Across-talker effects on non-native listeners' vowel perception in noise[a)]

Tessa Bent[b)] and Diane Kewley-Port
Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, Indiana 47405

Sarah Hargus Ferguson
University of Utah, Communication Sciences and Disorders, 390 South 1530 East, Room 1201, Salt Lake City, Utah 84112

(Received 5 January 2010; revised 25 August 2010; accepted 25 August 2010)

This study explored how across-talker differences influence non-native vowel perception. American English (AE) and Korean listeners were presented with recordings of 10 AE vowels in /bVd/ context. The stimuli were mixed with noise and presented for identification in a 10-alternative forced-choice task. The two listener groups heard recordings of the vowels produced by 10 talkers at three signal-to-noise ratios. Overall the AE listeners identified the vowels 22% more accurately than the Korean listeners. There was a wide range of identification accuracy scores across talkers for both AE and Korean listeners. At each signal-to-noise ratio, the across-talker intelligibility scores were highly correlated for AE and Korean listeners. Acoustic analysis was conducted for 2 vowel pairs that exhibited variable accuracy across talkers for Korean listeners but high identification accuracy for AE listeners. Results demonstrated that Korean listeners' error patterns for these four vowels were strongly influenced by variability in vowel production that was within the normal range for AE talkers. These results suggest that non-native listeners are strongly influenced by across-talker variability perhaps because of the difficulty they have forming native-like vowel categories. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3493428]

PACS number(s): 43.71.Hw, 43.71.Es, 43.71.Gv [MSS] Pages: 3142–3151

I. INTRODUCTION

Talkers' productions of a target sound or word can vary dramatically depending on a number of indexical factors, including gender, age, dialect, native language background, and emotional state. Across-talker differences have consequences for communication even when both the talker and the listener are normal-hearing, native users of the target language. The intelligibility of individual talkers varies not only in noisy conditions but also under ideal, quiet listening conditions (Hood and Poole, 1980; Bond and Moore, 1994; Bradlow *et al.*, 1996; Hazan and Markham, 2004). Despite an extensive literature on the influence of a listener's first language on the perception of second-language consonants and vowels, less work has been conducted on the influence of across-talker differences on non-native speech perception. Across-talker differences can affect several aspects of speech perception including overall intelligibility, identification accuracy for certain consonant or vowel contrasts, and perceptual assimilation patterns of second language (L2) phonemes to first language (L1) phoneme categories for non-native listeners.

In their investigation of inadvertently clear speech, Bond and Moore (1994) tested whether across-talker differences in native speaker intelligibility were maintained across native and non-native listeners. That is, are native talkers who are easy (or difficult) for *native* listeners to understand also easy (or difficult) for *non-native* listeners to understand? In both their word and sentence intelligibility tests, patterns of relative talker intelligibility were the same across native and non-native listeners. However, their study included a relatively small number of talkers (n=5), and the language backgrounds of the non-native listeners were not specified. Further, although they tested the listeners in quiet and in two different signal-to-noise ratios (SNRs), their results were averaged across these listening conditions. The current experiment used twice as many talkers to investigate the stability of across-talker intelligibility differences for native listeners and non-native listeners and employed listeners from a single language background to examine in detail the influence of L1 on L2 perception.

One aim of work on talker intelligibility has been to determine which acoustic-phonetic features promote intelligibility. The results obtained in a number of studies suggest that a variety of acoustic-phonetic parameters are important for highly intelligible speech, including increased vowel and word durations (Bond and Moore, 1994; Hazan and Markham, 2004; Ferguson and Kewley-Port, 2007); expanded vowel space (Bond and Moore, 1994; Bradlow *et al.*, 1996; Ferguson and Kewley-Port, 2007); frequent pauses (Bond and Moore, 1994); increased F0 range (Bradlow *et al.*,

[a)]Portions of this work were previously presented at the 154th meeting of the Acoustical Society of America, New Orleans, LA, November 2007 and the 156th meeting of the Acoustical Society of America, Miami, FL, November 2008.

[b)]Author to whom correspondence should be addressed. Electronic mail: tbent@indiana.edu

1996); more energy in the 1–3 kHz region (Hazan and Markham, 2004); and greater F2 differences between /i/ and /u/ (Hazan and Markham, 2004). Neel (2008) recently conducted a study focused specifically on across-talker differences in vowel intelligibility in quiet using a large database of upper Midwest English talkers who produced /hVd/ words (Hillenbrand *et al.*, 1995). For phonetically trained listeners from the upper Midwest, Neel reported weak associations between talker intelligibility and vowel space area, F1 and F2 range, duration ratio between long and short vowels, and the dynamic ratio between relatively dynamic and static vowels. All of these fine-grained acoustic characteristics accounted for less than a quarter of the variance in intelligibility. Neel concluded that acoustic properties that separate acoustically similar vowel pairs may be more useful in describing intelligibility than acoustic measures over the entire vowel space. We follow her recommendation here regarding "the need to examine specific vowel errors made by listeners, attempting to relate perceptual confusion to the acoustic characteristics of the vowels involved" (p. 584).

Talker characteristics have been shown to influence perceptual accuracy for at least one difficult non-native consonant contrast, /r/ versus /l/ (Lively *et al.*, 1993). Lively *et al.* (1993) trained native Japanese listeners to identify English /r/ and /l/ in a high-variability paradigm in which listeners were exposed to a large set of exemplars of the contrast through the inclusion of multiple talkers and varying phonetic contexts. The success of the high-variability training paradigm has been attributed to its ability to tune listeners' attention to relevant acoustic dimensions and reduce their attention to irrelevant variation, as proposed in Nosofsky's (1986) model of category learning. Although this approach has been quite successful in improving the identification of difficult non-native contrasts, listeners' accuracy for identifying targets during training still differed significantly depending on the talker, even with the inclusion of a large exemplar set. Similarly, Ingram and Park (1998) found talker differences in the perception of /r/ and /l/ by Japanese and Korean listeners. Identification and discrimination of the /r/ and /l/ tokens differed between the two male talkers in their study. The perceptual identification and discrimination patterns were related to the acoustics of the tokens presented. Spectrographic analysis showed that the tokens which had clearer acoustic cues to the identity of the sound, such as higher F3 and F4 for /l/, were more accurately identified and discriminated. These results suggest that the ease or difficulty of identifying non-native phoneme contrasts is not only influenced by the interaction of phoneme categories in the L1 and L2, but also by the indexical characteristics of the talker.

The present study addresses issues of how acoustic-phonetic information about speech and indexical information about talkers are processed and stored in memory. We will focus on vowels and consider two contrasting theories of speech perception: (1) the traditional, segmental approach (Jakobson *et al.*, 1952; Stevens, 2002, 2005); and (2) an exemplar-based model of speech perception promoted by Pisoni and his colleagues [for an overview, see Pisoni and Levi (2007)]. Stevens' (2005) Feature-Based Model of Speech Perception is a well-known example of traditional theories wherein perceptual processing removes indexical properties of talkers and vowel categories are represented by abstract feature bundles. In exemplar models, talker characteristics are integrated with the vowel and are stored with segmental cues. Understanding if and how indexical properties influence the processing of specific vowel tokens by native versus non-native listeners should shed light on vowel processing and category representation.

The two most prominent theories of cross-language speech perception, the Speech Learning Model (SLM) (Flege, 1999, 2002) and the Perceptual Assimilation Model (PAM) (Best, 1994; Best, 1995; Best *et al.*, 2001; Best and Tyler, 2007), state their main postulates in terms of (phonetic) *categories*. For example, in PAM, predictions regarding the accuracy with which naive listeners will discriminate non-native phonemes are based on the assimilation patterns between phonemic categories in the native and non-native languages. There is evidence that these assimilation patterns can vary depending on talker characteristics (Ingram and Park, 1997; Strange *et al.*, 2001). These results suggest that gradient phonetic detail, including that stemming from across-talker differences, is necessary to fully explain non-native assimilation and perceptual patterns. In fact, both PAM and SLM also consider fine-grained, gradient phonetic details, including concepts such as phonetic similarity, as important for explaining both naïve listeners' perception of non-native phones and second language learners' perception of L2 phones. However, how gradient phonetic details should be incorporated into these models is not fully specified.

The perception of American English vowels by native Korean listeners was chosen for this study because there is a substantial literature on vowel perception with this population, including studies in our own laboratory (e.g., Nishi and Kewley-Port, 2008). As described below, studies have compared the vowel systems of the two languages (Yang, 1996) and investigated assimilation patterns between Korean and English vowel categories (Ingram and Park, 1997). The design of the current study and interpretation of its results benefited from this previous literature. Further, focusing on the perception of vowels allowed the contribution of both spectral and temporal production differences across talkers to be assessed. The relationship between these acoustic variables—specifically F1, F2, and vowel duration—and listeners' perception of the vowels across talkers was then investigated. There are many variables that can affect across-talker differences in intelligibility, as discussed above, but focusing on vowels allowed us to take a detailed look at well-known acoustic properties that are likely to strongly influence across-talker differences in intelligibility.

A number of studies have looked at Korean listeners' perception of American and Australian English vowels. Standard Korean has been described as having 10 vowels including /i e ɛ a ʌ o ø y ɨ u/ (Yang, 1996). Younger speakers may not have a distinction between /e/ and /ɛ/ due to a merger between these two sounds. Further, while older speakers may have a distinction between short and long vowels, these distinctions are disappearing for younger speakers. Due to the differences between the English and Korean vowel systems, Korean listeners tend to have difficulty producing and perceiving certain English vowel contrasts. Two of the English vowel contrasts Korean listeners have difficulty perceiving and producing are /i/–/ɪ/ and /ɑ/–/ʌ/. These difficulties can be traced, at least partially, to differences between the Korean and English vowel systems. The distinction between /i/–/ɪ/ is difficult because both are perceptually assimilated to Korean /i/ (Ingram and Park, 1997). The Korean high front vowel /i/ is closer to the English /i/ than English /ɪ/, and is the only high front vowel in Korean (Yang, 1996). Native Korean talkers have difficulty both perceiving (Tsukada *et al.*, 2005) and producing the differences between /i/ and /ɪ/ (Flege *et al.*, 1997). The /ɑ/–/ʌ/ contrast is also highly confusable for Korean listeners (Nishi and Kewley-Port, 2008). The confusion pattern for these two vowels shows a marked asymmetry, with /ɑ/ being frequently identified as /ʌ/ but many fewer confusions in the other direction. American English /ʌ/ is more centralized than the Korean version (Yang, 1996). Korean does not have a low-back vowel like American English /ɑ/; the closest vowel in Korean is the mid-low vowel /a/. The present study explores the influence of talker differences on the confusability of these two vowel pairs.

The primary aim of this study was to assess how across-talker differences influence non-native vowel perception. Two specific issues were investigated. First, we explored whether relative intelligibility rankings of AE talkers are maintained between native and non-native listeners as listening conditions deteriorate. This issue was investigated by testing the intelligibility of 10 AE talkers for native and non-native listeners at three different signal-to-noise ratios. We expected that overall identification accuracy would be lower for the non-native listeners than for the native listeners, as a number of previous studies have shown that while non-native listeners may perform like native listeners in quiet, they tend to perform less accurately in noise compared to native listeners (Mayo *et al.*, 1997; Rogers *et al.*, 2006). However, the question explored here is whether *relative* differences in intelligibility among talkers will be maintained for native and non-native listeners at several signal-to-noise ratios. Second, we explored the acoustic-phonetic features of talkers' vowels that promote intelligibility for native and non-native listeners by conducting an in-depth analysis of two pairs known to be difficult for Korean listeners. The intelligibility of the vowels in these pairs was then related to the acoustics of the vowels.

II. METHODS

A. Participants

Fourteen listeners participated. All participants passed a hearing screening bilaterally at 20 dB HL at 500, 1000, 2000, and 4000 Hz and were paid for their participation. Seven listeners were native speakers of American English (all female) with an average age of 20 (range=19–22). Seven listeners were native speakers of Korean with a second language of English (6 female; 1 male) with an average age of 26 (range=20–35). The intention was to include both male and female listeners. However, listeners were accepted in order of responding to the recruitment materials without regard to gender. The Korean participants had been in the United States for an average of 3.7 years (range =3–4 years). One of the Korean participants did not report the specific length of their residence in the U.S. To recruit a homogeneous listener group in terms of length of residence, listeners were required to have lived in the U.S. for more than two years but less than five years. The rationale for this length of residence requirement was to recruit experienced listeners whose representations in the L2 would be fairly stable (Best and Tyler, 2007). Four additional participants were tested but their data were excluded. For two participants (one AE and one Korean), computer errors resulted in incomplete data. One of the AE participants did not pass the threshold of 85% correct identification for the easiest familiarization block (see details below under the procedures section). Lastly, one of the Korean participants did not pass the hearing screening.

B. Materials

Stimuli were taken from the Ferguson Clear Speech Database (Ferguson, 2004). The database includes 41 American English talkers (20 male; 21 female) producing /bVd/ words with 10 different vowels: /i, ɪ, e, ɛ, æ, ɑ, ʌ, o, ʊ, u/. Seven repetitions of each of these words were produced in sentence context in both clear and conversational styles. For the current experiment, 10 talkers were quasi-randomly selected (5 male and 5 female) to sample a wide range of performance for the AE listeners in Ferguson (2004). Only conversational tokens were used, two per vowel. All tokens were scaled to have the same peak RMS amplitude. Each token was presented 1 time in each signal-to-noise ratio condition (10 talkers × 10 vowels × 2 tokens × 3 SNRs=600 trials).

C. Procedure

1. General procedures

Participants sat in a sound-treated room in front of a computer monitor, keyboard, and mouse and listened to the stimuli over TDH-39 supra-aural earphones. The stimuli were output through TDT system II equipment. On each trial, unless otherwise noted, the test word was presented mixed with a segment of digitized 12-talker babble (Kalikow *et al.*, 1977) which was 1000 ms longer in duration than the test word. The test word and the babble were played out from separate channels of a 16-bit D/A converter (TDT DA1). The section of babble used in each trial was randomly selected from a 30 s babble file. The test word was attenuated (TDT PA4) to achieve a speech presentation level of 70 dB SPL and the babble was attenuated to achieve the desired signal-to-noise (SNR) ratio. The test word and babble segment were then mixed and low-pass filtered at 8500 Hz and delivered monaurally to the listener. Listeners identified the target word by clicking on one of 10 words displayed on the computer screen: bead, bid, bade, bed, bad, bod, bud, bode, book, booed. Several words that included the same vowel as the target word were also displayed next to each target word.

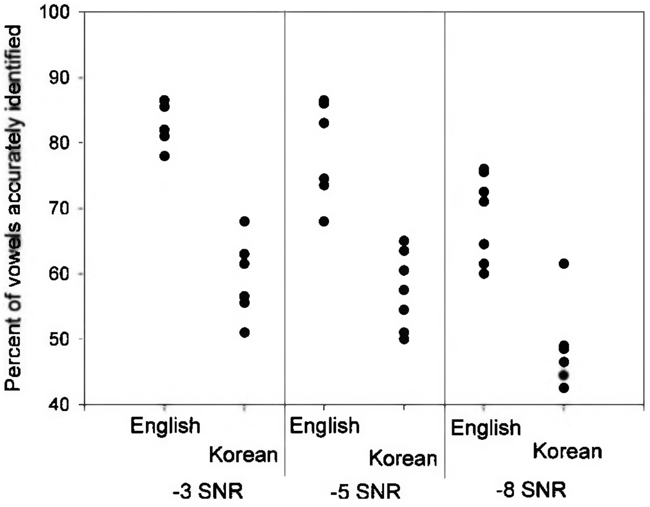

FIG. 1. Average identification accuracy scores for each of the 7 listeners in each group. Each point represents one listener's score averaged across talkers and across vowels. The three signal-to-noise ratios are shown on the *x*-axis.

2. Pilot procedures

A pilot study (Bent *et al.*, 2007) was conducted to determine the experimental parameters for the final study reported here. Only Korean listeners were recruited to participate in this study that followed the procedures for AE listeners as reported in Ferguson (2004). Thus, the pilot study included all 41 talkers in the Ferguson database. The SNR was set to 0 dB to equate overall vowel identification accuracy with the AE participants who were tested at −10 dB SNR. Results from this pilot indicated that the SNR for the final study should be between 0 and −10 dB to avoid ceiling performance for some AE listeners, and chance performance for some Korean listeners. Therefore, SNRs were set to −3, −5, and −8 dB for the final study (see below). Comparison of vowel identification accuracy for AE listeners from Ferguson (2004) and Korean listeners from this pilot study indicated that there was substantial variability for both listener groups such that a reduced set of 10 talkers would still span a wide range of performance and therefore permit a more detailed acoustic analysis of the resulting 200 tokens.

3. Experimental procedures

Each participant attended two sessions. In the first session, participants' hearing was screened, and they completed a language background questionnaire, and four familiarization blocks. The familiarization blocks were included to familiarize the participants with listening to speech in high levels of noise. Further, the familiarization blocks were used to screen AE participants, who were required to achieve at least 85% correct on the easiest SNR block (0 dB SNR) in order to be included. During the second session, participants completed the experimental blocks.

There were four familiarization blocks. The first familiarization block, which included 82 trials with two tokens from each of 41 talkers, was presented in quiet and participants received feedback. Blocks 2–4, which also included 82 trials each, were presented mixed with babble at three different SNRs over a wide range of values without feedback. The Korean listeners received SNRs of +10, 0 and −10 dB. For the AE listeners, the SNRs were 0, −3, and −10 dB. The three blocks presented with noise were ordered from easiest to most difficult to gradually accustom the listeners to the babble. The intent behind this gradual approach was to minimize learning effects during the test blocks. Therefore, each listener heard the familiarization blocks in same order.

There were six experimental blocks. For these experimental blocks, participants were presented with two tokens of each of the 10 vowels from 5 talkers, either all the males or all the females, for a total of 100 trials per block. The blocking by gender, which followed the protocol of Ferguson (2004), was intended to minimize the differences among talkers within a block. Within a block, trials were randomized. Participants were presented with three blocks of female talkers and three blocks of male talkers. Each block was presented at a different signal-to-noise ratio: −3, −5, or −8 dB SNR. Therefore, the participants heard the same materials three times but at different SNRs. For these blocks, the order of the blocks was pseudo-randomized for each participant. All participants received one of the −3 dB SNR blocks first and the remainder of the blocks was fully randomized. This ordering was selected so that listeners would have more time to fully adapt to the task in the first block before receiving the more difficult SNRs. The randomization of the remainder of the blocks was conducted so that differences across SNRs could not be attributed to practice or learning effects.

III. RESULTS

A. Results for overall intelligibility

Figure 1 displays the average performance for each listener in the two language background groups at the three different SNRs. A repeated-measures ANOVA with SNR (−3, −5, −8 dB) and talker (10 different talkers) as within-subject factors and listener (native AE, native Korean) as the between-subject factor showed highly significant effects of SNR [F(2,24)=51.90, p<0.001], talker [F(9,108)=37.29, p<0.001] and listener [F(1,12)=62.32, p<0.001]. The interaction between talker and SNR was significant [F(18,216)=2.68, p<0.001], but the other two- and three-way interactions were not. The main effect of SNR resulted from both groups performing more accurately in the easier SNRs. The main effect of talker was expected given that talkers were selected to represent a wide range of intelligibility scores for AE listeners. Lastly, the main effect of listener group was a result of the AE listeners performing significantly more accurately than the Korean listeners, by 22 percentage points on average. When tested at the same signal-to-noise ratio, as expected, the native listeners as a group always outperformed the non-native listeners. As can be seen in Fig. 1, there is almost no overlap in performance between the two listener groups within a single SNR, and the differences between the two groups are highly significant [–3 SNR:t(12)=9.76, p<0.001; –5 SNR:t(12)=6.15, p<0.001; –8 SNR:t(12)=5.91, p<0.001].

The interaction between talker and SNR is a result of different patterns of intelligibility for individual talkers across SNRs. For some talkers, intelligibility scores remained relatively flat across the three SNRs (averaged across the two listener groups). For example, talker M02's intelligibility only changed by 6% from the easiest to the most difficult listening condition (−3 SNR=74.6%; −5 SNR =74.3%; −8 SNR=68.2%). Other talkers showed very steep declines across the three noise conditions. For example, talker F18's intelligibility changed by over 20% from the easiest to the most difficult listening condition (−3 SNR =67.9%; −5 SNR=60.4%; −8 SNR=46.3%). Lastly, two talkers' intelligibility scores were slightly higher in the −5 SNR condition compared to the −3 SNR condition but lowest in the −8 SNR condition. This unexpected result was driven by the Korean listeners' data. For the AE listeners, the scores for these two talkers followed the expected pattern in which intelligibility scores decreased as the listening conditions became more difficult. However, the Korean listeners, who found the task more difficult, may have benefited from the specific ordering of the blocks, with one of the −3 dB blocks first. This ordering may have depressed the Korean listeners' performance in the −3 dB condition overall as they would have had more practice with the task by the time they were exposed to the −5 dB blocks. For the AE listeners, this additional practice did not appear to affect their performance pattern.

The lack of a significant talker-listener interaction suggests that the two groups of listeners found the same talkers most and least intelligible. In Figs. 2(a)–2(c), the relationship between the talker's scores at each SNR is shown for the two listener groups. Within an SNR condition, the correlations between the American English and Korean listeners' scores are very strong (−3 SNR:r=0.96, p<0.001; −5 SNR:r =0.86, p=0.001; −8 SNR:r=0.92, p<0.001). Therefore, when tested at the same signal-to-noise ratio, native and nonnative listeners usually find the same talkers most and least intelligible. It should be noted here that if the full set of 41 talkers from the Ferguson database were tested, we would expect to get correlations that are lower than those reported here because adding more talkers would presumably add more noise to the data.

B. Results for vowel pairs

Performance for specific vowels was also explored in this experiment. This approach was inspired by Neel (2008), but the present analysis is more descriptive in nature due to the smaller amount of data. Confusion matrices for the two listener groups were compiled. These confusion matrices were averaged across SNRs and talkers and are shown in Tables I and II for AE and Korean listeners, respectively. As can be seen in Table II, some vowel pairs are particularly confusable for the Korean listeners. Two of these vowel pairs, /i-ɪ/ and /ɑ-ʌ/, have been shown to cause perceptual difficulty for Korean listeners in previous studies (Flege *et al.*, 1997; Tsukada *et al.*, 2005; Nishi and Kewley-Port, 2008). Further, AE listeners identified these four vowels very accurately, but the Korean listeners identified the vowels much more poorly. Although these vowel pairs have been reported to cause considerable perceptual difficulty for Korean listeners, the focus of our analyses are the talker-specific acoustic-phonetic characteristics of the vowels that may make them more or less identifiable for the listeners. Therefore, these two vowel pairs were examined for all 10 talkers. The confusion matrices for the Korean listeners for the 10 talkers for these 4 vowels are shown in Table III. Figures 3 and 4 display the F1×F2 plots for these 4 vowels as produced by the 5 female and 5 male AE talkers, respectively. The purpose of the following analyses is to demonstrate

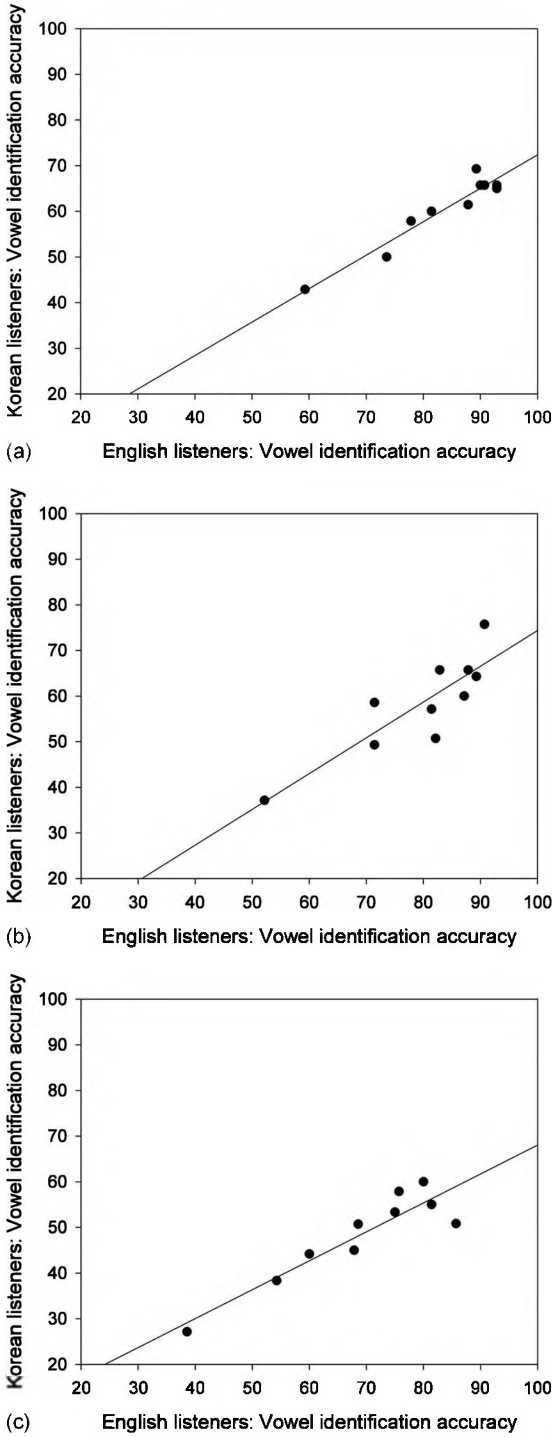

FIG. 2. [(a), (b), (c)] Intelligibility scores for 10 talkers at the three different signal-to-noise ratios of −3, −5 and −8 dB in Figs. 2(a), 3(b), and 2(c), respectively. American English listeners' scores are shown on the *x*-axis and Korean listeners on the *y*-axis. Each data point represents one talker.

TABLE I. Confusion matrix for the American English listeners. Percent identification scores are averaged across talkers and signal-to-noise ratios.

		Response									
		Bead	Bid	Bade	Bed	Bad	Bod	Bud	Bode	Book	Booed
Target	Bead	98	1	0	0	0	0	0	0	0	0
	Bid	5	73	2	5	1	0	3	2	6	4
	Bade	8	3	86	0	0	0	1	0	1	1
	Bed	1	24	2	55	9	0	2	1	3	3
	Bad	1	3	1	14	78	0	0	0	1	2
	Bod	1	0	1	1	1	90	4	2	0	0
	Bud	0	1	0	1	1	3	81	4	7	2
	Bode	0	1	0	0	0	1	5	79	6	6
	Book	2	2	1	2	2	1	23	7	54	6
	Booed	2	1	0	0	2	1	3	3	6	82

some acoustic bases for the more extreme talker differences observed for the Korean listeners.

Identification accuracy scores for the Korean listeners, averaged across talkers and SNRs, were 62% and 51% for the vowels /i/ and /ɪ/, respectively. When identification errors were made for these two vowels, they were most often misidentified as the other vowel in the pair. In contrast to the Korean listeners, the AE listeners identified /i/ very accurately (98% correct on average). AE listener identification accuracy for /ɪ/ was lower, but errors were broadly distributed across many vowels (Table I). For the Korean listeners, /i/–/ɪ/ identification accuracy varied widely depending on the specific talker (Table III) even though /i/-/ɪ/ formant values are widely separated for all talkers (see Figs. 3 and 4). Identification accuracy rates for /i/ ranged from 14% correct (talker F17) to 100% correct (talker M19). Similarly, identification accuracy varied for /ɪ/ from 12% correct (talker F09) to 79% correct (talkers F13 and M14). The rates of /i/-/ɪ/ identification accuracy for Korean listeners appeared to be strongly influenced by duration differences among talkers (symbol size). Longer duration /i/s lead to higher identification values, while short /i/s were frequently misidentified as /ɪ/. A correlation analysis was conducted using the duration values and identification scores (averaged across the three SNRs) for each token of each vowel (2 tokens per speaker for a total of 20 tokens). There was a significant, positive correlation between /i/ duration and identification accuracy ($r=0.87$, $p<0.001$). This analysis demonstrates that as the duration of /i/ increased, Korean listeners correctly identified more of the tokens. Likewise, there was a significant, inverse relationship between /ɪ/ duration and identification accuracy ($r=-0.56$, $p=0.01$). This analysis demonstrates that as the duration of /ɪ/ increased Korean listeners identified fewer tokens correctly. Korean listeners, therefore, relied more on vowel duration differences when identifying the members of this pair than on the spectral distinctions (Flege *et al.*, 1997).

For the vowels /ʌ/ and /ɑ/, identification accuracy scores for the Korean listeners were 60% and 51%, respectively. The accuracy rate for /ʌ/ ranged from 31% correct (talker M07) to 88% correct (talker F04), while for /ɑ/ accuracy ranged from 31% correct (talker F17) to 69% correct (talker F18). For /ʌ/, the best identification scores were seen for talkers who produced vowels with both short durations and relatively high F1 values, while low F1 resulted in more misidentifications. For /ɑ/, higher relative F1 values resulted in better identification scores. Thus for this vowel pair, Korean listeners relied on both vowel duration and spectral cues, primarily F1, in categorizing the vowels.

Although the acoustic measures examined here are a very small subset of those employed by Neel (2008), our

TABLE II. Confusion matrix for native Korean listeners. Bolded cells indicate vowel pairs that are further explored for individual talkers. Percent identification scores are averaged across talkers and signal-to-noise ratios.

		Response									
		Bead	Bid	Bade	Bed	Bad	Bod	Bud	Bode	Book	Booed
Target	Bead	**62**	**36**	0	0	0	0	1	0	0	0
	Bid	**15**	**51**	1	13	4	0	3	2	8	3
	Bade	15	18	55	6	3	0	0	0	1	1
	Bed	5	10	0	46	20	2	5	1	8	2
	Bad	1	0	0	36	59	0	1	0	0	1
	Bod	0	0	0	1	0	**51**	**41**	6	1	0
	Bud	0	2	0	3	0	**15**	**60**	4	9	6
	Bode	1	1	0	2	1	7	7	60	8	13
	Book	1	1	1	4	1	5	18	6	35	28
	Booed	2	1	0	3	1	3	4	2	13	71

TABLE III. Confusion matrices for each of the 10 talkers. Korean listeners' responses, in percent identification, are shown for two pairs of vowels that showed high degrees of confusability for the Korean listeners but relatively stable identification for the AE listeners.

		Response						Response			
F04		Bead	Bid	Bod	Bud	M02		Bead	Bid	Bod	Bud
Target	Bead	50	50	0	0	Target	Bead	55	45	0	0
	Bid	0	69	0	2		Bid	2	38	0	2
	Bod	0	0	62	31		Bod	0	0	45	55
	Bud	0	0	12	88		Bud	0	2	21	62
F13		Bead	Bid	Bod	Bud	M07		Bead	Bid	Bod	Bud
Target	Bead	38	60	0	2	Target	Bead	45	52	0	2
	Bid	7	79	0	2		Bid	0	48	0	7
	Bod	0	0	33	57		Bod	0	0	55	38
	Bud	0	0	19	45		Bud	0	5	7	31
F17		Bead	Bid	Bod	Bud	M14		Bead	Bid	Bod	Bud
Target	Bead	14	79	0	2	Target	Bead	57	40	0	0
	Bid	2	76	0	0		Bid	7	79	2	0
	Bod	0	0	31	57		Bod	0	0	48	43
	Bud	0	2	7	81		Bud	0	7	19	50
F18		Bead	Bid	Bod	Bud	M18		Bead	Bid	Bod	Bud
Target	Bead	90	10	0	0	Target	Bead	79	21	0	0
	Bid	0	26	0	14		Bid	21	52	0	0
	Bod	0	0	69	21		Bod	0	0	64	31
	Bud	0	0	19	57		Bud	0	0	12	86
F09		Bead	Bid	Bod	Bud	M19		Bead	Bid	Bod	Bud
Target	Bead	93	7	0	0	Target	Bead	100	0	0	0
	Bid	62	12	0	0		Bid	48	29	2	0
	Bod	0	0	57	36		Bod	0	0	43	40
	Bud	0	2	17	57		Bud	0	0	19	40

analyses lead to the following conclusion. Apparently, differences in acoustic properties among vowels produced by AE talkers were perceived quite differently by AE listeners, who identified the pairs with high accuracy, and Korean listeners, who selectively attended to more extreme spectral-temporal properties and misidentified vowels as a result.

IV. GENERAL DISCUSSION

The data presented here argue for a need to consider talker differences as a contributing factor to non-native speech perception. Our data extend the few extant studies on the contribution of talker variation to non-native speech perception. The contribution of across-talker differences was considered in two ways. First, across-talker differences in intelligibility were compared for native and non-native listeners. Second, the influence of across-talker differences in vowel characteristics on intelligibility was investigated.

The present study intentionally chose 10 AE talkers from Ferguson (2004) whose vowel intelligibility scores for AE listeners varied widely in the presence of noise. The present study compared the intelligibility of these 10 talkers for AE and Korean listeners at three SNRs, −3, −5, and −8

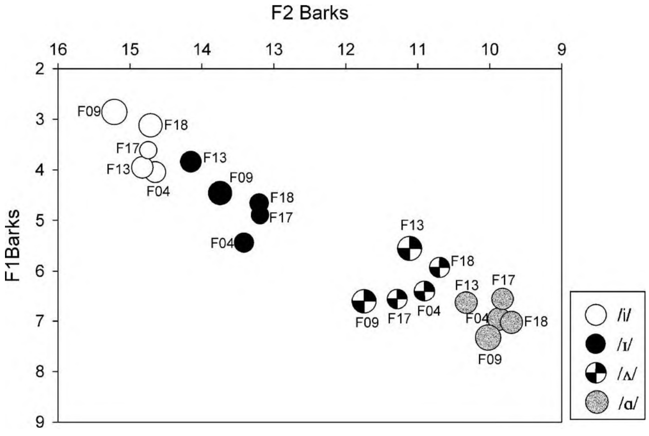

FIG. 3. F1 × F2 plot for two highly confusable vowel pairs for the female talkers. Duration is indicated by the size of the symbol with longer vowel durations having larger symbols.

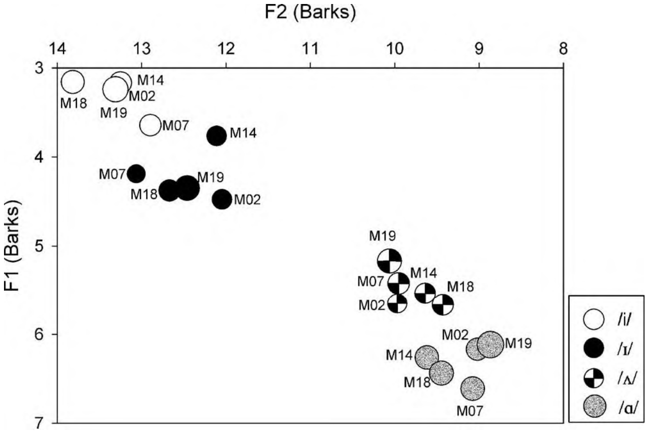

FIG. 4. F1 × F2 plot for two highly confusable vowel pairs for the male talkers. Duration is indicated by the size of the symbol with longer vowel durations having larger symbols.

dB. Talker intelligibility scores for Korean listeners were an average of 22 percentage points lower than for AE listeners, implying that Korean and AE listeners process acoustic information for vowels differently. However, remarkably high correlations between the talker intelligibility scores for the AE and Korean listeners (r=0.96, 0.86 and 0.92 for −3, −5 and −8 dB, respectively) indicated that both listener groups rank-ordered the talkers from least to most intelligible in the same way.

How do these results relate to traditional and exemplar theories of speech perception? Assume there are at least three speech processing modules: one that processes the incoming acoustic signal and produces a neural representation (acoustic module), one that stores a neural representation for each speech category (memory module), and one that compares a processed, new acoustic signal to the information in memory and identifies the speech category (decision module). Traditional theories, such as that of Stevens (2002, 2005), postulate that the neural representation consists of a small set of abstract features related only to segmental information and these features are not related to any properties of the talker. In contrast, exemplar theories postulate that the neural representation is much richer and includes indexical properties of the talker (Pisoni and Levi, 2007).

We argue that the talker variability effect observed here along with the high correlations between the talker intelligibility scores for the AE and Korean listeners are not compatible with traditional theories of speech perception. In Stevens' (2002, 2005) model, the acoustic module removes indexical information such that the neural representation in memory contains only segmental features. The AE listeners, whose speech processing systems are optimally tuned to perceive speech from AE talkers, nonetheless showed strong talker effects with spreads of 30–40 percentage point between the least and most intelligible talkers at each SNR. This result is hard to reconcile with the idea of feature-based neural representations with no talker properties. Given that the vowels were presented in noise, the result could possibly be accounted for in the acoustic module as, for example, an effect of strong interference from the noise that corrupts the process of correctly abstracting the features. Presumably for the Korean listeners all three processing modules (acoustic, memory, and decision) are less efficient and precise relative to the AE listeners, thereby introducing substantial error in vowel categorization. Yet, the Korean listeners ordered the talkers in the same way as AE listeners (r=0.92) in highly degraded listening conditions (−8 dB SNR) where they had severe difficulty identifying the vowels. This finding suggests the presence of talker-specific information that is encoded by both listener groups and affects processing similarly. Therefore, these results support a model in which talker information is incorporated into neural representations, and therefore provide strong counter-evidence against traditional theories of speech perception.

The overall intelligibility data suggest that certain acoustic-phonetic features of talkers across the vowel space will make their vowels more or less intelligible to both native and non-native talkers. For example, the talker with the lowest intelligibility scores for both listener groups in all three SNRs was also the talker with the lowest intelligibility for AE listeners in Ferguson (2004) in a larger set of 41 talkers. An inspection of the acoustics of this talker's vowels revealed that he had the smallest vowel space perimeter (the sum of the Euclidian distances between adjacent point vowels /i/ to /æ/, /æ/ to /ɑ/, /ɑ/ to /u/, /u/ to /i/) of all talkers in the Ferguson database. This talker's small vowel space may account for the difficulty both groups of listeners in this study had identifying his vowels. However, the talker with the highest (or lowest) identification accuracy score for a particular vowel was not necessarily the same talker with the highest (or lowest) accuracy score for another vowel. Therefore, in order to fully explain the differences in intelligibility across talkers, the production characteristics of each of the 200 vowels need to be examined.

Large acoustic variability has been observed in native talkers' vowel productions for different vowel categories (Peterson and Barney, 1952; Hillenbrand *et al.*, 1995; Neel, 2008). Although the differences across talkers in vowel production have some influence on native listeners' accuracy, these effects appear to be even greater for non-native listeners. The analysis relating the spectral and duration characteristics of particular talkers' vowels to identification accuracy and confusion patterns demonstrated that although the general patterns of identification accuracy and misidentifications can be explained with reference to abstract vowel categories, across-talker differences also have a strong influence on non-native vowel perception. With certain vowel pairs, such as /i/–/ɪ/, native listeners were relatively insensitive to differences across talkers and could reliably identify most of these vowels regardless of the spectral or duration differences across talkers. However, for the Korean listeners, a wide range of accuracy scores was observed across talkers (12%–100% correct for /i/ and /ɪ/). These results argue for the necessity of including gradient phonetic details, such as that stemming from talker differences, into models of cross-language speech perception.

The identification accuracy differences across talkers appear to be traceable to the specific acoustics of the vowel tokens and the attentional weight Korean listeners placed on the various acoustic dimensions, rather than to assimilation patterns between phonemic categories in the L1 and L2. If the interaction between L1 and L2 phoneme categories fully explained the perceptual identification patterns, then we would expect identification of L2 phones to be consistent across specific tokens. However, the Korean listeners seem to have based their identification judgments of /i/ and /ɪ/ on absolute vowel duration. This resulted in a wide range of identification accuracy scores, with longer /i/s and shorter /ɪ/s receiving higher identification scores than shorter /i/s and longer /ɪ/s. The Korean listeners thus appeared to be placing more weight on the temporal dimension than the spectral dimension, causing accuracy scores to range from poor to excellent across talkers depending on the temporal characteristics of the vowels produced by each talker. The difficulty that even experienced L2 learners have in spectrally differentiating these vowels may push them to attend to another acoustic dimension that is more perceptually salient to them, namely duration differences. This interpretation is consistent with earlier studies. Ingram and Park (1997) found that Korean listeners attended to absolute duration values during the perception of a different pair of vowels, Australian English /æ/ and /e/. Similarly, a number of previous studies have shown that L2 listeners attend to different stimulus dimensions than L1 listeners for both vowel and consonant contrasts (Fox *et al.*, 1995; Iverson *et al.*, 2003).

In addition to the two vowel pairs analyzed in depth in this study, there were several other vowel pairs that showed relatively high numbers of confusions for the Korean listeners including /ɛ/-/æ/ and /ʊ/-/u/. Confusions among these vowel pairs would be expected, as they are adjacent to one another in the F1 × F2 plane. Further, both represent vowel pairs in which one member of the pair is not present in the vowel inventory of Korean. The Korean language contains both /ɛ/ and /u/ but does not have /æ/ or /ʊ/. Nishi and Kewley-Port (2008) also found these two vowel pairs to be highly confusable for Korean listeners prior to perception training. However, the AE listeners in the current study also exhibited difficulty identifying the vowels in these two pairs when mixed with noise suggesting that factors other than the Korean listeners' language experience contributed to their lowered performance. Because the AE listeners also had difficulty identifying these vowels, the Korean listeners' identification patterns could not be specifically traced to factors about their language experience. Therefore, the current project focused on two vowel pairs for which the Korean listeners demonstrated difficulty in identification but the AE listeners showed high identification accuracy.

High-variability training paradigms (Logan *et al.*, 1991; Lively *et al.*, 1993) have been successful at shifting non-native listeners' attention to acoustic cues that are robust and reliable for phoneme identification. However, our participants, who had resided in the U.S. for 3–4 years and therefore had extensive naturalistic experience with multiple talkers and multiple phonetic environments, did not appear to attend to the most appropriate cues to vowel identity. Exemplar models of learning (Nosofsky, 1986) suggest that exposure to irrelevant variation helps the learner to focus attention on the relevant cues for category identity. The case of /i/–/ɪ/ is not contradictory to this model, because the duration differences between these two phonemes, while not fully reliable cues, are not irrelevant for vowel category identification. Learning to shift attention away from an acoustic parameter that is not at all reliable (such as F2 for /r/-/l/ identification) may be an easier task than shifting attention away from a cue that is partially reliable, as in the case of temporal differences for /i/ and /ɪ/.

Future research is needed to extend the current findings to a broader range of speech materials. As our study focused on vowels in /bVd/ frames, it is not clear if and how the outcomes of this study will generalize to other types of materials such as consonant contrasts or sentences. Additionally, the language background of the non-native listeners was limited to Korean. It remains possible that testing non-native listeners from different language backgrounds would lead to different patterns of relative intelligibility among talkers and/or different patterns of vowel confusions. Specifically, although Korean and English are quite distinct typologically and at higher levels of linguistic structure, there are similarities between their vowel inventories including the size of the inventory and the lack of contrastive vowel length (for younger speakers). However, despite the similarities between the Korean and English vowel systems, the Korean listeners performed significantly less accurately in the vowel identification task in noise compared to the AE listeners. It may be that assessing listeners from language backgrounds with more distinct vowel inventories compared to English, such as languages with much smaller inventories or those that contrastively use vowel length, may lead to even more degraded vowel identification accuracy and/or different patterns of vowel confusions. However, studies such as Nishi and Kewley-Port (2008), in which vowel identification by Korean and Japanese listeners was compared, have demonstrated that L2 listeners from L1s with very different vowel systems performed similarly on an AE vowel identification task before training. We hypothesize that the findings regarding the order of overall intelligibility scores across talkers would be similar across listeners from various language backgrounds, but that the details of identification of specific vowel tokens from some talkers would differ for listeners from different language backgrounds.

V. CONCLUSION

One of the most unexpected and important results of our study was the very high correlation in talker intelligibility between native and non-native listeners. These results provide very strong support for exemplar models of speech perception in which segmental and indexical information are integrated during speech processing. Moreover, the results provide strong counter-evidence against traditional speech processing theories that use abstract features. In addition, our results support the inclusion of across-talker influences on non-native speech perception into models of cross-language speech perception, since the specifics of talkers' production of certain vowels can result in performance that ranges from poor to excellent. Extensive across-talker differences in vowel production have been known for several decades, but this factor has not been included in theories of cross-language speech perception. As non-native listeners have less experience with the variability seen in speech in the real world, they may be less able to use multiple cues to identify vowels and perceptually compensate for the variability within and across talkers.

ACKNOWLEDGMENTS

We thank Jae Hee Lee, Morgan Dotts, and Kristin Knies for assistance with data collection, Meredith Poore, Katie Beam, and Kyung Ae Keum for assistance with acoustic analyses, and Mitch Sommers and three reviewers for their thoughtful comments on a previous version of this manuscript. This work was supported by the following grants from NIH-NIDCD: Grant Nos. T32 DC-00012, R03 DC-008886, R01 DC-02229, and R21 DC-010027.

Bent, T., Ferguson, S. H., and Kewley-Port, D. (**2007**). "Listener influences on vowel intelligibility," J. Acoust. Soc. Am. **122**, 3016.

Best, C. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Best, C. T. (**1994**). "The emergence of native-language phonological influences in infants: A perceptual assimilation model," in *The Development of Speech Perception: The Transition From Speech Sounds to Spoken Words*, edited by J. Goodman and H. C. Nusbaum (MIT, Cambridge, UK), pp. 167–224.

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," J. Acoust. Soc. Am. **109**, 775–794.

Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.-S. B. M. J. Munro (Benjamin, Philadelphia), pp. 13–34.

Bond, Z. S., and Moore, T. J. (**1994**). "A note on the acoustic-phonetic characteristics of inadvertently clear speech," Speech Commun. **14**, 325–337.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech. 1. Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Ferguson, S. H. (**2004**). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," J. Acoust. Soc. Am. **116**, 2365–2373.

Ferguson, S. H., and Kewley-Port, D. (**2007**). "Talker differences in clear and conversational speech: Acoustic characteristics of vowels," J. Speech Lang. Hear. Res. **50**, 1241–1255.

Flege, J. E. (**1999**). "Age of learning and second-language speech," in *Second Language Acquisition and the Critical Period Hypothesis*, edited by D. P. Birdsong (Lawrence Erlbaum, Hillsdale, NJ), pp. 101–132.

Flege, J. E. (**2002**). "Interactions between the native and second-language phonetic systems," in *An Integrated View of Language Development: Papers in Honor of Henning Wode*, edited by P. Burmeister, T. Piske, and A. Rohde (Wissenschaftlicher Verlag Trier, Trier, Deutschland), pp. 217–244.

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

Fox, R. A., Flege, J. E., and Munro, M. J. (**1995**). "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional-scaling analysis," J. Acoust. Soc. Am. **97**, 2540–2551.

Hazan, V., and Markham, D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility for adults and children," J. Acoust. Soc. Am. **116**, 3108–3118.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hood, J. D., and Poole, J. P. (**1980**). "Influence of the speaker and other factors affecting speech intelligibility," Audiology **19**, 434–455.

Ingram, J. C. L., and Park, S. G. (**1997**). "Cross-language vowel perception and production by Japanese and Korean learners of English," J. Phonetics **25**, 343–370.

Ingram, J. C. L., and Park, S. G. (**1998**). "Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners," J. Acoust. Soc. Am. **103**, 1161–1174.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47–B57.

Jakobson, R., Fant, G., and Halle, M. (**1952**). "Preliminaries to speech analysis: The distinctive features and their correlates," MIT, Technical Report No. 13.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech-intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/: The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–886.

Mayo, L. H., Florentine, M., and Buus, S. (**1997**). "Age of second-language acquisition and perception of speech in noise," J. Speech Lang. Hear. Res. **40**, 686–693.

Neel, A. T. (**2008**). "Vowel space characteristics and vowel identification accuracy," J. Speech Lang. Hear. Res. **51**, 574–585.

Nishi, K., and Kewley-Port, D. (**2008**). "Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training," J. Speech Lang. Hear. Res. **51**, 1480–1493.

Nosofsky, R. M. (**1986**). "Attention, similarity, and the identification-categorization relationship," J. Exp. Psychol. Gen. **115**, 39–57.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Pisoni, D. B., and Levi, S. V. (**2007**). "Some observations on representations and representational specificity in speech perception and spoken word recognition," in *The Oxford Handbook of Psycholinguistics*, edited by G. Gaskell (Oxford University Press, Oxford), pp. 3–18.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (**2006**). *Effects of Bilingualism, Noise, and Reverberation on Speech Perception by Listeners with Normal Hearing* (Cambridge University Press, Cambridge, UK), pp. 465–485.

Stevens, K. N. (**2002**). "Toward a model for lexical access based on acoustic landmarks and distinctive features," J. Acoust. Soc. Am. **111**, 1872–1891.

Stevens, K. N. (**2005**). "Features in speech perception and lexical access," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Malden), pp. 124–155.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., and Nishi, K. (**2001**). "Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners," J. Acoust. Soc. Am. **109**, 1691–1704.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H. Y., and Flege, J. (**2005**). "A developmental study of English vowel production and perception by native Korean adults and children," J. Phonetics **33**, 263–290.

Yang, B. G. (**1996**). "A comparative study of American English and Korean vowels produced by male and female speakers," J. Phonetics **24**, 245–261.