Perceptual Integration and Differentiation of Spectral Information Across Intervocalic Stop Closure Intervals

Bruno H. Repp

ABSTRACT

Perceptual interactions between implosive and explosive transitions of intervocalic stop consonants were investigated by preceding stimuli from a synthetic /br/-/dr/ continuum by either /ab/ or /ad/. At a very short closure (interstimulus) interval (15 msec), a single stop consonant is heard, with its perceived place of articulation determined primarily by the CV portion of the composite stimulus. However, the vowel consonant (VC) portion exerts a significant positive bias on perception. The strength of this bias varies with the acoustic structure of the VC precursor, which indicates that the bias is due to auditory temporal integration across the closure period. At a longer closure interval (140 msec), two stop consonants are heard if the VC and CV portions convey different places of articulation; a single stop consonant, otherwise. Here a contrast effect is obtained: for example, an /ab/ precursor shifts the /br/-/dr/ boundary towards /dr/, increasing the probability of hearing /ab-dr/. This contrast effect is equally pronounced in the backward direction, when stimuli from an /ab/-/ad/ continuum are followed by either /br/ or /dr/. This effect presumably arises at the phonetic level, where differentiation of the speech signal precedes integration.

INTRODUCTION

The present experiment supplements and extends Experiment III of Repp (1977). The earlier experiment investigated whether there are perceptual interactions in the perception of the spectral information preceding and following the closure period of an intervocalic stop consonant (that is, of the implosive and explosive formant transitions), or whether this temporally separated information is perceived independently. Synthetic stimuli from a /br/-/dr/ continuum were preceded by either /ab/ or /ad/. When the closure period separating the two stimulus portions (VC and CV) was very short (25 msec), a single consonant was perceived, with its place of articulation determined primarily by the CV portion (the explosive transitions). However, the VC precursors did exert a significant positive bias on perception: the /b/-/d/ phoneme boundary was shifted away from the precursor, relative to a control condition in which only the CV portions were identified. At a very

Acknowledgment: This research was supported by NICHD Grant HD01994. Thanks are due to Georgann Witte for her help in analyzing the data.

[HASKINS LABORATORIES: Status Report on Speech Research SR-51/52 (1977)]
long closure interval (265 msec) that permitted the listeners to hear two stop consonants, the VC precursors had no significant differential effect on the perception of the CV portion.

In two other conditions of the experiment, stimuli from an /ab/-/ad/ continuum were followed by either /bc/ or /dc/, in order to investigate whether perceptual interactions exist in a backward direction as well. At a closure period (115 msec) thought to be sufficient to hear two stop consonants when the VC and CV portions conveyed different places of articulation, there was no differential effect of the two CV postcursors on perception of the VC portion, although the mere presence of a postcursor biased perception towards /b/, relative to a control condition where only the VC portions were identified. In other words, perception of /ad-bc/ (or /ab-c/) with the /bc/ postcursor was just as likely as perception of /ad-c/ (or /ab-dc/) with the /dc/ postcursor. At a very long closure period (265 msec), however, there was a small but significant contrastive effect, such that the /ab/-/ad/ phoneme boundary was shifted towards the postcursor, relative to the control condition.

The assimilative perceptual interaction of implosive and explosive formant transitions at very short closure periods was interpreted as evidence of a form of auditory temporal integration, that is, perceptual integration of auditory information before phonetic categorization. That the effect was not merely due to occasional perceptual dominance of the implosive transitions was suggested by separate analyses of the subjects' rating responses within each response category. However, it is theoretically conceivable that the implosive transitions were covertly categorized and influenced the perception of the explosive transitions at the phonetic level, before the temporary category was lost from memory or integrated with the perceptually more dominant category derived from the explosive transitions. In order to investigate this possibility, two acoustically different versions of each precursor were used in the present experiment, one of which was closer to the phoneme boundary on the synthetic VC continuum than the other. If implicit categorization is involved, these within-category acoustic differences should have little effect. However, the auditory integration hypothesis predicts that the acoustically more extreme /ab/ (/ad/) will exert a stronger positive bias than the less extreme precursor from the same category.

A second condition of the present experiment replicated the earlier condition with a 115-msec closure period that had not yielded any differential postcursor effect. Again, stimuli from an /ab/-/ad/ continuum were followed by either /bc/ or /dc/, but the closure duration was increased to 140 msec. Informal evidence suggested that the 115-msec closure interval perhaps had been too short, so that the task was too difficult for most listeners. It was expected that the 140-msec closure interval would enable the subjects to perceive two different stop consonants without difficulty whenever the implosive and explosive formant transitions specified different places of articulation. A similar condition was included in which stimuli from a /bc/-/dc/ continuum were preceded by either /ab/ or /ad/. The latter condition investigated forward interaction (the influence of the VC precursor on the perception of the CV portion), while the former investigated backward interaction (the effect of the CV postcursor on the perception of the VC portion). If there was any interaction at all, it was expected to be contrastive,
because it would presumably take place at the phonetic (categorical) level, since the closure interval was long enough to permit categorization of the implosive transitions. It was of interest whether forward and backward interaction conditions would differ in the magnitudes of the interaction effects obtained. Intuitively, forward influences seemed more likely than backward influences; however, the earlier experiment showed the opposite at very long closure durations (265 msec), although the effects were very small.

METHOD

For details of stimulus construction and procedure, the reader is referred to Repp (1977, Exp. III).

Subjects

Nine paid volunteers who had not been subjects in the previous experiment participated.

Stimuli

The stimuli were the same as in the earlier experiment, except for the changes noted below. The first stimulus series was a duplicate of the isolated CV series of Experiment V. It was followed by a series of 180 stimuli grouped into three blocks of 60. Each block contained one randomization of the stimuli from the 7-member CV continuum (with the replication structure described in Repp, 1977) preceded by each of four VC precursors. The closure period was 15 msec. The four precursors were stimuli 1, 3, 6, and 8 from the 7-member VC continuum, stimulus 8 being an additional syllable especially constructed for this purpose and having even more extreme formant transition offset frequencies than stimulus 7. (It was assumed that stimulus 8 would have been consistently identified as /ad/, given that stimuli 6 and 7 were so identified.) The next stimulus series was a duplicate of the isolated VC series of Experiment V. It was followed by a series analogous to the earlier 115-msec series (where the stimuli from the VC continuum were followed by either /bɛ/ or /dɛ/), except that the closure duration was extended to 140 msec. Finally, another series with a 140-msec closure duration was recorded, in which the stimuli from the CV continuum were preceded by either /ab/ or /ad/.

Procedure

All subjects listened to the conditions in the order described above. The initial CV series and the following 15-msec series were repeated once. (Two subjects omitted the repetition of the CV series.) The instructions were the same as in Experiment V. A 6-point rating scale was used except in the two 140-msec series, where the task was to indicate whether one or two medial consonants were heard.

RESULTS

The results of the 15-msec condition are shown in Figure 1a. The dashed line represents the results for CV syllables in isolation. The function was somewhat less steep than in the earlier experiment, but this was mainly due to
the subjects' use of less extreme ratings; the two phoneme categories were well separated. The precursor functions looked somewhat different from those in the earlier study. In particular, the precursors seemed to have no differential effect at the /b/-end of the CV continuum, except for shifting the ratings slightly towards /b/. Towards the /d/-end of the continuum, however, a strong assimilative precursor effect emerged, in agreement with the earlier results. In addition, those precursors that were closer to the phoneme boundary produced a weaker effect than the precursors from the ends of the VC continuum, as predicted under the auditory integration hypothesis.

The statistical analysis showed a significant between-category effect of VC precursors (precursors 1 and 3 vs. 6 and 8: $F_{1,8} = 8.9$, $p < .025$), an interaction of this effect with position on the CV continuum ($F_{4,32} = 4.0$, $p < .025$; this significance level was relatively low only because of high between-subject variability), a significant interaction of the within-category precursor effect (precursors 1 and 6 vs. 3 and 8) with position ($F_{4,32} = 4.6$, $p < .01$), and a significant triple interaction ($F_{4,32} = 5.5$, $p < .01$). The main effect within precursor categories did not reach significance.

In order to clarify these results, weighted average ratings were computed conditionally on the two response categories (ratings 1-3 = B; ratings 4-6 = D). These data are plotted in Figure 1b. Separate analyses of variance were conducted on B and D responses, including (the first or last) four positions on the CV continuum. (These analyses were conducted on unweighted conditional ratings, while Figure 1b represents weighted means; therefore, the statistical results may deviate slightly from the graphical representation.) B responses showed a highly significant effect of position ($F_{3,24} = 43.3$, $p << .01$), obviously due to the increase in ratings between positions 3 and 4. The between-category precursor effect just fell short of significance ($F_{1,8} = 5.2$, $p < .06$), indicating a tendency in the expected direction. All other effects were far from significant. The D responses showed significant effects of position ($F_{3,24} = 3.9$, $p < .025$)—Figure 1b shows the effect to be very small—between precursor categories ($F_{1,8} = 15.6$, $p < .01$) and within precursor categories ($F_{1,8} = 6.0$, $p < .05$). The between-category effect also interacted with position ($F_{3,24} = 3.5$, $p < .05$), but this effect was negligible.

The results for D responses support the prediction that within-category acoustic differences between VC precursors would affect the degree of the assimilative precursor effect. Such a gradual effect most likely reflects auditory integration of implosive and explosive transitions; implicit categorization of the implosive transitions seems unlikely. It should be noted that the VC precursors were very consistently identified in isolation (cf. Figure 1c): stimuli 1 and 3 received 100 and 98 percent B responses, respectively; stimuli 6 and 7 received 98 and 100 percent D responses, respectively, so that it was safe to assume that stimulus 8 (the actual precursor) would have received 100 percent D responses, too. The between-category precursor effect was somewhat larger than the within-category effect, but this probably reflected the fact that the acoustic difference between categories was larger than that within categories.

The results for the 140-msec VC (backward interaction) series are shown in Figure 1c. The labeling function for VC stimuli in isolation and the

134
Figure 1:
(a) Mean "D-ness" ratings of syllables from a /be/-/de/ continuum in isolation (CV) and when preceded by a 15-msec silent interval and one of four VC precursors. The subscripts denote the position of the VC precursors on the /ab/-/ad/ continuum. (b) The same data, analyzed separately for the two response categories (B = ratings 1-3; D = ratings 4-6). (c) Average percentage of "D" responses to syllables from an /ab/-/ad/ continuum in isolation (VC) and when followed by one of two CV precursors after a 140-msec silent interval. (d) Average percentage of "D" responses to syllables from a /be/-/de/ continuum in isolation (CV) and when preceded by a 140-msec silent interval and one of two VC precursors.
general shape of the postcursor functions replicated the 115-msec condition of Repp (1977, Exp. III, Figure 10a). However, in contrast to the earlier results, there was a clear difference between the two postcursor functions in the predicted direction, viz., a contrast effect \( (F_{1,8} = 13.6, p < .01) \) which did not interact with position on the VC continuum. Seven of the nine subjects showed a contrast effect, one showed no clear effect, and one a slight assimilative effect. Thus, the increase in closure duration from 115 to 140 msec apparently facilitated the task, so that a contrast effect could emerge. At the 115-msec closure duration, some subjects still seemed to have a tendency to integrate implosive and explosive transitions; this tendency was no longer present at the 140-msec closure duration.

The results for the new 140-msec CV (forward interaction) series are shown in Figure 1d. It can be seen that this series, too, showed a pronounced effect of the VC precursors on the perception of the CV portions \( (F_{1,8} = 11.5, p < .01) \). Contrast effects were shown by six subjects; the remaining three showed no clear effects. The precursor functions followed the isolated-VC function more closely than the postcursor functions in Figure 1c followed the isolated-VC function: VC syllables sounded more labial when followed by a CV postcursor, but not vice versa. The former effect may reflect merely a peculiarity of the present stimuli, but it may also indicate that implosive transitions are perceptually less stable than explosive transitions.

**DISCUSSION**

The implosive and explosive formant transitions surrounding the closure period of an intervocalic stop consonant provide an extreme example of temporally distributed information for a single phoneme. In order to perceive this information as a single event, an integrative perceptual process is needed. The present experiment, together with its predecessors (Repp, 1976, 1977) suggests that there are at least two levels at which acoustic cues may be integrated. One level is auditory, the other phonetic in nature; this distinction follows common usage in speech perception research (for example, Studdert-Kennedy, 1976).

Auditory temporal integration is apparent in the situation where conflicting implosive and explosive transitions separated by a very brief closure period lead to the perception of a single phoneme (Dorman, Raphael, Liberman, and Repp, 1975). This "backward masking" effect was replicated here using an identification-judgment task. The analysis of conditional ratings, as well as the within-category precursor effect, specifically support the notion of auditory temporal integration with higher weighting of more recent information; they make an interpretation in terms of true recognition backward masking (that is, interruption of processing—Massaro, 1975) seem unlikely.

The temporal limits of this auditory integration may have just been reached in the inconclusive 115-msec condition of the earlier study. There seemed to be a region of uncertainty beyond the region over which the "backward masking" effect primarily operates. An estimate of the total integration period may be obtained by adding the duration of the spectral information being integrated to the maximal temporal separation. Thus, a value of approximately 200 msec is obtained, which is of the same magnitude as Huggins' (1975) estimate from the perception of temporally segmented speech.
Huggins hypothesized that "the ear tries to integrate into a single percept any two relatively similar events that coexist in echoic storage, and only becomes able to treat them as separate events if they do not coexist in echoic storage" (p. 156). If it is accepted that any implosive and explosive transitions are "relatively similar," the 200-msec auditory integration period obtained here may well reflect the duration of an echoic store. The perceptual predominance of the more recent information (explosive transitions) may be due to the fact that it resides longer in auditory memory (adopting Huggins' analogy of storage with a delay line). There are a number of related findings in the auditory literature suggesting a temporal integration period of about 200 msec (see Huggins, 1974, 1975, for a brief discussion). This may or may not be a coincidence; certainly, stimulus and task factors play a role in determining the precise interval over which integration occurs.

A second process of temporal integration must be postulated to account for the perceptual distinction between single and geminate stop consonants. When implosive and explosive transitions signalling the same place of articulation are separated by less than about 200 msec of silence, they are perceived as a single phoneme (Pickett and Decker, 1960; Repp, 1976). This integrative process, whose "time window" of about 300 msec clearly exceeds that of auditory temporal integration, most likely takes place at the phonetic level. Presumably, a phonemic percept is established on the basis of the implosive transitions plus silence, but if an identical phonemic percept is arrived at within a certain time span (on the basis of explosive transitions occurring 100-200 msec later), a higher-order perceptual rule integrates the two into a single percept. This perceptual rule may well represent the listener's implicit knowledge about articulatory processes, and therefore it may have no parallel outside speech perception. Moreover, Pickett and Decker (1960) have shown its extreme sensitivity to contextual variables, particularly rate of speech. Similar variability of temporal parameters is rarely found in auditory (nonspeech) perception.

The process of phonetic integration is paralleled by the contrast effects observed at the 140-msec closure duration in the present experiment. These contrast effects seem logically prior to integration: integration is possible only if implosive and explosive transitions are perceived as signalling the same place of articulation. However, the nonindependence of the perception of these spectral cues, both forward and backward in time, is in itself evidence of a perceptual process that operates on fairly long temporal segments of the speech signal. What has been called here phonetic integration may be considered merely a consequence of this more general process of "coperception." Coperception—the process that enables the listener to interpret a substantial stretch of the acoustic information in parallel—serves both to integrate and to differentiate acoustic cues. Perceptual contrast effects reflect the tendency of the perceptual system to differentiate; integration follows if differentiation does not occur. The perceptual interactions and individual differences observed by Repp (1977) at closure durations beyond the single-geminate boundary need further corroboration before it is concluded that perceptual interactions between implosive and explosive transitions and the phonetic integration responsible for the single-geminate distinction are not, in fact, due to the same underlying process. Coperception subsumes constraints imposed on perception by a short-lived auditory memory, as well as higher-order phonetic processes that are not
subject to similar limitations. While the former are characterized by a more or less fixed time constant, the latter are better described in terms of flexible perceptual rules. Both processing levels cooperate in differentiating and integrating the temporally distributed information in the speech signal.

REFERENCES


