Perception of Vowel Quality in a Phonologically Neutralized Context

Robert Allen Fox

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

Introduction

Linguistic and psycholinguistic studies completed in the last two decades have shown that the perception of speech can be affected by a number of variables including syntactic, semantic and pragmatic (Connine and Clifton, 1987; Elman and McClelland, 1986; Ganong, 1980; Grosjean, 1980; Morton, 1969; Samuel, 1986; Warren and Sherman, 1974). In turn, researchers developing models of speech perception such as TRACE (Elman and McClelland, 1986; McClelland and Elman, 1986) and Cohort Theory (Marslen-Wilson and Tyler, 1980) have specifically allowed the on-going integration of bottom-up, data-driven acoustic processing with top-down, knowlege-driven contextual information. The relevant acoustic information is generally considered to be the result of auditory processing (both peripheral and central) as well as more specialized categorical (i.e., phonetic) processing. The relevant top-down information may be derived from a variety of sources (and levels) of linguistic knowledge including the lexicon, syntactic and semantic expectations and/or predictability, and prosodic structure (Nakatani and Schaffer, 1978).

In studies examining the perception of vowels, it has been clearly demonstrated that the *phonetic* context surrounding a vowel may significantly affect its identification, especially if the vowel is ambiguous (e.g., Cowan and Morse, 1986; Crowder and Repp, 1984; Fox, 1985; Nearey, 1989; Repp, Healy and Crowder, 1979; Shigeno and Fujisaki, 1979; Shigeno and Fujisaki, 1980; Shigeno, 1986; Akagi, 1990; see also studies by Shigeno; Huang; and Akagi in the present volume). However, of interest in this study is the situation in which information from the phonological context might operate systematically to *reduce* phonetic/phonemic contrast. In particular, we will examine the extent to which phonological information—in the form of phonotactic constraints—may affect linguistic processing and subsequent perceptual decisions.

Two separate lines of research indicate that such contextual constraints may directly influence a listener's perception of phonetic segments. First, Ganong (1980) and later Fox (1984) demonstrated the effect of lexical context on the identification of word-initial stops. In each case, these researchers developed a phonetic continuum such as a VOT or place of articulation continuum which ranged from a word to a nonword (e.g., *tash—dash* or *task—dask*). Each found that the perceived categorical boundary of the phonetic continuum shifted in the direction of the nonword end of the continuum which meant that listeners were making more "word" than "nonword" responses, especially to the relatively ambiguous stimuli. Ganong suggested that the top-down effect influenced actual phonetic processing, although Fox argued that that effect could have resulted from post-perceptual biasing. Ignoring the question of the site of such effects, these data argue that the outcome of acoustic/phonetic processing may be affected by considerations of presence or absence of particular linguistic structures in English.

Second, Massaro and Cohen (1983) reported a study which was designed to examine the contribution of the phonological context *alone* to the perception of an initial glide. In one experiment Massaro and Cohen generated a phonetic continuum from [ri] to [li]. Tokens from this continuum were then placed in one of four phonetic contexts: following [p], [t], [s] or [v]. In English, both [ri] and [li] are permissible after [p] and neither are permissible after [v]. Following [t] only [ri] is permissible and following [s] only [li] is permissible. Massaro conducted an identification test to determine the extent to which the phonotactic constraints in English would affect listeners' labeling of the [ri]-[li] continuum. Their results were in line with interactive processing models of perception in which top-down information in the form of knowledge about permissible syllabic structures influenced the phonetic/phonological processing. In particular, listeners heard more [r] sounds following [t] and more [l] sounds following [s]. The effect was most pronounced when the [r-l] segment was most ambiguous. In addition, although listeners were run on the experiment on two separate days, the context effect *did not* appear to decrease with experience on the experimental task (Massaro and Cohen, 1983: 342) and thus does not represent simple unfamiliarity with the stimulus token.

The present study examined the extent to which a similar perceptual effect may occur in vowel rather than consonant perception and when the constraining context occurs following the stimulus rather than preceding it. Our study utilized the well-known neutralization of the tense/lax vowel distinction in English in syllables closed by a post-vocalic [r] (e.g., there is no significant [bir]-[bɪr] distinction in most dialects of American English). When making a vowel in this position, speakers will often produce a vowel intermediate between the two vowel qualities (Ladefoged, 1982). However, on the perceptual side, it is unknown as to whether this well-known contextual constraint limits the phonetic/auditory capabilities in the listener. Three experiments were directed at this question.

Experiment 1

The first experiment was designed to determine the extent to which listeners' ability to make phonetic and/or auditory distinctions was affected by the post-vocalic [r] context. This experiment utilized basic identification and AX discrimination tasks.

Method

Subjects. Thirty students from the Ohio State University participated in this experiment and received $3.00 for their participation. All were native speakers of a midwestern dialect of American English. The subjects were randomly assigned to one of the two listener groups. Each group of listeners completed both a forced-choice identification task and an AX (same/different) discrimination task for one of two stimulus continua ([hV] or [hVr]). To avoid possible response bias, listeners were not required to listen to both sets of stimuli.

Stimuli. Two different 13-step [ɪ-ɛ-æ] vowel continua were created. One was embedded within a [hV] context and one was within a [hVr] context. The stimuli were generating using the Klatt cascade/parallel synthesis program (Klatt 1980). The stepwise variations within the continua were produced by varying the frequencies of the first three formants.

The basic formant frequency values used in the synthesis of stimuli are shown in Table 1. These values were based on the formant values suggested by Klatt (1979) for [ɪ], [ɛ] and [æ] which represent continuum steps 1, 7, and 13, respectively. The formant values for the intervening vowel tokens were based on linear interpolation between either steps 1 and 7 or steps 7 and 13. Note that in order to produce more human-like tokens, the formant frequen-

[1] With one exception: step 13 was based on an earlier monophthongal version of [æ]. This created a somewhat exaggerated (and artifactual) difference in the AX discriminations between steps 11 and 13 which will have no effect on the basic conclusions reached.

cies changed over time[1] producing slightly diphthongized vowels. In particular, the formant values began at a given frequency and remained there for 150 ms at which time the frequency values changed linearly to the second formant value given in Table 1 at 350 ms. The [hV] tokens were 350 ms in duration while the [hVr] tokens were 470 ms in duration. The additional 120 ms duration of the [hVr] tokens represented a linear change from the final formant values given in Table 1 toward those appropriate for [r], that is, 420, 1310, and 1540 Hz, respectively. The frequencies of F4 and F5 were held steady at 3400 and 3900 Hz, respectively. The F0 frequency of each token began at 130 Hz and fell linearly to a value of 110 Hz at 350 Hz (where it remained steady for the duration of the [hVr] tokens). The [h] was created by shunting aperiodic noise through the cascade filters for the first 90 ms of each token.

Two identification tapes were made. Each identification tape contained 10 examples of each stimulus token from one of the stimulus continua in

Table 1. Formant frequency values used in the synthesis of both continua; all values are in Hz. Formants began at the first value given and remained at that value until 150 ms into the token at which time the frequency values changed linearly until reaching the second value 350 ms into the token. Each of the[hVr] tokens were 120 ms longer (470 ms in total duration). The frequencies of F1, F2 and F3 changed from the second frequency value noted below (at 350 ms) to 420, 1310, and 1540 Hz, respectively, at 470 ms.

Continuum Step	Formant 1	Formant 2	Formant 3
1	400-470	1800-1600	2570-2600
2	422-495	1780-1588	2558-2588
3	443-520	1760-1577	2547-2577
4	465-545	1740-1565	2535-2565
5	486-570	1720-1553	2523-2552
6	506-595	1700-1542	2512-2542
7	530-620	1680-1530	2501-2530
8	545-625	1675-1525	2520-2490
9	560-630	1672-1518	2578-2510
10	575-635	1669-1511	2466-2500
11	590-640	1666-1504	2454-2490
12	605-645	1663-1497	2442-2480
13	620-620	1660-1660	2430-2430

random order for a total of 130 test tokens. Each identification tape had a 10-item practice prior to the start of the test tokens. Two discrimination tapes were also constructed, one for each continuum. The discrimination tapes utilized a standard AX format. The tokens within each AX pair represented either an identical ("same") pair (e.g., step 1—step 1) or a pair of tokens differing by two steps (a "different" pair, e.g., step 1—step 3). The interstimulus interval (ISI) between the two tokens within a pair was 500 ms with a 2500 ms intrapair interval. Each tape contained 84 "different" and 84 "same" test pairs in random order for a total of 168 stimulus pairs.

Procedure. In the identification task, listeners were required to concentrate on the medial vowel of each stimulus token they heard. Listeners were told either that the identification tape contained speech tokens that began with an [h] and ended with a vowel, or that the identification tape contained speech tokens that consisted of the [h] followed by a vowel which was followed by [r]. Listeners identifying the [hVr] continuum were instructed to ignore the final [r] if at all possible. Listeners were asked to identify eách vowel as the vowel that occurs in either *hid, head* or *had.* Listeners completed the 10 practice items before starting the identification test proper.

In the discrimination tests, listeners were told to concentrate on the vowel in each pair and to indicate on the response form whether the vowels in the two tokens were the "same" or were "different" (in order to be the "same" listeners were instructed that two tokens had to be identical). Again, listeners in the [hVr] group were again warned to ignore the final [r] as much as possible. Both groups were told to expect approximately the same number of "same" and "different" pairs.

Results and Discussion

Identification . Panel a in Figure 1 shows the identification responses for all steps of the vowel continuum in the context [hV]. As shown, steps 1-3 are mostly identified as [ɩ], steps 5-7 are mostly [ɛ] and steps 11-13 are primarily identified as [æ]. Panel b shows the identification responses for all steps of the vowel continuum in the context [hVr]. As can be seen, steps 1-3, 5-7, and 11-13 are again identified the majority of the time as [ɩ], [ɛ], and [æ], respectively. Note, however, that the identification functions for the [hVr] group are somewhat "flattened" or "broadened" compared to those of the [hV] group, meaning that the change from one vowel category to another is not as sharp in the [hVr] context as it is in the [hV] context.

Differences in the identification responses between the two contexts can be more easily seen in Figure 2. As can be seen in panel a, the number of [ɩ] responses from the post-vocalic [r] group is smaller, even for the endpoint stimuli (step 1). Panel b shows the percentage of [ɛ] responses to all stimulus tokens. There are noticeable differences between the two groups in their responses to steps 1-7 and steps 9-13 of the vowel continua. It is clear that many more of the stimuli are being identified by the [hVr] group as [ɛ] than by the [hV] group. It thus appears that some of the distinctiveness between the vowel qualities has been lost due to the addition of the post-vocalic [r] (though no other change in the stimulus vowel was made). Panel c in Figure 2 shows the percent of [æ] responses to all stimulus tokens. The primary difference between the two context groups appears between continuum steps 9-13. Many vowels which were identified as [æ] in the [hV] context are being identified as [ɛ] in the context of a postvocalic [r]. Overall the figures support the contention that the presence of the post-vocalic [r] is significantly affecting listener's perception of vowel quality. In particular, the two groups show clear difference between the [ɩ] and [ɛ] categories, but the [hVr] group shows a reduced distinction between [ɛ] and [æ].

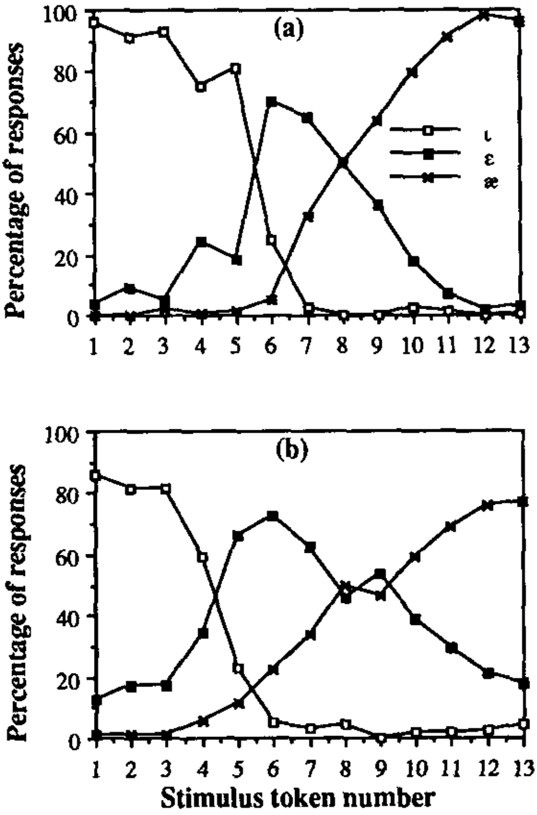

Figure 1. Identification responses of vowels in the [hV] continuum (panel a) and [hVr] continuum (panel b).

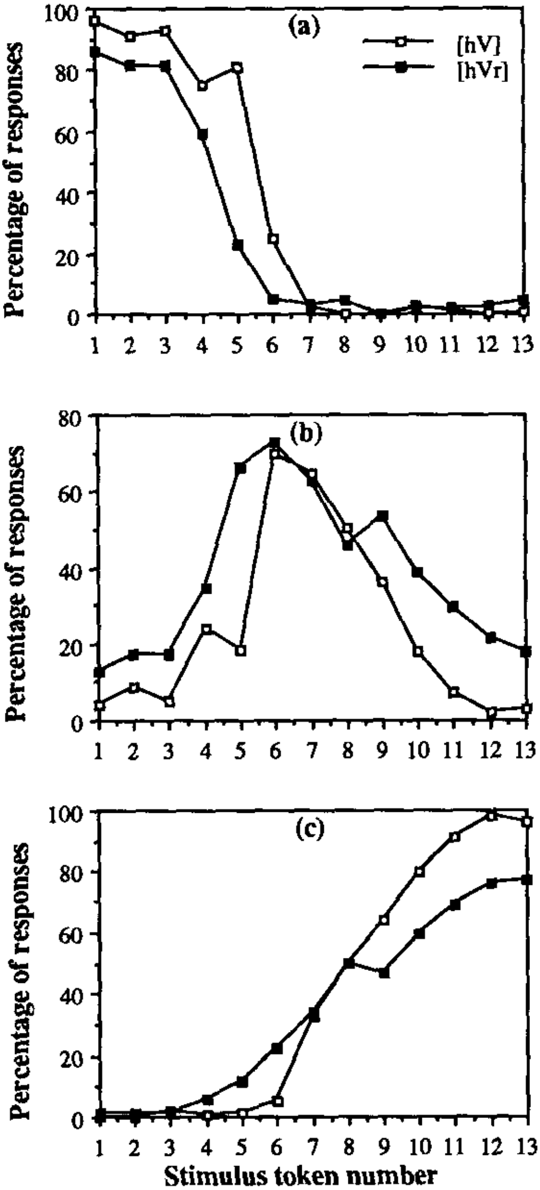

Figure 2. Identification responses of the [hV] and [hVr] continuua broken down into [ɪ] (panel a), [ɛ] (panel b) and [æ] responses (panel c).

Discrimination. The presence of a post-vocalic [r] thus affects the identification of the vowels, but will it also affect the ability of listeners to discriminate between different vowel qualities? The results of the discrimination tests are shown in Figure 3. Shown are the percentage of "same" responses for identical stimulus pairs. The percentage of same responses are high for both groups and there are few discernible differences.

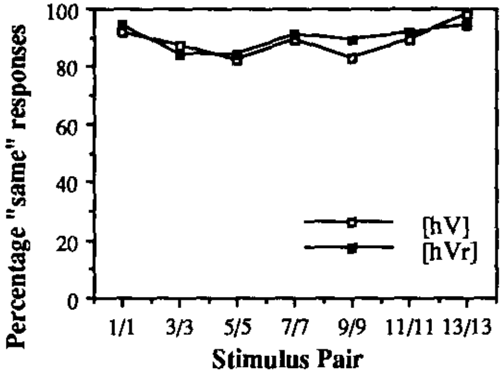

Figure 3. Discrimination responses to "same" pairs.

Shown in Figure 4 are the percentages of "different" responses for all 2-step different pairs. The responses are very similar for both groups to pairs 1/3 and 3/5. The 1/3 pair represent the [ɩ]-[ɩ] within-category comparison and the 3/5 pair represents the [ɩ]-[ɛ] between-category difference. As expected, listeners do very poorly on the within-category pair and very well on the between-category difference. However, the number of "different" responses for the remainder of the token pairs[2] is smaller in the [hVr] group than the [hV] group. In particular, the listeners in the post-vocalic [r] group seem to have more difficulty in making accurate discriminations. To emphasize this, Figure 5 is a plot of the difference between the discrimination accuracy of the [hV] group and the [hVr] group. A two-way analysis of variance with the factors stimulus pair and listener group of the discrimination responses showed a main effect of stimulus pair (F(5,149)=34.9, p<.001) but no main effect of listener group (F(1,149)=2.35, p>.13). However, there was a significant stimulus pair x listener group interaction (F(5,149)=6.0, p<.001) which demonstrated that although the two listener groups had the same pattern of responses for the [ɩ]-like tokens, they had a significantly different pattern for the [ɛ]- and [æ]-like tokens. This pattern of discrimina-

[2] The 11/13 pair is not shown; see footnote 1.

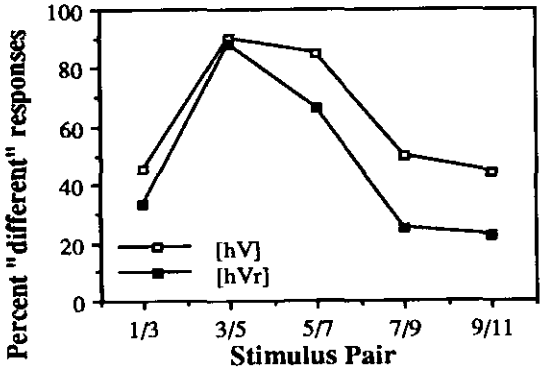

Figure 4. Discrimination responses to "different" pairs.

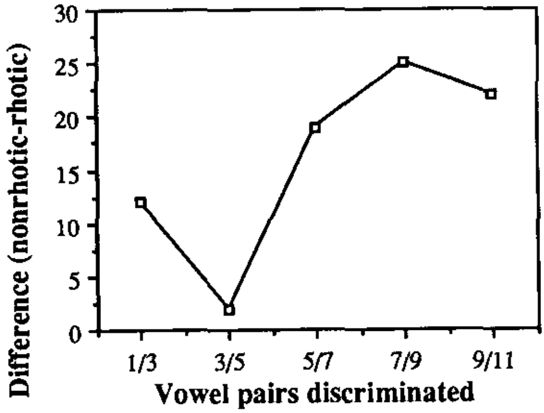

Figure 5. Difference in discrimination responses ([hV]-[hVr]).

tion would support the contention that phonological neutralization is, in fact, affecting listeners' abilities to make phonetic and/or auditory distinctions.

However, we still have not adequately addressed the question of whether the effect is at the categorical (e.g., phonological) level alone (which would include the syllabic level favored by Massaro and Cohen, 1983), at a subphonemic level alone, or both. In addition, although listeners were specifically told the nature of the context and were also told to concentrate on the vowel quality alone, the effect could represent some post-perceptual process, such as simple response bias (see Fox, 1984). Experiment 2 was designed, in part, to address these questions using a multidimensional scaling procedure, a methodology which does not require subjects to explicitly categorize (or phonetically identify) the stimulus tokens and which has been used successfully elsewhere in examining subphonemic differences among vowel stimuli (e.g., Fox, 1985; Kewley-Port and Atal, 1989). In Experiment 2, although listeners were required to evaluate the similarity or dissimilarity between different vowel pairs, they were never required to identify the vowels nor to categorize them in any way.

Experiment 2

Experiment 2 utilized a basic multidimensional scaling procedure to first obtain estimated perceptual distances between every vowel pair in two sets of vowel stimuli. These perceptual distances were then analyzed to determine the optimal perceptual space accounting for the distances. Of particular concern here was the extent to which a post-vocalic [r] would disrupt or modify the "normal" vowel space.

Method

Subjects. Thirty-three listeners who spoke a midwestern version of American English participated in Experiment 2. All were undergraduate or graduate students at the Ohio State University and received $3.00 for their participation.

Stimuli. The basic stimuli consisted of 2 sets of vowel tokens in the context [hVd] and [hVrd]. Each set of tokens included the vowels [i ɩ eɩ ɛ æ a ʌ oɷ ɷ u] and were produced (and evaluated) by a trained phonetician. Care was taken to produce similar initial vowel quality on all tokens. These tokens were then low-pass filtered at 4.8 kHz, digitally sampled at 10 kHz (with a 12-bit resolution) and stored on a computer disk. The formant frequencies of each vowel was measured using a Kay DSP 5500 spectrograph and a formant 1 by formant 2 plot of the initial formant values for each set of stimuli (measured at a point 50 ms past initial voicing) appears in Figure 6. As can be seen, the acoustic structure for the two sets of vowels is remarkably similar. The mean duration of the syllabic nuclei in each token was 397 ms in the [hVd] context and 440 ms in the [hVrd] context, including the post-vocalic [r]. In the [hVrd] tokens, the initial portion of all vowels was nonrhotic (third formant was not lowered) and subsequently not r-colored for an average of 211 ms (approximately 48% of the token duration of the syllable nucleus). The mean amplitudes of the tokens were equalized within a 2 dB range.

In creating the stimulus tape for each of the two contexts, each stimulus

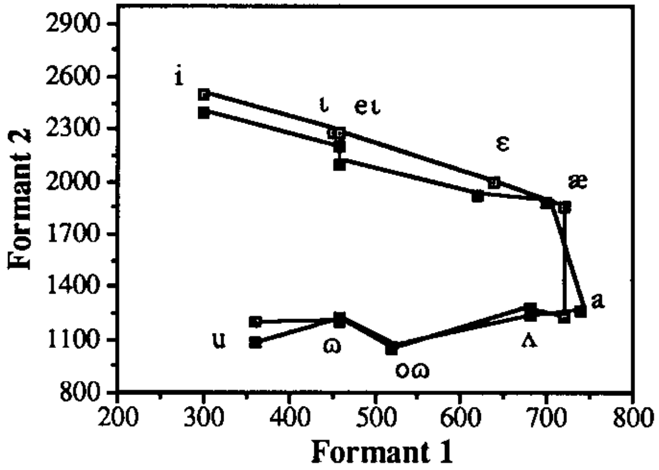

Figure 6. Formant 1 x Formant 2 plot of the stimulus vowels used in Experiment 2.

token was paired with all other tokens (excluding itself) producing a total of 90 different vowel pairs. Two different blocks of 90 vowel pairs (separately randomized) were produced for each stimulus set. Thus in each context condition, a listener made similarity judgments on 180 vowel pairs. The generation of the stimulus tape was under computer control with low-pass filtering (at 4.8 kHz) at a 10 kHz sampling rate. The vowel pairs were output with a 500 ms interstimulus interval and approximately 4 s between each different vowel pair.

Procedure. This study required listeners to hear two sequentially presented stimulus tokens and to judge their similarity/dissimilarity on a 9-point scale (Fox, 1983). Listeners indicated their judgments on the prepared response sheets by circling the appropriate point on the prepared rating response sheets. They were instructed (orally and in writing) to ignore all but vowel quality differences between the two vowels in each trial. In the post-vocalic [r] context they were further instructed to ignore the presence of the [r] in each stimulus token and to make their judgments based on the quality of the initial portions of each token. Listeners were also cautioned to use the entire range of the 9-point scale. Listeners practiced on an introductory block of 10 stimulus trials. After determining that the listeners were indicating their responses in the correct manner (and were using most or all of the range of the rating scale), the experimenter answered any questions that the listener might have.

Each listener's similarity/dissimilarity judgments were checked for consistency before his/her data were included in the multidimensional scaling analysis. This was done to ensure that excess "noise" was not included in the perceptual distances submitted to ALSCAL analysis. As Terbeek (1977) cautioned, excessive noise in the data may cause a solution of too many dimensions to appear more formally acceptable (using various criteria) than a (more correct) solution of fewer dimensions. In addition, excess noise in the data may obscure any disruptive effect of the post-vocalic [r]. This consistency check involved comparing a listener's responses on one random block of trials to his/her responses on the second—for both sets of data. A listener's data were used only if a Pearson's *r* was significant at the 0.01 level or beyond for both sets of stimuli. Eight listeners failed to meet this criterion and their data were eliminated from ALSCAL analysis.

ALSCAL Analysis. These perceptual data were analyzed using the flexible multidimensional scaling program ALSCAL (Young and Lewychkyj, 1979). Like most multidimensional scaling programs ALSCAL assumes that the scaled similarity judgments can be viewed as representative of the "perceptual distances" among the stimuli in an underlying *n*-dimensional perceptual space. The goal of the multidimensional scaling program is to determine this underlying space, producing a coordinate value for each scaled object (here vowels) for each separate perceptual axis or dimension. In addition, many versions of multidimensional scaling can produce so-called "subject weights" for each listener for each perceptual dimension which indicates the relative salience of that dimension for that particular listener's responses.

Each of the sets of response data (one for each context) were analyzed. Two separate types of analysis were used in each case. First, solutions in 1-4 dimensions were obtained using the INDSCAL mode of ALSCAL, which allows one to determine the extent to which each perceptual dimension was utilized by each particular listener. In this way we could evaluate whether the perceptual space was an accurate reflection of all listeners' data. Second, mean perceptual distances were calculated across listeners for each of the two sets of data and inserted into a single perceptual distance matrix (referred to in the following as the "mean" data). In turn, these two matrices were analyzed using the Euclidean mode of ALSCAL and may provide a better overall representation of the underlying perceptual space in that it represents even further elimination of intersubject variability. For both types of scaling, both interval (i.e., metric) and ordinal (i.e., nonmetric) solutions were obtained. Since the ordinal solutions accounted for more of the variance than did the interval solutions and since the perceptual distances that listeners produce can be considered to represent ordinal- rather than interval-level data, only the nonmetric solutions will be discussed here.

Results and Discussion

The two-dimensional space was chosen as the solution to evaluate for both the nonrhotic and rhotic contexts. This decision was based on s-stress, variance accounted for and interpretability of the solution. The two-dimensional solution for the mean [hVd] data produced an s-stress of 0.063 and accounted for 97.8% of the data. In addition, this solution was easily interpretable in terms of well-known phonetic/phonological characteristics of the stimulus vowels. The two-dimensional solution for the mean [hVrd] data produced an s-stress value of 0.026 and accounted for 99.7% of the variance. Examination of the INDSCAL-mode versions of the two-dimension solutions showed that no listener had a squared subject weight lower than 0.48 for dimension 1 or 0.16 for dimension 2 (most were far higher). The subject weights obtained are shown in Figure 7. The solutions to be discussed seem to accurately represent the perceptual judgments to both sets of data of the listeners as a group as well as the listeners individually.

Shown in Figure 8 are D1 by D2 plots of the 2-dimensional ALSCAL solutions for the [hVd] and [hVrd] stimulus sets. As can be seen, the vowel space for the [hVd] stimuli can be interpreted and labeled easily in terms of the common vowel distinctions of high/low and front/back. D1 (front/back) was the first and most salient dimension to emerge and it, along with D2, easily recreates the vowel quadrangle. This solution is perfectly compatible with other multidimensional scaling studies using a similar set of stimuli (e.g., Fox, 1983, 1985; Fox and Trudeau, 1988; Singh and Woods, 1970). The D1 by D2 plot of the [hVrd] stimulus solution, however, shows some disruption in the placement of the vowels. In particular, note the decreased distance between [i] and [ι], between [ɛ] and [æ], and between [ω] and [u]. In addition, D2 might better reflect a round/nonround dimension rather than a front/back dimension as both the rounded and unrounded

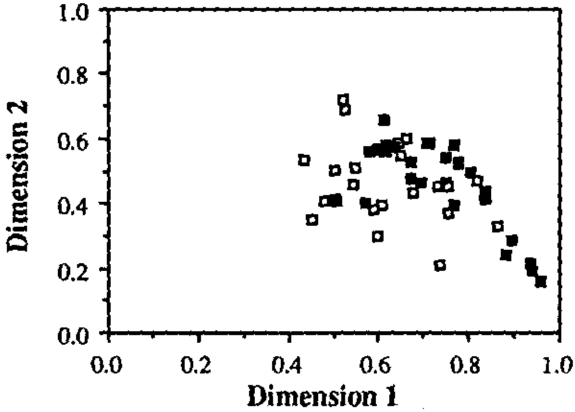

Figure 7. Plot of the ALSCAL subject weights obtained for the nonrhotic (open squares) and rhotic (closed squares) 2-dimensional solutions.

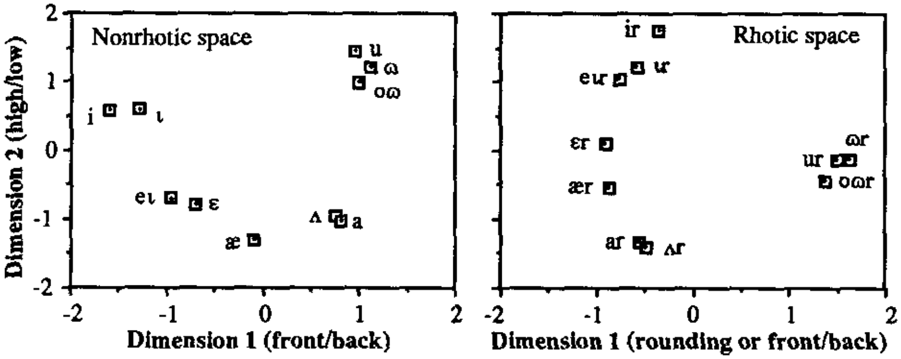

Figure 8. Dimension 1 x Dimension 2 plots of the ALSCAL solutions for the nonrhotic ([hVd]) and rhotic ([hVrd]) stimulus sets.

vowels cluster more closely together on D2 than in the [hVd] vowel space. Thus we can suggest that the front/back distinction is noticeably reduced in this particular context.

These results again suggest that the phonotactic constraint on vowel quality imposed by the presence of the post-vocalic [r] did exert an influence on listener's perceptual judgments. Since it is possible that listeners *can* base their similarity judgments on subphonemic differences between vowel tokens (Fox, 1985; Kewley-Port and Atal, 1989), we may thus have a perceptual effect which occurs without the *explicit* use of phonetic labels (as in the identification task) or *implicit* use of such labels (in the discrimination task, the traditional explanation of categorical perception). This may argue against a simple response-bias explanation of the results from Experiment 1.

However, one possibility that we have not addressed, in terms of explaining the perceptual disruption of the vowel space in the post-vocalic [r] context, is that although the initial formant values of the [hVd] and [hVrd] tokens are similar, the resulting vowel quality—even at vowel onset—may be sufficiently different so as to produce the perceptual space differences obtained. Given the results from Experiment 1, we do expect that the identifications of the vowels *will* be modified by the post-vocalic [r], but Experiment 3 will address both issues.

Experiment 3

Experiment 3 used a standard identification procedure to examine the abilities of listeners to identify the vowel quality of the vowel nuclei used in Experiment 2. We would like to determine (1) whether the onset vowel quality of the two sets of vowels (i.e., *prior* to any lowering of the third formant for the production of post-vocalic [r]) differe significantly and (2) whether presence of the post-vocalic [r] when the whole entire syllable nucleus is heard will again significantly affect a listener's ability to identify vowel quality—given natural human speech.

Method

Subjects. Ten listeners who spoke a midwestern dialect of American English participated Experiment 3 and were paid $3.00 for their participation.

Stimuli. The stimulus tokens in this experiment were digitally edited versions of the tokens used in Experiment 2. One set of tokens represented excised syllable nuclei from each [hVd] and [hVrd] token. These nuclei were created by excising all energy prior to the onset of voicing (which represents the [h]) and all energy following the onset of the final voiced stop. These syllable nuclei will be referred to as the "long" stimuli. A second set of tokens was created which eliminated all [r]-coloring from the stimuli. In particular, each of the syllable nuclei (the rhotic vowels) from the [hVrd] tokens were analyzed (using FFTs) to determine the spectral structure of the token over time. We determined the location of the start of rhotacization (i.e., lowering of F3) and created a stimulus consisting of the initial, nonrhotic portion of the syllabic nucleus. On the average, the nonrhotic portion of the vowel represented the first 48% of the syllable nucleus. These will be referred to as the "short" stimuli. Since there could be an effect due to vowel shortening alone, short versions of the nonrhotic syllable nuclei were also created. These consisted of the initial portion of the nucleus whose duration was the same as the short version of the vowel's short rhotic counterpart.

A stimulus tape was created which included 10 examples of both sets of long and short tokens in random order. Prior to the start of the test tokens, there were 10 practice tokens.

Procedure. The experimental task was a forced-choice identification task. Listeners were required to identify each stimulus token as the vowel in the words *heed, hid, hayed, head, had, hod, hoed, hood, who'd* or*hud.* Listeners were cautioned that approximately half of the vowels they were to hear were long while the other half were short. In addition, they were warned that 25% of the vowel tokens ended in a [r] sound which they were to ignore as well as they could.

Results and Discussion

The results of the identification tests are shown in Figure 9. As can be seen, the perception of the rhotic short vowels is extremely similar to both

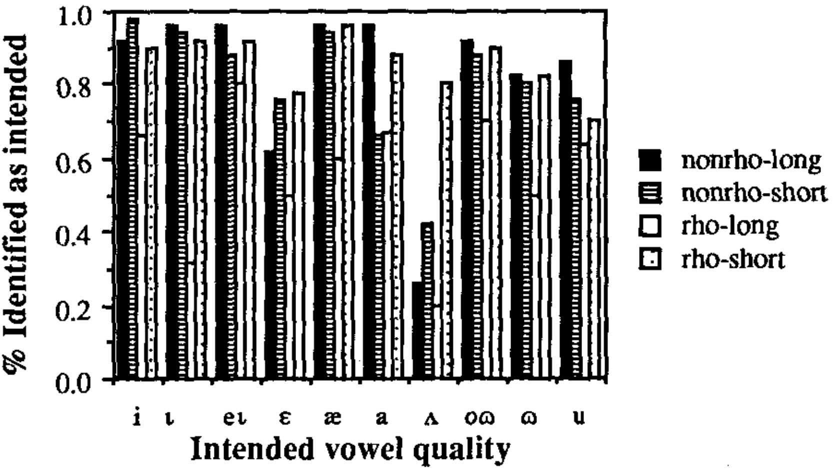

Figure 9. Identification results obtained from Experiment 3.

the nonrhotic short vowels and the nonrhotic long vowels. Only when the final [r] occurs in the stimulus token are listeners' patterns of identification responses different. This demonstrates two things: First, the onset vowel quality of the two sets of tokens used in the scaling task are equivalent and should not have contributed significantly to the perceptual space differences found in Experiment 3. Second, as in Experiment 1, listeners' identification of vowel quality is affected by the presence of a post-vocalic [r] even when listeners are explicitly warned to expect it.

Although the first three experiments have demonstrated a significant effect of context, two questions have not as yet been addressed: First, what is the time course of such an effect? Is is a continuous effect, subject to the continuous uptake of acoustic information as noted by Warren and Marslen-Wilson (1988) or categorical all/none? Second, and more problematic, is this effect merely auditory/phonetic? That is, perhaps perceived vowel quality is been affected merely because of the addition of an acoustic signal in the offglide (in the form of the [r]) not normally associated with a given vowel. To address these questions, Experiment 4 examines the perception of vowel quality in both a post-vocal [r] and post-vocalic [l] contexts. In addition, it examines the time-course of any obtained contextual effect.

Experiment 4

Method

Subjects. Fifteen listeners who were native speakers of a midwestern dialect of American English participated in Experiment 4. These listeners were recruited from an introductory phonetics class and had been introduced to the basic nature of phonetic transcription and the concept of vowel quality distinctions in English. The nature of the stimulus offsets were described to each listener and they were warned to try to ignore the contextual changes in making their vowel identifications ("try to ignore any influence of the partial or complete /l/ or /r/ offsets"). Each listener received class credit for their participation.

Stimuli. The stimulus tokens used in the final experiment were constructed from a set of tokens consisting of the vowels [i ι e ε æ] in the contexts [h_], [h__r] and [h__l]. A trained phonetician produced several sets of these triads for each vowel attempting to make the fundamental frequency, overall length, and beginning vowel quality the same. These stimuli were then analyzed using the Kay 5500 and a set of matched tokens selected for each vowel quality. For each token selected, the onset vowel quality did not change for at least 150 ms.

A "baseline" token for each vowel quality was created which represented the first 150 ms of the static [hV] token. Next, a set of [Vr] and [Vl] offsets were created which represented the last portion of the steady state vowel until the offset of either the [hVr] or [hVl] tokens. Through the original controlled production of these tokens and careful digital editing, the durations of these "token endings" were the same within each vowel quality (though, of course, not across vowel quality). A set of stimulus tokens was then created by appending 20, 40, 60, 80 or 100% of these token endings to the appropriate baseline token. The resulting tokens had perceptibly smooth, uninterrupted transitions from the baseline to the appended ending. Finally, to eliminate clicks associated with rapid offsets of acoustic energy, the amplitude of the final 10 ms of each token was linearly ramped to zero. There were a total of 55 different tokens.

A stimulus tape was created which contained 8 examples of each token in random order. There were 10 practice items. In addition, at the beginning of the tape were 20 example tokens illustrating the different contexts that would be heard.

Procedure. Listeners heard each stimulus token and were required to identify its vowel quality by circling one of the following on the answer sheet: *hi hι heι hε hæ.* The listeners were familiar with these IPA symbols and had had 5 weeks of practice in phonetic transcription.

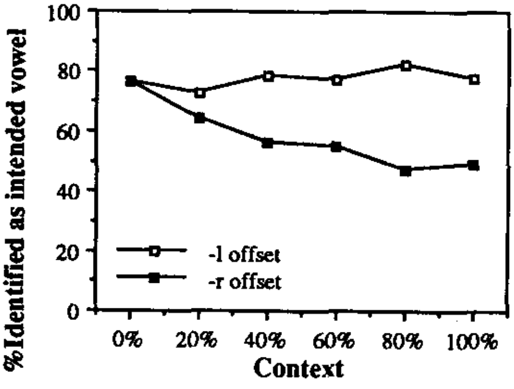

Figure 10. Mean vowel identification obtained in Experiment 4 with the [Vl] (open squares) and [Vr] (closed squares) offsets.

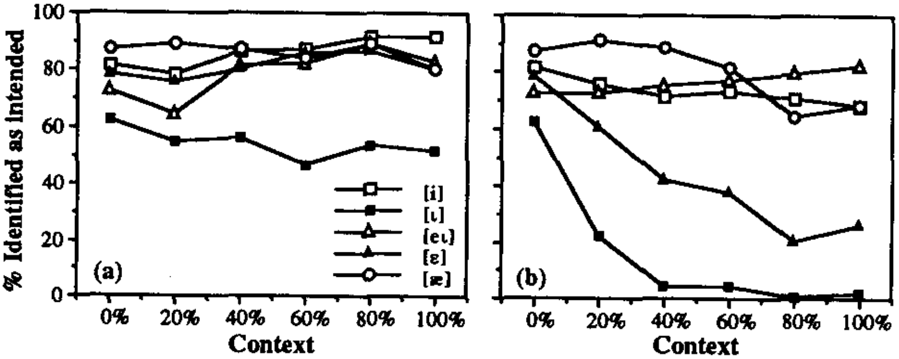

Figure 11. Identification of separate vowel qualities with the addition of the [Vl] offset (panel a) and [Vr] offset (panel b).

Results and Discussion

Figure 10 shows the mean vowel identifications in the two different contexts (plus the baseline token without any appended ending). These data show that, overall, the [Vl] context (which does not phonotactically constrain allowable vowel quality) did not have a significant effect on the perceived vowel quality of the stimulus token as did the presence of postvocalic /r/. This reduces the possibility that the effect described in Experiment 1-3 is merely auditory or phonetic in nature (e.g., a product of backward masking).

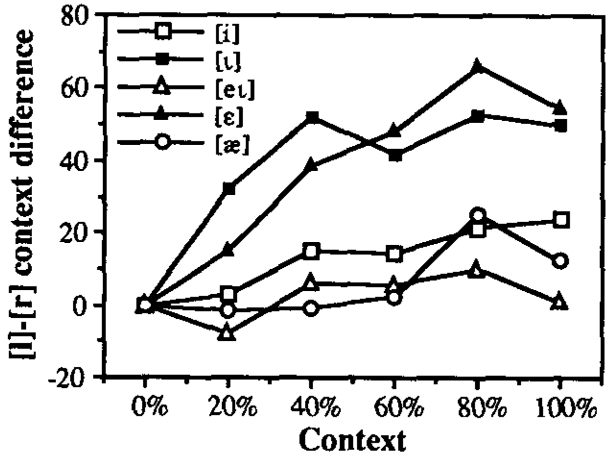

Figure 12. Differences in vowel identification between the [h__l] and [h__r] contexts.

Shown in Figure 11 are the mean identifications of each individual vowel in both [h__l] and [h__r] contexts. In the [h__l] context although some vowels are identified more often as the intended vowels than others (e.g., compare the identification of [i] with [ι] in Figure 11, panel a), the gradual addition of the final [l] offset seems to affect the identification functions very little. On the other hand, the identification of the vowels is significantly affected by the gradual addition of the [r] offset (as shown in panel b). The identification of the tokens with intended [ι] and [ɛ] was affected the most. For each of the vowels the effect of context was noticeable following the addition of even the shortest token ending (20%) and this effect increased as more of the ending was added.

Figure 12 shows the difference between the identifications in the [h_l] contexts and those in the [h_r] contexts. The differences were calculated by subtracting the percentage of intended identification of the [h_r] context from the [h_l] context. As can be seen, the [h_r] context produces fewer intended identifications in practically all cases—especially for the vowels [ι] and [ɛ].

General Discussion

On the basis of these data one can claim that knowledge of a language's morphophonemic rules—for example, in the form of phonotactic constraints—may affect the ability of listeners to make perceptual decisions about vowel quality. The next logical problem is to determine the site or processing level of this effect. The identification data could be explained on the basis of top-down effects directly interacting with the categorization process or on the basis of post-perceptual biasing (Fox, 1984; McQueen, 1989). However, the discrimination data—which did not require listeners to explicitly identify the data in making the same/different judgments—are less compatible withthe biasing explanation. The multidimensional scaling data would suggest that the top-down effects occur at least at the identification level and perhaps earlier (with auditory images or sub-phonemic representations of the input stimuli).

In this study I have taken the position that the inability of listeners to ignore the effect of the postvocalic [r]—despite detailed instructions and a relatively limited set of stimulus tokens—reduces the likelihood that these contextual effects are a result of "simple response bias" of the sort claimed by Sieb Nooteboom during a discussion of the oral version of this paper. It does not seem to be the case that listeners' response patterns are derived from an desire to respond in a manner compatible with English (a *response* bias). Rather, it seems that listeners are actually unable to perceive the vowel distinctions being made (which could be termed a *perceptual* bias).

One auditory processing explanation for the obtained results which has not been explored here might claim that the phonetic disruption produced by the presence of the post-vocalic [r] could have stemmed from a backward masking effect. That is, the presence of the [r]-offglide inhibited processing of the preceding vowel. This possibility should be explored experimentally and evaluated, but the ability of English listeners to make vowel distinctions in the the context of a post-vocalic [l] would argue against it.

In summary, the data suggest that knowledge of a language's phonological rules—for example, in the form of phonotactic constraints—may affect the ability of listeners to make perceptual decisions about vowel quality. The next logical problem is to determine the site or processing level of this effect. The identification data could be explained on the basis of top-down effects directly interacting with the categorization process or on the basis of post-perceptual biasing (Fox, 1984; Luce and Pisoni, 1985; McQueen, 1989). However, the discrimination data—which did not require listeners to explicitly identify the data in making the same/different judgments—are less compatible with a biasing explanation. The multidimensional scaling data would suggest that the top-down effects occur at least at the identification level and perhaps earlier (with auditory images or subphonemic representations of the input stimuli.

References

Akagi, M. (this volume, 1991). Psychoacoustic evidence for contextual effect models.

Connine, C. & Clifton, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, **13**, 291-299.

Cowan, N. and Morse, P.A. (1986). The use of auditory and phonetic memory in vowel discrimination. *Journal of the Acoustical Society of America*, **79**, 500-507.

Crowder, R.G. and Repp, B.H. (1984). Single formant contrast in vowel identification. *Perception and Psychophysics*, **35**, 372-378.

Elman, J.L. & McClelland, J.L. (1986). Exploiting lawful variability in the speech wave. In J.S. Perkell & D.H. Klatt (eds.) *Invariance and Variability in Speech Processes*. Hillsdale, NJ: Erlbaum. Pp. 360-385.

Elman, J.L. & McClelland, J.L. (1989). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. Unpublished manuscript.

Fowler, C.A. (1984). Segmentation of coarticulated speech and perception. *Perception & Psychophysics*, **36**, 359-368.

Fowler, C.A. & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In: J. Perkell & D.H. Klatt (eds.), *Invariance and Variability of Speech Processes*, Hillsdale NJ: Erlbaum. 123-160.

Fox, R.A. (1983). Perceptual structure of monophthongs and diphthongs in English. *Language & Speech*, **26**, 21-60.

Fox, R.A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 526-540.

Fox, R.A. (1985). Multidimensional scaling and perceptual features: Evidence of stimulus processing or memory prototypes? *Journal of Phonetics*, **13**, 205-217.

Fox, R.A. & Trudeau, M. (1988). A multidimensional scaling study of esophageal vowels. *Phonetica*, **45**, 30-42.

Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, **6**, 110-125.

Grosjean, F. (1980). Spoken word recognition and the gating paradigm. *Perception & Psychophysics*, **28**, 267-283.

Kewley-Port, D. & Atal, B.S. (1989). Perceptual differences between vowels located in a limited phonetic space. *Journal of the Acoustical Society of America*, **85**, 1726-1740.

Klatt, D.H. (1980). Synthesis of consonant-vowel syllables. Unpublished manuscript, MIT.

Ladefoged, P. (1982). *A Course in Phonetics, Second Edition*. New York: Harcourt Brace Jovanovich.

Marslen-Wilson, W. & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition*, **8**, 1-17.

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

Massaro, D.W. & Cohen, M.M. (1983). Phonological context in speech perception. *Perception & Psychophysics*, **34**, 338-348.

McClelland, J.L. & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.

McQueen, J.M. (1990). The influence of phonetic categorization: The critical case of word-final stimuli. Manuscript in review.

Morton, J. (1969). Interaction of information on word recognition. *Psychological Review*, **76**, 165-178.

Nakatani, L.H. & Schaffer, J.A. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America*, **63**, 234-245.

Nearey, T. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of* *the Acoustical Society of America*, **85**, 2088-2113.

Repp, B.H., Healy, A.F., and Crowder, R.G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, **5**, 129-145.

Repp, B. & Liberman, A. (1984). Phonetic categories are flexible. *Status Report on Speech Research Haskins Laboratories*, SR-77/78, 31-53.

Samuel, A.G. (1986). The lexicon in speech perception. In E.C. Schwab & H.C. Nusbaum (eds.), *Pattern Recognition by Human and Machines: Volume 1, Speech Perception*, Orlando, FL: Academic Press.

Singh, S. & Woods, G. (1971). Perceptual structure of 12 American English vowels. *Journal of the Acoustical Society of America*, **52**, 1698-1713.

Shigeno, S. (1986). The auditory tau and kappa effects for speech and nonspeech stimuli. *Perception and Psychophysics*, **40**, 9-19.

Shigeno, S. and Fujisaki, H. (1979). Effect of a preceding anchor upon the categorical judgment of speech and nonspeech stimuli. *Japanese Psychological Research*, **21**, 165-173.

Shigeno, S. and Fujisaki, H. (1980). Context effects in phonetic and non-phonetic judgments. *Annual Bulletin RILP*, **14**, 217-224.

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

Warren, P. and Marslen-Wilson, W. (1988). Cues to lexical choice: Discriminating place and voice. *Perception and Psychophysics*, **43**, 21-30.

Warren, R.M. & Sherman, G. (1974). Phonemic restorations based on subsequent context. *Perception & Psychophysics*, **16**, 150-156.

Young, F.W. & Lewychkyj, R. (1979). *ALSCAL.4 User's Guide: A Guide for Users of ALSCAL.4—A Nonmetric Multidimensional Scaling and Unfolding Program with Several Individual Differences Options*. Carrboro NC: Data Analysis and Theory Associates.