Phonetica
Editor: K. Kohler, Kiel

Separatum
Publisher: S. Karger AG, Basel
Printed in Switzerland

Phonetica 1989;46:97–116

Dynamic Information in the Identification and Discrimination of Vowels[1]

Robert Allen Fox

Division of Speech and Hearing Science, Ohio State University, Columbus, Ohio, USA

Abstract. The dynamic theory of vowel perception emphasizes the importance of the consonant transitions in CVC syllables to the identification of the vowel itself. However, the dynamic theory is relatively vague in terms of what perceptual processes may be involved and it says nothing about the role of auditory and/or phonetic memory for such 'dynamic events'. The present study created three synthetic [ɪ]–[ɛ] continua consisting of a [bVb] full-syllable continuum, a silent-center continuum (which had over 72% of the medial vowel replaced by silence) and an isolated vowel continuum to specifically examine (1) the use of auditory and phonetic memory for dynamic acoustic cues and (2) the ability of listeners to track the trajectories of very brief formant transitions. Experiment 1 demonstrated that there were no significant differences among the three continua in terms of listener identifications but that the silent-center tokens demonstrated significantly lower within-category discriminations, perhaps because of degraded auditory representation. Experiment 2 required listeners to make cross-continuum vowel discriminations and showed that some degree of acoustic similarity was important – listeners were particularly poor at discriminating between the silent-center and vowel-only tokens. This suggests that listeners were not able to make discriminations on the basis of abstract vowel labels alone. Experiment 3 demonstrated that listeners could make quite accurate vowel identifications even when as little as 1 pitch period of acoustic energy signaled the consonant transitions in the silent-center tokens. Experiment 4 showed that listeners' identifications of the silent-center stimuli were based on the formant changes in the transitions rather than their endpoint frequencies.

Vowel quality (particularly for monophthongs) is traditionally described as being determined on the basis of the frequencies of formants 1 and 2, (F_1, F_2) during a relatively steady-state portion of the vowel [Joos, 1948; see a brief history and critique of this viewpoint in Jenkins, 1987]. This steady-state section of the vowel represents the time at which the formants have

[1] A preliminary version of this paper was presented at the 1988 Regional Meeting of the Chicago Linguistic Society, Chicago, Ill., April 28–30, 1988.

reached their 'target' positions. F_1 is inversely correlated with the tongue height distinction while F_2 (or F_2–F_1 difference) is directly correlated with the tongue advancement distinction [Ladefoged, 1982; Shriberg and Kent, 1985]. Several perceptual studies utilizing multidimensional scaling [Singh and Woods, 1971; Shepard, 1972; Terbeek, 1977; Fox, 1982, 1983, 1985a] have demonstrated that such target formant values can successfully account for listeners' perception of vowel quality.

However, any strict version of such a 'target theory' in which vowels are identified in terms of a set of steady-state formant values is difficult to maintain in the face of several different lines of research. First, several researchers – particularly Winifred Strange and her colleagues [Strange et al., 1976; Verbrugge et al., 1976] – have found that vovels produced in a CVC context were identified as well as or better than isolated vowels, despite the fact that the consonantal context introduces coarticulatory variations in the vowel's formant frequencies [Lindblom, 1963; Stevens and House, 1963]. Such formant variations would tend to disrupt the formant targets and should, therefore, *reduce* vowel identification scores. Following the publication of these studies, a number of papers appeared which demonstrated that isolated vowels could be identified as well as vowels in context [Macchi, 1980; Diehl et al., 1981; Assmann et al., 1982] and suggested that the failure to find good identification of isolated vowels resulted from a failure to control for factors such as orthographic interference and dialect [but see Rakerd et al., 1984]. However, no study has found that isolated vowels are consistently identified more accurately than vowels in a consonantal context as might be expected from a strict vowel target theory [Strange, 1987].

Second, although it has long been recognized that inherent spectral change was important for the identification of diphthongs [Joos, 1948; Gay, 1970], studies by Assmann et al. [1982] and Nearey and Assmann [1986] support the contention that even in the case of isolated 'monophthongal' vowels, inherent spectral change (i.e., changes in formant frequencies over time associated with specific vowel qualities and independent of consonant context) may play a significant role in vowel identification. For example, Nearey and Assmann [1986] presented listeners with two 30-ms acoustic sections gated from naturally produced vowels (extracted at positions corresponding to 24 and 64% of the total duration). Listeners were able to identify vowel quality on the basis of these two short sections as well as when the full vowel was heard – provided that the two sections were presented in their natural order. If the first section was repeated or the order of the sections reversed, identification scores decreased significantly.

A third, even more dramatic, set of data involves the identification of the medial vowels in so-called 'silent-center' speech tokens [Jenkins et al.,1983; Strange et al., 1983; Strange, 1987]. In these studies, they demonstrated that vowels of naturally produced CVC syllables (such as [bab] and [bib] could be reliably identified even when most of the medial vowel (up to 65%) had been replaced by silence (the silent-center stimuli). Furthermore, these identification rates were comparable to, if not lower than, the error rates obtained when the medial vowel alone was presented. These data are problematic for the target theory since there is, in fact, no steady-state vowel whatsoever

in the silent-center tokens, only a consonant transition from the initial consonant into the medial vowel (the CV transition) or the transition from the medial vowel into the final consonant (the VC transition).

Strange and Jenkins and their colleagues have formulated what they have termed a 'dynamic theory' of vowel perception [Strange, 1987]. This theory suggests that one important source of information for vowel quality is the formant movements into and out of the vocalic nucleus of the syllable. Such dynamic information reflects the changes in vocal tract shape caused by movements of the articulators and cannot be easily described in terms of a set of steady-state formant 'targets' or frequencies at any single point in time during the production of the vowel. In their view, vowels (particularly within consonantal contexts) may be best considered in terms of *articulatory gestures* or *events* and their corresponding acoustic consequences [Fowler, 1977, 1980, 1983, 1987; Fowler et al., 1980; Browman and Goldstein, 1986, in press]. From this point of view, the articulatory gesture associated with the vowel would overlap, to some extent, with the articulatory gesture of the consonant. The resulting acoustic information for the vowel would then be 'specified in the temporospectral structure of the entire acoustic syllable' [Strange, 1987, p. 556].

Although the dynamic theory is quite attractive, at this point it is relatively vague in terms of specific perceptual processes involved in identifying vowels from dynamic acoustic changes or the particular kinds of acoustic information which may be crucial to the listener's judgments. For example, in discussing the identification of a 'hybrid' silent-center token (which combines CV and VC transitions from different speakers), Verbrugge and Rakerd [1986; Rakerd and Verbrugge, 1988] suggest that listeners integrate the relevant vowel information from the two separate transitions and define some type of function (or relation) over them. The nature of the function itself is left unspecified except that it is compatible with the articulatory event hypothesis [Fowler, 1980, 1987; Fowler et al., 1980], which assumes that 'the early and late stages of an event should bear a principled relation to one another' [Verbrugge and Rakerd, 1986, p. 52]. A somewhat more specific hypothesis about the perception of coproduced consonant and vowel segments was suggested by Fowler and Smith [1986], which they termed 'vector analysis' [based to some extent on work by Johansson, 1973]. In vector analysis, it is assumed that the relatively brief consonant gesture is superimposed on the longer vowel gesture. The listener must identify the vowel gesture (and establish the common vector) before he/she can recognize the consonant gesture. However, even this point of view provides few details about the *acoustic* parameters that might be used in the perceptual process.

We also know little about the role of auditory and/or phonetic memory for such dynamic events in such event-driven models which tend to concentrate upon the organization of the articulatory gestures rather than their acoustic consequences. In particular, vowels – perhaps because of their relatively steady-state nature – are generally considered to create relatively strong traces in auditory memory [Pisoni, 1973, 1975; Crowder, 1981, 1982]. The information contained in auditory and/or phonetic memory for such events is critical

since memorial representations play such an important role in many vowel perception models [Fujisaki and Kawashima, 1971; Repp et al., 1979; Crowder, 1981; Crowder, 1982] and are of great interest in accounting for the effects of phonetic context upon the identification of vowels [Sawusch and Nusbaum, 1979; Crowder and Repp, 1984; Fox, 1985b, c]. However, the role of auditory memory may not be as significant for the perception of *dynamic* events in vowel identification. For example, Tartter [1981] found that although listeners could identify vowels reliably from a 40-ms CV transition, discrimination of those transitions was close to categorical. Studies by Sachs [1969] as well as Sawusch et al. [1980] have also shown that vowels are perceived somewhat more categorically within a word context than in an isolated condition, which might suggest a reduction of auditory information for dynamic events relevant to vowel recognition.

The present paper will describe four different perceptual experiments designed to investigate the role of formant transitions in the process of vowel perception. The first two experiments involve the identification and discrimination of silent-center vowels as compared to full CVC tokens and isolated vowels. The second two experiments consider the amount of transition needed by listeners to accurately identify medial vowels removed from a set of silent-center tokens and the nature of formant trajectory tracking. The stimuli used in each of the experiments were created from a synthetic *bib-beb* continuum (i.e., a [ɪ–ɛ] vowel continuum in the context [b__b]). The use of these continua will allow us to examine finer vowel quality distinctions than have other relevant studies in the available literature. In addition, the use of synthetic speech will severely reduce the number of acoustic parameter variations present in the stimulus set compared to the other silent-center vowel studies, which have used waveform-edited natural speech exclusively.

Experiment 1

Methodology

Stimulus Material

A natural sounding, 7-step *bib-beb* continuum was created using a cascade/parallel software synthesizer [Klatt, 1980]. Each of the seven tokens of each continuum was 300 ms in total duration. The total length of the vowel (including transitions) was 255 ms. The steady-state portion of the medial vowel was 195 ms in duration [appropriate for these two short, lax vowels in English, Peterson and Lehiste, 1960] while both the CV and VC transitions were 30 ms in duration. The VC transition was a mirror image of the CV transition.There was a short 5-ms burst frame corresponding to a stop closure release with an aperiodic noise source at the start of each token. At the end of each token was a low-frequency, low-energy murmur for a final, unreleased 'voiced stop', which was 40 ms in duration.The murmur was synthesized using a sinusoidal voicing source with the formant frequency values unchanged from the VC transition offset values.

The formant frequency values used in the synthesis of the tokens are shown in table 1 [based on Klatt, 1978]. They include the formant frequency values at the start of the burst frame, the onset and offset of the consonant transitions (at which point normal voicing either starts or stops) and the formant frequencies of the steady-state portion of the vowel. These formant frequency values were calculated according to the formulae and target values developed by Klatt [1978]. In particular, the formant frequency values for the transition onsets (and end of VC transition offset) are calculated as follows *(vowel target* equals the formant values for the following steady-state vowel):

Table 1. Formant frequency values used in the synthesis of the original *bib-beb* continuum

Stimulus	Release burst			Transition onset/offset			Steady-state vowel		
	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
1	285	1,297	2,267	370	1,494	2,393	400	1,800	2,570
2	291	1,291	2,266	381	1,481	2,381	422	1,780	2,558
3	296	1,284	2,265	391	1,468	2,380	443	1,760	2,547
4	302	1,277	2,264	403	1,454	2,378	465	1,740	2,535
5	307	1,264	2,263	414	1,441	2,376	487	1,720	2,523
6	312	1,258	2,262	424	1,428	2,374	508	1,700	2,511
7	318	1,251	2,261	435	1,415	2,372	530	1,680	2,500

Shown are the F_1, F_2, and F_3 frequency values at the consonant burst, the onset and offset of the consonant transitions and the steady-value vowel.

$F_1 = 340+0.50 \times$ [vowel target $- 340$]
$F_2 = 900+0.66 \times$ [vowel target $- 900$]
$F_3 = 2{,}350+0.15 \times$ [vowel target $- 2{,}350$]

The frequencies of F_1, F_2, and F_3 at the burst frame are calculated in a similar manner using a slightly different formula. These formulae allow the formant pattern of the following vowel to determine, in part, the acoustic characteristics of the preceding (and following) stop consonant. Thus, the synthetic tokens also include both a vowel and consonant component in the consonant transition itself.

The transitions for F_1, F_2, and F_3 were synthesized on the basis of linear interpolation between the onset or offset frequency of the transition and the steady-state portion of the vowel. The fundamental frequency (F_0) for each token began at 115 Hz where it remained for 100 ms. F_0 then fell linearly to 100 Hz at the end of the token. The frame rate for all acoustic changes in the synthesis was 5 ms.

Note that, similar to humanly produced tokens, these stimuli have consonant transitions which are affected both by the nature of the stop itself (i.e., the formant values are typical of a bilabial place of articulation) and the adjacent vowel (i.e., the starting transitions are affected by the frequencies of the eventual steady-state targets). The unedited stimulus continuum will be called the *full-syllable tokens*.

Two other stimulus types were created using the same basic continuum. One set of tokens, called the *silent-center tokens*, was created in which most of the medial vowel was removed. In particular, all of the vowel except for four pitch periods (approximately 35 ms) of the CV and VC transitions was replaced by silence. Four pitch periods corresponded to approximately 35 ms of the initial part of the vowel, while four pitch periods corresponded to approximately 38 ms of the ending part of the vowel. In each case, these four pitch periods included all of the CV or VC formant transitions. Removal of these pitch periods resulted in approximately 72% of the vowel being replaced by silence in the token – comparable to the most severe reductions of Strange et al [1983]. The silent-center tokens were 300 ms in total duration. Spectrographic analysis confirmed that the silent-center tokens contained no steady-state information.

The third set of tokens, called the *vowel-only tokens*, represented the steady-state medial vowel which had been removed from the silent-center tokens. These tokens contained no transitional information. This was also confirmed by spectrographic analysis. The total duration of the vowel-only tokens was approximately 185 ms. Figure 1 shows a spectrogram of the *bib* endpoint for each of the three stimulus types.

For each of these three sets of stimulus tokens, three audio tapes were constructed: one single-item identification test and two discrimination tests. The identification tape contained a random sequence of

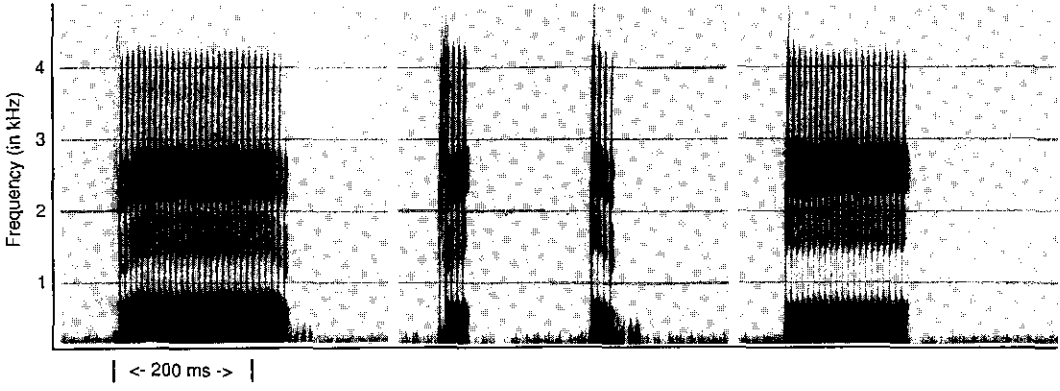

Fig. 1. Spectrogram of [ɪ] endpoint of the three different stimulus continua. From left to right, these represent the full token, silent-center, and vowel-only stimuli.

112 stimuli (16 repetitions of each of the 7 continuum steps) with an interstimulus interval of 3 s. Preceding the test stimuli was a short practice of 10 items. The practice items were randomly selected examples of the test tokens to follow. The discrimination tapes each contained a randomized set of stimulus pairs (as used in the AX discrimination procedure). Each of the discrimination tapes included identical pairs – each stimulus step paired with itself (6 each) – and a set of two-step 'different' pairs (6 repetitions of 1/3, 3/5 and 5/7 in both orders). The different pairs included both within-category (1/3 and 5/7) and between-category (3/5) pairs. (The stimulus set in the first experiment also included 1-step discrimination pairs [1/2, 3/4, 4/5, and 6/7 in both orders]. However, given the variability in the location of the phoneme boundary among listeners, none of these pairs accurately represent a 'between-category' stimulus pair. Therefore, we will present and discuss only the 2-step discriminations.)

There were two separate discrimination tapes for each stimulus group. In one tape the interval between stimulus tokens in a pair (the intrapair interval) was 500 ms, in the second tape, the interval was 2,000 ms. Following Repp et al. [1979] and Crowder [1981, 1982], the longer intrapair interval was included to produce a condition in which the amount of auditory memory for the first token of a pair might be reduced. This experimental manipulation tends to make vowel discrimination somewhat more 'categorical'. The same random orders for the single identification tape and the two discrimination tapes were used for each the three stimulus continua to reduce extraneous between-subject factors.

Listeners

There were 24 listeners. All were undergraduate or graduate students at the Ohio State University and were native speakers of American English. None had any known speech or hearing impairment and all were unfamiliar with the goal of the experiment.

Procedure

The listeners were randomly divided into three groups of 8 listeners each. Each listener was assigned to either the full token, silent-center, or vowel-only group. The listeners in each group completed both an identification test and a discrimination test. During the identification test listeners heard the identification tape and identified the tokens as having the vowel [ɪ] or [ɛ] by circling either the responses *bib* or *beb* on their response sheets. The instructions for the silent-center group explained that the medial vowel had been removed from either *bib* or *beb* and had been replaced with silence. They were to identify the original token (and therefore vowel). The instructions for the vowel-only group were similar in that they stated

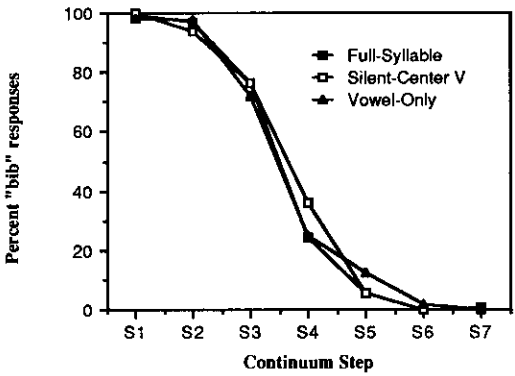

Fig. 2. Identification functions (percent [ɪ]responses) for the full-syllable, silent-center, and isolated vowel tokens.

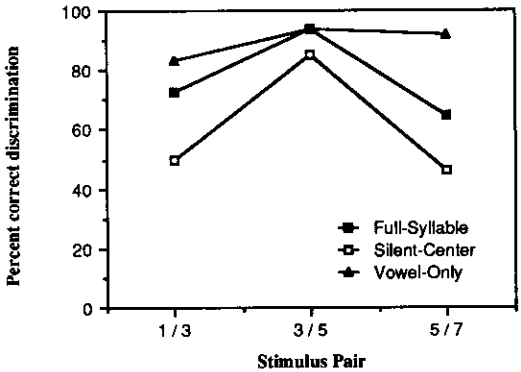

Fig. 3. Mean discrimination scores (percent correct) for the full word, silent-center, and isolated vowel tokens with a 500-ms interpair interval.

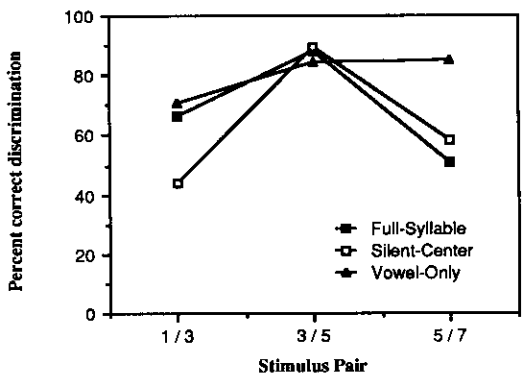

Fig. 4. Mean discrimination scores (percent correct) for the full-syllable, silent-center, and isolated vowel tokens with a 2,000-ms interpair interval.

that the vowels they were to hear had been removed from the tokens *bib* or *beb*.

The discrimination tests used the standard AX paradigm. Listeners heard sequences of stimulus pairs and were instructed to indicate whether the first stimulus in each pair was the same (i.e., identical) as or different from the second stimulus. One half of each group completed the 500-ms discrimination tape before the 2,000-ms tape while the other half heard the tapes in the reverse order. For all three groups the identification test was completed before the discrimination tests.

Results and Dicussion

The identification results obtained for each of the three stimulus sets are shown in figure 2. The identification functions for the three groups are remarkably similar. Phoneme boundaries for each separate listener were calculated (using linear interpolation) to determine the 50% *bib-beb* cross-over point. These phoneme boundaries were then analyzed using a one-way analysis of variance with the factor stimulus group (full-syllable, silent-center, and vowel-only). There was no significant effect of stimulus group [$F(2, 21) = 0.91$, $p > 0.40$].

The pattern of discriminations is not so similar across groups. Shown in figure 3 are the discrimination responses (in terms of percent correct) when the interval duration between tokens in an AX pair was 500 ms. Figure 4 shows the discrimination responses when the interval was 2,000 ms. Several things can be noted about these responses. First, the between-category dis-

Table 2. Two-step discrimination results from experiment 1

	Interval duration					
	500 ms			2,000 ms		
	within-category	between-category	difference	within-category	between-category	difference
Full-syllable	68.7	93.8	25.1	58.9	88.5	29.6
Silent-center	47.9	85.4	37.5	51.0	89.6	38.6
Vowel-only	87.5	93.8	6.3	78.1	84.4	6.3

The numbers in the table represent percent correct discrimination responses for the "different" pairs. The "difference" column indicates the percentage difference between within-category and between-category responses.

criminations are more accurate, in general, than are the within-category discriminations. This is a pattern commonly obtained in discrimination tests involving speech stimuli that are, to some extent, categorically perceived [Repp, 1982]. Second, the within-category scores are much higher for the vowel-only stimuli than for either the silent-center or the full-syllable tokens. Therefore, the discriminations involving the silent-center tokens have a greater difference between the between- and within-category discriminations than do the vowel-only tokens. This would suggest that the silent-center tokens are more categorically perceived than are the vowel-only tokens. The full-syllable tokens seem to be midway between the silent-center and vowel-only results in the 500-ms condition, although this pattern changes somewhat during the 2,000-ms condition. Similar decrements in the discrimination of vowels in segmental contexts are compatible with the results obtained by Sawusch et al. [1980], Sachs [1969], and Stevens [1968].

To concentrate on the difference between the within- and between-category discriminations, table 2 shows the mean discrimination results for all three groups with the two within-category responses (1/3 and 3/5) collapsed together. Note that the differences between the within-category responses are much greater for the silent-center group than for the vowel-only group. Large differences are often associated with more 'categorically perceived' stimuli such as consonants, whereas small differences are associated with stimuli that are more 'continuously perceived'.

The percentage of correct responses from each listener was next analyzed using a three-way analysis of variance with the factors stimulus group (full-syllable, silent-center, and vowel-only), discrimination type (within- vs. between-category), and interval duration (500 vs. 2,000 ms). Since the means of proportional data (such as percentages) are correlated with the variances, these percentages were arcsine-transformed [using Studebaker's, 1985, rationalized arc-

sine method] prior to analysis. The analysis of variance indicated that there was a significant effect of stimulus group [$F(2, 86) = 6.31$, $p < 0.003$]. A Duncan multiple-range test showed that there were fewer correct responses for the silent-center data (68.5%) than for either the vowel-only data (86.0%) or the full-syllable data (77.4%). The responses to the vowel-only and full-syllable groups were not significantly different.

There was also a significant effect of discrimination type [$F(1, 86) = 35.21$, $p < 0.001$] demonstrating that there were more correct between-category discriminations (89.3%) than within-category discriminations (65.3%). Contrary to the results of Repp et al. [1979], there was no significant main effect of interval duration [$F(1, 86) = 2.08$, $p < 0.15$], although there was a tendency for the listeners to have fewer correct responses when the interval was long (75.0%) than when it was short (79.5%). Of the interaction effects, only the group X discrimination type interaction was significant [$F(2, 86) = 5.01$), $p < 0.001$]. This interaction effect stems from the fact that although the mean difference between the within- and between-category discriminations for the full-syllable and silent-center data were 27.4 and 38.1%, respectively, the mean difference for the vowel-only data was 6.3%.

As stated above, it seems as though the silent-center tokens are perceived 'more categorically' than are the vowel-only tokens. In general, categorical perception is the phenomenon in which the discrimination scores can be predicted from the identification scores [Repp, 1982]. It is normally the case that these predictions are better for consonant stimuli (especially stops) than for vowels. Usually for vowels, the actual discriminations are much better than are the predicted discriminations. Compare the predicted discrimination scores [shown in table 3 and based on the identification results, Pollack and Pisoni, 1971] with the actual discrimination scores. Each group is more accurate in making between-category distinctions than would be predicted. However, the within-category predictions are very accurate for the silent-center tokens. On the other hand, the within-category discriminations for the vowel-only data are considerably better than would be predicted. The actual full-syllable discriminations are better than the predictions, but not as much better than the vowel-only tokens. This is despite the fact that the full-syllable tokens have more acoustic data upon which to make a discrimination than do either the vowel-only or silent-center tokens.

Table 3. Predicted percent correct discrimination scores for experiment 1

	Within-category	Between-category
Full-syllable	51.7	74.2
Silent-center	51.9	73.1
Vowel-only	52.5	67.5

These predictions were based on the identification responses in terms of formulae suggested by Pollack and Pisoni [1971]. Predicted within-category discrimination scores are mean predicted scores for steps 1 and 3 and steps 5 and 7.

The near-chance level of within-category discriminations obtained for the silent-center tokens is a result compatible with one of the more common explanations of the distinction between within- and between-category responses. This explanation suggests that beweeen-category discriminations re-

sult from comparisons primarily in terms of internal phonetic labels (with this information stored in phonetic or working memory) while within-category discriminations may utilize auditory information which is subject to relatively rapid decay [Pisoni, 1973, 1975; Repp, 1982]. It is normally assumed that a speech sound with relatively steady-state acoustic characteristics (such as a vowel) will leave a relatively strong auditory memory trace even following categorization (identification). On the other hand, a sound characterized by rapidly changing acoustic information (such as a stop) normally does not [Repp, 1982]. Given the full-syllable discriminations, it is also evident that the presence of the consonant transitions in the full token may decrease the listener's ability to make within-category discriminations, perhaps by reducing the amount of auditory memory available for the vowel.

The present study does not directly compare the role of auditory memory in contextual contrast as opposed to its role in vowel discrimination nor does it address the possible distinction between a fast-decaying auditory memory [as proposed by Crowder] and a slower-decaying memory [Massaro, 1975; see discussion in Repp, 1982]. Rather, the present study demonstrates a possible difference in the memorial representation of dynamic as opposed to static vowel information. Further research will examine these other questions in greater detail.

Experiment 2

The next experiment is an attempt to *directly* compare the perceptual 'images' (e.g., memorial representations) of silent-center tokens with both the full-syllable and vowel-only tokens. In this experiment, listerners completed an AX discrimination task with *cross-continuum* stimulus pairs. That is, one of the tokens in each stimulus pair was from one continuum while the second token was from another continuum. This should give insight into the comparability of the vowel qualities extracted in the different stimulus conditions.

Methodology

Stimuli

The stimuli were constructed using the tokens described above. Three different stets of discrimination tapes were constructed. Each set of tapes compared two different stimulus types: full-syllable vs. silent-center, full-syllable vs. vowel-only and silent-center vs. vowel-only. That is, one of the tokens in each stimulus pair was from one stimulus group (e.g., full-syllable) while the second token was from another stimulus group (e.g., silent-center). Each set of discrimination tapes was composed of two different tapes differing in terms of stimulus order (e.g., tape 1a had full-syllable tokens followed by silent-center tokens, while tape 1b had the silent-center tokens followed by full-syllable tokens). Each tape contained 100 stimulus pairs: 10 repetitions of the 'same' pairs (using continua steps 1, 3, 5, and 7) and 10 repetitions each of the 2-step 'different' pairs in both orders (e.g., 1/3, 3/1, 3/5, 5/3, 5/7, 7/5). The intrapair interval was 500 ms for all tapes. Note that the correct response for a 'same' pair was a judgment that the two simuli had the same vowel and not a judgment of being physically identical. The correct response to a 'different' pair was a judgment that the two stimuli had different vowels.

Listeners

There were 30 listeners. All were undergraduate and graduate students at the Ohio State University who were native speakers of American English with no known speech or hearing impairment; all were unfamiliar with the goals of the experiment.

Procedure

The 30 listeners were randomly assigned to one of three groups. Within each group, 5 listeners heard the two tapes in one order (e.g., 1a/1b) while the other 5 listeners heard the tapes in the reverse order (e.g., 1b/1a).

Results and Discussion

Figure 5 shows the percent correct responses for the 'different' pairs. The same pattern of more accurate between-category discriminations is obtained, although the percentage of correct scores is somewhat lower than was obtained in experiment 1. These responses were analyzed using a two-way analysis of variance with the factors-continua combinations (full-syllable/silent-center, full-syllable/vowel-only, and silent-center/vowel-only) and discrimination type (between-category and within-category). As in experiment 1 these ratio scores were arc-sine-transformed before analysis [Studebaker, 1985].

The data from one of the listeners in the full-syllable/silent-center group had to be eliminated because she did not complete the test in the required manner. The cells are, therefore, unbalanced, which accounts for the unexpected degrees of freedom. The General Linear Model procedure from SAS [Ray, 1982] was used to analyze these data.

There was a significant main effect of continua combination [$F(2, 52) = 4.29$, $p < 0.02$]. A Duncan multiple-range rest showed that the silent-center/vowel-only condition had significantly fewer correct responses (46.5%) than did either the full-syllable/silent-center (55.6%) or full-syllable/vowel-only (53.5%) conditions (at the 0.05 level). There was also a significant main effect of discrimination type [$F(1, 52) = 196.33$, $p < 0.001$], showing that there were fewer correct responses to the within-category stimuli (33.4%) than the between-category stimuli (70.8%). There was also a significant condition X discrimination type interaction [$F(2, 52) = 11.65$, $p < 0.001$]. This result was obtained because the difference between the within- and between-category responses for the silent-center/vowel-only condition (20.6%) was significantly less than the differences for the full-syllable/silent-center (48.3%) and full-syllable/vowel-only (48.0%) conditions.

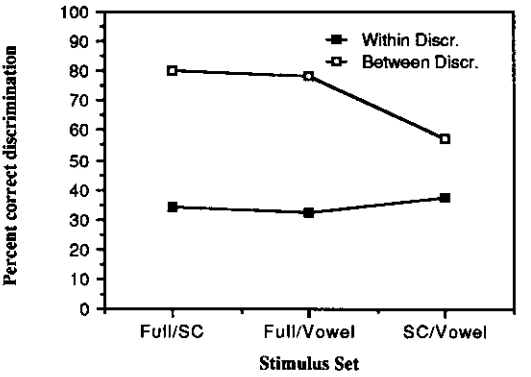

Fig. 5. Mean cross-continuum discrimination scores for 'different' pairs (percent correct). SC = Silent center.

The results for the 'same' comparisons are shown in figure 6. Again, a two-way analysis of variance using the factors condition and stimulus pair was done on the arc-sine-transformed data. There was a significant main effect of condition [$F(2, 104) = 17.18$, $p < 0.001$]. A Duncan multiple-range test showed that the percentage of correct responses was greater for the full-syllable/silent-center (85.1%) and the full-syllable/vowel-only (90.4%) conditions than for the silent-center/vowel-only condition (72.3%).

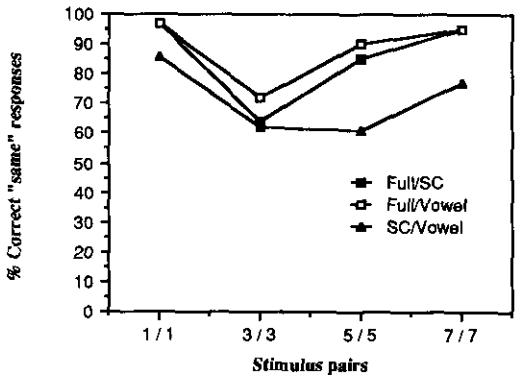

Fig. 6. Mean cross-continuum discrimination scores for 'same' pairs (percent correct). SC = Silent center.

There was also a significant main effect of stimulus pair [$F(3, 104) = 22.6$, $p < 0.01$]. A Duncan multiple-range test showed that the percentage of correct responses was significantly greater (at the 0.05 level) for both the 1/1 (94.1%) and 7/7 (89.6%) pairs than for the 3/3 (67.3%) and 5/5 (79.3%) pairs. In addition the difference between the 3/3 and 5/5 pairs was significant (at the 0.05 level). These data show that when the tokens represent the endpoints of a continuum, and are thus relatively unambiguous, it is easier for listeners to compare vowel quality. However, when the vowel is somewhat ambiguous, listeners find the comparison much more difficult – perhaps because they have categorized the two vowels differently.

These results suggest that whatever percepts or cognitive representation are being compared in the cross-continuum discrimination task, the representations corresponding to the silent-center tokens are relatively different from those corresponding to the vowel-only tokens. While this is not unexpected in the within-category discriminations, one would not necessarily expect the significant decline in the accuracy of the between-category responses, since these responses may have been based on category labels following vowel identification [assuming the dual-code model of Fujisaki and Kawashima, 1971; Pisoni, 1975; but see Crowder, 1981]. The pattern of discriminations indicates differences in the memory representations of vowel quality between the continua and supports the contention that listeners are able to make vowel comparisons based on the transitions *or* the steady-state vowels, but that a single abstract, categorical percept is not used.

Experiment 3

Although Strange and colleagues have emphasized the importance of formant movement in the transitions themselves, how much formant transition is necessary? Are there observable and significant differences in the identification responses as the amount of formant transition is reduced? Such information might give insight into the calculation of formant trajectories and, possibly, how listeners estimate hypothetical vowel targets. The question addressed by experiment 3 is simply, how little transitional information is needed for listeners to make reliable identifications of vowels?

Methodology

Stimuli

The stimuli for experiment 3 were four seven-step *bib-beb* silent-center continua. The basic continuum was the silent-center continuum used in experiments 1 and 2. The formant transitions in this continuum were contained in four glottal pulses in both

the CV and VC portions of the tokens. Three other continua were created using a waveform editor which had 1, 2, or 3 pitch periods in the transitional portions of the token (cut progressively farther away from the vowel target). All waveforms were cut at a zero crossing. The overall length of all tokens was maintained at 300 ms (including stop burst and final stop murmur) although the length of transitions and duration of the medial silence covaried with the number of pitch periods included.

A single identification tape was made. Since in the original continuum steps 1 and 2 were consistently identified as *bib* (100% and 93.7%, respectively) and steps 6 and 7 were consistently identified as *beb* (100% and 99.2%, respectively), only steps 2–6 of the four continua were used in the identification test. The tape contained 10 examples of each of the five steps of each continua. There were 10 practice items before the start of the identification test itself and the practice set included at least two tokens from each of the continua.

Listeners

There were 10 listeners. All were undergraduate or graduate students at the Ohio State University and were native American English speakers with no known speech or hearing impairment and all were unfamiliar with the goal of the experiment.

Procedure

The subjects were given the same basic instructions as were given to the silent-center group in the identification test of experiment 1. They had to identify each token as being an edited version of either *bib* or *beb*. They were warned that although all of the tokens had been made by removing the vowel from either *bib* or *beb*, some might be very difficult to identify since so much of the vowel had been removed. They were encouraged to guess if unsure of their answer.

Results and Discussion

The results of experiment 3 are shown in figure 7. This figure shows the identification functions obtained for each of the four continua plotted in terms of percent *bib* responses. Note that the functions, even for the 1- and 2-pulse transitions, are remarkably similar. For all four continua, steps 2 and 3 are identified as *bib* more than 70% of the time. For the 2-, 3-, and 4-pulse continua, steps 5 and 6 are identified as *beb* more than 70% of the time. The slope for the 1-pulse continuum is flatter than for the other three continua (in general, the identification function becomes flatter as the transitions get progressively shorter), but given the little amount of relevant acoustic information available, one might have expected chance-level responses for all steps.

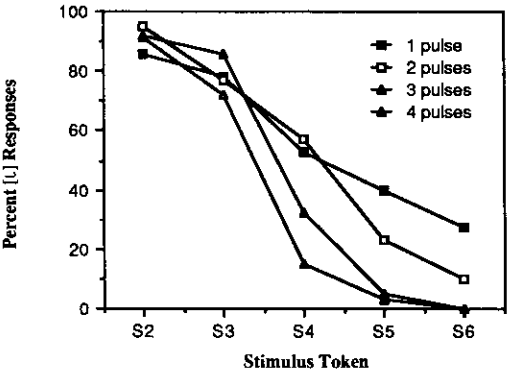

Fig. 7. Identification functions (percent [ɪ] responses) for the 1-, 2-, 3-, and 4-pulse silent-center stimuli.

There does seem to be at least one significant trend in the data, namely, that as the number of pitch periods included in the stimulus decreases, the phoneme boundary moves toward the *beb* end of the continuum. To examine this trend, the number of *bib* responses given by each listener to steps 3, 4, and 5 (the more ambiguous tokens near the phoneme boundary) for each continuum were calculated. (Phoneme boundaries were not used in the analysis because they often could not be reliably calculated

Table 4. Formant frequencies (in Hz) of CV transitions after 1, 2, 3, or 4 glottal pulses for steps 2–6 of the silent-center tokens used in experiment 3 (these represent the formant transition "endpoints").

Vowel continuum step	F_1	F_2	F_3
After 1 glottal pulse			
Step 2	391	1,416	2,427
Step 3	404	1,406	2,423
Step 4	420	1,396	2,418
Step 5	433	1,381	2,414
Step 6	446	1,372	2,409
After 2 glottal pulses			
Step 2	402	1,542	2,472
Step 3	417	1,529	2,466
Step 4	435	1,515	2,459
Step 5	451	1,499	2,452
Step 6	467	1,485	2,444
After 3 glottal pulses			
Step 2	413	1,668	2,518
Step 3	431	1,651	2,509
Step 4	451	1,634	2,499
Step 5	470	1,616	2,489
Step 6	489	1,599	2,480
After 4 glottal pulses			
Step 2	422	1,780	2,558
Step 3	443	1,760	2,547
Step 4	465	1,740	2,535
Step 5	487	1,720	2,523
Step 6	508	1,700	2,511

for individual listeners for the 1-pulse tokens.) These data were then submitted to a one-way analysis of variance with the factor continuum. There was a significant effect of continuum [$F(3, 36) = 4.09$, $p < .02$]. A Duncan multiple-range test showed that the responses to steps 3–5 of the 1-pulse continuum had significantly fewer *bib* responses than did the 3- and 4-pulse continua (at the 0.05 level). The 1- and 2-pulse continua were not significantly different, nor were the 2- and 3-pulse continua nor the 3- and 4-pulse continua.

It is thus clear that relatively little relevant acoustic information is required to allow listeners to make vowel identifications. This would suggest that listeners can reliably use the acoustic information available in the burst frame and the very onset of the consonant transition to identify the vowel (perhaps by tracking the formant trajectory and extrapolating to a hypothetical formant target).

There is at least one possible explanation of these data which would not involve dynamic information or require the listeners to track and predict formant trajectories. In particular, Nearey and Assmann [1986] suggested that the endpoints of the CV and VC formant trajectories might serve to specify the dynamic vowel targets (we assume this means interpolating between the two endpoints across the 'silent' vowel). Thus one might suggest that in experiment 3, listeners are making their vowel judgments on the basis of the formant frequency values at the CV transition offsets (and VC transition onsets). This would claim that listeners identify the formant frequency values at the transition onset/offset as the formant frequency values of the missing medial vowel.

Shown in table 4 are the frequencies of the F_1, F_2, and F_3 transition endpoints after 1, 2, 3, or 4 pitch periods (these values were estimated on the basis of the synthesis parameters and corroborated with spectrographic analysis). Note that the frequency of F_2 decreases as the number of pitch periods decreases (because F_2 is rising in frequency during the 30-ms transition). One might argue that the obtained increase in

the number of [ɛ] responses as the number of pitch periods decreases supports the transitional endpoint hypothesis, since the frequency of F_2 for [ɛ] is usually lower than for [ɪ]. A problem for this explanation is that the F_1 and F_2 values after the offset of one or even two pitch periods are *not* representative of either [ɪ] or [ɛ] for either male or female speakers of American English [Peterson and Barney, 1952]. However, experiment 3 utilized a *forced-choice* task in which listeners had to identify the stimulus as either [ɪ] or [ɛ]; perhaps listeners would have labeled these incompatible vowels in the obtained patterns even if they were not actually perceiving the vowel qualities [ɪ] or [ɛ].

Experiment 4

The fourth experiment was conducted to specifically test the transition endpoint hypothesis. This experiment required listeners to identify a set of vowels that had formant frequencies equivalent to the formant values shown in table 4. If listeners are making vowel identifications on the basis of the endpoint frequencies of the formant transitions in the silent-center stimuli, then the pattern of responses obtained in experiment 4 should be similar to those obtained in experiment 3.

Methodology

Stimuli

The stimuli for experiment 4 were four sets of synthetic vowels created using a cascade/parallel software synthesizer. Each of the vowels was 200 ms in duration – similar to the duration of the vowel-only tokens in experiment 1. All vowels were steady state (no change in formant frequency over time). Four sets of vowels were created and each set contained 5 vowels. The formant frequencies of these 20 vowels were matched with the transition endpoint frequencies of the 20 stimuli used in experiment 3 and shown in table 4. Thus one set of vowels corresponded to the 1-pulse offsets, one set of vowels corresponded to the 2-pulse offsets, etc. For ease of comparison to the results of experiment 3, we will refer to these vowel sets as the 1-, 2-, 3- or 4-pulse vowel continua, although all vowels had periodic energy throughout their length.

Each vowel was synthesized with five formants. The frequencies of formants 4 and 5 were 3,300 and 3,850 Hz, respectively, for all vowels. F_0 for each token began at 115 Hz, where it remained for 50 ms; F_0 then fell linearly to 100 Hz at the end of the token. Overall amplitude rose linearly over the first 60 ms, where it remained until a symmetrical fall over the last 60 ms. One identification tape was created. The tape contained 8 examples of all 20 vowels in random order. There were 10 practice items before the start of the test tokens and the practice set included at least two tokens from each of the four vowel sets.

As expected, formant values given in table 4 produced many vowels that did not sound like [ɪ] or [ɛ]. Impressionistically (based upon the author's transcription of the stimulus vowels presented in random order), the 1-pulse vowels represented various gradations of [ɔ] and the 2-pulse vowels were more centralized versions of [ɔ] and [ʌ]. The 3-pulse vowels sounded like a very centralized continuum somewhat more fronted than the 1- and 2-pulse continua and only the 4-pulse vowels represented a clear [ɪ]-[ɛ] continuum. Out of a forced-choice context, it seems that only the 4-pulse vowels would give rise to the [ɪ] and [ɛ] responses.

Listeners

There were 10 listeners. All were undergraduate or graduate students in Speech and Hearing Science at the Ohio State University and were native speakers of American English with no known speech or hearing impairment. None had participated in experiment 3 and all were unfamiliar with the goals of the experiment.

Procedure

The listeners were instructed that they were to identify a set of vowels as either [ɪ] or [ɛ] (all were

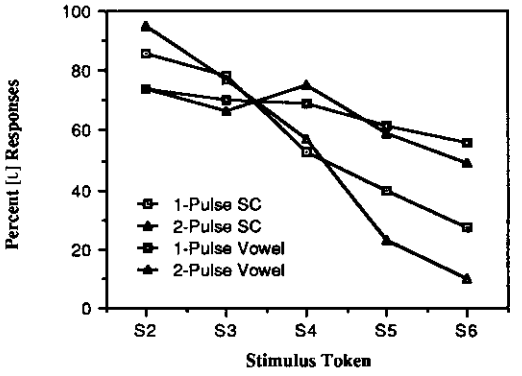

Fig. 8. Identification functions (percent [ɪ] responses) for the 1- and 2-pulse vowels as compared with the 1- and 2-pulse silent-center token (from experiment 3).

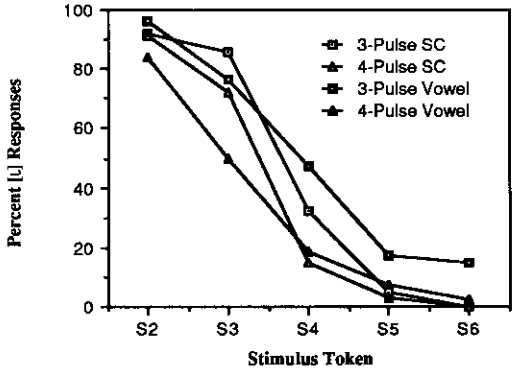

Fig. 9. Identification functions (percent [ɪ] responses) for the 3- and 4-pulse vowels as compared with the 3- and 4-pulse silent-center tokens (from experiment 3).

familiar with the phonetic symbols). They were warned that some of the vowels on the tape could sound relatively different from either [ɪ] or [ɛ], but since this was a *forced-choice* experiment they could only use one of the two available alternatives in their identifications. They were encouraged to guess whenever they were unsure.

Results and Discussion

Shown in figure 8 are the identification responses obtained for the 1- and 2-pulse vowels as compared to the 1- and 2-pulse silent-center tokens (obtained in experiment 3). As one can note, there is a significant amount of difference between the two sets of responses. In particular, none of the vowel stimuli were identified as [ɪ] more than 75% of the time nor as [ɛ] more than 51.3% of the time. The identification functions for the two vowel continua tend to be much flatter (closer to chance level responses) than their silent-center counterparts. The mean data shown in figure 8 do not show the high degree of variability between listeners in response to the vowel stimuli. For example, 2 listeners identified the 1-pulse vowels as [ɪ] 97.3% of the time while 2 other listeners identified these vowels as [ɛ] 99.1% of the time. Every listener commented that most of the vowels which they heard in this test were *very* different from either [ɪ] or [ɛ] and that the identification task was rather difficult. Most listeners stated that they simply decided on a response category to use for these 'odd' vowels. This phenomenon did not occur in experiment 3. The listeners in experiment 3 did not express any difficulty with the task and there were no complaints or observations that vowels other than [ɪ] or [ɛ] occurred.

Shown in figure 9 are the identification responses for the 3- and 4-pulse vowels as compared to the 3- and 4-pulse silent-center tokens. As can be seen, the responses for these sets of vowels are now quite similar to their silent-centre counterparts, although the identification function for the 3-pulse vowel shows some tendency to be flatter than its silent-center counterpart. In addi-

tion, there was much less variability between listeners. For example, in all but one case the step 2 and 3 tokens in both the 3- and 4-pulse continua were identified more often as [ɪ] than were the step 5 and 6 tokens. Note that the 4-pulse vowel is equivalent to the vowel-only stimuli of experiment 1 and actually represents the vowel target.

If listeners had identified the vowels in the silent-center syllables from the formant frequencies at the endpoints of the CV and VC transitions, then we should not have found a difference in the identification functions between the silent-center syllables and the vowel syllables synthesized with the same formant frequencies. However, we did find a difference in the identification functions. Consequently, these results do not support the hypothesis that the silent-center vowels are identified as the formant frequencies represented by the endpoints of the CV and VC transitions. Rather, the data are compatible with the hypothesis that listeners are utilizing the dynamic information contained in the formant trajectories or the stop closure release to identify the vowels.

General Discussion

In summary, experiment 1 demonstrated that one could replicate the phenomenon of the identification of vowels in 'vowelless' syllables using synthetic stimuli. No significant difference in listener's identification was found as a function of whether listeners heard the full unmodified token, its silent-center counterpart, or the steady-state vowel alone. The use of an AX discrimination task did show some differences in the perception of these three types of tokens. A reduced ability to make within-category discriminations in the silent-center token and the observation that silent-center tokens were more categorically perceived suggests that listeners do not have a stable auditory trace of the vowel when only transitional information is present. In addition, just the presence of the transition tends to make the listeners respond more categorically, as was shown in the responses to the full token stimuli.

Experiment 2 demonstrated aptly that when required to compare across these three continua, acoustic similarity was important. Listeners were particularly poor in discriminating between the silent-center and vowel-only tokens, which suggests these same/different judgments were not easily made on the basis of abstract categories (i.e., phonemic/phonetic labels) alone. These data also support the notion that the memorial representation of vowels developed on the basis of dynamic information is somewhat different from that developed from static information. A complete model of vowel perception will need to account for this difference.

Experiment 3 demonstrated that listeners could make quite accurate [ɪ–ɛ] distinctions even when the amount of relevant acoustic information (i.e. stop burst and formant transitions) was severely restricted. This suggests that the perceptual mechanism is well-suited to extract vowel information from what must be primarily consonant acoustic information. Listeners do *not* need the endpoints of the consonant transition to reach the formant targets for the following vowel in order to make reliable vowel identifications, nor, as shown by experiment 4, are the endpoint frequencies themselves sufficient to account for the vowel identifications.

The identification data fit well into a theoretical approach which emphasizes the importance of dynamic information in the perceptual process such as the articulatory event/gesture approach [Browman and Goldstein, 1986, in press; Strange, 1987; Fowler, 1987]. This viewpoint states that movement trajectories (and their acoustic counterparts) can be analyzed into a set of discrete, concurrently active underlying gestures. In addition, these gestures can overlap such that a given stretch of acoustic information can concurrently contain information about more than a single gesture. The goal of the listener is to extract the information relevant for each of these overlapping gestures, perhaps in some type of vector analysis [Fowler and Smith, 1986]. From a more acoustic point of view, the relevant vowel information needs to be separated from the consonant information in the formant transition [see discussion in Nearey and Assmann, 1986]. Lindblom's [1963] undershoot model provides some suggestion as to how formant trajectories could be analyzed into separate additive components. More recently, Broad and Clermont [1987] have begun to develop a more exhaustive mathematical descriptive model for modeling formant contours in a CVC context – again in terms of additive components separating the vowel target from the consonantal influence. The data obtained here suggest that listeners are not only able to track formant trajectories, but are able to extrapolate its probable contour when it has been drastically shortened.

Clearly the results presented here need to be expanded before they can be broadly generalized. For example, similar experiments should be completed with a wider range of vowel qualities, formant transitions (in terms of manner and/or place distinctions) and, possibly, speech rates. In addition, as Macchi [personal commun.] has pointed out, the perceptual contribution of the short release burst has not been addressed in this series of experiments. This issue will be addressed in studies to be conducted soon in our lab which will include tokens without burst releases and/or final stop murmurs. The importance of dynamic information in the perception of vowels seems obvious (from this and earlier studies cited above), but we are only beginning to develop the detailed perceptual models needed to adequately deal with the relevant experimental results.

Acknowledgments

The author would like to thank Marian Macchi and Brad Rakerd for many suggestions on the improvement of the article. I would also like to give special thanks to John Ohala for his suggestions and his tireless efforts in correcting my list of references. I take full responsibility for any remaining problems.

References

Assmann, P.; Nearey, T.; Hogan, J.: Vowel identification: orthographic, perceptual, and acoustic aspects. J. acoust. Soc. Am. *71:* 975–989 (1982).

Broad, D.; Clermont F.: A methodology for modeling vowel formant contours in CVC context. J. acoust. Soc. Am. *81:* 155–165 (1987).

Browman, C.; Goldstein, L.: Towards an articulatory phonology. Phonol. Yb. *3:* 219–252 (1986).

Browman, C.; Goldstein, L.: Tiers in articulatory phonolgy, with some implications for casual speech; in Beckman, Kingston, Papers in Lab. Phonol. I: Between the grammar and the physics of speech (in press).

Crowder, R.: The role of auditory memory in

speech perception and discrimination; in Myers, Laver, Anderson, The cognitive representation of speech, pp. 167–179 (North-Holland, Amsterdam 1981).

Crowder, R.: Decay of auditory memory in vowel discrimination. J. exp. Psychol.: Learn. Mem. Cog. *8:* 153–162 (1982).

Crowder, R.; Repp, B.: Single formant contrasts in vowel identification. Perception Psychophysics *35:* 372–378 (1984).

Diehl, R.; McCusker, S.; Chapman, L.: Perceiving vowels in isolation and in consonantal context. J. acoust. Soc. Am. *69:* 239–248 (1981).

Fowler, C.: Timing control in speech production; PhD diss. University of Connecticut (1977).

Fowler, C.: Coarticulation and theories of extrinsic timing. J. Phonet. *8:* 113–133 (1980).

Fowler, C.: Converging sources of evidence on spoken and perceived rhythms of speech: production of vowels in monosyllabic stress feet. J. exp. Psychol.: Gen. *112:* 386–412 (1983).

Fowler, C.: Perceivers as realists, talkers too: commentary on papers by Strange, Diehl et al., and Rakerd and Verbrugge. J. Mem. Lang *26:* 574–587 (1987).

Fowler, C.; Rubin, P.; Remez, R.; Turvey, M.: Implications for speech production of a general theory of action; in Butterworth, Language production I, pp. 373–420 (Academic Press, London 1980).

Fowler, C.; Smith, M.: Speech perception as 'vector analysis': an approach to the problems of invariance and segmentation; in Perkell, Klatt, Invariance and variability in speech processes, pp. 123–136 (Erlbaum, Hillsdale 1986).

Fox, R.: Individual variation in the perception of vowels: implications for a perception-production link. Phonetica *39:* 1–22 (1982).

Fox, R.: Perceptual structure of monophthongs and diphthongs in English. Lang. Speech *26:* 21–60 (1983).

Fox, R.: Multidimensional scaling and perceptual features: evidence of stimulus processing or memory prototypes? J. Phonet. *13:* 205–217 (1985a).

Fox, R.: Auditory contrast and speaker quality variations in vowel perception. J. acoust. Soc. Am. *77:* 1552–1559 (1985b).

Fox, R.: Within- and between-series contrast in vowel identification: full-vowel versus single-formant anchors. Perception Phychophysics *38:* 223–226 (1985c).

Fujisaki, H.; Kawashima, T.: A model of the mechanisms for speech perception: quantitative analysis of categorical effects in discrimination. Annu. Rep. Engineering Res. Inst., Faculty of Engineering, University of Tokyo *30:* 59–68 (1971).

Gay, T.: A perceptual study of American English diphthongs. Lang. Speech *13:* 65–88 (1970).

Jenkins, J.: A selective history of issues in vowel perception. J. Mem. Lang *26:* 542–549 (1987).

Jenkins, J.; Strange, W.; Edman, T.: Identification of vowels in 'vowelless' syllables. Perception Psychophysics *34:* 441–450 (1983).

Johansson, G.: Visual perception of biological motion and a model for its analysis. Perception Psychophysics *14:* 201–211 (1973).

Joos, M.: Acoustic phonetics. Lang. Monogr., No. 23 (Waverly, Baltimore 1948).

Klatt, D.: Analysis and synthesis of CV syllables in English (Massachusetts Institute of Technology, Cambridge, unpublished, 1978).

Klatt, D.: Software for a cascade/parallel formant synthesizer. J. acoust. Soc. Am. *67:* 971–995 (1980).

Ladefoged, P.: A course in phonetics; 2nd ed. (Harcourt Brace Jovanovich, New York 1982).

Lindblom, B.: Spectrographic study of vowel reduction. J. acoust. Soc. Am. *35:* 1773–1781 (1963).

Macchi, M.: Identification of vowels spoken in isolation versus vowels spoken in consonantal context. J. acoust. Soc. Am. *68:* 1636–1642 (1980).

Massaro, D.: Language and information processing; in Massaro, Understanding language, pp. 3–28 (Academic Press, New York 1975).

Nearey, T.; Assmann, P.: Modeling the role of inherent spectral change in vowel identification. J. acoust. Soc. Am. *80:* 1297–1308 (1986).

Peterson, G.; Barney, H.: Control methods used in a study of the vowels. J. acoust. Soc. Am. *24:* 175–184 (1952).

Peterson, G.; Lehiste, I.: Duration of syllable nuclei in English. J. acoust. Soc. Am. *32:* 693–703 (1960).

Pisoni, D.: Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception Psychophysics *13:* 253–260 (1973).

Pisoni, D.: Auditory short-term memory and vowel perception. Mem. Cogn. *3:* 7–18 (1975).

Pollack, I.; Pisoni, D.: On the comparison between identification and discrimination tests in speech perception. Psychonomic Sci. *24:* 299–300 (1971).

Rakerd, B.; Verbrugge, R.: Evidence that the dynamic information for vowels is talker independent in form. J. Mem. Lang. *26:* 558–563 (1988).

Rakerd, B.; Verbrugge, R.; Shankweiler, D.: *Monitoring for vowels in isolation and in a consonantal context.* J. acoust. Soc. Am. *76:* 27–31 (1984).

Ray, A.A.: SAS User's Guide: Statistics (SAS Institute, Cary 1982).

Repp, B.: Categorical perception: issues, methods, findings. Haskins Laboratory, Status Rep. Speech Res. *SR 70:* 99–183 (1982).

Repp, B.; Healy, A.; Crowder, R.: Categories and contexts in the perception of isolated steady-state vowels. J. exp. Psychol.: Hum. Perception Performance *5: 129–145 (1979).*

Sachs, R.: Vowel identification and discrimination in isolation vs. word context. Q. Progr. Rep., No. 93, *pp. 220–229* (MIT, Research Laboratory of Electronics, Cambridge 1969).

Sawusch, J.; Nusbaum, H.: Contextual effects in vowel perception. I. Anchor-induced contrast effects. Perception Psychophysics *25:* 292–302 (1979).

Sawusch, J.: Nusbaum, H.; Schwab, E.: Contextual effects in vowel perception. II. Evidence for two processing mechanisms. Perception Psychophysics *27:* 421–434 (1980).

Shepard, R.: *Psychological representation of speech sounds; in David, Denes, Human communication:* a unified view, pp 67–113 (McGraw-Hill, New York 1972).

Shriberg, L.; Kent, R.: Clinical phonetics (Wiley & Sons, New York 1985).

Singh, S.; Woods, G.: Perceptual structure of 12 American English vowels. J. acoust. Soc. Am. *49:* 1861–1866 (1971).

Stevens, K.: *On the relations between speech movements and speech perception.* Z. Phonet. Sprachw. KommunForsch. *21:* 102–106 (1968).

Stevens, K.; House, A.: Perturbations of vowel articulations by consonantal context: an acoustical study. J. Speech Hear. Res. *6: 111–128 (1963).*

Strange, W.: Information for vowels in formant transitions. J. Mem. Lang. *26:* 550–557 (1987).

Strange, W.; Jenkins, J.; Johnson, T.: Dynamic specification of coarticulated vowels. J. acoust. Soc. Am. *74:* 695–705 (1983).

Strange W.; Verbrugge, R.; Shankweiler, D.; Edman, T.: Consonant environment specifies vowel identity. J. acoust. Soc. Am. *60:* 213–224 (1976).

Studebaker, G.: A 'rationalized' arcsine transformation. J. Speech Hear. Res. *28:* 455–462 (1985).

Tartter, V.: A comparison of the identification and discrimination of synthetic vowel and stop consonant *stimuli with various acoustic properties.* J. Phonet. *9:* 477–486 (1981).

Terbeek, D.: A cross-language multidimensional scaling study of vowel perception. UCLA Working Papers Phonet. *37:* 1–271 (1977).

Verbrugge, R.; Rakerd, B.: Evidence of talker-independent information for vowels. Lang. Speech *26:* 39–55 (1986).

Verbrugge, R.; Strange, W.; Shankweiler, D.; Edman, T.: What information enables a listener to map a talker's vowel space? J. acoust. Soc. Am. *60:* 198–212 (1976).

Received: July 19, 1988
Accepted: April 21, 1989

Robert Allen Fox
Division of Speech and Hearing Science
Ohio State University
324 Derby Hall, 154 N. Oval Mall
Columbus, OH 43210-1372 (USA)