JOURNAL OF MEMORY AND LANGUAGE 26, 564–573 (1987)

Vowels as Islands of Reliability

RANDY L. DIEHL, KEITH R. KLUENDER, DONALD J. FOSS, ELLEN M. PARKER, AND MORTON A. GERNSBACHER

University of Texas at Austin

Vowel nuclei of syllables appear to provide a relatively stable (although not stationary) frame of reference for judging consonant events. We offer evidence that reliable consonant identification demands prior or simultaneous evaluation of this "vocalic frame." Listeners were presented a list of /bVs/, /dVs/, and /gVs/ syllables and were instructed to press a response key immediately upon recognizing a particular initial consonant target. Three groups of subjects monitored for /b/, /d/, and /g/, respectively. The test syllables contained 10 English vowels varying substantially in intrinsic duration. Response times to the initial consonants correlated positively with the duration of the following vowels, even when the effect of consonant-vowel formant transition duration was partialed out. The results suggest that consonant recognition is vowel dependent and, specifically, that a certain amount or proportion of the vowel formant trajectory must be evaluated before consonants can be reliably identified. © 1987 Academic Press, Inc.

A recurrent finding in speech research is that the perceptual interpretation of consonant cues may depend critically on information about the following vowel. To take an example from the early literature, Liberman, Delattre, and Cooper (1952) reported that a brief filtered noise burst (centered at about 1440 Hz) was typically perceived as /p/ in front of a steady state /i/ or /u/, but as /k/ in front of a steady state /a/. More recently, Mann and Repp (1980) found an effect of the following vowel on the identification of a noise segment that was perceptually intermediate between /s/ and /ʃ/. The segment was more likely to be labeled /s/ before /u/ than before /a/.

These perceptual effects may be explained in terms of the listener's implicit knowledge of the normal acoustic consequences of speech production. Adjacent phonetic segments tend to be *coarticulated,* that is, produced in a temporally overlapping manner. For example, in the production of /su/ or /ʃu/, lip rounding appropriate for the vowel occurs during the preceding fricative segment, causing the fricative noise to be lowered in frequency. Apparently to compensate for this, listeners in the Mann and Repp (1980) study more readily accepted a lower frequency noise as /s/ (rather than /ʃ/) before the vowel /u/.

In this paper we are concerned with a rather different (although perhaps related) effect of vowel context on the identification of a preceding consonant. This effect was serendipitously discovered by Foss and Gernsbacher (1983), while testing a model of phoneme recognition called the "dual code" hypothesis (Foss, Harwood, & Blank, 1980). The model states conditions under which phonemes are identified either in a "bottom-up" mode of direct acoustic-phonetic analysis or in a "top-down" mode involving prior lexical access. The subjects' task was to press a button as quickly as possible whenever a specified word-initial target phoneme occurred in a sentence or in a list of words and nonwords (in the latter case the target was always in utterance-initial position). According to the model, conditions favoring a top-down processing mode should have yielded faster response times for target consonants appearing in words rather than in nonwords. On the other hand, lexical status of the target-bearing item should not have mattered under conditions favoring a bottom-up mode of processing. In general, the results were not consistent with the dual-code model. One particularly troubling finding was that, in some conditions, response times were actually faster to initial consonants appearing in nonwords.

This work was supported by NIH Grant HD18060 to the first author.

E. M. Parker's current address is Computer Cognition, 6696 Mesa Ridge Road, San Diego, CA 92121. M. A. Gernsbacher's current address is Department of Psychology, University of Oregon, Eugene, OR 97403. Requests for reprints should be sent to Dr. Randy L. Diehl, Department of Psychology, University of Texas, Austin, TX 78712.

0749-596X/87 $3.00

Given the apparent inconsistencies in their data, Foss and Gernsbacher sought to learn whether any purely phonetic properties among the stimulus items had confounded the results. One of us (Diehl) noted that the reaction times to the target consonant were generally slower when the following vowels were of greater intrinsic duration. For example, monitoring latencies to the initial /d/ phoneme were greater for items such as "dan" (/dæn/), and "dine" (/daɪn/) than for items such as "din" (/dɪn) and "den" (/dɛn/). Foss and Gernsbacher proceeded to measure the correlation between initial-consonant monitoring latencies and the typical durations of the following vowel categories, as determined by Peterson and Lehiste (1960). They found a significant positive relation between these two measures ($r = .63$), and they concluded that vowel duration differences could explain many of the apparent discrepancies among the results of their own study as well as related studies by other investigators.

Why should consonant identification time be affected by the duration of the following vowel? We can think of one explanation that is at least superficially plausible. One of the most important variables correlated with intrinsic vowel duration is degree of mouth opening (House, 1961; House & Fairbanks, 1953; Lindblom, 1967; Peterson & Lehiste, 1960). This relation is often attributed to an inertial constraint on the movement of the tongue and especially the jaw (Lehiste, 1970). The greater duration of open vowels such as /æ/ is seen as a direct consequence of the longer articulatory trajectories involved in their production. (See Lindblom, 1967, for supporting evidence from a modeling study of articulatory dynamics in vowel production.) Accordingly, one would expect consonant-vowel transitions to be longer for open vowels, and the greater consonant monitoring latencies in the Foss and Gernsbacher (1983) study might simply be a reflection of these longer transitions.

As plausible as this account may seem, there is reason for skepticism. Based on his analysis of data from Lehiste and Peterson (1960), Lisker (1974) concluded that greater intrinsic durations for open vowels are attributable to longer quasi steady state intervals rather than to longer consonant-vowel or vowel-consonant transition durations. In fact, there appears to be little correlation between overall vowel duration and transition duration.[1] This obviously calls into question the mechanico-inertial explanation of intrinsic vowel duration differences, and it raises anew the question of why consonant monitoring latencies are dependent on following vowel duration.

Before attempting to answer this question, we decided to try to replicate and extend the findings of Foss and Gernsbacher (1983), using more than one talker and three different initial consonant targets.

A Further Test of the Relation between Consonant Identification Time and Following Vowel Length

Listeners were presented a randomized

[1] That greater articulatory displacements of open vowels do not generally result in longer formant transitions is explained by some results of Sussman, MacNeilage, and Hanson (1973), which slow that velocity of jaw movement varies directly with displacement in consonant-vowel and vowel-consonant syllables.

list of /bVs/, /dVs/, and /gVs/ syllables produced in citation form by two male talkers and were asked to press a response key immediately upon recognizing a particular initial consonant. Three groups of 12 subjects monitored for /b/, /d/, and /g/, respectively. The test items contained 10 English monophthong vowels that varied substantially in duration. The durations corresponding to the different vowel categories correlated highly between the two talkers and also with values reported previously in the literature. Before listening to the actual test stimuli, the subjects practiced the monitoring task on a list of 40 syllables spoken by a different male talker from those used for the test stimuli. Following the practice session, subjects in each monitoring condition listened to 640 stimulus tokens, half of which contained the target consonant, and the other half of which were equally divided between the two nontarget categories.

Only response times for correct monitoring responses were included in the analyses. In addition, response times greater than two standard deviations above each subject's mean were excluded, because these were assumed to represent failures to respond "as quickly as possible." In the /b/, /d/, and /g/ monitoring conditions, errors amounted to 1.0, 1.4, and 1.2% of the responses, and outliers represented 3.3, 3.9, and 4.2% of the responses, respectively.

Figure 1 displays the scatterplots for the /b/ monitoring condition. Response time is plotted on the ordinate, and vowel duration, measured from the consonant release burst to the offset of periodicity preceding the final /s/, is registered on the abscissa. The separate scatterplots (T1 and T2) correspond to the two talkers. Notice that for both talkers, the correlation coefficients of .39 and .28 are substantially smaller than the .63 value obtained by Foss and Gernsbacher (1983). Neither correlation is significant by a two-tailed test, although the r value of .39 for Talker 1 is nearly so ($.05 < p < .10$).

The results for the /d/ monitoring condition are shown in Fig. 2. The correlations

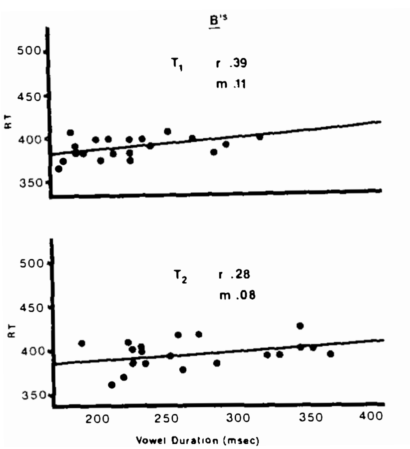

FIG. 1. Results for the /b/ monitoring condition. Response times (RT) are plotted as a function of vowel duration for the syllables produced by Talker 1 (T1) and Talker 2 (T2); r refers to the Pearson product–moment correlation coefficient; m is the slope of the regression line.

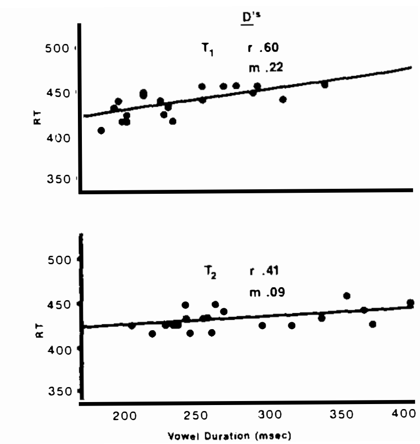

FIG. 2. Results for the /d/ monitoring condition. Response times (RT) are plotted as a function of vowel duration for the syllables produced by Talker 1 (T1) and Talker 2 (T2); r refers to the Pearson product–moment correlation coefficient; m is slope of the regression line.

between response time and vowel duration are considerably larger than those in the /b/ condition: The r value of .60 for Talker 1 is significant ($p < .01$), whereas the value of .41 for Talker 2 just failed to reach significance ($.05 < p < .10$). Notice that the correlation for Talker 1 is very close to that reported by Foss and Gernsbacher (1983), who always used /d/ as the target consonant.

Finally, Fig. 3 shows the results for the /g/ monitoring condition. Again, the correlations are greater than in the /b/ condition, and both r values are significant (Talker 1: $r = .49$, $p < .01$; Talker 2: $r = .67$, $p < .05$). Thus, except for the marginal relation found in the /b/ monitoring condition, our results essentially replicated those of Foss and Gernsbacher (1983). Monitoring latencies were positively related to the duration of the following vowel.

Earlier, we discussed the possibility that such a correlation might stem from longer consonant-vowel formant transitions being associated with more open vowels. Because formant transitions are directly informative about consonant place of articulation, that is, /b/ versus /d/ versus /g/ (Liberman Delattre, Cooper, & Gerstman, 1954), longer transitions could provide a simple and rather trivial explanation of longer monitoring latencies. Recall that we discounted this possibility on the grounds that Lisker (1974) had noted little correlation between intrinsic vowel duration and formant transition duration. However, to check whether Lisker's conclusion held true for our stimulus set, two of the experimenters independently estimated the transition durations, measuring from the release burst to the end of the rapid spectral change corresponding to the opening gesture. These estimates correlated .90 across the two judges. We then took the average of each estimate pair as our measure of transition duration. Consistent with Lisker's analysis, the correlations between overall vowel duration and transition duration were rather small: .22 for Talker 1 and .30 for Talker 2, with only that for Talker 2 being significant. Moreover, when we partialed out the effect of transition duration from our earlier analysis, the correlations between monitoring latencies and vowel duration varied only slightly from the simple correlations and the pattern of significance levels was unchanged. In short, the observed correlations appear to depend on vowel length per se and not on variations in the durational extent of consonant cues in the formant transitions.

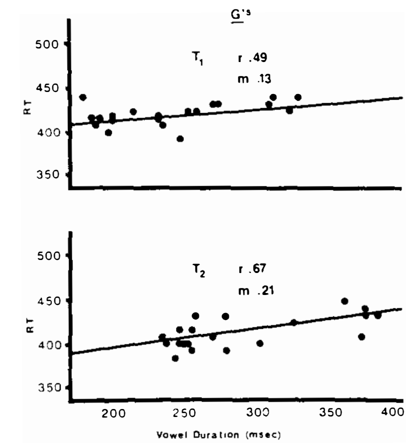

FIG. 3. Results for the /g/ monitoring condition. Response times (RT) are plotted as a function of vowel duration for the syllables produced by Talker 1 (T1) and Talker 2 (T2); r refers to the Pearson product–moment correlation coefficient; m is the slope of the regression line.

Reliability of the Vowel Length Effect

Although the present findings are consistent with the earlier results of Foss and Gernsbacher (1983), a recent study by Cutler, Mehler, Norris, and Segui (1987) has raised questions about the reliability of the vowel length effect on consonant monitoring latencies. In several experiments, Cutler et al. found a significant vowel length effect when /b/ was the initial consonant target, but not when /d/ was the target, a reversal of our own results.

There are at least two possible reasons for this discrepancy. In our experiment, all of the test syllables ended with /s/, whereas in the study by Cutler et al., variation in vowel category was confounded with variation in final consonant category, including differences in manner class (viz., stops, nasals, liquids, fricatives, and affricates) and voicing. Both manner and voicing are known to influence the duration of the preceding vowel (Lehiste, 1970). For example, vowels are generally longer in front of fricative consonants than in front of stop consonants at the same place of articulation. Such length effects are most naturally explained in terms of adjustments in vowel trajectory in the vicinity of the final consonant. Accordingly, one would expect them to have little bearing on initial-consonant monitoring latencies, which we suggest are influenced primarily by that portion of the vowel trajectory proximal to the initial consonant (i.e., that portion that extends through roughly the first half of the syllable). Thus, for the purposes of explaining initial-consonant monitoring latencies, it is useful to distinguish between *relevant* vowel length variation due, for example, to variation in tongue height or jaw opening (which changes the entire vowel trajectory) and *irrelevant* vowel length variation due to local adjustments in the region of the final consonant. The fact that these two sources of vowel length variation were confounded in the study by Cutler et al. may help to explain the difference between their results and ours.

A second possible reason for the discrepancy concerns the onset characteristics of the stimuli used in the two studies. English voiced stops such as /b/ and /d/ are produced either with the onset of voicing (i.e., vocal fold vibration) occurring at or shortly after the closure release of the articulators or else with the voicing onset preceding the release by intervals of up to 150 or more. (The latter case is referred to as *prevoicing*.) Acoustically, prevoicing is a low-frequency periodic signal that contains very little information about consonant place of articulation.

In our study, approximately 75% of the test syllables produced by both talkers originally contained significant prevoicing. We digitally removed these prevoicing segments, so that the onset of each stimulus corresponded to the moment of closure release. Since stimulus onset marked the beginning of the measured response time, the excision of prevoicing ensured that response time reflected only the processing of information relevant to consonant place of articulation. Among our (unedited) stimulus items, prevoicing varied in duration from 0 ms to about 200 ms. Obviously, including these varying intervals in the measure of response time would have obscured any effect of following vowel length on consonant monitoring latency.

In the study by Cutler et al., response time was measured from stimulus onset as determined by a voice activated signal detector, and prevoicing was not removed from the stimulus items (Cutler, personal communication). To the extent that there was prevoicing present on some of the items (particularly the /d/ stimuli) and to the extent that it was sufficiently intense to activate the timing device, one would expect a marked discrepancy between the results of the Cutler et al. study and our own findings. We are satisfied, therefore, that the converging results of the present experiment and the earlier study by Foss and Gernsbacher (1983) are reliable.

Perception of Coarticulated Segments

In attempting to explain the relation between consonant monitoring latencies and following vowel duration, we will (a) outline two alternative theories of consonant-vowel coarticulation; (b) use one of these theories (or at least one of its ancillary assumptions) to motivate some claims about

the different perceptual status of consonants and vowels, claims that we think provide a rationale for our results; and (c) describe more generally how well our results comport with each of these theories of coarticulation. We will not attempt to argue, however, that our perceptual results provide decisive support for either theory.

In the traditional phonetics literature, coarticulation was usually viewed as a kind of merging or assimilation of contextual properties with a segment, such that articulatory (and hence acoustic) trajectories from segment to segment were shortened and smoothed. Lindblom (1963), for example, proposed an assimilatory account of vowel "undershoot" effects in consonant context. (Formant patterns of consonant-vowel-consonant syllables often do not reach the "target" frequencies characteristic of sustained isolated vowels.) He suggested that the undershoot observed with lower stress levels and faster articulatory rates is a consequence of the earlier arrival of the final consonant control signal, which deflects the articulator away from its target trajectory. (See Daniloff & Hammarberg, 1973, and Hammarberg, 1976, for a somewhat more general assimilatory account of coarticulation.)

Recently, Fowler (1980, 1983) has offered an alternative perspective on coarticulation, based partly on earlier empirical and theoretical work of Kozhevnikov and Chistovich (1965), Ohman, (1966, 1967), and Perkell (1969). According to Fowler, consonants and vowels are *coproduced* rather than assimilated. Coproduction implies temporally overlapping gestures that are, for the most part, separate and context-independent events. Consider again our earlier example of /su/. Fowler would reject the usual assimilatory account according to which the lip-rounding feature of the vowel spreads to the /s/ segment; instead, she would argue that the entire lip-rounding gesture belongs exclusively to the /u/ segment and that the vowel simply overlaps in time with the fricative gesture. To a first approximation, the consonant and vowel segments are viewed as distinct, orthogonal, and noninterfering events.

An ancillary claim of the theory of coproduction is that consonants and vowels are fundamentally different kinds of events. Vowels involve relatively slow and continuous changes in overall vocal tract shape, whereas consonants are produced as abrupt local constrictions of the vocal tract. Ohman (1966, 1967) and Perkell (1969) suggested that (to a rough approximation) consonant constrictions are carried out by intrinsic tongue muscles that control tongue shape, whereas vowel production is regulated by extrinsic tongue muscles responsible for positioning the tongue body. By this view, consonant gestures are superimposed on a continuous underlying vowel trajectory. Not only are consonants typically briefer and more local events than vowels, they are also usually produced with much less acoustic energy.

Although the theory of coproduction is faced with some difficulties (Diehl, 1985), we think that the above characterization of the differences between vowels and consonants is theoretically useful. We suggest that, because of differences in the way they are produced, vowels and consonants have a quite different perceptual status. Vowel information is distributed over a longer temporal interval, at least the entire length of a syllable and perhaps beyond (Ohman, 1966). Owing to their greater degree of openness, vowels are usually much more intense than consonants. Finally, vowel information is more stable in the sense that vowel trajectories appear to be less perturbed by consonant gestures than vice versa. All this suggests that vowels may be more easily and consistently detected than consonants. Relatively speaking, vowels are islands of reliability in the speech stream.

We also claim that vowels provide a more or less stable (although not stationary) frame of reference for judging consonants. The acoustic correlates of a consonant are largely structured by the vowel context (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), and it is reasonable to suppose that reliable consonant identification demands prior or simultaneous evaluation of what Fowler (1983) calls the "vocalic frame."[2]

Support for the "islands of reliability" claim comes from several lines of research. Jenkins, Strange, and their colleagues (Jenkins, Strange, & Edman, 1983; Strange, Jenkins, & Johnson, 1983) found that listeners' identification of vowels in consonant-vowel-consonant syllables was little affected when between 50 and 65% of the syllable, including all of the central quasi steady state portion, was deleted. In a related experiment, Parker and Diehl (1984) showed that vowel recognition remained fairly accurate even when 90% of the syllable was deleted, leaving only a pitch period or two at the syllable margins. Moreover, most of the errors that occurred represented confusions with the spectrally most similar vowel category.[3]

These deletion studies did not directly compare recognition performance on vowels and consonants, but such a comparison was made by Winitz, Scheib, and Reeds (1972). When end portions of consonant-vowel syllables were removed, leaving just the initial release burst and aspiration, listeners were generally able to identify the vowel more accurately than the consonant. (See also Blumstein & Stevens, 1980, and Jusczyk, Smith, & Murphy, 1981.)

Finally, Tallal and Piercy (1974) found that, compared to normal control subjects, aphasic children show substantially greater deficits in the perception of consonants than in the perception of vowels.

Indirect support for our claim that consonant recognition may require prior or parallel evaluation of the vocalic frame is provided by two studies mentioned earlier. Recall that Liberman et al. (1952) and Mann and Repp (1980) demonstrated a clear effect of the following vowel on syllable-initial consonant identification. However, more direct evidence for a "vowel-first" recognition strategy comes from the present study. Our results indicate that a certain amount or proportion of the vowel trajectory must apparently be detected before consonant cues can be reliably interpreted. The longer the following vowel, the longer it takes to evaluate the vowel trajectory, and, thus, the greater are the consonant monitoring latencies.

At first glance, this account may seem implausible. If formant transitions are followed by a steady state vowel pattern, then any temporal portion of the post-transitional signal should be fully predictive of the rest. The problem with this analysis is that, unlike many stylized synthetic speech syllables, actual syllables rarely contain formant patterns that are truly steady state. Even disregarding the very rapid spectral change associated with the initial opening and final closing gestures, the vowel itself is a dynamic event with formant movement throughout its trajectory. Accordingly, it is reasonable to suppose that evaluation of the vocalic frame will take longer in the case of vowels with greater intrinsic duration. As we suggested, the relevant variable in vowel identification may be something like the proportion of the overall vowel trajectory that has elapsed.

As we noted, our hypothesis about the different perceptual status of vowels and consonants derives in part from Fowler's theory of coproduction, particularly, her claim that vowel and consonant gestures are fundamentally different kinds of events. It is worth considering whether our consonant monitoring results have any additional bearing on the issue of assimilation versus coproduction.

The overall pattern of our results is consistent with an assimilatory account of coarticulation, which assumes that phonetic segments are produced in a context-sensitive manner. Such context sensitivity in production would presumably reveal itself in the kind of context sensitivity that we found in perception. It would be a mistake, however, to conclude that the observed correlation between consonant monitoring time and following vowel duration is incompatible with the theory of coproduction. Borrowing from the work of Johannson (1950), Fowler and Smith (1986) described the recognition of coproduced consonants and vowels as a kind of perceptual "vector analysis." The separate but temporally overlapping consonant and vowel gestures both obviously contribute to the time-varying acoustic signal. The listener's task, according to Fowler and Smith, is to decompose the information in the signal into both a vowel and a consonant component. Once the comparatively large vowel component has been extracted from the signal, it can be "factored out," leaving only the information specific to the consonant gesture.

Following Fowler and Smith (1986), we will describe this in a slightly different way. By the theory of coproduction, the relatively brief consonant gesture is superimposed on the longer underlying vowel trajectory. Although the consonant is assumed to be orthogonal with respect to the vowel, it will nevertheless share a common vector of motion with the vocalic frame. The listener must identify this common vector (i. e., the vowel gesture) before he or she can recognize the consonant-specific vector of motion. We suggest that the most reliable way to ascertain the common vector is to have access to vowel trajectory information when the signal is relatively unperturbed by the presence of the consonant, that is, after the consonant-vowel formant transitions. For this reason, we think that the theory of coproduction (together with Fowler and Smith's corresponding perceptual theory of vector analysis) can neatly explain the correlation between consonant monitoring latency and following vowel duration.

There is, however, one aspect of our results that may be more naturally explained in terms of assimilation than in terms of coproduction. From an assimilatory perspective, the relatively low correlations between response time and vowel duration in the /b/ monitoring condition are perhaps attributable to a lower degree of context dependency among cues for labial stops. Alveolar consonants such as /d/ and velar consonants such as /g/ may be assumed by assimilationists (although not by coproductionists) to involve a substantial degree of mechanical interaction with tongue movements of the following vowel, while labial gestures are more or less independent of the vowel tongue movements. This greater degree of articulatory independence for labials may result in acoustic cues that are more nearly context independent than those of alveolar and velar consonants. If so, a vowel-first recognition strategy may be less efficient or necessary in the case of labial consonants. Further evidence that this is indeed the case is provided by the generally lower reaction times to labial targets.

In the final analysis, perceptual data alone are not likely to be decisive in the de-

[2] This account contrasts with the theory of consonant place perception advocated by Blumstein and Stevens (1979). These authors argued that there are invariant and distinctive place cues available to the listener in the first few 10s of milliseconds following stop release. If such context-independent cues are available, one would expect practiced subjects in a speeded classification task to use them to reduce response times to a minimum. That subjects rely on something more is suggested by the present results.

[3] That listeners can achieve reasonably high rates of vowel recognition for such stimuli means only that there is vowel information in the syllable margins. It does *not* mean that listeners ignore information from the full vowel trajectory when that information is available. Our present results suggest that even when subjects are performing a speeded classification task, they tend to rely on syllabic information beyond the consonant-vowel transitions.

bate between assimilationists and coproductionists (Diehl, 1985). As we have seen in the present case, each of the two theories of coarticulation has some unique advantages in describing certain aspects of the data, but, generally speaking, assimilation and coproduction appear to be about equally compatible with most of the basic facts of speech perception.

Although, as Fowler (1987) points out in her comments on this paper, additional experiments are needed before conclusions may be confidently drawn, we think the present results provide reasonable evidence for our stated hypothesis: consonant recognition depends on prior or simultaneous evaluation of the following vowel trajectory.

REFERENCES

BLUMSTEIN, S. E., & STEVENS, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, **66**, 1001–1017.

BLUMSTEIN, S. E., & STEVENS, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, **61**, 648–662.

CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, **19**, 141–177.

DANILOFF, R., & HAMMARBERG, R. (1973). On defining coarticulation. *Journal of Phonetics*, **1**, 239–248.

DIEHL, R. L. (1985). Coproduction and direct perception of phonetic segments: A critique. *Journal of Phonetics*, **13**, 61–66.

FOSS, D. J., & GERNSBACHER, M. A. (1983). Cracking the dual code: Toward a unitary model of phoneme identification. *Journal of Verbal Learning and Verbal Behavior* **22**, 609–632.

FOSS, D. J., HARWOOD, D. A., & BLANK, M. A. (1980). Deciphering decoding decisions: Data and devices. In R. A. Cole (Ed.) *Perception and production of fluent speech*. Hillsdale, NJ: Erlbaum.

FOWLER, C. A. (1980). Coarticulation and theories of extrinsic timing control. *Journal of Phonetics*, **8**, 113–133.

FOWLER, C. A. (1983). Converging sources of evidence on spoken and perceived rhythms of speech. *Journal of Experimental Psychology: General*, **112**, 386–412.

FOWLER, C. A. (1987). Perceivers as realists, talkers too: Commentary on Strange, Rakerd and Verbrugge, and Diehl et al. *Journal of Memory and Language*, **26**, 574–587.

FOWLER, C. A., & SMITH, M. R. (1986). Speech perception as vector analysis: An approach to the problem of invariance and segmentation. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123–136). Hillsdale, NJ: Erlbaum.

HAMMARBERG, R. (1976). The metaphysics of coarticulation. *Journal of Phonetics*, **4**, 353–363.

HOUSE, A. S. (1961). On vowel duration in English. *Journal of the Acoustical Society of America*, **33**, 1174–1178.

HOUSE, A. S., & FAIRBANKS, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, **25**, 105–113.

JENKINS, J. J., STRANGE, W., & EDMAN, T. R. (1983). Identification of vowels in 'vowelless' syllables. *Perception & Psychophysics*, **34**, 441–450.

JOHANSSON, G. (1950). *Configurations in event perception*. Uppsala, Sweden: Almqvist & Wiksell.

JUSCZYK, P. W., SMITH, L. B., & MURPHY, C. (1981). The perceptual classification of speech. *Perception & Psychophysics*, **30**, 10–23.

KOZHEVNIKOV, V. A., & CHISTOVICH, L. A. (1965). *Speech: Articulation and perception* (Nauka, Moscow and Leningrad). [Trans. Clearinghouse for Federal Scientific and Technical Information, 30, 543. U.S. Department of Commerce, Washington, DC.]

LEHISTE, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

LEHISTE, I., & PETERSON, G. E. (1961). Transitions, glides, and diphthongs. *Journal of the Acoustic Society of America*, 33, 268–277.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431–461.

LIBERMAN, A. M., DELATTRE, P. C., & COOPER, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, **65**, 497–516.

LIBERMAN, A. M., DELATTRE, P. C., COOPER, F. S., & GERSTMAN, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, **68**, (Whole No. 379) 1–13.

LINDBLOM, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, **35**, 1773–1781.

LINDBLOM, B. (1967). Vowel duration and a model of lip-mandible coordination. Royal Institute of Technology, Speech Transmission Laboratory, Quarterly Progress and Status Rep. 4.

LISKER, L. (1974). On 'explaining' vowel duration variation. *Glossa*, **8**, 223–246.

MANN, V. A., & REPP, B. H. (1980). Influence of vocalic context on perception of the /ʃ/–/s/ distinction. *Perception & Psychophysics*, **28**, 213–228.

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

OHMAN, S. E. G. (1967). Numerical model of coarticulation. *Journal of the Acoustic Society of America*, **41**, 310–328.

PARKER, E. M., & DIEHL, R. L. (1984). Identifying vowels in CVC syllables: Effects of inserting silence and noise. *Perception & Psychophysics*, **36**, 369–380.

PERKELL, J. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Cambridge, MA: MIT Press.

PETERSON, G. E., & LEHISTE, E. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, **32**, 693–703.

STRANGE, W., JENKINS, J. J., & JOHNSON, T. L. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America*, **74**, 695–705.

SUSSMAN, H. M., MACNEILAGE, P. F., & HANSON, R. J. (1973). Labial and mandibular movement dynamics during production of bilabial stop consonants: Preliminary observations. *Journal of Speech & Hearing Research*, 16, 397–420.

TALLAL, P., & PIERCY, M. (1974). Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia*, **12**, 83–93.

WINITZ, H., SCHEIB, M. E., & REEDS, J. A. (1972). Identification of stops and vowels from the burst portion of /p,t,k/ isolated from conversational speech. *Journal of the Acoustical Society of America*, **51**, 1309–1317.

(Received January 26, 1987)
(Revision received May 27, 1987)