Identification and Discrimination of Silent-Center Vowels

Robert Allen Fox
The Ohio State University

1.0 Introduction

According to the most commonly encountered viewpoint in phonetics, vowel quality is determined primarily on the basis of the frequencies of formants 1 and 2. The first formant is inversely correlated with the high/low distinction while the second formant is directly correlated with the front/back distinction. A strict interpretation of this viewpoint (here called the "target theory" of vowel perception) contains at least two fundamental assumptions (Johnson, 1987): (1) that there is a target or canonical representation of the acoustic information for a given vowel quality and (2) that the characterization of that information can be described in terms of a relatively steady-state portion of the produced vowel, however brief.

However, a 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 (e.g., Strange, Verbrugge, Shankweiler & Edman, 1976; Verbrugge, Strange & Edman, 1976)--have found that vowels 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. 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 (e.g., Macchi, 1980; Diehl, McCuster & Chapman, 1981; Assmann, Nearey & Hogan, 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 (see also Rakerd, Verbrugge, & Shankweiler, 1984). However, the important thing to note is that *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).

A second, even more dramatic, set of data, involves the identification of the medial vowels in so-called "silent-center" speech tokens (e.g., Jenkins, Strange & Edman, 1983; Strange, Jenkins & Johnson, 1983, Strange, 1987). In these studies, a set of talkers produce a CVC token such as [bab] and [bib]. These tokens are then waveform edited such that most of the medial vowels (up to 65%) is removed and replaced by silence. Another type of token constructed from the original stimulus include the isolated vowel alone--that is, that portion of the vowel which was removed from the CVC token. In subsequent identification tasks, sets of listeners are required to identify the vowel contained in these different tokens: control (the original, unedited token), silent-center, and vowel-only (other stimulus tokens created from the control token commonly include the CV or VC transitions alone). Results obtained show that the silent-center tokens are identified with remarkably low error rates, comparable to, if not lower than, the error rates for the vowel alone. 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 (see 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. Data collected from other lines of research--including investigations into the role of vowel-intrinsic spectral change in vowel identification (Assman, Nearey & Hogan, 1982; Nearey & Assmann, 1986) and the identification of initial consonants as

a function of the quality of the following vowel (Diehl, Kluender, Foss, Parker, and Gernsbacher, 1987)--support this viewpoint. Such dynamic information reflects the movement of the vocal tract and cannot be easily described in terms of a set of formant "targets" or frequencies at any single point in time during the production of the vowel. In their view, vowels must be described in terms of *gestures*. These hypothesized gestures have intrinsic metrical characteristics and are compatible with the recent theoretical claims of Browman & Goldstein (1986, 1988) in terms of a model of articulatory phonology.

However attractive the dynamic theory, at this point it is relatively vague in terms of the nature of the perceptual processes involved in identifying vowels from dynamic acoustic changes. We also know little about the role of auditory and/or phonetic memory for such dynamic events. In particular, vowels--perhaps because of their relatively steady-state nature--are generally considered to create relatively strong traces in auditory memory. However, this may not be true for *dynamic* events related to vowel perception. For example, Tartter (1981) found that although listeners could identify vowels reliably from a 40 msec CV transition, that discrimination of those transitions was close to categorical. The information concerning memory for such events is critical since memorial representations play such an important role in many vowel perception models (e.g., Fujisaki & Kawashima, 1971; Repp, Healey & Crowder, 1979; Crowder, 1981; Crowder, 1982) and are of great interest in accounting for the effects of phonetic context upon the identification of vowels (e.g., Sawusch & Nusbaum, 1979; Crowder & Repp, 1984; Fox, 1985a; Fox, 1985b).

In addition, Nearey & Assmann (1986) have suggested that the *endpoints* of the formant trajectories may serve to specify the *dynamic* targets. To what extent do these CV or VC transitions allow the listener to predict what the hypothetical "formant target" might be? If listeners can accurately "compute" the formant trajectory to determine the eventual target, a version of the target model may, in fact, be compatible with the silent-center data.

The present paper will describe three different perceptual experiments. The first two experiments involve the identification and discrimination of silent-center vowels as compared to full CVC tokens and isolated vowels. The third experiment concerns the amount of transition needed by listeners to accurately identify medial vowels removed from a set of silent-center tokens. The stimuli used in each of the experiments represent a *bib-beb* continuum (i.e., a [I-E] 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. After the data have been presented, there will be a discussion concerning the implications of the experimental results for perceptual models.

2.0 Experiment 1

2.1 *Methodology*

2.1.1 Stimulus Material

A natural sounding, 7-step *bib-beb* continuum was created using the Klatt software synthesizer. Each of the seven tokens of each continuum was 300 msec in duration. The total length of the vowel (including transitions) was 255 msec. The steady-state portion of the medial vowel was 195 msec in duration while both the CV and VC transitions were 30 msec in duration. There was a 10 msec burst frame, with an aperiodic noise source, at the start of each token. The formant frequency values used in the synthesis of the tokens are show in Table 1 (based on Klatt, 1978). These values[1] 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. At the end of each token was a low-energy murmur for a "voiced stop". This was synthesized using a sinusoidal voicing source with the formant frequency values remaining at the VC transition offset values. The fundamental frequency for each token began at 115 Hz where it remained for 100 msec. F0 then fell linearly to 100 Hz at the end of the token. Note that, similar to humanly

Table 1. Formant frequencies values used in the synthesis of the original *bib-beb* continuum. Shown are the F1, F2, and F3 frequency values at the consonant burst, the onset and offset of the consonant transitions and the steady-value vowel.

	Release Burst			Transition onset/offset			Steady-State Vowel		
Stimulus	F1	F2	F3	F1	F2	F3	F1	F2	F3
1	285	1297	2267	370	1494	2383	400	1800	2570
2	291	1291	2266	381	1481	2381	422	1780	2558
3	296	1284	2265	391	1468	2380	443	1760	2547
4	302	1277	2264	403	1454	2378	465	1740	2535
5	307	~~1356~~	~~2320~~ 2262	414	1441	2376	487	1720	2523
6	312	1264	2262	61 424	1428	2374	508	1700	2511
7	318	1258	2261	60 435	1415	2372	530	1680	2500

Table 2. Two-step discrimination results from Experiment 1. 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. Predicted within-category discrimination scores are mean predicted scores for steps 1 and 3 and steps 5 and 7.

	Interval Duration					
	500 msec			2000 msec		
	Within-Category	Between-Category	Difference	Within-Category	Between-Category	Difference
Full Word	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

Table 3. Predicted percent correct discrimination scores for Experiment 1. These predictions were based on the identification responses in terms of formulae suggested by Pollack & Pisoni (1971).

	Within-category	Between-category
Full Token	51.7	74.2
Silent-Center	51.9	73.1
Vowel-Only	52.5	67.5

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 following vowel (i.e., the starting transitions are affected by the frequencies of the eventual steady-state targets--see footnote 1 for more details).

From this basic continuum, two other continua were created. One set of tokens, called the *silent-center tokens*, were created in which most of the medial vowel was removed. In particular, all but 4 glottal pulses (approximately 30 msec) of the CV and VC transitions were replaced by silence. This resulted in approximately 75% of the vowel being replaced by silence in the token--comparable to the most severe reductions of Strange et al. 1983. 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 very little, or no transitional information. This was also confirmed by spectrographic analysis.

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 112 stimuli (16 repetitions of each of the 7 continuum steps) with an interstimulus interval of 3 sec. Preceding the test stimuli was a short practice of 10 items.

The discrimination tapes each contained pairs of stimulus tokens (as used in the AX discrimination procedure). Each of the discrimination tapes contained a completely randomized sequence of the following stimulus pairs: each stimulus step paired with itself (6 each); a set of 1-step discrimination pairs (6 repetitions of 1/2, 3/4, 4/5 and 6/7 in both orders); and a set of 2-step discrimination pairs (6 repetitions of 1/3, 3/5 and 5/7 in both orders). This includes both within-category pairs (e.g., 1/3, 5/7) and between-category pairs (3/5). Due to length constraints, only the 2-step discriminations will be discussed.

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 msec, in the second tape, the interval was 2000 msec. Following Repp et al. (1979) and Crowder (1981, 1982), the longer intrapair interval was included to produce a condition in which the amount of memory for the first token of a pair might be reduced. This experimental manipulation tends to make vowel discrimination somewhat more "categorical."

2.1.2 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.

2.1.3 Procedure

The listeners were randomly divided into three groups of eight 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 [I] 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 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 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 msec discrimination tape before the 2000 msec 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.

2.2. *Results and Discussion*

The identification results obtained for each of the three groups are shown in Figure 1. This is a graph of the identification function coded in terms of percentage of [I] responses. Obviously the identification functions for the three groups are remarkably similar. Phoneme boundaries for each separate listener were calculated (using linear interpolation). These boundaries indicate the point on the stimulus continuum corresponding to the *bib-beb* 50% cross-over point. These phoneme boundaries were then analyzed using a one-way analysis of variance with the factor stimulus group (full token, silent-center, and vowel-only). There was no significant effect of stimulus group ($F(2,21)=0.91$, $p>.40$). The pattern of discriminations, however, are not so similar across groups.

Shown in Figure 2 are the discrimination responses (in terms of percent correct) when the interval duration between tokens in an AX pair was 500 msec. Figure 3 shows the discrimination responses when the interval was 2000 msec. Several things can be noted about these responses. First, the between-category discriminations 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 (cf. Repp, Healy & Crowder, 1979). Second, 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, control tokens seem to be midway between the silent-center and vowel-only results in the 500 msec condition, although this pattern changes somewhat during the 2000 msec condition. This decrement in the discrimination of vowels in context was also obtained by Sawusch, Nusbaum, and Schwab (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 responses from each listener were next analyzed using a three-way analysis of variance with the factors stimulus group (full token, silent-center, vowel-only), discrimination type (within- vs. between-category), and interval duration (500 vs. 2000 msec). Since the means of ratio data (such as percentages) are correlated with the variances, these percentages were arcsine-transformed (using Studebaker's, 1985, rationalized arcsine method) prior to analysis. The analysis of variance indicated that there was a significant effect of stimulus group ($F(2,86)=6,31$, $p<.003$). 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 token data (77.4%). The vowel-only and full token groups were not significantly different.

There was also a significant effect of discrimination type ($F(1,86)=35.21$, $p<.001$) 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 ($F(1,86)=2.08$, $p<.15$). However, similar to the Repp, et al. data, 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 ($F(2,86)=5.01$), $p<.001$). This interaction effect stems from the fact that although the mean difference between the within- and between-category discriminations for the full token 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 (Pollack & Pisoni, 1971). It is normally the case that these predictions are better for consonant

Figure 1. Identification Responses--Experiment 1

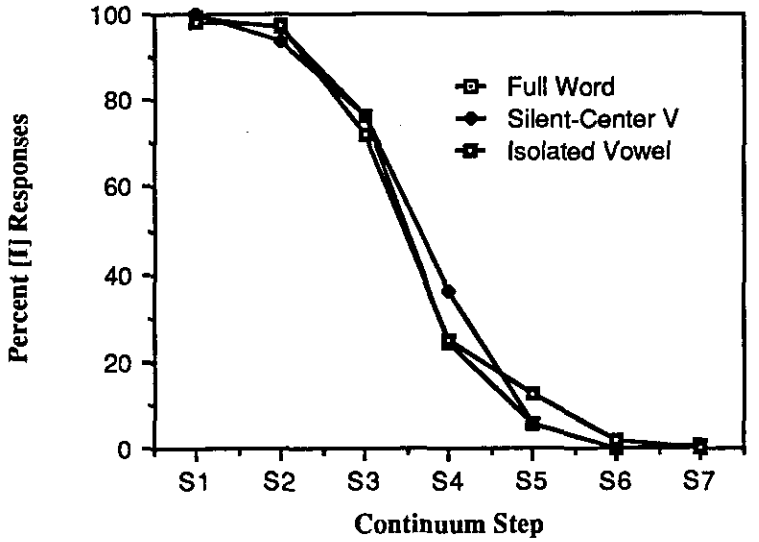
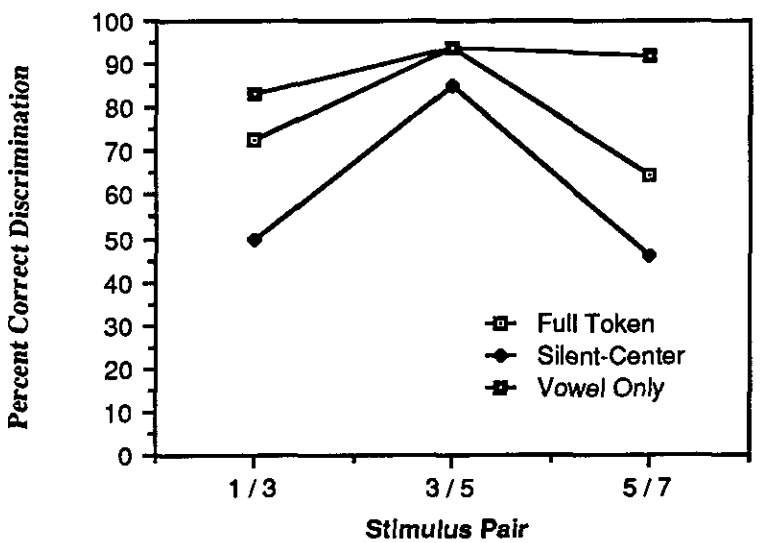


Figure 2. Discriminations, 500 msec Interval, Exp. 1



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) 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 token discriminations are better than the predictions, but not as much better as the vowel only tokens. This is despite the fact that the full word tokens had more acoustic data upon which to make a discrimination than did either the vowel-only token or the silent-center token.

It is not surprising that the within-category discriminations are near chance level for the silent-center tokens. One of the more common explanations of the distinction between between- and within-category distinctions is that the between-category discriminations result from comparisons in terms of internal phonetic labels (with the information stored in some type of phonetic memory) while the within-category discriminations result from comparisons based on auditory-level memory traces. It is normally assumed that a vowel will leave a relatively strong auditory memory trace even following categorization (identification). A consonant, normally does not. This would explain the differences between the silent-center and vowel-only tokens well. Evidently, 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.

3.0 Experiment 2

Most, if not all, of the published experiments on silent-center tokens have compared the identifications of the silent-center tokens with responses to related tokens (such as the full tokens, and vowel-only tokens). The next experiment is an attempt to directly compare the silent-center tokens with both the full and vowel-only tokens. In this experiment, listeners 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 tokens was from another continuum. This should give insight into the comparability of the vowel qualities extracted in the different stimulus conditions.

3.1 *Methodology*

3.1.1 Stimuli

The stimuli were constructed using the tokens described above. Three different sets of discrimination tapes were constructed. Each set of tapes compared two different continua--full token vs. silent-center, full token vs. vowel-only, and silent-center vs. vowel-only. Each set of discrimination tapes was composed of two different tapes differing in terms of stimulus order (e.g., tape 1a had full tokens followed by silent-center tokens, while tape 1b had the silent-center tokens followed by full tokens). Each tape contained 100 stimulus pairs: 10 repetitions of the "identical" pairs (using continua steps 1, 3, 5, and 7) and 10 repetitions each of the 2-step difference comparisons in both orders (e.g., 1/3, 3/1, 3/5, 5/3, 5/7, 7/5). The intrapair interval was 500 msec for all tapes.

3.1.2 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.

3.1.3 Procedure

The 30 listeners were randomly assigned to one of three groups. Within each group, 4 listeners heard the two tapes in one order (e.g., 1a/1b) while the other 4 listeners heard the tapes in the reverse order (e.g., 1b/1a).

3.2 *Results and Discussion*

Figure 3. Discriminations--2000 msec Interval, Exp. 1

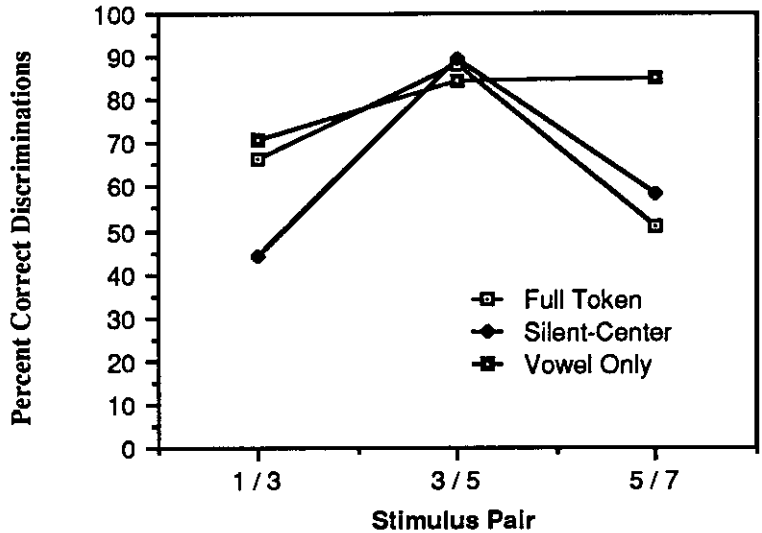


Figure 4. Discriminations--Experiment 2

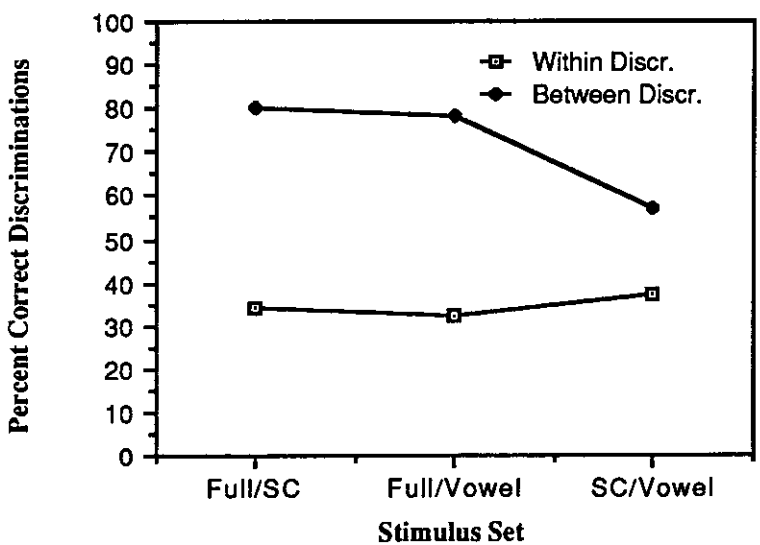


Figure 4 shows the percent correct responses for the "different" pairs. The same pattern of more accurate between-category discriminations is obtained, although the percentage of correct scores is somewhat lower than in Experiment 1. These responses were then analyzed using a two-way analysis of variance[2] with the factors condition (continua combinations) and discrimination type (between vs. within). As in Experiment 1, these ratio scores were arcsine transformed before analysis (Studebaker, 1985).

There was significant main effect of condition ($F(2,52)=4.29$, $p<.02$) A Duncan multiple-range test showed that the silent-center/vowel-only condition had significantly fewer correct responses (46.5%) than did either the full token/silent-center (55.6%) or full token/vowel-only (53.5%) conditions (at the .05 level). There was also a significant main effect of discrimination type ($F(1,52)=196.33$, $p<.001$) showing that there were fewer correct responses to the within-category stimuli (33.4%) than the between-category stimuli. There was also a significant condition X discrimination type interacation ($F(2,52)=11.65$, $p<.001$). This result was obtained because the difference between the within- and between-category responses for the silent-center/vowel-only condition (20.6%) was significant less than the differences for the full token/silent-center (48.3) and full token/vowel-only (48.0%) conditions.

The results for the "same" comparisons are shown in Figure 5. Again, a two-way analysis of variance using the factors condition and stimulus pair was done on the arcsine transformed data. There was a significant main effect of condition ($F(2,104)=17.18$, $p<.001$). A Duncan multiple-range test showed that the percentage of correct responses was greater for the Full token/silent-center (85.1%) and the full token/vowel-only (90.4%) conditions than for the silent-center/vowel-only condition (72.3%).

There was also a significant main effect of stimulus pair ($F(3,104)=22.6$, $p<.01$). A Duncan multiple-range test showed that the percentage of correct responses was significantly greater (at the .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 .05 level). These data show that when the tokens represent the endpoints of a continuum, are are thus relatively unambiguous, that 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 are compared in the cross-continuum discrimination task, the percepts used in the comparisons in the silent-center condition are relatively different from those used in the vowel-only condition. This may demonstrate that somewhat different phonetic processes are utilized in the perception of the silent-center tokens than the vowel-only tokens. It may also indicate differences in the auditory memory traces developed for the different continua. The pattern of the full tokens demonstrate compatibility with both silent-center and vowel-only tokens. These seem to show that listener may make comparisons based on the transitions *or* the steady-state vowels but that a single abstract, categorical percept is not used in the discriminations.

4.0 Experiment 3

Although Strange and colleagues have emphasized the importance of formant movement in the transitions themselves, how much movement is necessary? Also, a great many researchers have suggested that the formant movements in the transition contain information both for the consonant as well as for the vowel (e.g., Fowler, 1983, 1987; Neary & Assmann, 1986). One view of the perceptual system would require that it partition formant trajectories into consonant and vowel components (e.g., Neary & Assmann, 1986). The question addressed by Experiment 3 is simply, how little transitional information is needed for listeners to make reliable identifications of vowels?

4.1 *Methodology*

4.1.1 Stimuli

Figure 5. "Same" Pair Responses--Experiment 2

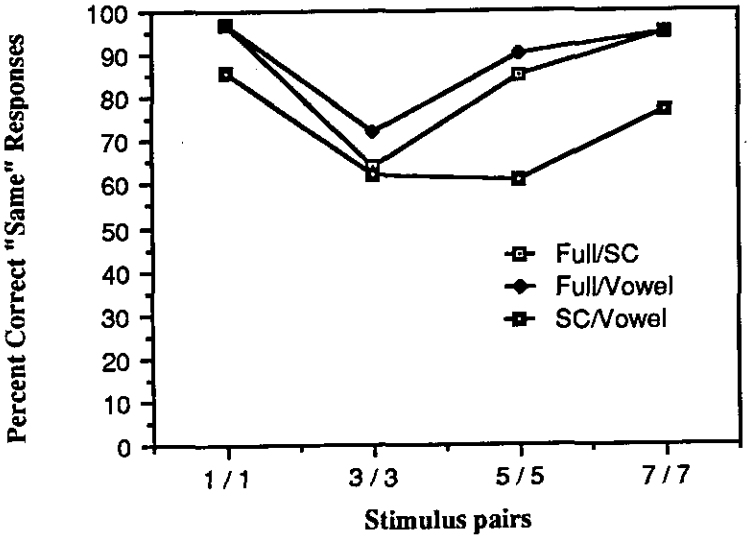
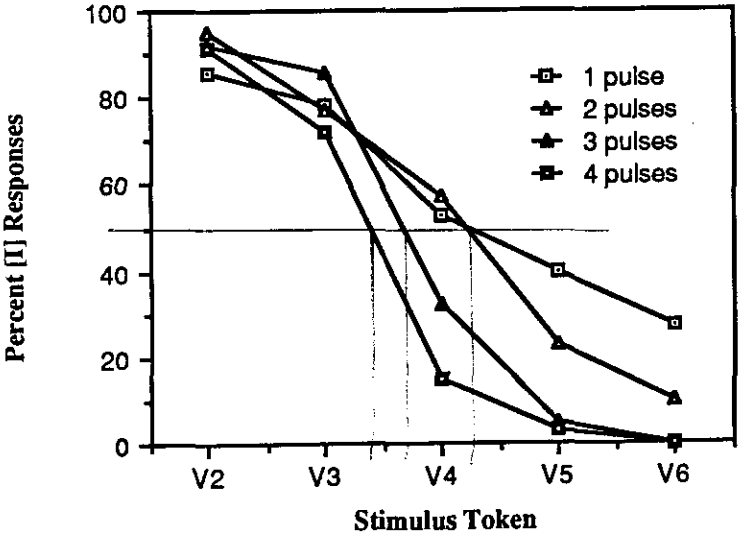


Figure 6. Identification data from Experiment 3



The stimuli for Experiment 3 were four 7-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 one, two, or three glottal pulses in the transitional portions of the token. All waveforms were cut at a zero crossing.

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.

4.1.2 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.

4.1.3 Procedure

The subjects were given the same basic instructions as were given to the silent-center group in the Experiment 1 identification test. 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.

4.2 *Results and Discussion*

The results of Experiment 3 are shown in Figure 6. 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 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, but given the little amount of relevant acoustic information available, one might have expected chance-level responses for all steps. (Impressionistically speaking, the 1-pulse tokens sounded like pairs of high-pitched clicks).

There does seem at least one trend in the data, namely, that as the number of glottal pulses included decreases, the more 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 for each continuum were calculated (phoneme boundaries were not used because they often could not be reliably calculated for the 1-pulse continuum). 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 1-pulse continuum was different from both the 3- and 4-pulse continua (at the .05 level). The 1-, 2-, and 3-pulse continua were not significantly different, nor were 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.

5.0 General Discussion

In summary, Experiment 1 demonstrated that one could replicate the phonomenon of the identification of vowels in "vowelless" syllables using synthetic stimuli. No significant difference in listener's identifications was found as 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 listener 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, that acoustic similarity was important. Listeners were particularly poor in discriminating between the silent-center and vowel-only tokens which suggests that these same/different judgments were not easily made on the basis of abstract, categories (i.e., phonemic/phonetic labels) alone. Experiment 3 demonstrated that listeners could make quite accurate [I-ɛ] distinctions even when the amount of relevant acoustic information (i.e., stop burst and formant transtions) 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.

The identification data fit well into a theoretical approach which is essentially *gestural* in approach such as Browman & Goldstein's articulatory phonology. 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 contained 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 & Smith, 1986). However, this model does not seem to predict the pattern of discrimination results obtained. One might question whether any "phonological" model should necessarily account for such "low-level" phonetic detail, but the trend of some phonological models (such as that being developed by Browman & Goldstein, 1986, 1988) is to be more sensitve to actual articulatory contraints, movements, and perhaps listener responses. Clearly we are at the very beginning in terms of such an approach and we will need to know more about the nature about both articulation and perception (and acoustic processing) and how they interact.

6.0 Footnotes

1. 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 following (*vowel target* equals the formant values for the following steady-state vowel):

$$F1=340+0.50*[\text{Vowel target}-340].$$
$$F2=900+0.66*[\text{Vowel target}-900].$$
$$F3=2350+0.15*[\text{Vowel target}-2350].$$

   The frequencies of F1, F2, and F3 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, as mentioned in the text, the synthetic tokens also include both a vowel and consonant component in the consonant transition itself.
2. The data from one of the listeners in the full token/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 was used to analyze these data.

7.0 References

Assmann, P., Nearey, T., & Hogan, J. (1982). Vowel identification: Orthographic, perceptual,

and acoustic aspects. *Journal of the Acoustical Soceity of America*, 71:975-989.
Browman & Goldstein (1986). Towards and articulatory phonology. *Phonology Yearbook 3.*
Browman & Goldstein, (1988). Tiers in articulatory phonology, with some implications for casual speech. To be published in *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech.*
Crowder, R. (1981). The role of auditory memory in speech perception and discrimination. In T. Myers et al. (eds.), *The Cognitive Representation of Speech*, Amsterdam: North-Holland.
Crowder, R. (1982). Decay of auditory memory in vowel discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8:153-162.
Crowder, R. & Repp, B. (1984). Single formant contrasts in vowel identification. *Perception & Psychophysics*, 17:48-52.
Diehl, R., McCuster, & Chapman (1981). On the identifiability of syntehsized steady-state vowels in isolation and in consonantal context. *Journal of the Acoustical Society of America*, 68:239-248.
Diehl, R., Kluender, K., Foss, D., Parker, E. & Gernsbacher, M. (1987). Vowels as islands of reliability. *Journal of Memory and Language*, 26:564-573.
Fowler, C. (1983). Converging sources of information for spoken and perceived rhythms of speech. *Journal of Experimental Psychology: General*, 112:386-412.
Fowler, C. (1987). Perceivers as realists, talkers too: Commentary on papers by Strange, Diehl et al., and Rakerd and Verbrugge. *Journal of Memory and Language*, 26:574-587.
Fowler, C. & Smith, M. (1986). Speech perception as vector analysis: An approach to the problem of invariance and segmentation. In J. S. Perkell and D. H. Klatt (eds.), *Invariance and Variability in Speech Processes.* 123-136.
Fox, R. (1985a). Auditory contrast and speaker quality variations in vowel perception. *Journal of the Acoustical Society of America*, 77:1552-1559.
Fox, R. (1985b). Within- and between-series contrast in vowel identification: Full-vowel versus single-formant anchors. *Perception & Psychophysics*, 38:223-226.
Fujisaki & Kawashima (1971). A model of the mechanisms for speech perception: Quantitative analysis of categorical effects in discrimination. *Annual Report of the Engineering Research Institute*, Faculty of Engineering, University of Tokyo, 30:59-68.
Jenkins, J., Strange, W., & Edman (1983). Identification of vowels in 'vowelless' syllables. *Perception & Psychophysics*, 34:441-450.
Johnson, N. F. (1987). A tutorial symposium on dynamic conceptions of vowel perception: An introduction. *Journal of Memory and Language*, 26:539-541.
Klatt, D. (1978). *Analysis and Synthesis of CV Syllables in English.* Unpublished manuscript. Cambridge MA: Massachusetts Institute of Technology.
Macchi, M. (1980). Identification of vowels spoken in isolation versus vowels spoken in consonantal context. *Journal of the Acoustical Soceity of America*, 68:1636-1642.
Nearey, T. & Assmann, P. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acosutical Society of America*, 80:1297-1308.
Pollack, I. & Pisoni, D. (1971). On the comparison between identification and discrimination tests in speech perception. *Psychonomic Sciences, 24: 299-300.*
Rakerd, B., Verbrugge, R., & Shankweiler, D. (1984). Monitoring for vowels in isolation and in a consonantal context. *Journal of the Acoustical Soceity of America*, 76:27-31.
Repp, B. (1982). Categorical perception: Issues, methods, findings. *Status Report on Speech Research*, SR70:99-183.
Repp, B., Healy, A., & Crowder, R. (1979). Categories and contexts in the perception of steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 5:129-145.
Sachs, R. (1969). Vowel identification and discrimiation in isolation vs. word context.

Quarterly Progress Report No. 93, Cambridge, MA: MIT, Research Laboratory of Electronics, pp. 220-229.

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

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

Stevens, K. (1968). On the relations between speech movements and speech perception. *Zeitschrift für Phonetik, Sprachwissenschaft, und Kommunikationsforschung*, 21:102-106.

Strange, W. (1987). Information for vowels in formant transitions. *Journal of Memory and Language*, 26:550-557.

Strange, W., Jenkins, J., & Johnson, T. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America*, 74:695-705.

Strange, W., Verbrugge, R., Shankweiler, D. & Edman, T. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America*, 60:213-224.

Studebaker, (1985). A "rationalized" arcsine transformation. *Journal of Speech and Hearing Research*, 28:455-462.

Tartter, V. (1981). A comparison of the identification and discriminatio of synthetic vowel and stop consonant atimuli with various acoustic properties. *Journal of Phonetics*, 9:477_486.

Verbrugge, R., Strange, W., Shankweiler, D. & Edman, T. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America*, 60:198-212.