Journal of Phonetics (1980) 8, 185–194

Accessing phonetic information during perceptual integration of temporally distributed cues

Bruno H. Repp

Haskins Laboratories, 270 Crown Street, New Haven, CT 06510, U.S.A.

Received 4th September 1978

Abstract: A group of untrained subjects, as well as several experienced listeners, participated in two forced-choice reaction time experiments requiring discrimination between two different stop consonants in VCV context. The duration of the silent closure interval in the test utterances was varied to derive functions relating response latencies to closure duration. These functions indicated that untrained subjects responded in part, and experienced listeners responded exclusively, to the VC formant transitions, which preceded the closure. Given that, at closure durations below 200 ms, the phonetic information conveyed by pre- and post-closure cues (VC and CV formant transitions) is integrated into a single phonemic percept – that of an intervocalic stop consonant at a certain place of articulation – the present results indicate that phonetic information can be accessed during perceptual integration of temporally distributed cues, and by implication, that phonetic information is continuously extracted from the speech signal.

Introduction

The perception of speech involves the perceptual integration of temporally distributed information at a number of levels. At the level of the sentence, listeners must integrate the meanings of its individual words. Similarly, at the level of word recognition, listeners must integrate the individual syllables and phonemes into a perceptual whole. Finally, in order to for a single phoneme to be identified, the temporally distributed acoustic cues conveying its nature must be integrated into a single percept.

We may ask whether the integration and interpretation of information at a given level is a continuous or a discrete process. A continuous model assumes that the incoming acoustic signal is immediately and continuously interpreted at all levels of processing, and successive pieces of information automatically group themselves into perceptual units if they are perceived as belonging together. As a consequence, decisions as to the identity or meaning of a phoneme, word, or sentence could be made at any point in time, based on the information extracted so far. A discrete model, on the other hand, assumes that all the information for a unit of processing at a higher level must accumulate at a lower level before it is interpreted and integrated. Thus, the listener must make some decision about when to stop accumulating and begin processing a chunk of information. Before such a decision is made, no higher-level interpretation of the accumulated input would be available to the listener.

The possibility of discrete processing has been seriously raised, at least with regard to the syntactic level (e.g. Bever & Hurtig, 1975; Jarvella, 1971). Nevertheless, careful

0095–4470/80/020185+10

consideration of the discrete hypothesis reveals a serious logical problem: What tells the perceptual system that all the information for a given processing unit has accumulated? The boundaries of such units are generally poorly marked in the speech signal. Recently, strong evidence has been obtained that the syntactic and semantic processing of sentences is continuous and interactive (Marslen-Wilson, 1975, 1976), and a similar conclusion seems warranted for the phonetic, lexical, and semantic processing of words in fluent speech (Cole & Jakimik, 1977; Marslen-Wilson & Welsh, 1978).

The question of continuous versus discrete processing may also be raised with regard to phoneme recognition, considering the fact that the acoustic cues for a given phoneme are generally spread out in time. The question here is whether phoneme recognition requires the accumulation of all relevant acoustic cues in a sensory buffer and subsequent integration and phonetic processing of these cues (the discrete hypothesis), or whether phonetic information is continuously extracted from the acoustic signal, so that the identity of a given phoneme could be decided as soon as the first sufficient cues have been processed (the continuous hypothesis).

Although the continuous hypothesis is more plausible on logical grounds, it would be useful to obtain more direct evidence in its favor. Empirical evidence bearing on this hypothesis is harder to come by than in the case of larger units such as words or phrases, simply because listeners normally do not consciously identify phonemes when listening to speech. In order to incite listeners to make conscious decisions at the phonemic level (which they are certainly able to do), a rather constrained experimental task is required. Moreover, it is necessary to obtain very fast decisions, or else the listener's responses will convey no information about whether decisions were made before or after the end of the cue integration period, which may last no longer than 200 - 300 ms. To these ends, a simple forced-choice reaction time task was employed in the present studies.

It must be recognized that such an experimental task is somewhat removed from real life. It is designed to reveal not so much what listeners normally do when perceiving speech, but what they *can* do under certain optimal conditions. It is to be hoped, however, that the range of what is possible in speech perception will permit inferences about the processes underlying perception in less constrained situations.

Another point to be considered is that, in a task that presses subjects to their limits, individual skill, practice, and experience with the stimulus material will play a role. Therefore, the present experiments included several experienced listeners, in addition to a group of experimentally naive subjects. This rather informal comparison proved sufficiently informative to make laborious training procedures dispensable for the purpose of the present experiments.

Experiment I

The experimental task required subjects to decide as quickly as possible whether the medial stop consonant in a synthetic vowel-consonant-vowel (VCV) utterance was a /b/ or a /d/. VCV stimuli are particularly well suited for asking the question of continuous vs. discrete processing, since the phonetic information specifying the (place of articulation of the) medial stop consonant is conveyed by two sets of temporally separated cues: the formant transitions into the closure (VC transitions) and the formant transitions out of the closure (CV transitions). The two sets of transitions are separated by the silent (or nearly silent) closure interval. Unless the closure interval exceeds about 200 ms, VC and CV transitions are integrated into a single phonemic percept, that of a stop consonant embedded between two vowels. At closures longer than 200 ms, double or geminate stop consonants are heard (Repp, 1978*a*). Given that these temporally distributed cues are perpetually integrated at closure durations shorter than 200 ms, we may ask whether decisions about the identity of the stop consonant can be made before the end of the integration period. The discrete hypothesis would predict that listeners have to wait for the CV transitions in order to make a phonemic decision, unless the closure period exceeds 200 ms, so that the VC and CV transitions give rise to separate phonemic percepts. The continuous hypothesis, on the other hand, predicts that listeners can make decisions on the basis of the VC transitions alone, even when CV transitions follow after a short interval.

It would be difficult to conclude from average reaction times to stimuli with a fixed closure duration that subjects' decisions were based on the VC transitions alone. Latencies would have to be so short as to rule out decisions made on the basis of the CV transitions–a goal difficult to attain with untrained listeners. Therefore, reaction times were examined as a function of systematic (but unpredictable) variations in closure duration. If phonemic decisions require processing of the CV transitions, reaction times (measured from the end of the VC transitions) should increase linearly with closure duration until the latter approaches 200 ms (the single-geminate boundary), whereupon latencies should remain stable or even decrease. On the other hand, if decisions can be made on the basis of the VC transitions alone, reaction times should be unaffected by variations in closure duration.

A very similar question was asked by Repp (1978*b*) in a series of same-different reaction time experiments using similar stimuli. These experiments revealed little ability on the part of the subjects to ignore the CV transitions and base their decisions on the VC transitions alone. However, this negative outcome was ascribed to the more complex paradigm which required the listener to hold one stimulus in memory before matching it to a second stimulus. Also, all subjects were unpracticed. It was hoped that the simpler forced-choice reaction time task used here and the inclusion of some experienced subjects would lead to a better assessment of the listeners' processing capabilities.

Method

Subjects

Nine paid volunteers (college students) participated. Some of them had participated in earlier experiments using synthetic speech, but all were relatively inexperienced in reaction time tasks. Three experienced listeners (the author and two colleagues) also served as subjects; their data were treated separately.

Stimuli

Four VCV utterances – /abɛ/, /abi/, /adɛ/, and /adi/ – were created on the Haskins Laboratories parallel resonance synthesizer. The stimuli consisted of two acoustic segments, 185 and 300 ms long, respectively, that were separated by a completely silent interval of variable duration. The first (VC) segment consisted of an initial steady state followed by formant transitions that were independent of the final vowel. The second (CV) segment began with formant transitions and ended in a steady state. The closure durations used were 65, 165, 265, and 365 ms, resulting in total stimulus durations of between 550 and 700 ms. All stimuli had the same flat pitch contour. For a more detailed description of the stimulus parameters, see Repp (1978*a*).

The experimental tape contained first a randomized sequence of 50 VC syllables (25 each of /ab/ and /ad/) with onset-to-onset intervals of 3,850 ms (VC sequence). This sequence was followed by another which contained 90 stimuli made up of five different successive randomizations of all 18 stimuli (4 VCVs with 4 closure durations, plus 2 VCs). The onset-to-onset interval was again 3,850 ms (VCV sequence).

Procedure

Each subject listened first to the VC sequence, with instructions to press one response key for /ab/ and the other key for /ad/. The response-hand assignment was counterbalanced across subjects. For the VCV sequence, the task remained the same, except that the subjects were urged to respond as quickly as possible and to ignore whatever followed the initial VC portion of each utterance. The subjects were warned that there would be some VC stimuli in the VCV sequence. Both sequences were preceded by examples selected randomly from the tape.

The tape was played back on an Ampex AG-500 tape recorder, and the subjects listened over Telephonics TDH-39 earphones. Presentation was binaural, with the intensity set at a comfortable level (about 80 dB SPL). The onset of a stimulus triggered a Hewlett-Packard 522B electronic counter which was stopped by the depression of one of the two response keys. The kind of response given was indicated to the experimenter by two lights of different color. The reaction times were printed out by a Hewlett-Packard 560A digital recorder.

The data analysis was conducted on mean latencies (omitting errors, latencies shorter than 165 ms, and obvious outliers). An appropriate constant was subtracted from the reaction times, so that they were measured from the offset of the VC segment.

Results

The two different VC syllables in the initial VC sequence were easily discriminated by all subjects. The average reaction time was 444 ms, and the error rate was 1·6%.

Figure 1(a) shows the results for the VCV sequence, separately for the nine naive subjects ($N = 9$) and the three experienced subjects ($N = 3$). Average choice reaction times are plotted as a function of closure duration. The data points on the very right represent the mean latencies for the VC stimuli (infinite closure duration) included in the VCV sequence.

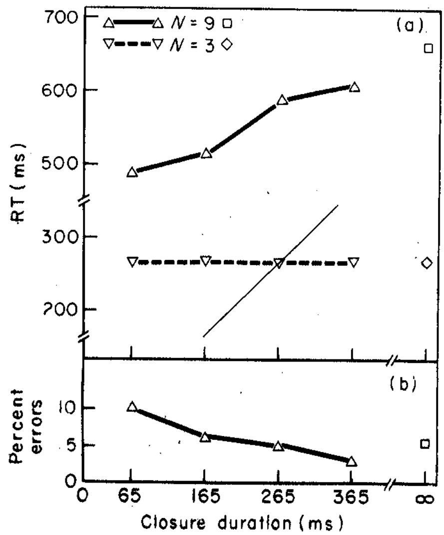

Figure 1 Reaction times and error percentages for VCV stimuli as a function of closure duration. Data for a group of inexperienced listeners (N=9) and three experienced listeners (N=3) are shown separately. The thin diagonal line with a slope of 1 indicates the end of the closure interval on the reaction time scale (ordinate). The closure duration of infinity represents VC stimuli.

It can be seen that the average reaction times of the naive subjects increased with closure duration, $F(3,24) = 8{\cdot}1$, $P < 0{\cdot}01$. Individual functions (not shown in Fig. 1) varied a great deal, in part due to the small number of observations, but only a single subject showed a latency function with a slope close to 1, as predicted by the discrete hypothesis. (A thin line with this slope is also drawn in Fig.1.) All other subjects had flatter functions, indicating that they made partial use of the VC transitions in deciding the identity of the consonant. According to the discrete hypothesis, the functions should have levelled off at the longer closure durations (beyond 200 ms), but there was no clear trend in that direction. However, none of the naive subjects had a completely flat latency function.

For all naive subjects, VC latencies were slower than the average VCV latencies and, for most subjects, even slower than the VCV latencies at the longest closure period. It is interesting to note that all subjects gave much slower responses to the VC syllables in the VCV sequence than in the homogeneous VC sequence, $F(1,8) = 18{\cdot}0$, $P < 0{\cdot}001$. This large difference (661 vs. 444ms) obviously reflects a strong tendency to wait for the CV transitions before making a decision.

The three experienced subjects, on the other hand, had completely flat latency functions–reaction time was not influenced by closure duration. Also, VC latencies were of the same magnitude as VCV latencies. The stimulus context effect on VC latencies was present, however, for all three listeners: The average reaction time was 270 ms to VCs in the VCV sequence, but only 245 ms in the VC sequence. The diagonal line with a slope of 1 in Fig. 1 represents the end of the closure period (the onset of the CV transitions) on the reaction time scale. Data points lying below this line represent responses that, on the average, were made before the CV portion of the stimulus had even begun. Thus, these responses must have been completely independent of closure duration.

For the naive subjects, the average percentage of errors in the VCV sequence was 5·9, with individual rates ranging from 2·2 to 13·3%. Figure 1(b) shows that errors decreased as closure duration (and latencies) increased. The performance of the experienced listeners was virtually error-free; errors were not recorded for them.

Discussion

The low error rates and relatively fast latencies for the stimuli in the VC sequence demonstrate that the VC transitions were sufficient for indentification of the stop consonants. Thus, it is unlikely that poor intelligibility of the VC transitions forced listeners to wait for the CV transitions before making their decisions. Nevertheless, the CV transitions clearly influenced the responses of inexperienced subjects to VCV stimuli. They even had a profound effect in stimuli where they were not present at all, viz., in those VC syllables that were embedded in the VCV sequence. A similar effect has recently been reported by Shand (1976): Choice reaction times reflect expectancies induced by the context in which the stimuli are presented. These "global" context effects appear to be of a more general sort than the "local" influence on latencies exerted by the CV transitions in VCV stimuli. This is indicated by the fact that the experienced listeners showed no local effects whatsoever, but they did show the global effect of list structure on VC reaction times.

Although the latencies of the naive subjects increased with closure duration, the slope of the function was far less than 1, indicating that the subjects did not make their decisions solely on the basis of the CV transitions (or the combined VC and CV transitions). Also, the average VCV reaction times at the 365-ms closure interval approached the minimal reaction times that could conceivably still derive from responses to the CV transitions (closure duration + duration of CV transitions + approximate minimal decision and response time = 365 + 50 + 200 = 615 ms), which suggests that these decisions must have been based largely on the VC transitions (*cf.* Fig. 1). It is difficult to make direct inferences from the reaction times of the naive subjects at shorter closure durations, because these latencies were generally long enough to reflect decisions based either on VC or CV transitions. The simplest assumption is that the listeners alternated between decisions made on the basis of the VC transitions (VC decisions) and decisions made on the basis of the CV (or VC+CV) transitions (CV decisions), so that the average latencies represented a mixture of two distributions.

(The present data were clearly not sufficient to test this hypothesis directly by examining latency distributions.) Alternatively, the listeners may have reached decisions at various points during the integration process. In any case, the data provide evidence that the naive listeners were able at least partially to utilize the information provided by VC transitions in VCV utterances, thus refuting the discrete hypothesis.

The three experienced listeners were completely unaffected by the variations in closure duration. Thus, it is possible in principle to selectively attend to the VC portion in VCV stimuli, at least at closure periods exceeding 65 ms. This task seems to require certain perceptual skills available only to listeners who have experience with synthetic speech and are able to make very fast decisions. Nevertheless, the results clearly indicate that phonemic decisions can be made before the process of cue integration is complete, thus supporting the continuous hypothesis.

The decline in errors with increasing closure duration (naive subjects) requires a separate explanation. A speed-accuracy tradeoff seems highly unlikely, in view of the relatively long latencies. It is possible that CV decisions led to more errors than VC decisions; this would explain the decline in errors, since the proportion of CV decisions presumably decreased as closure duration increased. Alternatively, a more complex perceptual interaction between pre- and post-closure cues may have occurred at short closure durations and may have affected the intelligibility of the stimuli, perhaps because the closure durations were too short compared to natural speech (Westbury, 1977), or because coarticulation effects across the closure interval (Öhman, 1966) had not been taken into account in synthesizing the stimuli.

Experiment II

Experiment II attempted to encourage naive listeners to focus on the VC transitions by increasing the proportion of VC stimuli in mixed (VCV and VC) stimulus lists. In addition, a number of stimuli were included in which the VC transitions had been removed (V-CV stimuli). In order to make decisions about the consonants in these stimuli, the subjects had to wait for the CV transitions, so that V-CV reaction times were expected to increase linearly with closure duration. These functions were to be contrasted with those for VCV (structurally, VC-CV) stimuli which were expected to be more nearly flat.

The experiment took advantage of the fact that VCV and V-CV stimuli are difficult to distinguish as long as the closure durations are short (Repp, 1978*a*). By restricting the design to relatively short closure intervals, it was possible to intermix VCV and V-CV stimuli in a perceptually fairly homogeneous sequence. Nevertheless, the two types of stimuli were expected to generate very different latency functions.

Method

Subjects

The subjects were nine new, relatively inexperienced subjects and the author. The author's data were included in the average data; differences will be pointed out below.

Stimuli

The same four VCV utterances as in Experiment I were used with five closure durations: 65, 90, 115, 140, and 165 ms. For each of the resulting 20 VCV stimuli, a corresponding V-CV stimulus was synthesized by replacing the VC formant transitions with steady-state frequencies appropriate for the initial vowel /a/, leaving all other parameters unchanged. These 40 stimuli were randomized together with 20 VC stimuli (10 each of /ab/ and /ad/) in five blocks, resulting in a total of 300 stimuli. The interstimulus interval was constant at 3 s.

Procedure

This experiment was preceded by another task using similar stimuli; its results are reported elsewhere (Repp, 1978*a*). Thus, the subjects had had some exposure to the stimuli when they began the choice reaction time task. A practice sequence of 20 VC syllables (10 /ab/ and 10 /ad/ stimuli) was presented first. Subsequently, the subjects were told to focus on the initial VC portions of the stimuli and to ignore whatever followed. They were told that many VC stimuli were included in order to remind them of their targets and to reveal whether they followed the instructions. However, the subjects (except for the author) were *not* informed about the presence of stimuli without VC transitions. Only a single subject commented spontaneously after the experiment that he had heard utterances that sounded like /a-bi/, etc. Most subjects remained unaware of the presence of two different kinds of VCV stimuli. All other details of procedure were as in Experiment I, except that median reaction times were computed for the repetitions of each individual stimulus before calculating averages.

Results

The results are shown in Fig. 2. Figure 2(a) shows the average median latencies for VCV and V-CV stimuli as a function of closure duration, as well as regression lines fitted to these points. The numbers on the right indicate the slopes of the regression lines. It can be seen

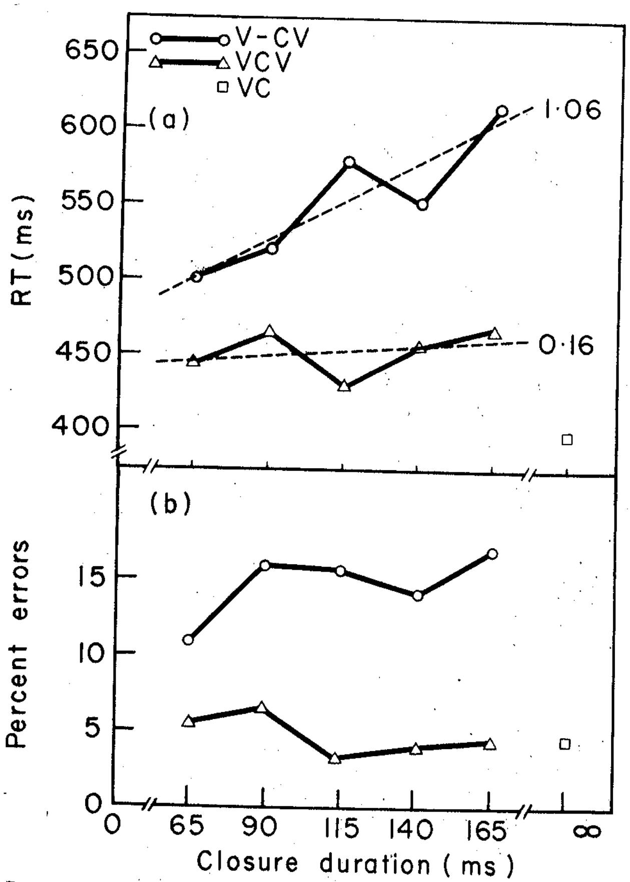

Figure 2

Reaction times and error percentages for VCV and V-CV stimuli as a function of closure duration. Data for VC stimuli are shown at the very right.

that the results conformed quite well to the expectations. Reaction times to V-CV stimuli increased with closure duration, and although the increase was somewhat irregular, the best-fitting regression line had a slope close to 1·0, as expected. Reaction times to VCV stimuli, on the other hand, hardly increased with closure duration; the slope of the regression line was only 0·16. Statistical analysis confirmed this impression: Reaction times to V-CV stimuli were significantly slower than reaction times to VCV stimuli, $F(1,9) = 181{\cdot}6, P < 0{\cdot}001$, and in addition to a significant effect of closure duration, $F(4,36) = 10{\cdot}9, P < 0{\cdot}001$, there was a significant interaction of closure duration with presence VS absence of transitions, $F(4,36) = 9{\cdot}7, P < 0.001$. When the VCV reaction times were anlayzed separately, the main effect of closure duration did not reach significance, $F(4,36) = 2{\cdot}3, P > 0.05$.

Despite the relative flatness of the VCV latency function, average reaction times to VC syllables were faster than to VCV syllables, $F(1,9) = 12{\cdot}4, P < 0.01$, as can be seen in Fig. 2(a). This was not true, however, for the author and one other very skilled subject; they showed equal latencies to VCV and VC stimuli. The average latencies for the practice VC syllables were 66 ms faster than those for the VC syllables in the mixed series, $F(1,9) = 28{\cdot}1$, $P < 0{\cdot}001$, which replicates the effect of list structure observed in Experiment I.

The average error rate was 7·8%, with individual rates varying from 0·7 to 28·7 %. An additional 1·5% of the latencies were considered anticipations (mostly with V-CV stimuli) and excluded. Fig. 2(b) shows that there were many more errors with V-CV utterances than with VCV or VC utterances. However, there was little change in the error pattern with closure duration (except between the two shortest closure durations), in contrast to the V-CV latencies. The error rate for the practice VC stimuli was 4·5 %.

Discussion

Judging from the flatness of the VCV latency functions, the results of this experiment seem to show that even inexperienced listeners can make decisions on the basis of the VC transitions in VCV utterances. However, this interpretation is challenged by the fact that VC latencies were markedly faster than VCV latencies (excepting the two best subjects, as noted above). The difference between VCV and VC latencies was most pronounced for the poorest subjects (i.e. those with the slowest latencies and the highest error rates).

The finding of fast VC reaction times contrasts with Experiment I, where VC latencies were generally slower than VCV latencies. These conflicting results may have to do with the frequency of occurrence of the VC stimuli: In Experiment I, only one ninth of the stimuli in the mixed (VCV and VC) sequence were VCs, but in Experiment II they constituted one third. In order for stimulus frequency to have this selective effect, however, it must be the case that the CV portion of VCV stimuli did affect reaction time. Since, in the present experiment, this effect seemed to be inhibitory rather than facilitative, the question arises why VCV reaction times did not *decrease* as closure duration increased. After all, the latencies for VCVs must eventually approach those for VCs, given that the closure period is sufficiently long.

The most obvious possibility is that, for most subjects, the VCV reaction times again represented a mixture of two latency distributions: a faster VC distribution and a slower CV distribution. As closure duration increased, the mean of the CV distribution increased by the same amount, but simultaneously the proportion of latencies from this distribution decreased because the longer interval facilitated reactions to the VC portion alone. If these two factors were in approximate balance, they could have led to a flat latency function.

Perhaps the most surprising feature of the present experiment was the high error rate for V-CV stimuli, despite the fact that they were difficult to tell apart from VCV stimuli, at least at the shorter closure durations. (Repp. 1978*a*, showed that, at the shortest closure duration used here, discrimination scores in an AXB task were near chance.) The errors may have resulted from poor intelligibility of the CV portions of the stimuli. This possibility cannot be ruled out with certainty, since the CV portions were never presented in isolation, but it seems unlikely; one might have expected higher error rates for VCV stimuli with short closure durations in this case. It is more tempting to attribute the errors to anticipatory responses – responses made on the basis of the initial vowel portions of the V-CV stimuli, which did not contain any information about the consonant at all. The author's own data contained a number of clear instances of such anticipations – reaction times that were too fast to be based on the CV transitions and whose accuracy was at chance level. Since the reaction times of most other subjects (especially of those with high error rates) were much slower, anticipations could only rarely be identified on the basis of fast latencies. However, if the subjects really tried to follow the instructions and focussed on the initial portions of the stimuli, some anticipatory responses to V-CV stimuli should have occurred, and half of them must have been errors, on the average.

General discussion

Experiment II supplements Experiment I in furnishing evidence that listeners can (and do) utilize the information conveyed by VC transitions in VCV stimuli. However, in agreement with Experiment I, inexperienced listeners are unable to ignore the CV transitions completely – the responses remain a function of both sets of spectral cues, with their relative perceptual weights depending on closure duration. Nevertheless, we may infer from the results that phonetic information is continuously extracted from the speech signal, and that phonemic decisions can be made before the process of cue integration is complete. Even though fast phonemic decisions are not required in everyday speech perception, the conclusion that phonetic information accumulates continuously may legitimately be generalized to more natural situations. This conclusion agrees well with the findings of other investigators concerning the continuous nature of the syntactic and semantic levels of processing (Cole & Jakimik, 1977; Marslen-Wilson, 1975, 1976; Marslen-Wilson & Welsh, 1978).

One objection that might be raised against the present results is that the listeners, particularly the experienced ones, may have employed an auditory criterion to distinguish between the VC transitions in /ab/ and /ad/. However, it seems unlikely that auditory discrimination of rapid formant transitions would lead to decisions as fast and as accurate as those observed. Also, the author's subjective experience as a listener was one of making phonemic judgments.

It remains to be seen whether the ability of listeners to make perceptual use of the VC transitions in VCV utterances extends to closure durations shorter than the ones used here (<65 ms). There is some reason to believe that the perception of the VC transitions may be interfered with by the CV transitions at such short intervals. Certainly, such interference is obtained when the two sets of transitions specify different places of articulation (Dorman, Raphael & Liberman, 1979; Repp, 1978*a*). Closure durations that are too short may also lead to a general decrease in intelligibility. The range of closure durations in the present study was chosen with these possibilities in mind.

Finally, it should be mentioned that several recent studies support the conclusions of the present research. Porter & Castellanos (1977) and Castellanos (1978), in a shadowing task with VCV stimuli, have obtained latencies short enough to indicate that the listeners directly responded to the VC transitions. These studies replicated and extended earlier findings by Chistovitch and her colleagues (described in Kozhevnikov & Chistovitch, 1965, 223pp.). Note, however, that rapid shadowing does not require conscious identification of phonemes; a continuous translation from perception to articulation could take place at an earlier level. Thus, the present experiments, while in agreement with the shadowing experiments, did investigate a different aspect of the problem. A same-different reaction time experiment by Remington (1977), though aimed at a somewhat different issue, also yielded results consonant with the hypothesis that phonetic information is extracted continuously. Streeter & Nigo (1979) have reported faster recognition latencies for words containing VC transitions than for the same words with the VC transitions removed. As in the present Experiment II, the absence of VC transitions may have delayed recognition of the corresponding consonant and, thus, the access to a phonetically organized lexicon. Taken together, these and other results provide convincing evidence for the continuous nature of speech perception.

This research was supported by NICHD Grant HD01994 and BRS Grant RR05596, both to Haskins Laboratories. I would like to thank Alice Healy, David Isenberg, Virginia Mann, and Alvin Liberman for helpful comments on earlier drafts of this paper.

References

Bever, T. G. & Hurtig, R. (1975). Detection of a non-linguistic stimulus is poorest at the end of a clause. *Journal of Psycholinguistic Research* **4**, 1–7.

Castellanos, F. X. (1978). Shadowing and consonantal release. *Journal of the Acoustical Society of America* **63** (Supplement No. 1), S45–S46. (Abstract.)

Cole, R. A. & Jakimik, J. (1977). *Experiments with Mispronunciations.* Paper presented at Psychonomic Society Meeting, Washington, D.C., November 1977.

Dorman, M. F., Raphael, L. J. & Liberman, A. M. (1979). Some experiments on the sound of silence in phonetic preception. *Journal of the Acoustical Society of America* **65**, 1518–1532.

Jarvella, R. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior* **10**, 409–416.

Kozhevnikov, V. A. & Chistovich, L. A. (1965). *Speech, Articulation and Perception.* Washington D.C.: Joint Publications Research Office JPRS-30543.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science* **189**, 226–228.

Marslen-Wilson, W. D. (1976). Linguistic descriptions and psychological assumptions in the study of sentence perception. In *New Approaches to Language Mechanisms* (R. J. Wales & E. C. T. Walker, eds). Amsterdam: North Holland.

Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* **10**, 29–63.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* **39**, 151–168.

Porter, R. J., Jr. & Castellanos, F. X. (1977). Speech production measured of speech perception: replications and extensions. *Journal of the Acoustical Society of America* **61** (Supplement No. 1), S91–S92. (Abstract.)

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

Repp, B. H. (1978*a*) Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception and Psychophysies* **24**, 471–485.

Repp. B.H. (1978*b*). "Coperception": influence of vocalic context on same-different judgments about intervocalic stop consonants. Unpublished manuscript (available from the author).

Shand, M. A. (1976). Syllable vs. segmental perception: On the inability to ignore *"irrelevant"* stimulus parameters. *Perception and Psychophysics* **20**, 430–432.

Streeter, L. A. & Nigro, G. (1979). The role of medical consonent transitions in word perception. *Journal of the Acoustical Society of America* **65**, 1533–1541.

Westbury, J. R. (1977). Temporal control of medial stop consonant clusters in English. Paper presented at 93rd Meeting of Acoustical Society of America, University Park, Pennsylvania, June 1977. Abstracted in *Journal of the Acoustical Society of America* **61**, (Supplement No. 1), S91.