, 102, 142, and 182 msec—those used in the experiment—the percentages of CHIP responses were 100, 73, 16, and 6, respectively. Thus, as we had every reason to expect, duration of the noise is a powerful cue for the fricative–affricate distinction. The SHIP-CHIP boundary was estimated to be at 119 msec of noise duration. In contrast to the stimuli of Experiment 1, whose noise durations all fell below this boundary and therefore were predominantly heard as affricates, those of the present experiment

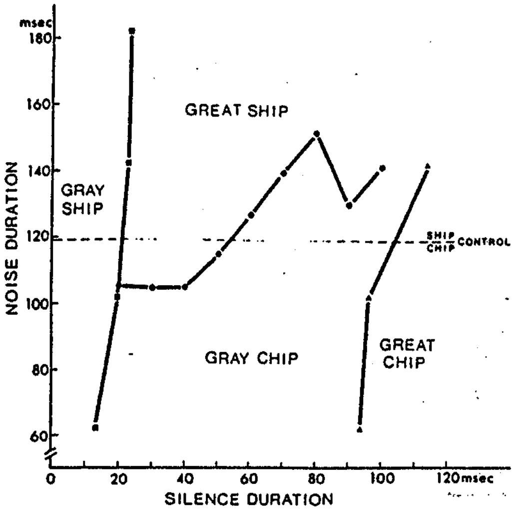

Figure 4. Boundaries that divide the several response categories, represented as joint functions of duration of silence and duration of fricative noise.

spanned the entire range from affricate to fricative.

The more important results of the experiment are seen by examining the graphs that tell us how the stimuli were perceived in the sentence context. We note first that when the silence was of short duration (less than 20 msec), the subjects perceived primarily GRAY SHIP. At those very short durations of silence, no stoplike effect was evident, either as an affricate at the beginning of the second syllable (CHIP) or as a stop consonant at the end of the first syllable (GREAT). With increasing durations of silence, a stoplike effect emerged. As in Experiment 1, somewhat more silence was required at the longer noise durations for this stoplike effect to occur, $F(3, 27) = 6.93$, $p < .01$.

Perhaps the most interesting result was that once a stop was heard, its perceptual placement in the utterance depended crucially on the duration of the fricative noise: At short noise durations, the listeners reported predominantly GRAY CHIP; at longer noise durations, GREAT SHIP. This resulted in a significant response category by noise duration interaction, $F(9, 81) = 71.52$, $p < .001$.

We also see that the response percentages were in fair agreement with the results for isolated words. When the critical word was heard as CHIP in isolation, it was generally heard as (GRAY or GREAT) CHIP in sentence context, too—provided, of course, that it was preceded by at least 30 msec of silence—and words heard as SHIP in isolation were generally heard as (GREAT) SHIP. Responses in the GREAT CHIP category occurred at the longer silence durations when the noise was short, but even at the longest silence duration and shortest noise such responses reached only about 50%.

A more concise representation of the results, showing perceptual boundaries as determined by the probit method, is to be found in Figure 4. There we see three functions, each of which links those combinations of silence duration and noise duration that are precisely balanced between certain response alternatives, as we specify below. The dashed horizontal line represents the SHIP–CHIP boundary for isolated words.

Consider first the nearly vertical function at the left (squares). This function characterizes the boundary between GRAY SHIP and all other responses. In other words, at each combination of silence and noise durations on this function, listeners were just as likely to hear a stoplike effect as they were to hear no stop at all. The lower part of this function, which represents the boundary between GRAY SHIP and GRAY CHIP, corresponds directly to the SAY SHOP—SAY CHOP boundary functions of Experiment 1 (cf. Figure 2). As in Experiment 1, this part of the function is slanted and thus reflects a trading relation between silence and noise durations. Moreover, again in agreement with Experiment 1, the trading relation can be described as a constant ratio of silence to noise. However, this ratio (about .20) is considerably smaller than that obtained in Experiment 1 at a comparable speaking rate (.44). This is presumably due to the fact that in the present experiment less silence was needed to obtain a stoplike effect. The reason for that was suggested by listening to the words preceding the silence when taken out of context. The SAY of Experiment 1 actually sounded like SAY (not SAYT) in isolation, but the excised word GRAY of the present experiment, although correctly pronounced in the original sentence, sounded much more like GREAT. Thus, the vocalic portion preceding the silence contained stronger stop-manner cues in the present experiment than in Experiment 1, so that less silence was required to hear a stoplike effect. These observations provide indirect evidence for yet another trading relation between two cues for stop manner: the (spectral and temporal) characteristics of the vocalic portion preceding the silence, and silence duration itself.

Returning to the boundary function at the left of Figure 4, we note that the function changes from slanted at short noise durations to completely vertical at longer noise durations. In other words, the trading relation between silence and noise durations which characterizes the GRAY SHIP versus GRAY CHIP distinction disappears as the distinction changes to GRAY SHIP versus GREAT SHIP. This phonetic contrast, located in the first syllable, is apparently not affected by further increases in noise duration in the second syllable but depends only on silence duration.

We turn now to the second function in Figure 4, that connecting the circles. This function represents the boundaries between GREAT SHIP, on the one hand, and GRAY CHIP or GREAT CHIP on the other. (GRAY SHIP responses did not enter into the calculation of these boundaries.) Since GREAT CHIP responses occurred primarily at long silence durations, the major part of the boundary function represents the distinction between GREAT SHIP and GRAY CHIP, that is, the perceived location of juncture. It is clear that noise duration was the major juncture cue, as we should have expected given earlier observations of Lehiste (1960) and Nakatani and Dukes (1977). Had it been the only cue, the boundary function would have been perfectly horizontal. As we see, however, the function shows a clear rise at intermediate silence durations (40–80 msec): GREAT SHIP responses were more frequent at short silence durations, and GRAY CHIP responses were more frequent at longer silence durations. Thus, silence duration was a secondary cue for the location of the word boundary (cf. Christie, 1974, for a related result).

The third function in Figure 4—that connecting the triangles—represents the boundary between GRAY CHIP and GREAT CHIP, excluding other responses. There was no obvious dependency of this boundary on noise duration; the uppermost data point, which may suggest such a dependency, was based on only a few observations, since at this noise duration (142 msec) GREAT SHIP responses predominated (cf. Figure 3). We note that a fairly long period of silence (about 100 msec) was required to hear both a syllable-final stop and an affricate.

Experiment 2b. By using the sentence containing the word GRAY as the "source" for half of the stimuli, Experiment 2b partially replicated Experiment 2a. These results are shown in the top panels of Figure 5. They may be contrasted with the results obtained with the new GREAT source, shown in the bottom panels. For each source, the effects of noise and silence duration were similar to those observed in Experiment 2a; therefore, they need no further comment. The change in the response pattern as a function of noise duration was again highly significant, $F(3, 27) = 58.95$, $p < .001$.

The effect of primary interest was that of source. It can be seen that more GREAT (both GREAT SHIP and GREAT CHIP) responses occurred when the source was GREAT, as shown by a significant Source × Response Categories interaction, $F(3, 27) = 10.11$, $p < .01$. However, this effect did not substantially change the overall response pattern. At silence durations of less than 20 msec, the listeners still reported GRAY SHIP; at longer silence durations GRAY CHIP was heard when the noise was short, even though the original utterance had been GREAT. Thus, the cues for stop manner in the word GREAT were readily integrated with the intial consonant of the next word if the short noise biased perception toward hearing an affricate.

As in Experiment 2a, we calculated three kinds of perceptual boundaries (cf. Figure 4).[7] These are shown in Figure 6, where they

[7] The GREAT SHIP versus GRAY CHIP (+ GREAT CHIP) boundary estimates were based on only two data points (noise durations). To obtain probit

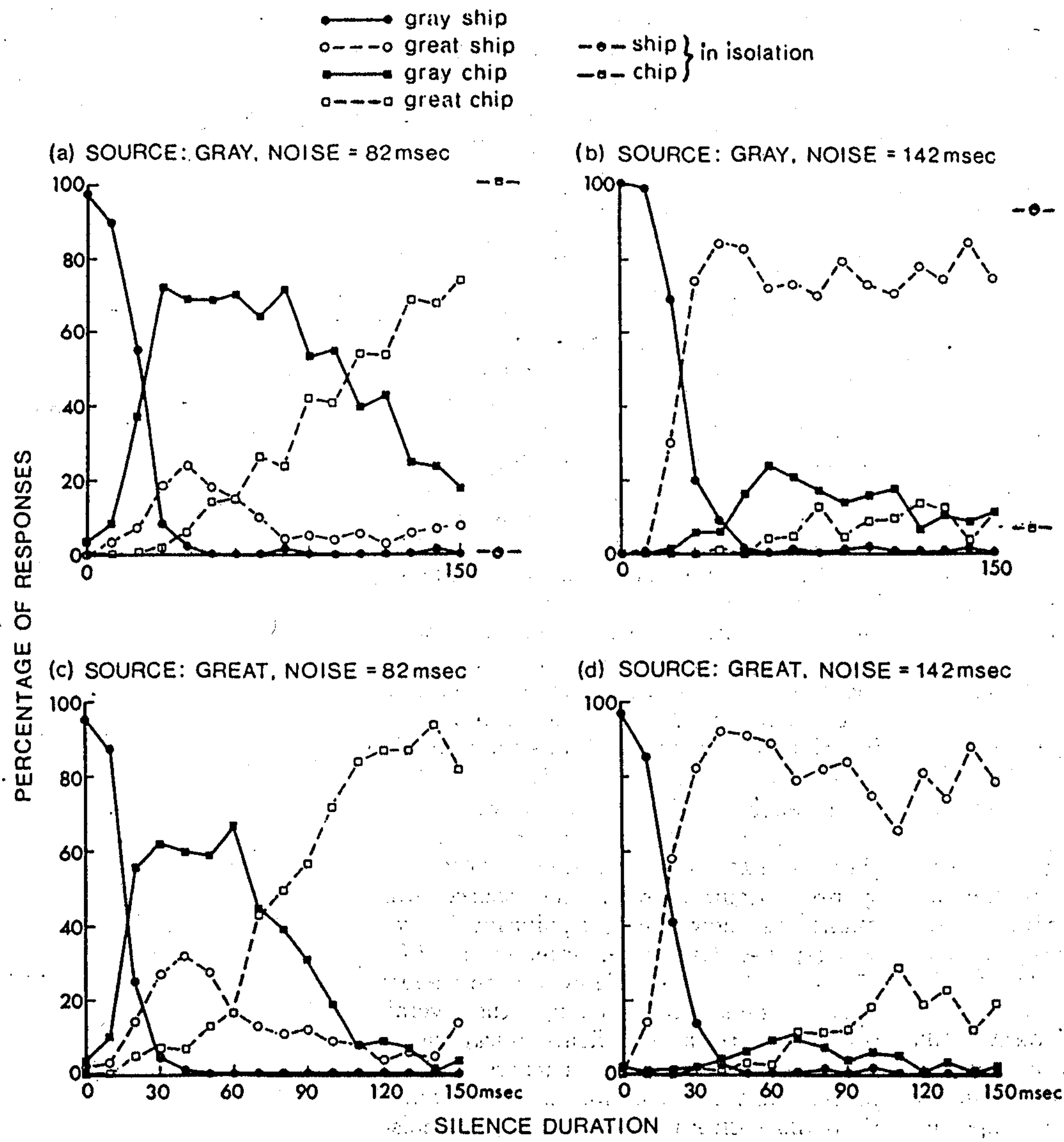

Figure 5. Effects of varying the "source" (original pronunciation as GRAY or GREAT) on the perception and placement of stop (or affricate) manner. (These are shown at each of two durations of noise and represented as the percentage of occurrence of the several responses plotted against the duration of silence.)

are plotted, separately for each "source," as joint functions of silence duration and noise duration. We see that the boundary between GRAY SHIP and the other responses (squares) shifted significantly to the left as the source changed from GRAY to GREAT, $F(1, 9) = 33.66$, $p < .01$. In other words, less silence was needed to hear a stoplike effect (regardless of whether it was placed at the end of the first or at the beginning of the second syllable) when the original utterance had contained the word GREAT. Note that the stop-manner cues preceding

estimates, we added two hypothetical anchor points: 22 msec (of noise) with 0% GREAT SHIP responses, and 202 msec (of noise) with 100% GREAT SHIP responses.

a relatively short silence were readily integrated with those following the silence: Within the range of silence (and noise) durations in which the subjects' responses were either GRAY SHIP or GRAY CHIP, the frequency of GRAY CHIP responses actually was increased when the source was changed from GRAY to GREAT.

The second boundary function—that separating GREAT SHIP from GRAY CHIP and GREAT CHIP responses (circles)—also showed an interesting pattern of source effects. At shorter silence durations, in which the distinction was mainly between GREAT SHIP and GRAY CHIP, the change in source from GRAY to GREAT increased GREAT SHIP responses and decreased GRAY CHIP responses. This is reasonable, although it provides a counterexample to the recent conclusion by Nakatani and Dukes (1977) that cues in the first word have no effect on the perceived location of the word boundary. At long silence durations (beyond 100 msec), on the other hand, the phonetic distinction was primarily between GREAT SHIP and GREAT CHIP; there, source ceased to have any effect. Thus, when the silent interval exceeded about 100 msec, stop-manner cues preceding the silence were no longer integrated with those that followed it.

The third boundary, GRAY CHIP versus GREAT CHIP (triangles), showed by far the largest source effect. Since the phonetic contrast was located here in the word that was actually changed in pronunciation and since, because of the relatively long silence duration, the stop-manner cues preceding the silence were perceived independently of the cues following it, the large effect is readily understandable. On the other hand, the effect is not trivial, since, as we pointed out earlier, the word GRAY from the GRAY source actually sounded like GREAT in isolation. That the stimuli derived from the GRAY source received any GREAT CHIP responses at all was probably due to the presence of relatively strong stop-manner cues in the word GRAY.

General Discussion

The most interesting aspect of the data, in our view, is that whether a syllable-final stop consonant was perceived (GRAY vs. GREAT) depended on the duration of the noise following the silence—an acoustic event occurring much later in time. There

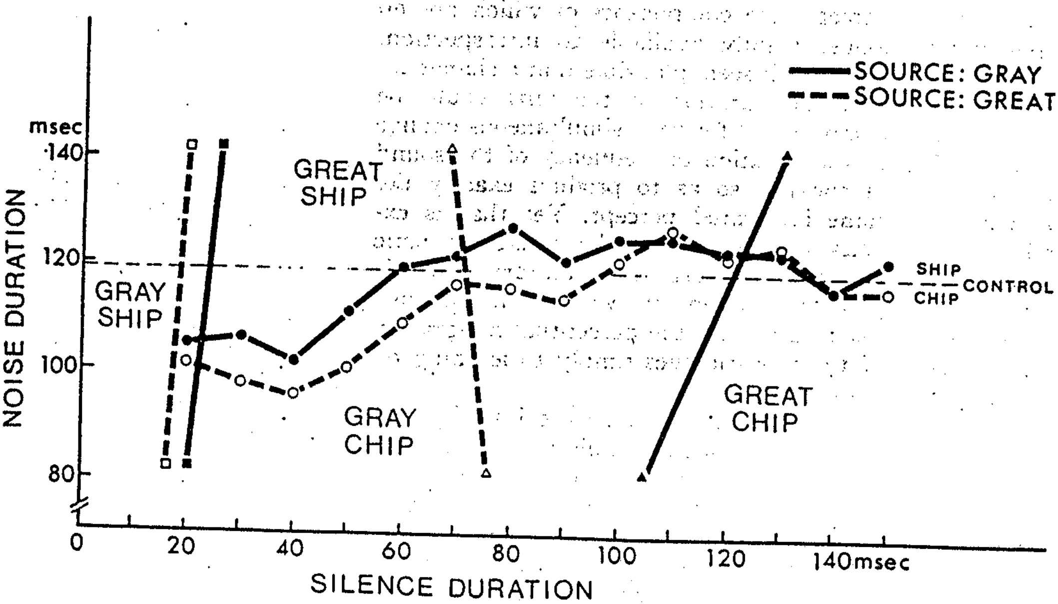

Figure 6. Effects of varying the "source" (original pronunciation as GRAY or GREAT) on the boundaries that divide the several response categories.

are three questions we may ask about this temporal integration: Why does it occur? What are its limits? And when does the listener reach a decision about what he has heard? We consider these questions in turn.

Why does temporal integration occur? We have seen that cues as diverse and as widely distributed as (a) the spectral and temporal properties of the vocalic portion preceding the silence, (b) the silence duration itself, and (c) the spectral and temporal properties of the noise portion following the silence are all integrated into a unitary phonetic percept. Can we explain such integration on a purely auditory basis? Auditory integration does occur—for example, it is responsible for the perceptual coherence of homogeneous events such as the fricative noise—and surely we have much more to learn about such integration, especially in the case of complex acoustic signals. But it seems to us quite implausible to suppose that purely auditory principles could ever account for perceptual integration of acoustic cues as heterogeneous and temporally spread as those we have dealt with here.

We encounter similar problems when we seek to explain our results in terms of feature detectors as they have been postulated by several contemporary theorists (e.g., Blumstein, Stevens, & Nigro, 1977; Eimas & Corbit, 1973; Miller, 1977). Consider again the case in which the perception of a syllable-final stop consonant (GREAT vs. GRAY) depends on whether the fricative noise following the silence extends beyond a certain duration. If a single phonetic feature detector were responsible for the syllable-final stop, then its integrative power and complexity would have to be so great as to remove from the concept of feature detector the simplicity that is its chief attraction. Alternatively, there might be many simple auditory feature detectors, each responsive to elementary properties of the signal, whose outputs are integrated by a higher level phonetic decision mechanism (cf. Massaro & Cohen, 1977). But that view fails to provide any principled reason why the outputs of certain feature detectors feed into a single phonetic decision in the way they do. Without reference to the articulatory system that produced the speech signal, the rules by which the detector outputs might be integrated would seem entirely arbitrary.

One might suppose, of course, that the diverse cues have become integrated into a unitary percept as a result of learning. Surely, the cues have frequently been associated in the production and perception of speech. But would such association be sufficient to cause them eventually to sound alike, as the integration (and various trading relations) indicate that they do? Common experience and common sense suggest that it would not. Consider, for example, a listener who has for many years heard a bell and a buzzer, each of a particular kind, always sounded in close temporal and spatial contiguity. It is reasonable to expect that these very different stimuli (or rather their corresponding percepts) would become associated in his or her mind: On hearing one he or she would expect to hear the other, and either would presumably become a sufficient sign for whatever it was that the two, taken together, normally signified. But it seems implausible that they would ever be integrated into a unitary percept, the components of which are no longer readily available to introspection. Nor does it seem plausible that a change in, say, the duration of the buzz could be compensated for by a simultaneous change in the duration or frequency of the sound of the bell so as to produce exactly the same integrated percept. Yet that is exactly what is true of the diverse acoustic cues that converge on a unitary phonetic percept. At all events, we think it implausible to attribute the perceptual integration of the acoustic cues simply to learning by association.

As we pointed out in the introduction, we believe that the guiding principle of temporal integration in phonetic perception is to be found in the articulatory act that underlies the production of the relevant phonetic segment. By an *articulatory act* we mean not a particular articulatory gesture but all articulatory maneuvers that result from the speaker's "intention" to produce a given segment, for example, a stop consonant. Thus, our definition of the articulatory act is intimately tied to the hypothesis that units of phoneme size are physiologically real at some early level in speech production. At the later articulatory level, we can distinguish individual gestures (such as closing and opening the jaw, raising the tongue tip) that form the components of the articulatory act. It is, of course, these several gestures that produce the several (and sometimes even more numerous) acoustic cues. The perceptual process by which the acoustic cues are integrated into a unitary phonetic percept somehow recaptures the gestures and also mirrors the processes by which they unfolded from a unitary phonetic intention (or motor program). We find it plausible to suppose that speech perception, as a unique biological capacity, has in fact evolved to reflect the equally species-specific capacity for speech production. The consequence is that, in a very real sense, the listener perceives directly the speaker's "intent"—the phonetically significant articulatory act (for views related to ours in their emphasis on the perception of articulatory events but different from ours in other respects, see Fowler, 1977; Bailey & Summerfield, Note 6; Summerfield, Note 7).

We turn now to our second question, that about the limits of temporal integration. From the data of our experiments, we obtain an estimate according to the following considerations. The boundary between GRAY CHIP and GREAT CHIP indicates the longest period of time over which the stop-manner cues preceding the silence are still integrated with the cues following the silence into a single stoplike percept (affricate). Although the exact temporal interval varied with the strength of the stop-manner cues preceding the silence (cf. Figure 6), a silence duration of 100 msec is a reasonably typical value. To this must be added the approximate temporal extent of the relevant cues preceding and following the silence—at least 100 msec for the duration of both the vocalic portion and the fricative noise. We thus arrive at a temporal range of 300—350 msec for the integration of stop-manner cues. This estimate is in good agreement with results on the single-geminate distinction for intervocalic stop consonants (e.g., TOPIC vs. TOP PICK), since, as Pickett and Decker (1960) and Repp (Note 1) have shown, that boundary occurs around 200 msec of silence at normal rates of speech. Inasmuch as the manner cues following the closure interval (the formant transitions of the second vocalic portion) are shorter in this case (perhaps 50 msec), we arrive again at an integration period of about 350 msec. This coincidence is not surprising since the articulatory gesture underlying an intervocalic stop consonant is similar to that for a stop consonant embedded between a vowel and a fricative. In our view, the range of temporal integration in perception reflects not an auditory limitation—such as the duration of a preperceptual auditory store (Massaro, 1975) —but the longest acceptable duration of the underlying articulatory act. Different articulatory acts may well be associated in perception with different ranges of temporal integration.

We thus arrive at our third question: When do the listeners decide what they have heard? Before we can answer that question, we must point out that there are two logically distinct decisions the listener must make: (a) *What* phoneme has occurred? (b) *Where* does it belong? Thus, in the case of the GREAT SHIP—GRAY CHIP distinction, the listener must decide first that a stop consonant has occurred and, then, whether it belongs with the first or the second syllable. We see three possibilities for the temporal organization of the listener's decisions: (a) Both the What and Where decisions occur after all relevant cues have been integrated; (b) the What decision occurs as soon as sufficient cues are available, but the Where decision is delayed until the end of the integration period; and (c) both a What decision and a Where (default) decision are made as soon as sufficient cues are available, but the Where decision may be revised in the light of later information. We discuss these hypotheses in turn.

The first hypothesis implies, in the case of GREAT SHIP, that listeners do not know whether they have heard a stop consonant until they have processed at least the first

120 msec of the fricative noise. This seems implausible on intuitive grounds. It is more likely that phonetic information accumulates continuously from the speech signal and that What decisions can be made, in principle at least, before all cues have been processed (cf. Remington, 1977; Repp, Note 1). If this were not so, we should have to assume that the relevant cues are integrated at a prephonetic level and thus are held in a temporary auditory memory—precisely the argument we do not wish to make. On the other hand, if temporally separate cues (such as those preceding and following the silence in GRAY CHIP) are immediately translated into phonetic representations, temporal integration merely combines identical phonetic codes within a certain time span and thus is not dependent on auditory limitations. In terms of our experiment, this means that the listener already "knew" at the end of the vocalic portion of GRAY (which, as the reader may remember, contained sufficient stop-manner cues to be perceived as GREAT in isolation) that a stop had occurred; the silence duration cue (if less than about 100 msec) and the noise duration cue (if less than about 120 msec) merely confirmed this perceptual knowledge.

The remaining two hypotheses differ in their assumptions about when the Where decision occurs. According to one hypothesis, listeners do not know whether they have heard GRAY or GREAT until they have processed the fricative noise; in other words, the Where decision is postponed until all relevant cues have been integrated. The alternative hypothesis assumes that listeners group the stop consonant automatically with the preceding syllable until later information leads them to revise that decision. This leads to the paradoxical prediction that in an utterance heard as GRAY CHIP, listeners actually perceive GREAT for the brief moment that extends from the end of the vocalic portion to the end of the fricative noise, as they would have if CHIP had never occurred. We hope to conduct experiments in the future that will shed more light on these questions.

Reference Notes

1. Repp, B. H. *Perception of implosive transitions in VCV utterances* (Status Report on Speech Research, SR-48). New Haven, Conn.: Haskins Laboratories, 1976.
2. Bastian, J. *Silent intervals as closure cues in the perception of stop consonants.* (Quarterly Progress Report No. 33, Appendix 1.) New Haven, Conn.: Haskins Laboratories, 1959.
3. [Bastian, J.] *Silent intervals and formant transitions as cues in the perception of stop phonemes* (Quarterly Progress Report No. 35, Appendix 2.) New Haven, Conn.: Haskins Laboratories, 1960.
4. Dorman, M. F., Raphael, L. J., & Liberman, A. M. *Further observations on the role of silence in the perception of stop consonants* (Status Report on Speech Research, SR-48). New Haven, Conn.: Haskins Laboratories, 1976.
5. Erickson, D., Fitch, H. L., Halwes, T. G., & Liberman, A. M. *A trading relation in perception between silence and spectrum.* Unpublished manuscript, 1978.
6. Bailey, P. J., & Summerfield, A. Q. *Some observations on the perception of [s] + stop clusters* (Status Report on Speech Research, SR-53, Vol. 2). New Haven, Conn.: Haskins Laboratories, 1978.
7. Summerfield, A. Q. *On articulatory rate and perceptual constancy in phonetic perception.* Unpublished manuscript, 1977.
8. Dorman, M. F., Raphael, L. J., & Liberman, A. M. *Some experiments on the sound of silence in phonetic perception.* Unpublished manuscript, 1978.

References

Ainsworth, W. A. The influence of precursive sequences on the perception of synthesized vowels. *Language and Speech*, 1974, *17*, 103–109.

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

Blumstein, S. E., Stevens, K. N., & Nigro, G. N. Property detectors for bursts and transitions in speech perception. *Journal of the Acoustical Society of America*, 1977, *61*, 1301–1313.

Christie, W. M., Jr. Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America*, 1974, *55*, 819–821.

Cutting, J. E., & Rosner, B. S. Categories and boundaries in speech and music. *Perception & Psychophysics*, 1974, *16*, 564–570.

Eimas, P. D., & Corbit, J. D. Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 1973, *4*, 99–109.

Finney, D. J. *Probit analysis* (3rd ed.). Cambridge, England: Cambridge University Press, 1971.

Fowler, C. A. *Timing control in speech production.* Unpublished doctoral dissertation, University of Connecticut, 1977.

Fujisaki, H., Nakamura, K., & Imoto, T. Auditory perception of duration of speech and non-speech stimuli. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech.* London: Academic Press, 1975.

Gay, T. Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, 1978, *63*, 223–230.

Gerstman, L. *Cues for distinguishing among fricatives, affricates, and stop consonants.* Unpublished doctoral dissertation, New York University, 1957.

Kozhevnikov, V. A., & Christovich, L. A. *Speech, articulation, and perception* (NTIS No. JPRS-305430) Washington, D.C.: U.S. Dept. of Commerce, 1965.

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

Liberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology: Vol. 8. Perception.* Heidelberg: Springer-Verlag, in press.

Massaro, D. W. Preperceptual images, processing time, and perceptual units in speech perception. In D. W. Massaro (Ed.), *Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics.* New York: Academic Press, 1975.

Massaro, D. W., & Cohen, M. M. Voice onset time and fundamental frequency as cues to the /zi/–/si/ distinction. *Perception & Psychophysics*, 1977, *22*, 373–382.

Miller, J. L. Properties of feature detectors for VOT: The voiceless channel of analysis. *Journal of the Acoustical Society of America*, 1977, *62*, 641–648.

Nakatani, L. H., & Dukes, K. D. Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, 1977, *62*, 714–719.

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

Remington, R. Processing of phonemes in speech: A speed–accuracy study. *Journal of the Acoustical Society of America*, 1977, *62*, 1279–1290.

Summerfield, Q., & Haggard, M. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, 1977, *62*, 435–448.

Received February 2, 1978 ■