Article

Language *and* Speech

Perception of English Vowels by Bilingual Chinese–English and Corresponding Monolingual Listeners

Language *and* Speech
2014, Vol. 57(2) 215–237

DOI: 10.1177/0023830913502774
las.sagepub.com
SAGE

Jing Yang and Robert A Fox
The Ohio State University, USA

Abstract

This study compares the underlying perceptual structure of vowel perception in monolingual Chinese, monolingual English and bilingual Chinese–English listeners. Of particular interest is how listeners' spatial organization of vowels is affected either by their L1 or their experience with L2. Thirteen English vowels, /i, ɪ, e, ɛ, æ, u, ʊ, o, ɔ, ɑ, ɔɪ, aɪ, aʊ/, embedded in /hVd/ syllable produced by an Ohio male speaker were presented in pairs to three groups of listeners. Each listener rated 312 vowel pairs on a nine-point dissimilarity scale. The responses from each group were analyzed using a multidimensional scaling program (ALSCAL). Results demonstrated that all three groups of listeners used high/low and front/back distinctions as the two most important dimensions to perceive English vowels. However, the vowels were distributed in clusters in the perceptual space of Chinese monolinguals, while they were appropriately separated and located in that of bilinguals and English monolinguals. Besides the two common perceptual dimensions, each group of listeners utilized a different third dimension to perceive these English vowels. English monolinguals used high-front offset. Bilinguals used a dimension mainly correlated to the distinction of monophthong/diphthong. Chinese monolinguals separated two high vowels, /i/ and /u/, from the rest of vowels in the third dimension. The difference between English monolinguals and Chinese monolinguals evidenced the effect of listeners' native language on the vowel perception. The difference between Chinese monolinguals and bilingual listeners as well as the approximation of bilingual listeners' perceptual space to that of English monolinguals demonstrated the effect of L2 experience on listeners' perception of L2 vowels.

Keywords

Multidimensional scaling, vowel perception, bilingualism, second-language acquisition

Corresponding author:
Jing Yang, 110 Pressey Hall, 1070 Carmack Rd, Department of Speech and Hearing Science, The Ohio State University, Columbus, OH 43210, USA.
Email: yang.1198@osu.edu

Introduction

Cross-language speech perception has been widely examined on the basis of selected phonetic segments or segmental contrasts, which vary in the degree of similarity to segmental properties in the native language (L1). Focusing on individual segments, research has shown that adult listeners often perceive non-native speech sounds differently from native speakers of that language (Flege, MacKay, & Meador, 1999; Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Levy & Strange, 2008; Polka, 1995). However, only a few studies have examined a detailed perceptual profile of complete segmental inventories and compared them cross-linguistically in order to draw broader conclusions about systemic perceptual organization of speech sounds. The purpose of this study is to gain more insight into the perceptual organization of native versus non-native (L2) vowel systems. We examine the underlying perceptual spaces utilized by bilingual Chinese–English listeners and corresponding monolingual English and Chinese listeners in perceiving a relatively complete set of American English vowels. By comparing the underlying perceptual spaces of the three groups of listeners, we aim to understand how listeners' spatial organization of vowels is affected either by their L1 or their experience with L2.

Among all the factors affecting cross-linguistic vowel perception, the acoustic-phonetic features of vowels provide the physical basis for listeners' identification or discrimination. As previous studies have demonstrated, listeners' perception of vowels across different languages may be affected by phoneme inventory size, distances between vowels in the acoustic space or dynamic vowel properties such as the inherent spectral change over time (e.g., Flege, Munro, & Fox, 1994; Fox, Flege, & Munro, 1995; Jacewicz & Fox, 2012; Nishi, Strange, Akahane-Yamada, Kubo, & Trent-Brown, 2008; Rogers, Glasbrenner, DeMasi, & Bianchi, 2013; Strange, Bohn, Trent, & Nishi, 2004). In addition, speech perception, in general, and vowel perception in particular, is also influenced by listeners' L1 and experience with other languages. Earlier work found that listeners' perception of speech sounds was shaped in a language-specific manner in the first year of life (e.g., Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984). Long-term experience with L1 shapes a listener's perceptual space and predisposes him to "assimilate" acoustically similar L2 vowels into L1 categories (Bradlow, 1993; Butcher, 1976; Flege, Bohn, & Jang, 1997; Jusczyk, 1993; Scholes, 1967, 1968; Terbeek, 1977). Other than the native language background, the amount of exposure to L2 may also affect listeners' perceptual vowel space (e.g., Bohn & Flege, 1990; Cebrian, 2006; Frieda & Nozawa, 2007; Højen & Flege, 2006; Levy & Strange, 2008).

Multidimensional scaling (MDS) will be employed in the present study to determine the perceptual spaces of three groups of listeners. MDS has been used in speech perception research to uncover the basic perceptual dimensions underlying vowel categorization. This approach involves obtaining similarity/dissimilarity judgments of all non-identical pairs of vowel tokens. In turn, these judgments are viewed as a reflection of the underlying perceptual distances between the two stimuli in each pair. A model of the underlying perceptual space is then generated as a function of this set of pairwise perceptual distances. In particular, the stimuli are placed in an *n*-dimensional space such that the distances among the vowels optimally correspond to the obtained perceptual distances from the similarity judgments. In general, the axes of this *n*-dimensional space are interpreted as representing the strategies that the listeners use to identify or discriminate the stimuli.

MDS was initially used to discover listeners' perceptual dimensions in their L1 (Fox, 1978, 1982, 1983; Pols, van der Kamp, & Plomp, 1969; Shepard, 1972; Singh & Woods, 1971) and, subsequently, to examine cross-language perception (Fox et al., 1995; Gandour & Harshman, 1978; Takane & Sergent, 1983; Terbeek, 1977; Terbeek & Harshman, 1971). Two of these studies

are of particular relevance to the present work. First, Terbeek (1977) tested the perception of a set of 12 monophthongs by listeners from English, German, Thai, Turkish and Swedish backgrounds. The vowel stimuli represented a hybrid of selected vowels from the same five languages, which did not represent the vowel inventory of any particular language. The multidimensional analyses indicated that both language-universal and language-specific rules were involved in cross-language vowel perception. On one hand, all listeners utilized common perceptual dimensions that represent the three main vowel features (roundness, height and backness). On the other hand, listeners also used language-specific dimensions to supplement their perceptual judgment. In particular, Thai listeners used peripheralness/centralness, while English listeners used retroflexion as an extra dimension to perceive the vowel stimuli.

Secondly, Fox, et al. (1995) examined the perceptual structure of native English and native Spanish listeners in perceiving selected English and Spanish vowels. Four hundred and five vowel pairs composed of Spanish /i/, /e/ and /a/ and English /i/, /ɪ/, /e/, /ɛ/, /æ/, /ʌ/ and /ɑ/ were presented to native English and Spanish listeners, both experienced and inexperienced in English, for dissimilarity rating. MDS analysis revealed that English and Spanish listeners employed different perceptual strategies. Specifically, English listeners used high/low, front/back and central/non-central dimensions, while Spanish listeners used high/low and central/non-central dimensions to perceive the stimuli. In addition, comparison of the perceptual spaces between proficient and non-proficient Spanish–English bilinguals showed that the perceptual space of proficient listeners was more English-like than that of non-proficient listeners. However, this study was somewhat limited in that it did not sample the entire acoustic vowel space of either Spanish or English (e.g., there were no high back vowels included in the stimulus set).

Both of these studies used MDS to demonstrate that native language, at least in part, determines a listener's perception of vowels, and experience in L2 will modify that perceptual process. The present study aims to extend the previous studies by employing MDS with native Chinese listeners. Unlike these two studies, the vowel stimuli employed will represent an almost complete vowel inventory from a single language (English) that samples the entire acoustic vowel space. By doing this, we aim to obtain a more comprehensive profile of the listeners' perceptual vowel space.

In the present study, participants from Northern dialects and the Xiang dialect region of China were recruited for the vowel perception experiment. Chinese includes seven dialect families: Northern, Wu, Xiang, Gan, Yue, Min and Kejia (Hakka) (Yuan, 1983). Each dialect family encompasses many separate dialects. Among the seven dialects, the Northern refers to a group of related dialects spoken across northern and southwestern China. The Beijing dialect is representative of Northern dialects and also serves as the basis for Standard Chinese (also called Putonghua or Mandarin), the official language of China. According to Lin and Wang (2001), Standard Chinese includes eight monophthongal vowels, /i, ɿ, ʅ, y, u, a, ɤ, o/. Among these eight vowels, the two apical vowels /ɿ/ and /ʅ/occur in complementary phonetic contexts of /i/ and are actually allophonic variations of /i/. Therefore, Standard Chinese can be considered to have six basic vowel phonemes.[1] Xiang is the dialect family spoken in Hunan province, which can be divided into the new Xiang dialect and the old Xiang dialect. Compared to the old Xiang dialect, the new Xiang dialect is closer to Standard Chinese. It does not preserve the distinction between voiced and voiceless obstruents. Changsha dialect is the representative of the new Xiang dialect. It includes 10 monophthong vowels, /i, ɿ, ʅ, y, u, a, ə,[2] o, õ, ə̃/ (Shi, 2005). Compared to the vowel inventory of Standard Chinese, new Xiang dialect has two more nasalized vowels, /õ/ and /ə̃/. These two nasalized vowels have non-nasalized counterparts. Therefore, these two vowels are not counted as the basic vowel phonemes. As in Standard Chinese, /ɿ/ and /ʅ/ are also allophonic variants of /i/. Thus, the new Xiang dialect can also be described as having six basic vowel phonemes like Standard Chinese (shown in Figure 1). In terms of the consonants, the

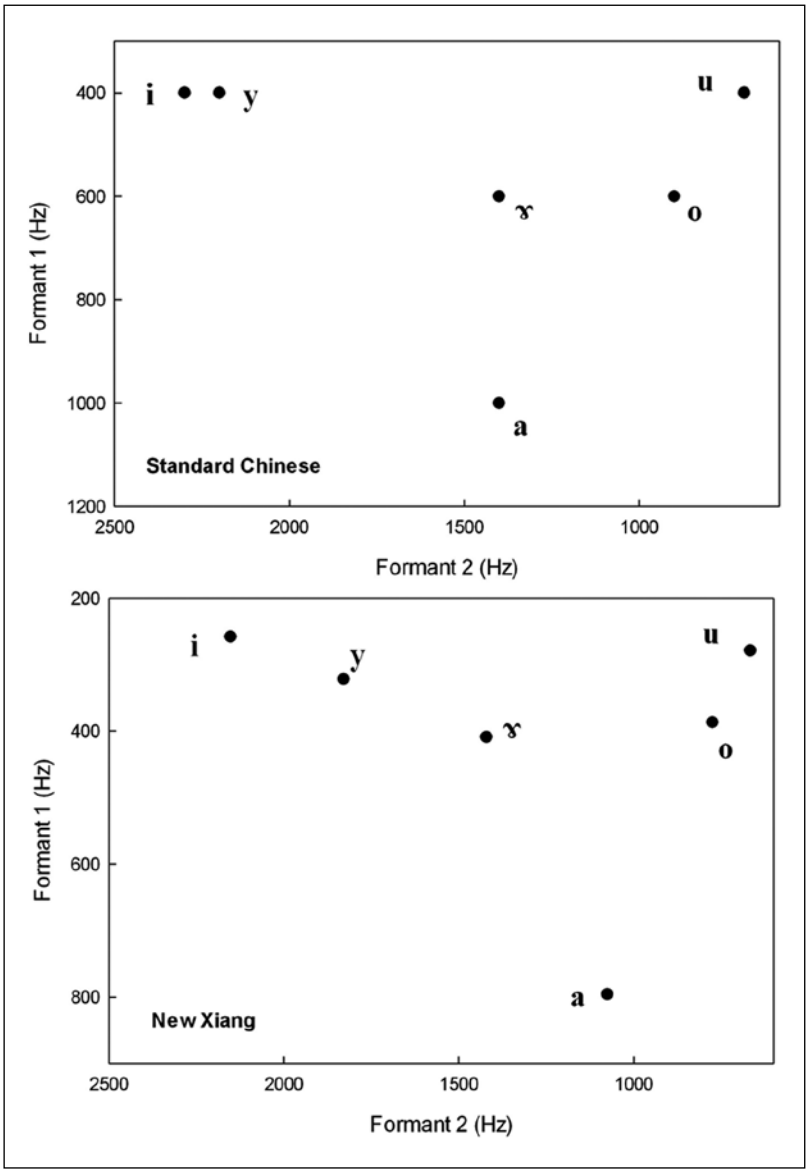

Figure 1. The acoustic vowel space of six vowels in the Standard Chinese and new Xiang dialect (the allophonic variants of /ɿ/ and /ʅ/ are not included). The data used to plot Standard Chinese vowel space is from Wu (1986); the data used to plot new Xiang dialect vowel space is from Shi (2005).

new Xiang dialect has 22 consonants. Compared to Standard Chinese, it has one more nasal sound, /n̥/. As for the tone system, compared to four tones in Standard Chinese, there are six tones in the new Xiang dialect. While there are some overall phonological differences between new Xiang and Northern dialect families, the specific differences between the new Xiang dialect (e.g., Changsha) and Standard Chinese are minor, and crucially for the present study, both dialects have smaller vowel inventories than that of American English.

We hypothesize that when native Chinese listeners listen to English vowels, the corner vowels /ɑ/, /i/, /u/ will be perceived with larger distances in the perceptual space because these vowels are distinct in terms of their acoustic distances in the F1–F2 space and appear as corner vowels in both English and Chinese (see Chung et al., 2012). We also expect that native Chinese listeners (especially those with less experience in English) will be more likely to assimilate less peripheral English vowels that are closer together in the acoustic space into similar L1 vowel categories, because both the Northern and new Xiang Chinese dialects have smaller vowel inventories (six contrastive monophthongs) than does American English (12 contrastive monophthongs). However, for those Chinese listeners who are more experienced and proficient in English, the immersion in English will gradually modify their perceptual structure and enable them to differentiate English vowels in a manner more like native speakers.

2 Methods

2.1 Stimuli

One male native American English speaker from Ohio produced 13 /h**V**d/ syllables that contained 13 English vowels, /i, ɪ, e, ɛ, æ, u, ʊ, o, ɔ, ɑ, ɔɪ, aɪ, aʊ/. Recordings were made directly to a computer's hard drive using a specially written Matlab program with a high-quality head-mounted microphone (Shure SM10A) positioned 2 inches from the mouth with a 44.1-kHz sampling rate and 16-bit quantization. The speaker produced three different exemplars of each token separately presented on the computer screen. The exemplar "most representative" of the vowel category was chosen by an experienced phonetician (R Fox) who is very familiar with the central Ohio dialect for use in this study. The tokens were then normalized for peak intensity. Each token was then paired with each of the other 12 tokens in both orders (AB and BA), which produced 312 (13*12*2) vowel pairs in total for the similarity identification test. The stimulus pairs were presented in pseudorandom order; that is, in random order except for the constraint that no vowel appeared twice in consecutive pairs.

The acoustic vowel space of these 13 vowels plotted in terms of the midpoint formant value is shown in Figure 2(a). Evident in the plot are several features of the central Ohio dialect. For example, the high back /u/ is relatively fronted, the low front vowel /æ/ is raised, the low back vowels /ɔ/ and /ɑ/ are relatively close in proximity (Central Ohio is in the process of a merger of those two vowels) and the diphthong /aʊ/ is fronted (Clopper, Pisoni, & de Jong, 2005; Labov, Ash, & Boberg, 2006). Based on previous research showing that naturally produced vowels are almost always characterized by inherent spectral change (Assmann, 1995; Hillenbrand & Nearey, 1999; Jacewicz & Fox, 2013; Morrison, 2013, Nearey & Assmann, 1986; Strange, Jenkins, & Johnson, 1983), the frequencies of F1, F2 and F3 were made at the 20%, 35%, 50%, 65% and 80% point of the vowel duration to estimate the formant movement over the duration of the vowel. The resulting formant tracks of all 13 vowels are shown in the Figure 2(b). The phonemic diphthongs /aɪ/, /ɔɪ/, /aʊ/ showed the greatest amount of spectral change over the vowel duration. The non-phonemic diphthongs /eɪ/ and /oʊ/ also showed a significant amount of formant movement. However, even the remaining monophthongs /i, ɪ, ɛ, æ, u, ʊ, ɔ, ɑ/ were naturally produced with measurable formant movement.

2.2 Listeners

Participants included 11 native Chinese listeners, 10 bilingual Chinese–English listeners and 10 native English listeners. All bilingual listeners were graduate students at The Ohio State University

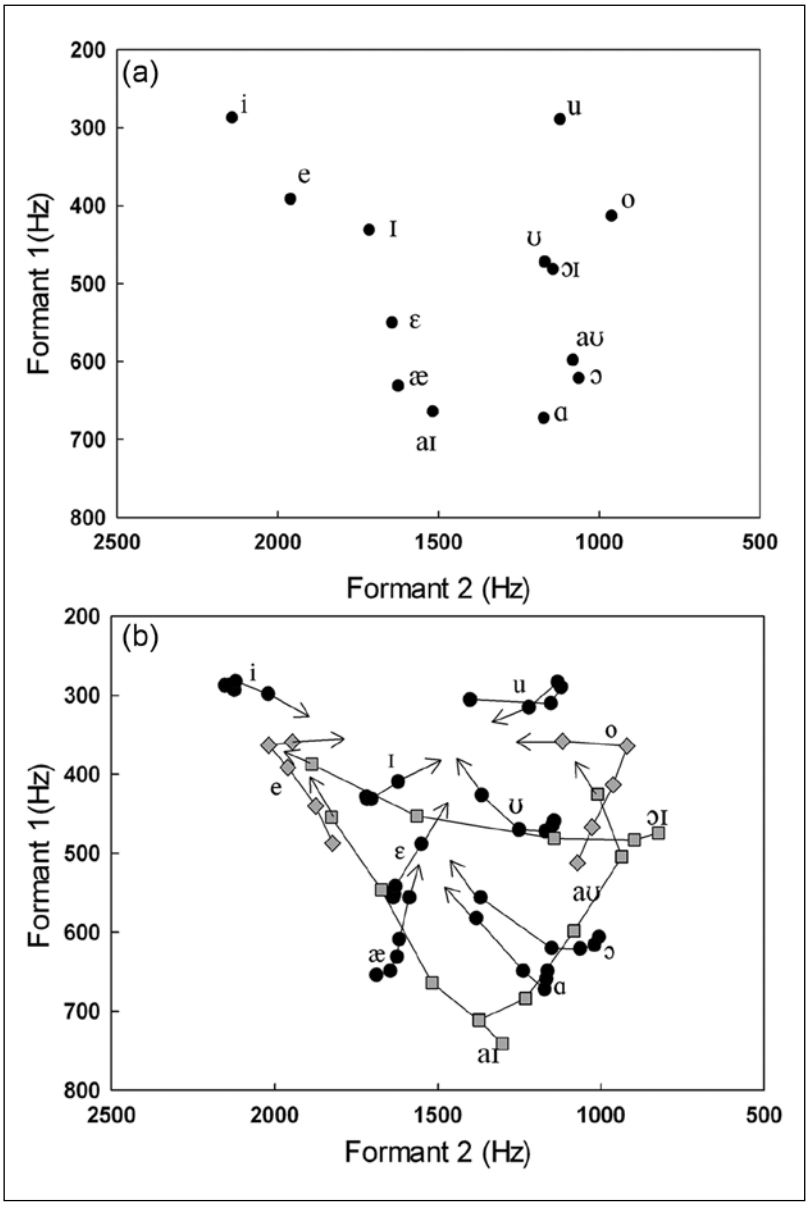

Figure 2. Acoustic vowel space (a) and formant movement pattern (b) of 13 vowels in the Ohio dialect. The acoustic vowel space was plotted on the basis of midpoint formant frequency value of individual vowels. In the plot of formant movement pattern, the diphthongs and diphthongized vowels are filled with gray, while the other vowels are filled with dark color in the plot of spectral change (b).

who were studying and living in the US at the time of testing with an average age of 27 years. All of the bilingual listeners had studied English for more than 10 years but arrived at the US at a late age (M = 24 years). All monolingual American English listeners were also graduate students at Ohio State with an average of 22 years. All monolingual Chinese listeners had lived in China with little or no exposure to English with an average age of 53 years.[3] In terms of the dialect

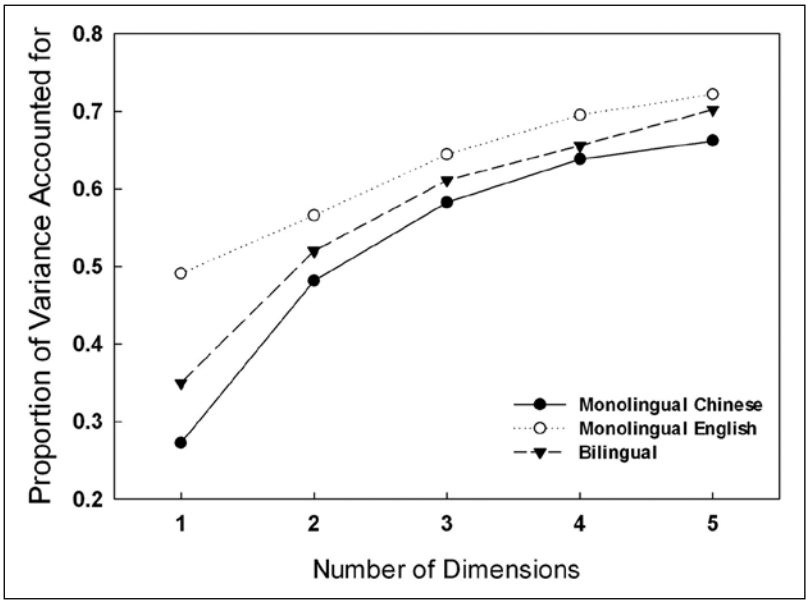

Figure 3. Fit curve for three groups of listeners showing the proportion of variance accounted for in one- to five-dimension solution in the ALSCAL analysis.

background, seven of the bilingual participants came from Northern dialect regions and three of them came from the new Xiang dialect region of China. Monolingual Chinese listeners were recruited from the new Xiang dialect regions. All monolingual Chinese listeners could speak both new Xiang and Standard Chinese natively. All monolingual English listeners were native speakers of the central Ohio dialect. All bilingual listeners spoke both a regional dialect of Chinese (either new Xiang or Northern) and Standard Chinese natively, and English as an L2 and were recruited in central Ohio of the United States. None of the listeners reported having a history of speech-language disorders or hearing problems.

2.3 Procedure

Each listener rated the 312 token pairs in a single session. All listeners were tested in a quiet room with a laptop set in front of them. Stimuli were delivered binaurally over AKG-K240 headphones and the volume was adjusted by listeners to a comfortable level for them. Each listener listened to two consecutive tokens and was required to judge the similarity/dissimilarity of the two vowels contained in each pair on a nine-point scale (from 1: very similar to 9: very different). All listeners were told to use the whole scale and to guess if they felt uncertain. Listeners were allowed to listen to the stimulus pairs no more than three times. Each pair was presented 1.0 s after the last response and the interstimulus interval (ISI) between the two stimuli in each pair was 250 ms. The whole experiment was implemented using a specially written Matlab program.

2.4 Analysis

The responses to all of the 312 vowel pairs for each listener were organized into a 13 by 13 full symmetric matrix[4] with each cell representing the sum of all rating scores for the same vowel pair

in both orders. The matrices of rating scores of all listeners in each group were examined using an alternating least squares scaling (ALSCAL) algorithm for non-metric individual differences MDS (called the INDSCAL model, see Takane, Young, & de Leeuw, 1977) as implemented in SPSS (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.). This model of ALSCAL was chosen because it develops not only the group space for all the listeners (with coordinate values for each stimulus vowel on each of the dimensions extracted), but also a set of subject weights indicating how salient each group dimension is for each separate listener. Because of this subject weighting, the model produces a unique orientation of the axes in the group vowel space so that no rotation of the axes is allowed (unlike older MDS programs), as any rotation will reduce the optimality of the solution.

A MDS analysis produces a spatial perceptual "map" in which vowels are located in an *n*-dimensional space. It is the job of the experimenter to determine the correct number of dimensions for the solution and to interpret these dimensions in terms of the phonetic and/or acoustic characteristics of the stimuli. In the present study, ALSCAL was first used to derive the perceptual space for each separate group and then the perceptual spaces were compared across each group of listeners. Finally, the subject weights for each separate group were examined.

3 Results

3.1 Dimensionality and interpretation

In a MDS analysis, the first decision involves choosing the appropriate dimensionality of the perceptual space. Generally, there are three criteria to rely on: the proportion of variance accounted for by the solution (a goodness-of-fit measure—older versions of MDS use values of "stress," which is a "lack-of-fit" measure), stability or uniqueness of solution, and interpretability (Fox & Trudeau, 1988; Kruskal & Wish, 1978).

The MDS output provides coordinate values for the vowels for each of the dimensionalities requested, as well as subject weights. These spatial models (or maps) provide a geometric representation of the perceptual relationships among the vowels. In this study, dimensionalities from 1 to 5 were obtained from each of group of listener responses. In addition to the coordinate values and the subject weights, ALSCAL provides a measure of how much variance in the input distance matrices is accounted for by each dimensionality. A plot of the variance-accounted-for values versus dimensionality represents a "fit curve." Since the INDSCAL model of ALSCAL only allows completion of an analysis with two or more dimensions, a basic Euclidean model was used to obtain the one-dimension solution. As shown in Figure 3, the one-dimension solution for monolingual English listeners accounted for approximately 50% of the variance. However, the one-dimension solution for monolingual Chinese listeners accounted for only 27% of variance. The proportion of variation accounted for in the one-dimension solution for the bilingual listeners was intermediate between these two groups. In all three groups, the proportion of variance accounted for increased as a function of the number of dimensions in the solution.

Generally, the "rule of thumb" in deciding the correct dimensionality of the solution is to find an "elbow" in the fit curve that represents a significant decrease in the proportion of variance accounted for between one dimensionality and the next (Borg & Gorenen, 2005; Jaworska & Chupetlovska-Anastasova, 2009). However, no clear "elbow" is apparent at first glance for any of these groups of listeners (which is not an uncommon result in MDS studies). Looking more carefully, one finds that for each of the three groups, the magnitude of increase in the fit curve values shows

Table 1. Correlation coefficients (*r* value) of split-half analysis for each dimension between two subgroups in each listener group.

	English monolingual	Bilingual	Chinese monolingual
D1	0.834**	0.974**	0.776**
D2	0.797**	0.917**	0.823**
D3	0.875**	0.812**	0.813**

**Correlation is significant at the 0.01 level (two-tailed).

a decrease when moving from three dimensions to four dimensions. Specifically for the monolingual English listeners, the one- to five-dimension solutions accounted for 49.1%, 56.6%, 64.5%, 69.5% and 72.2% of the variance, respectively. For the bilingual listeners, the one- to five-dimension solutions accounted for 34.9%, 52.0%, 61.2%, 65.6% and 70.2% of the variance, respectively. The proportion of variance accounted for in the bilingual listeners increased dramatically from one to three dimensions, but increased at a much lower rate from three to five dimensions. In the monolingual Chinese listeners, the one- to five-dimension solutions accounted for 27.2%, 48.2%, 58.2%, 63.8% and 66.2% of variance, respectively. Based on the interpretability of the three-dimensional map relative to the two-dimensional and four-dimensional maps, three-dimension solutions were selected for each of the three groups of listeners.

To further validate the choice of the three-dimension solution, one can also consider the stability of these spatial configurations. In order to ascertain whether or not the *n*-dimension solution is an appropriate configuration for each group, a split-half analysis (Fox & Trudeau, 1988; Fox et al., 1995; Gandour & Harshman, 1978; Harshman & Lundy, 1984; Kruskal & Wish, 1978) was completed to test reliability of the three-dimension solutions for each listener group. Split-half analysis involves dividing the perceptual data from each listener group into two separate subgroups and completing an ALSCAL analysis separately on each subgroup. The dimensions obtained for each subgroup are then compared using a correlation test. If the coordinates of each dimension obtained from one half of the responses are significantly correlated to those obtained from the second half, the perceptual dimensions are verified as being common to all subjects as a whole and thus are more likely to represent "true" perceptual dimensions and not "noise" in the data.

The most common approach to split-half analysis involves dividing subjects equally into two groups. However, since there is a relatively small number of subjects in the present study, an alternative split-half approach was used. Instead of dividing subjects into two halves, the rating scores of each subject were split into two halves. Because each vowel pair appeared four times in two different orders in the whole stimuli set, the 312 vowel pairs were divided into two subsets. Each subset contained rating scores of 156 vowel pairs with each vowel pair appearing once in each order (AB and BA; $N = 13*12 = 156$). These scores were again put into symmetricized 13 × 13 vowel matrices. For each dataset, these matrices served as proximity information for INDSCAL analysis. The matrices of these two ALSCAL analyses had no responses in common. ALSCAL analysis was then run on each separate dataset. The coordinates of *n*-dimension solutions from each half were then compared by calculating Pearson's *r*.[5]

As shown in Table 1, across all three groups of listeners, comparison of the coordinates in the two subsets for the three-dimension solutions revealed a high correlation in all three dimensions. These results provide evidence that the three-dimension solution reflected the significant, psychologically real dimensions for these groups of listeners.

3.2 Perceptual space of three groups of listeners

The final goal in MDS analysis is to interpret the output in a meaningful way. If a dimension obtained is uninterpretable, it provides insight into neither the data nor the performance of the listeners. As noted above, in addition to the fit curve and split-half analysis (which determines the uniqueness of the solution at a given dimensionality), another critical criterion in selecting the dimensionality of the solution is whether or not the axis of the dimensions (or clustering of the vowels within two-dimensional planes) can be readily interpreted in terms of the known phonetic (i.e., articulatory and/or acoustic) properties of the stimulus token.

As shown in Figure 4, for monolingual English listeners, in the D1 × D2 panel, D1 reflects the front/back distinction. D2 represents the high/low distinction even though /e/ is located in a lower position than /ɛ/, while /ɔɪ/ is located closer to /o/ rather than /ɔ/. In D3, the vowels /i/, /aɪ/ and /ɔɪ/ are located on one end of the dimension and are separated from the other vowels; this dimension reflects a contrast between vowels with a high-front offset and all other vowels, although the vowel /e/ also with high-front offset is not located in the same end as /i/, /aɪ/ and /ɔɪ/. These three dimensions are consistent with the findings in Fox (1983).

The D1 × D2 plot of bilingual listeners (shown in Figure 5) is similar to that of the monolingual English listeners. D1, again, represents a front/back distinction and D2 reflects the high/low distinction except for the vowel /ɔɪ/. D3 does not seem to match traditional phonetic features at first glance. However, examining D3 more carefully, it can be seen that the vowels with significant spectral change such as /aɪ/, /ɔɪ/ and /aʊ/ are clearly separated from the vowels with less spectral change, such as /i/ and /u/. Two diphthongized vowels, /e/ ([eɪ]) and /o/ ([oʊ]), are also located at the same side of the other three diphthongs. In addition, in the D1 × D3 panel, we can find that D3 also separates high vowels /i/, /ɪ/, /u/ and /ʊ/ relatively far away from non-high vowels. Therefore, D3 was associated with the formant movement and partially overlaps with D2 in separating high vowels from non-high vowels.

As shown in Figure 6 for monolingual Chinese listeners, unlike the monolingual English and bilingual listeners whose perceptual spaces were clearly separated by high/low and front/back distinction, the majority of the vowels were clustered into three large groups. The rounded vowels /u/, /ʊ/, /ɔɪ/, /o/ were clustered in one group. The vowels /ɔ/, /ɑ/, /aʊ/, /aɪ/ were clustered tightly in another group. The vowels /i/, /ɪ/, /e/, /ɛ/, /æ/ were clustered together. In contrast with the monolingual English and bilingual solutions, the vowels are not clearly separated in the D1 × D2 plane for the monolingual Chinese listeners. This provides support for the claim that monolingual Chinese listeners are less able to distinguish English vowels than listeners in the other two groups. In the D1 × D3 panel, it is clearly shown that the third dimension separated the vowels /i/ and /u/ distinctively from the other vowels. The third dimension supplemented D1 and D2 in separating the two long high vowels from all the other vowels, although it is unclear just what acoustic-phonetic property is represented by D3.

3.3 Acoustic interpretation of perceptual structure

Although the perceptual dimensions obtained using ALSCAL have been described, we have not addressed how these dimensions relate to the acoustic features of the vowels on which the perceptual decisions of the listeners are made. To address this issue, correlations between the vowel coordinates for each dimension and a set of acoustic measurements of the vowels were obtained. The acoustic measurements used for correlation analysis included vowel duration, F1, F2 and F3 at five points in vowel duration, F2–F1 and F3–F2 at five time points, the magnitude of change in frequency for each of the first three formants (ΔF1, ΔF2 and ΔF3),[6] the rate of change in frequency

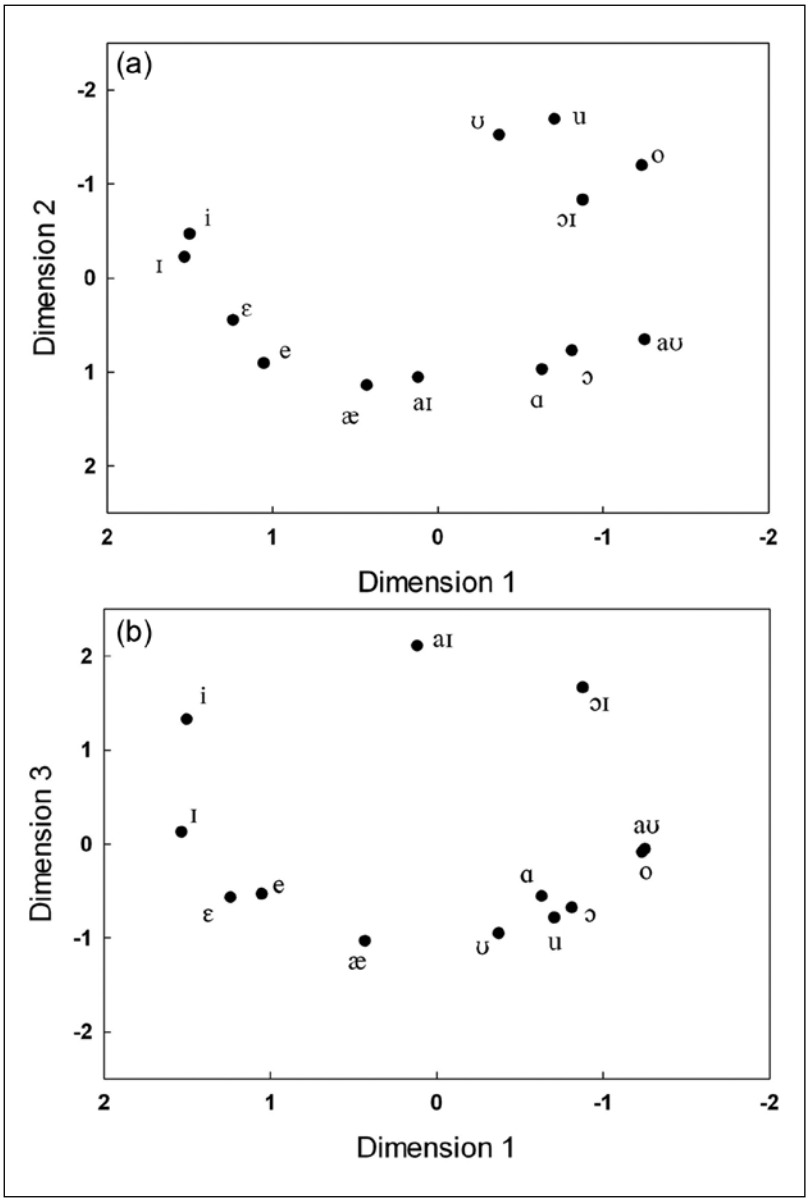

Figure 4. Three-dimensional INDSCAL solution for monolingual English listeners. (a) Plot of the D1 × D2 plane. (b) Plot of the D1 × D3 plane.

for each of these formants (F1_roc, F2_roc and F3_roc),[7] the formant frequency trajectory length for each section between each two consecutive time points (TL12, TL23, TL34 and TL45 representing trajectory length between the 20–35%, 35–50%, 50–65% and 65–80% points, respectively),[8] total trajectory length (TL_total, the sum of TLs in all four sessions) and the rate of total trajectory change (TL_total_roc)[9] (shown in Table 2).

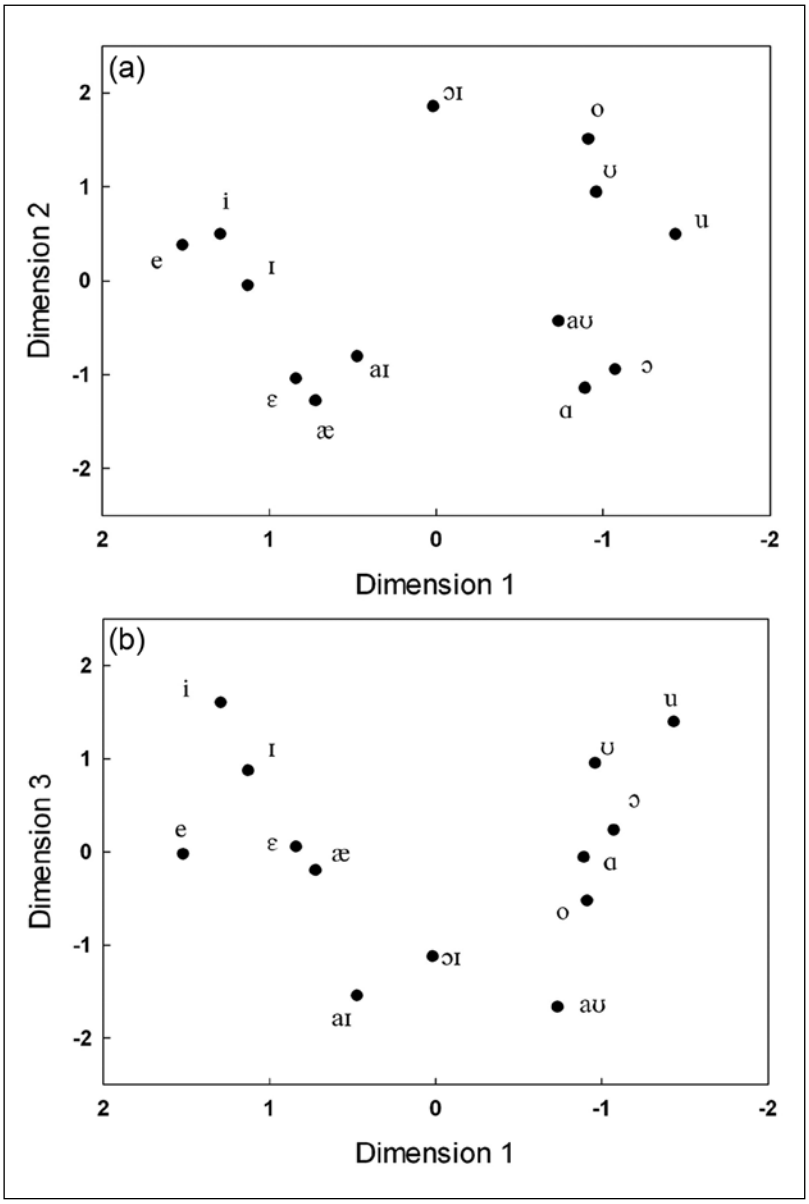

Figure 5. The three-dimension solution obtained for bilingual listeners. (a) Plot of the D1 × D2 plane. (b) Plot of the D1 × D3 plane.

As shown in Table 3, for the monolingual English listeners, D1 was significantly correlated with F2 and F2–F1 measures at each time point, indicating that D1 could be interpreted in terms of tongue advancement (the front/back dimension). D2 was significantly correlated with F1 at each time point. Since the frequency of F1 varies inversely with tongue height of the vowel, D2 represents the high/low distinction. There were three types of acoustic measures that correlated

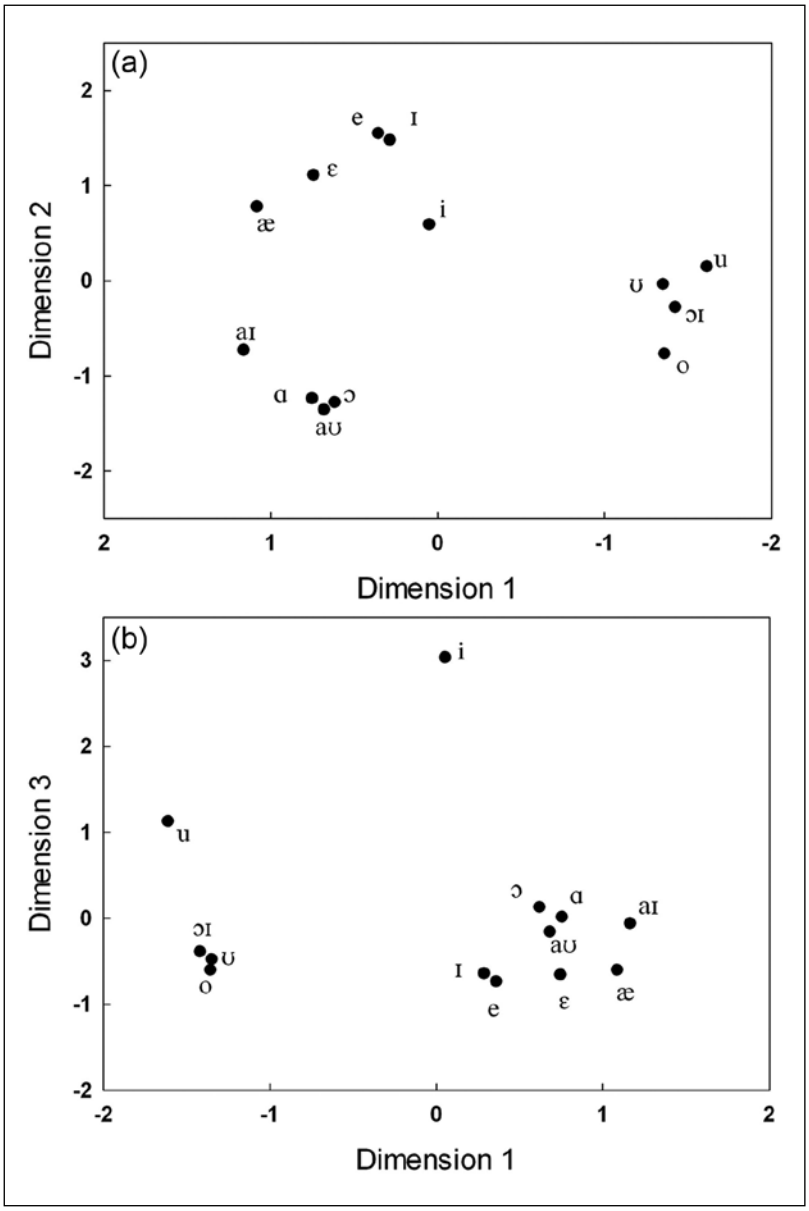

Figure 6. The three-dimension solution obtained for monolingual Chinese listeners. (a) Plot of the D1 × D2 plane. (b) Plot of the D1 × D3 plane.

significantly with D3: one type included formant frequency values at the fifth measurement point (S5_F2, S5:F2–F1, S5:F3–F2). A second type involved acoustic measurements associated with F3 (S1–F3, ΔF3 and F3_roc). A third type included acoustic measurements related to spectral change (TL23, TL34, TL_total, TL_total_roc). The correlation to the acoustic measurements at 80% point (S5) indicated that D3 was related to the offset of the vowels. The correlation to the acoustic

Table 2. Acoustic measurements of the 13 vowels produced by the native Ohio speaker.

	Heed	Hid	Heyd	Head	Had	Hod	Whod	Hood	Hoed	Hawed	Hide	Hoyd	Howed
s1_F1	293	428	487	556	654	649	305	459	512	606	741	474	711
s1_F2	2124	1719	1823	1637	1688	1163	1402	1144	1072	1006	1302	824	1374
s1_F3	2734	2298	2311	2359	2195	2473	2027	2319	2313	2522	2618	2702	2150
s2_F1	287	431	440	552	649	659	310	465	467	616	712	483	684
s2_F2	2153	1717	1874	1634	1646	1167	1153	1149	1027	1020	1374	897	1231
s2_F3	2579	2276	2256	2321	2224	2471	2020	2323	2313	2502	1698	2641	2262
s3_F1	287	431	391	550	631	672	289	472	413	621	664	481	598
s3_F2	2142	1716	1960	1645	1626	1173	1122	1170	962	1064	1518	1144	1082
s3_F3	2569	2285	2253	2325	2206	2457	2027	2305	2282	2483	2381	2399	2359
s4_F1	282	431	363	542	609	649	283	470	364	620	546	453	504
s4_F2	2120	1704	2018	1631	1618	1238	1133	1251	920	1151	1674	1565	936
s4_F3	2501	2321	2238	2341	2295	2438	2038	2288	2256	2421	2292	2263	2430
s5_F1	298	409	359	488	556	582	315	426	358	556	454	387	425
s5_F2	2020	1622	1946	1551	1588	1382	1221	1366	1118	1369	1827	1887	1009
s5_F3	2437	2347	2320	2394	2361	2387	2076	2285	2152	2436	2300	2362	2310
s1_F2–F1	1831	1291	1336	1081	1034	514	1097	685	560	400	561	350	663
s1_F3–F2	610	579	488	722	507	1310	625	1175	1241	1516	1316	1878	776
s2_F2–F1	1866	1286	1434	1082	997	508	843	684	560	404	662	414	547
s2_F3–F2	426	559	382	687	578	1304	867	1174	1286	1482	324	1744	1031
s3_F2–F1	1855	1285	1569	1095	995	501	833	698	549	443	854	663	484
s3_F3–F2	427	569	293	680	580	1284	905	1135	1320	1419	863	1255	1277
s4_F2–F1	1838	1273	1655	1089	1009	589	850	781	556	531	1128	1112	432
s4_F3–F2	381	617	220	710	677	1200	905	1037	1336	1270	618	698	1494
s5_F2–F1	1722	1213	1587	1063	1032	800	906	940	760	813	1373	1500	584
s5_F3–F2	417	725	374	843	773	1005	855	919	1034	1067	473	475	1301
ΔF1	5	−19	−128	−68	−98	−67	10	−33	−154	−50	−287	−87	−286
ΔF2	−104	−97	123	−86	−100	219	−181	222	46	363	525	1063	−365
ΔF3	−297	49	9	35	166	−86	49	−34	−161	−86	−318	−340	160
dur	209	157	238	164	217	226	226	149	225	229	213	248	253
F1_roc	0.039872	−0.2017	−0.89636	−0.69106	−0.75269	−0.4941	0.073746	−0.36913	−1.14074	−0.3639	−2.2457	−0.58468	−1.88406
F2_roc	−0.82935	−1.02972	0.861345	−0.87398	−0.76805	1.615044	−1.33481	2.483221	0.340741	2.641921	4.107981	7.143817	−2.40448
F3_roc	−2.36842	0.52017	0.063025	0.355691	1.274962	−0.63422	0.361357	−0.38031	−1.19259	−0.62591	−2.48826	−2.28495	1.054018
TL12	29.61419	3.605551	69.35416	5	42.29657	10.77033	249.0502	7.81025	63.63961	17.20465	77.62087	73.5527	145.5266
TL23	11	1	98.9798	11.18034	26.90725	14.31782	37.44329	22.13594	84.50444	44.28318	151.7893	247.0081	172.0378
TL34	22.56103	12	64.40497	16.12452	23.4094	68.94926	12.52996	81.02469	64.53681	87.00575	195.6016	421.9301	173.6433
TL45	101.2719	84.89994	72.11103	96.51943	60.90156	158.8238	93.6376	123.13	198.0909	227.2004	178.5301	328.6944	107.5639
TL_total	164.4471	101.5055	304.85	128.8243	153.5148	252.8612	392.661	234.1009	410.7717	375.6939	603.5419	1071.185	598.7717
TL_total_roc	1.311381	1.077553	2.134804	1.30919	1.179069	1.864758	2.89573	2.618578	3.042754	2.734308	4.72255	7.198826	3.944477

Table 3. Pearson correlations between selected acoustic parameters of the vowels and the coordinates of each perceptual dimension.

	English monolinguals			Bilingual			Chinese monolinguals		
	D1	D2	D3	D1	D2	D3	D1	D2	D3
s1_F1	−.313	.747**	.013	−.108	−.611*	−.801**	.659*	−.496	−.525
s1_F2	.844**	.217	−.032	.748**	−.253	.446	.391	.723**	.398
s1_F3	.166	.140	.730**	.295	.129	−.113	.119	−.163	.334
s2_F1	−.322	.735**	−.009	−.149	−.648*	−.755**	.656*	−.516	−.507
s2_F2	.911**	.293	.078	.855**	−.242	.390	.449	.738**	.350
s2_F3	−.031	−.145	−.091	−.008	.278	.234	−.206	−.033	.173
s3_F1	−.268	.720**	−.039	−.150	−.685**	−.642*	.654*	−.488	−.480
s3_F2	.938**	.305	.216	.925**	−.178	.339	.424	.768**	.314
s3_F3	.080	.323	.468	.162	−.121	−.134	.348	−.354	.338
s4_F1	−.193	.679*	−.194	−.153	−.712**	−.451	.622*	−.393	−.473
s4_F2	.871**	.262	.383	.918**	−.048	.256	.318	.727**	.264
s4_F3	.167	.482	.179	.201	−.381	−.103	.565*	−.269	.251
s5_F1	−.138	.662*	−.342	−.161	−.758**	−.292	.615*	−.318	−.433
s5_F2	.682*	.217	.557*	.803**	.084	.112	.206	.544	.236
s5_F3	.426	.597*	.174	.479	−.435	−.042	.625*	.079	.084
S1:F2–F1	.816**	−.056	−.031	.668*	−.018	.636*	.119	.772**	.506
s1:F3–F2	−.623*	−.114	.373	−.481	.271	−.425	−.268	−.678*	−.172
s2:F2–F1	.881**	.023	.070	.780**	−.006	.569*	.181	.793**	.457
s2:F3–F2	−.772**	−.320	−.113	−.713**	.348	−.199	−.482	−.629*	−.198
s3:F2–F1	.894**	.045	.199	.847**	.054	.489	.168	.813**	.419
s3:F3–F2	−.904**	−.188	−.048	−.862**	.134	−.385	−.297	−.888**	−.192
s4:F2–F1	.835**	.041	.398	.865**	.160	.357	.107	.763**	.371
s4:F3–F2	−.818**	−.115	−.328	−.855**	−.068	−.286	−.145	−.808**	−.187
s5:F2–F1	.660*	.026	.599*	.777**	.275	.179	.027	.581*	.329
s5:F3–F2	−.614*	−.027	−.565*	−.731**	−.254	−.142	−.004	−.585*	−.236
ΔF1	.327	−.463	−.345	.004	.172	.895**	−.378	.424	.357
ΔF2	−.255	−.029	.517	−.053	.328	−.351	−.214	−.254	−.194
ΔF3	.053	.191	−.805**	−.076	−.427	.116	.234	.251	−.366
dur	−.535	.243	.202	−.223	.090	−.499	.016	−.479	.131
F1_roc	.270	−.468	−.353	−.028	.206	.880**	−.400	.384	.390
F2_roc	−.292	−.042	.492	−.103	.323	−.352	−.222	−.299	−.200
F3_roc	.059	.184	−.803**	−.062	−.410	.109	.217	.278	−.406
TL12	−.423	−.312	.011	−.387	.221	−.089	−.366	−.174	.223
TL23	−.471	.038	.582*	−.070	.386	−.783**	−.158	−.367	−.218
TL34	−.453	−.028	.618*	−.115	.392	−.663*	−.211	−.402	−.185
TL45	−.567*	−.167	.484	−.363	.403	−.427	−.336	−.571*	−.099
TL_total	−.600*	−.130	.571*	−.273	.450	−.652*	−.326	−.485	−.112
TL_total_roc	−.595*	−.188	.586*	−.298	.473	−.630*	−.366	−.486	−.137

*Correlation is significant at the 0.05 level (two-tailed).
**Correlation is significant at the 0.01 level (two-tailed).

measurements related to F2 reflected a common phonetic feature of high F3 among /ɪ/, /aɪ/ and /ɔɪ/. As for the spectral change correlation, it is likely a function of the locations of /aɪ/ and /ɔɪ/ along the D3 axis as they show the greatest magnitude of formant movement.

In the bilingual listener vowel space, the pattern of correlations between D1, D2 and the acoustic measures were similar to that of the monolingual English vowel space. Specifically, D1 was significantly correlated to F2, F2–F1 and F3–F2. D2 was significantly related to F1 at all five points. Similar to English listeners, D1 reflected the front/back distinction and D2 reflected high/low distinction. The correlation between D3 and the selected acoustic measurements showed that D3 was most significantly correlated with F1_roc, ΔF1, S1_F1 and TL23. The correlation with the acoustic measurements associated with F1 indicated that D3 was related to vowel height. The correlation with the measurements associated with spectral change indicated that D3 reflected the phonetic feature of monophthong versus diphthong.

In the monolingual Chinese vowel space, the pattern of correlations between D1, D2, D3 and the acoustic measurements found for the other two vowels spaces was not obtained. In general, the size of the correlations obtained was noticeably reduced. Among all the selected acoustic measurements, D1 had relatively strong correlation with F1 at all five time points. D2 had most significant correlation with F3–F2 at the 50 and 65% points and F2–F1 at the 20, 35, 50% points. In addition, D2 was consistently highly correlated with F2 at multiple time points during the vowel duration. As discussed before, the variation of F1 reflects the change of tongue height and the change in F2 is associated with tongue advancement. Thus, D1 could be generalized as high/low distinction. D2 represented the front/back distinction. D3 did not exhibit a significant correlation with any of the acoustic measures examined here, although it showed a relatively higher correlation with F1 than with other acoustic measures. This probably resulted from the separation of the high vowel /i/ and /u/ from other vowels in the D1 × D3 plane.

3.4 Subject weights

In addition to providing coordinate values, INDSCAL also provides a "subject weight" for each listener on each perceptual dimension. The subject weight indicates the salience of a given dimension for a given listener. The subject weights provide insight into how much listeners in each group are utilizing their own group's dimension.

As shown in Figure 7, for monolingual English listeners, D1 (the front/back distinction) provided the primary and most salient cue for all listeners except for subject 5. This group of listeners places less weight on D2 and D3 ($M_{D1} = 0.574$, $SD_{D1} = 0.054$; $M_{D2} = 0.425$, $SD_{D2} = 0.069$; $M_{D3} = 0.349$, $SD_{D3} = 0.074$). A repeated-measure analysis of variance (ANOVA) was conducted to determine if there was a significant difference in subject weights across the three dimensions. The results showed a significant difference, $F(2, 18) = 24.230$, $p < 0.05$, partial $\eta^2 = 0.729$, in monolingual English listeners' subject weights. Post-hoc tests (using Bonferroni-corrected paired-sample t-tests) indicated a significant difference between the D1 and D2 subject weights ($t = 4.467$, $df = 9$, $p < 0.0167$) as well as D1 and D3 subject weights ($t = 8.922$, $df = 9$, $p < 0.0167$). However, no significant difference was found between D2 and D3 subject weights. In addition, the coefficient of variation was calculated to examine the consistency of subject weights on each dimension. The coefficient of variation (CV) in D1 subject weight was 9.4%, much smaller than that in D2 subject weight (16.3%) and D3 subject weight (21.1%). This demonstrated that native English listeners were more consistent in their dependence on D1 than on D2 and D3.

As shown in Figure 8, like the English listeners, bilinguals also relied primarily on D1 (the dimension most closely associated with the front/back distinction) ($M_{D1} = 0.515$, $SD_{D1} = 0.067$; $M_{D2} = 0.415$, $SD_{D2} = 0.077$; $M_{D3} = 0.401$, $SD_{D3} = 0.059$) except for subjects 7 and 8. Another observation is that, compared to monolingual English listeners' coefficient of variation ($CV_{D1} = 9.4\%$, $CV_{D2} = 16.3\%$, $CV_{D3} = 21.1\%$), bilingual listeners showed a greater amount of variation in the

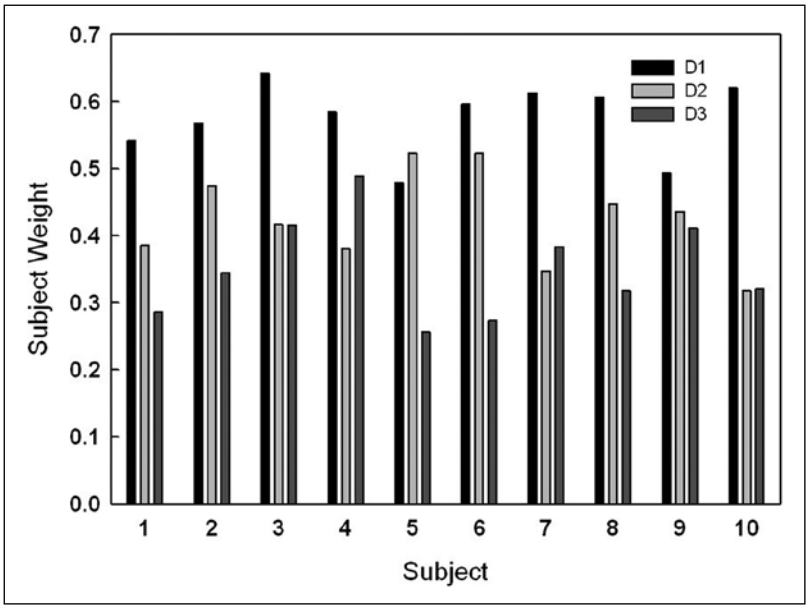

Figure 7. Subject weights for monolingual English listeners on each perceptual dimension.

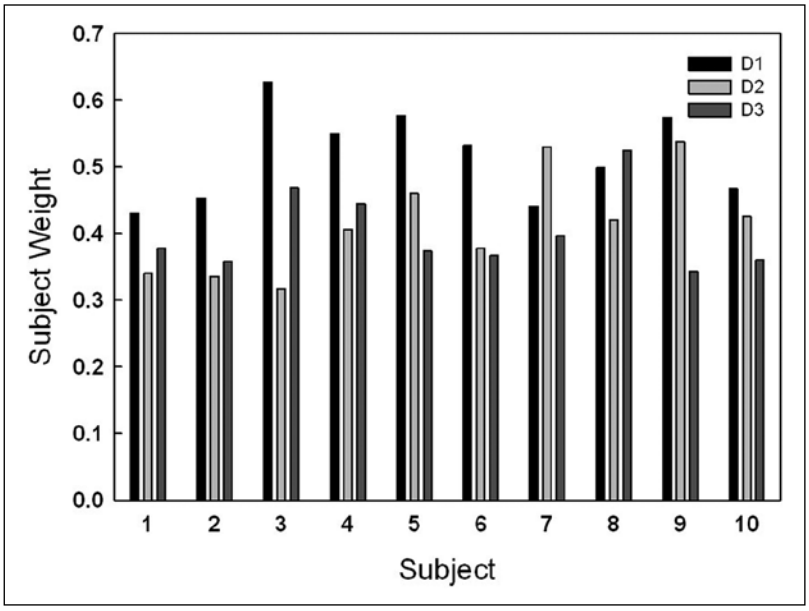

Figure 8. Subject weights for bilingual listeners on each perceptual dimension.

salience of D1 and D2 (CV_{D1} = 13.0%, CV_{D2} = 18.6 %) but more consistency in that of D3 (CV_{D3} = 14.6%), even though it was the least important perceptual dimension for the majority of bilingual listeners. Again, a repeated-measure ANOVA revealed that there were significant difference of

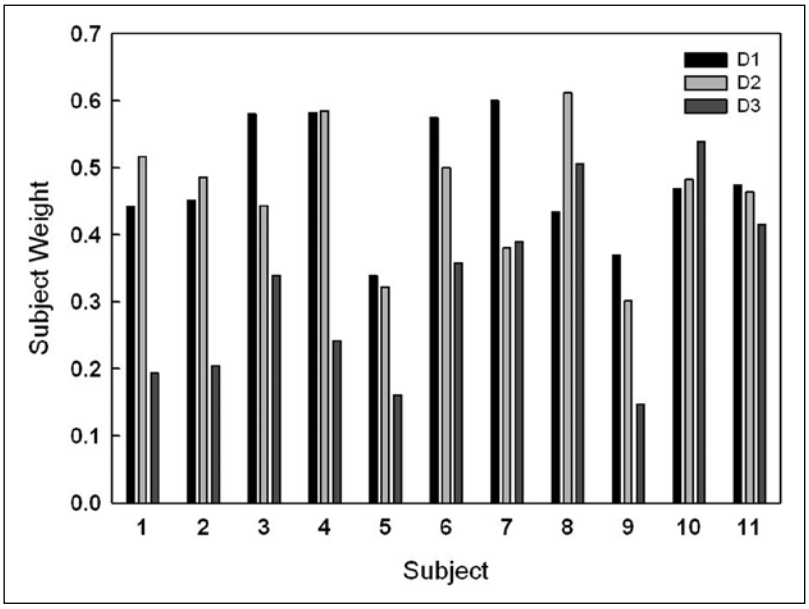

Figure 9. Subject weights for monolingual Chinese on each perceptual dimension.

subject weights for the three dimensions in bilingual listeners too, $F(2, 18) = 8.309$, $p < 0.05$, partial $\eta^2 = 0.480$. Specifically, paired sample *t*-tests showed that D1 subject weights were significantly larger than D2 subject weights ($t = 3.101$, $df = 9$, $p < 0.0167$) and D3 ($t = 4.614$, $df = 9$, $p < 0.0167$), but there was no statistical difference between D2 and D3 subject weights.

As shown in Figure 9, although most monolingual Chinese listeners put slightly more weight on D1, compared to the other two groups, they placed similar reliance on D1 and D2 ($M_{D1} = 0.483$, $SD_{D1} = 0.090$; $M_{D2} = 0.463$, $SD_{D2} = 0.098$). The mean difference between D1 and D2 subject weights in monolingual Chinese listeners was 0.020, much smaller than that in either monolingual English listeners (0.149) or bilingual listeners (0.100). The monolingual Chinese group showed less reliance on D3 ($M_{D3} = 0.318$, $SD_{D1} = 0.137$) than on D1 and D2, in general, although some listeners (such as subject 8 and 10) showed more dependence on D3. A repeated-measure ANOVA demonstrated a significant difference among these three-dimension subject weights in monolingual Chinese listeners, $F(2, 20) = 11.746$, $p < 0.05$, partial $\eta^2 = 0.540$. In particular, the paired sample *t*-tests showed that D1 subject weights were significantly larger than D3 subject weights ($t = 4.083$, $df = 10$, $p < 0.0167$). The D2 subject weights were also significantly larger than D3 subject weights ($t = 3.738$, $df = 10$, $p < 0.0167$) but the D1 and D2 subject weights were not significantly different. The coefficient of variation of subject weights in each dimension of monolingual Chinese listeners displayed greater extent of dispersion in both D1 and D2 ($CV_{D1} = 18.5\%$, $CV_{D2} = 21.1\%$) and especially D3 ($CV_{D3} = 43.3\%$) than that in the other two groups of listeners. Thus, compared to the other two groups of listeners, monolingual Chinese listeners were less consistent in their reliance on these perceptual dimensions to rate vowel similarity.

Comparing the subject weights among these three groups of listeners, we noticed that monolingual English and bilingual listeners showed less variability in each perceptual dimension. This indicates that monolingual English and bilingual listener groups were more homogeneous in their

use of the group vowel space than were the monolingual Chinese listeners. The relatively low subject weights in monolingual Chinese listeners and greater individual within-group variation in subject weights across all three perceptual dimensions show that monolingual Chinese listeners do not have a well-developed and consistent perceptual space for English vowels.

4 Summary and discussion

The present study investigated the underlying perceptual structure of bilingual Chinese–English and monolingual English and Chinese listeners using MDS. All three groups demonstrated some commonalities in their vowel spaces. That is, all three groups of listeners used front/back and high/low distinctions to perceive English vowels. These two perceptual dimensions have been found in numerous studies and likely represent the most basic criteria that humans use to perceive vowels in different languages (Fox, 1982, 1983; Fox & Trudeau, 1988; Fox et al., 1995; Pols et al., 1969). Thus, even though monolingual Chinese participants had little or no experience with English, they still used the same basic acoustic features as monolingual English and bilingual Chinese–English listeners to perceive and discriminate English vowels.

However, differences among these three groups of listeners were also evident, particularly with regard to the third perceptual dimension. In monolingual English listeners, the third dimension primarily represented the feature of high-front offset, although it was also acoustically related to the spectral change. However, in bilingual listeners, the third dimension was primarily associated with the monophthong/ diphthong distinction. For the monolingual Chinese listeners, the third dimension separated the two long high vowels /i/ and /u/ from the remaining vowels. In addition, although D1 and D2 also generally reflected the front/back and high/low distinction, the vowels in this plane were distributed into clusters in the perceptual space rather than being systematically separated as in the vowel spaces of native English and bilingual listeners.

The differences in the perceptual spaces of monolingual English and monolingual Chinese listeners demonstrate the effect of L1 on cross-language vowel perception. English has a larger number of vowel phonemes than most dialects of Chinese (including the Northern and Xiang dialects). The monolingual Chinese listeners were less likely to judge vowels as dissimilar compared to monolingual English listeners, producing the clustering seen in the monolingual Chinese vowel spaces. In addition, except for the differences in the inventory size, the acoustic-phonetic similarity between vowels in individual vowel pairs may also account for non-native listeners' ability to discriminate the L2 vowels, as posited by the Speech Learning Model (Flege, 1995) and the Perceptual Assimilation Model (Best, 1995; Best, McRoberts, & Goodell, 2001). However, since the current study does not directly test the magnitude of difficulty of discriminating English vowel pairs in Chinese listeners using a vowel discrimination task, it is unclear how this factor affects Chinese listener's perception of English vowels.

These observed effects might also support the native language magnet (NLM) theory proposed by Kuhl (Kuhl & Iverson, 1995; Kuhl et al., 2008). This theory predicts that the difficulty posed by a given L2 segment depends on its proximity to an L1 magnet. The closer the L2 segment is to the magnet, the more it will be assimilated to that L1 category, making it indistinguishable from the L1 sound. Thus, well-established instances of L1 categories (here the established Chinese vowel categories) act as magnets that shrink the perceptual space in the regions of these instances, which will lead to the clustering obtained in the monolingual Chinese vowel space. Thus, when unfamiliar speech sounds are presented, listeners assimilate new sounds to similar native sound prototypes. Monolingual Chinese listeners thus perceived English vowels in terms of relatively broad categories, conforming to the vowels in their native language.

In addition, the differences in perceptual space between bilingual and monolingual Chinese listeners affirmed the role of language experience in shaping listener's perception of non-native speech sounds. The exemplar model of speech perception (Johnson, 2006) represents one approach to explaining the role of language experience. In particular, this model posits that perception of an individual speech sound involves comparing the perceived sound with stored exemplars of each category. According to this model, when listeners have more experience in one language, they will have a richer bank of the exemplars for each sound category, which enables the listeners to have a better representation of the sound categories. As shown in the perceptual space of bilingual subjects, the English vowels were more evenly scattered than in the perceptual space of monolingual Chinese listeners. This indicates that bilingual listeners were more able to detect the differences among the English vowels. In the present study, bilingual listeners had learned English for years and resided in the U.S for a few years. This experience with English allowed them to establish separate categories for English vowels and thus allowed them to better discriminate these vowels.

In summary, all three groups of listeners showed a basic sensitivity to the acoustic structure of the vowels, which partially supported the presence of a universal auditory mode in vowel perception. However, listeners' auditory judgments of similarity/dissimilarity of non-native vowels are affected by their language experience. In particular, the monolingual Chinese listeners tended to assimilate English vowels into large clusters, which likely correspond to Chinese vowel categories, while bilingual listeners perceived English vowels more like monolingual English listeners.

In the future, clearly focused research is needed to advance knowledge of how the L2 experience (with different languages) modifies the underlying perceptual structures of native Chinese listeners and the extent to which change is gradient depending on the amount and type of experience in the L2. Since this current study only recruited two groups of Chinese listeners who either had no experience with English (the monolingual Chinese listeners) or a similar level of English proficiency (the bilingual listeners), we were unable to address this question. Recruiting non-native listeners differing in the age of L2 acquisition and/or the amount of L2 exposure would allow an exploration of how these two factors affect L2 vowel perception.

Acknowledgements

Thanks to all the volunteers participating in this study. We thank Dr Sameer ud Dowla Khan and Dr Kevin McGowan for their detailed and constructive comments. We also thank Dr Ewa Jacewicz for her valuable comments.

Funding

This research was supported by The Ruth Beckey Irwin and Harry Power Irwin Fund in the Department of Speech and Hearing Science at The Ohio State University.

Notes

1. The vowel /o/, considered by many to be a diphthong, is often not included in this set of basic vowel phonemes. For example, Duanmu (2000) described Standard Chinese as having a five-vowel phoneme system which consists of /a i u y ə/ with each phoneme representing several variants.
2. This vowel is transcribed as /ɤ/ by Yuan (1983). To facilitate the comparison between the Standard Chinese and new Xiang vowel systems in the acoustic vowel space, this vowel was labeled as /ɤ/ in Figure 1.
3. The significant age difference between the monolingual Chinese listeners and the other two listener groups is due to the fact that English instruction has been mandatory in Mainland China since the early 1980s. Younger adults are required to receive formal English education from middle school (or primary school in some regions).

4. A copy of these symmetricized matrices can be found at the http://las.sagepub.com/.
5. Spearman's *rho* was also calculated, but it showed the same patterns and is not described here.
6. The magnitude of spectral change for each formant was calculated with the formula $\Delta F_n = F_n\,(80\%) - F_n\,(20\%)$.
7. The rate of spectral change for each formant was calculated with the formula $F_n\_roc = \Delta F_n / (0.6*\text{duration})$.
8. The trajectory length between each two consecutive time points was calculated with the formula $TL_{n(n+1)} = \sqrt{(F1_n - F1_{n+1})^2 + (F2_n - F2_{n+1})^2}$ (Fox & Jacewicz, 2009).
9. The rate of total trajectory change was calculated with the formula TL_total_roc = TL_total /(0.6*duration) (Fox & Jacewicz, 2009).

References

Assmann, P. F. (1995). The role of formant transitions in the perception of concurrent vowels. *Journal of the Acoustical Society of America*, *97*, 575–584.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*, 775–794.

Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *2*, 303–328.

Borg, I., & Gorenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications*. New York, NY: Springer.

Bradlow, A. R. (1993). *Language-specific and universal aspects of vowel production and perception: A cross-linguistic study of vowel inventories*. Ithaca, NY: DMLL publications, Cornell University.

Butcher, A. (1976). The influence of the native language on the perception of vowel quality. *Arbeitsberichte Institut für Phonetik, 6*, University of Kiel.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*, 372–387.

Chung, H., Kong, E. J., Edward, J., Weismer, G., Fourakis, M., & Hwang, Y. (2012). Cross-linguistic studies of children's and adults' vowel spaces. *Journal of the Acoustical Society of America*, 131, 442–454.

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America*, *118*, 1661–1676.

Duanmu, S. (2000). *The phonology of Standard Chinese*. New York, NY: Oxford University Press.

Flege, J. E. (1995). Second language speech learning: theory, findings and problems. In W. Strange (Eds.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–273). Timonium, MD: York Press.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*, 2973–2987.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, *95*, 3623–3641.

Fox, R. A. (1978). Individual perception variation and a perception/production link in vowels. In F. Farkas, W. M. Jacobsen, & K. W. Todrys (Eds.), *Papers from the 14th Meeting of the Chicago Linguistic Society* (pp.98–107). Chicago, IL: Chicago Linguistic Society.

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

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

Fox, R. A., Flege, J. E., & Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America*, *97*, 2540–2551.

Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America*, *125*, 2603–2618.

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

Frieda, E. M., & Nozawa, T. (2007). You are what you eat phonetically: The effect of linguistic experience on the perception of foreign vowels. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 79–86). Amsterdam, The Netherlands: John Benjamins.

Gandour, J., & Harshman, R. (1978). Cross language differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, *21*, 1–33.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, *107*, 2711–2724.

Harshman, R. A., & Lundy, M. E. (1984). The PARAFAC model for three way factor analysis and multidimensional scaling. In H. G. Law, C. W. Snyder, J. A. Hattier, & R. P. MacDonald (Eds.), *Research methods of multimode data analysis* (pp. 122–215). New York, NY: Praeger.

Hillenbrand, J. M., & Neary, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *Journal of the Acoustical Society in America*, *105*, 3509–3523.

Højen, A., & Flege, J. E. (2006). Early learners' discrimination of second language vowels. *Journal of the Acoustical Society of America*, *119*, 3072–3084.

Jacewicz, E., & Fox, R. A. (2012). The effects of cross-generational and cross-dialectal variation on vowel identification and classification. *Journal of the Acoustical Society of America*, *131*, 1413–1433.

Jacewicz, E., & Fox, R. A. (2013). Cross-dialectal differences in dynamic formant patterns in American English vowels. In G. S. Morrison & P. Assmann (Eds.), *Vowel inherent spectral change* (pp. 177–198). New York, NY: Springer.

Jaworska, N., & Chupetlovska-Anastasova, C. (2009). A review of multidimensional scaling and its utility in various psychological domains. *Tutorials in Quantitative Methods for Psychology*, *5*, 1–10.

Johnson, K. (2006). Resonance in an exemplary-based lexicon: The emergence of social identify and phonology. *Journal of Phonetics*, *34*, 485–499.

Jusczyk, P. (1993). How word recognition may evolve from infant speech perception capabilities. In G. Altmann, & R. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 27–56). Hove, UK: Lawrence Erlbaum Associates.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. London, UK: Sage.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606–608.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Timonium, MD: York Press.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Paden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 979–1000.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English: Phonetics, phonology, and sound change*. Berlin, Germany: Mouton de Gruyter.

Levy, E., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language Experience. *Journal of Phonetics*, *36*, 141–157.

Lin, T., & Wang, L. (2001). *Yuyinxue Jiaocheng (Phonetics)*. Beijing: Beijing Daxue Chubanshe (Peking University Press).

Morrison, G. S. (2013). Theories of vowel inherent spectral change. In G. S. Morrison and P. Assmann (Eds.), *Vowel inherent spectral change (pp.*31–47). New York, NY: Springer.

Nearey, T. M., & Assmann, P. F. (1986) Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America*, *80*, 1297–1308.

Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., & Trent-Brown, S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *Journal of the Acoustical Society of America*, *124*, 576–588.

Pols, L., van der Kamp, L., & Plomp, R. (1969). Perceptual and physical space of vowels sounds. *Journal of the Acoustical Society of America*, *46*, 458–467.

Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America*, *97*, 1286–1296.

Rogers, C. L., Glasbrenner, M. M., DeMasi, T. M., & Bianchi, M. (2013). Vowel inherent spectral change and the second-language learner. In G. S. Morrison & P. F. Assmann (Eds.), *Vowel inherent spectral change* (pp. 263–282). New York, NY: Springer.

Scholes, R. (1967). Phonemic categorization of synthetic vocalic stimuli by speakers of Japanese, Spanish, Persian, and American English. *Language and Speech*, *10*, 46–68.

Scholes, R. (1968). Phonemic interference as a perceptual phenomenon. *Language and Speech*, *2*, 86–103.

Shepard, R. (1972). Psychological representation of speech sounds. In E. David & P. Denes (Eds.), *Human communication: A unified view (pp.*67–113). New York, NY: McGraw-Hill.

Shi, X. (2005). Hanyu Fangyan Yuanyin Geju de Shiyan Yanjiu (Phonetic experiment of vowel patterns in Chinese dialects). *(Dissertation)*, Tianjin, China: Nankai University.

Singh, S., & Woods, G. (1971). Perceptual structure of 12 American English vowels. *Journal of the Acoustical Society of America*, *49*, 1861–1866.

Strange, W., Bohn, O.-S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, *115*, 1791–1807.

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

Takane, Y., & Sergent, J. (1983). Multidimensional scaling models for reaction times and same-different judgments. *Psychometrika*, *48*, 393–425.

Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: Alternating least squares method with optimal scaling features. *Psychometrika*, *42*, 7–67.

Terbeek, D. (1977). A cross-language multidimensional scaling study of vowel perception. *UCLA Working Papers in Phonetics*, 37, 1–27.

Terbeek, D., & Harshman, R. (1971). Cross-language differences in perception of natural vowel sounds. *UCLA Working Papers in Phonetics*, 19, 26–38.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.

Wu, Z. J. (1986). *The spectrographic album of monosyllables of standard Chinese*. Beijing, China: Chinese Academy of Social Sciences.

Yuan, J. (1983). *Hanyu Fangyan Gaiyao* (*An outline of Chinese dialects*). Beijing, China: Wenzi Gaige Chubanshe.