The influence of talker and foreign-accent variability on spoken word identification

Tessa Bent^{a)} and Rachael Frush Holt

Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, Indiana 47405

(Received 5 October 2011; revised 14 December 2012; accepted 2 January 2013)

In spoken word identification and memory tasks, stimulus variability from numerous sources impairs performance. In the current study, the influence of foreign-accent variability on spoken word identification was evaluated in two experiments. Experiment 1 used a between-subjects design to test word identification in noise in single-talker and two multiple-talker conditions: multiple talkers with the same accent and multiple talkers with different accents. Identification performance was highest in the single-talker condition, but there was no difference between the single-accent and multiple-accent conditions. Experiment 2 further explored word recognition for multiple talkers in single-accent versus multiple-accent conditions using a mixed design. A detriment to word recognition was observed in the multiple-accent condition compared to the single-accent condition, but the effect differed across the language backgrounds tested. These results demonstrate that the processing of foreign-accent variation may influence word recognition in ways similar to other sources of variability (e.g., speaking rate or style) in that the inclusion of multiple foreign accents can result in a small but significant performance decrement beyond the multiple-talker effect. © 2013 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4776212>]

PACS number(s): 43.71.Bp, 43.71.Hw [BRM]

Pages: 1677–1686

I. INTRODUCTION

Variability in the speech signal is pervasive. At the segmental level, there are many-to-many mappings between acoustic signals and perceptual phoneme categories (Lieberman *et al.*, 1967; Peterson and Barney, 1952). Extralinguistic, indexical sources of variability—such as talker characteristics and speaking rate—add additional sources of variability to all levels of sound structure. A substantial body of evidence suggests that much of the detail introduced by these sources of variability is encoded and retained in memory (Goh, 2005; Goldinger, 1997, 1998; Luce and Lyons, 1998; Palmeri *et al.*, 1993). Variability in the acoustic speech signal can cause performance decrements in spoken word recognition (McLennan and Luce, 2005). However, not all types of variability negatively influence perceptual processing (Bradlow *et al.*, 1999; Magnuson and Nusbaum, 2007; Nygaard *et al.*, 1995; Sommers and Barcroft, 2006; Sommers *et al.*, 1994). In the phonetic relevance hypothesis, Sommers and Barcroft (2006) proposed that variability in acoustic dimensions that affect the mapping of acoustic information to phonemic categories (e.g., speaking rate), as opposed to those that do not influence this mapping process (e.g., amplitude), are responsible for decrements in speech perception performance. One source of speech variability is the presence of a foreign accent. This investigation examines whether listeners incur additional processing costs when listening to speech from multiple talkers with different foreign accents over that experienced when listening to multiple talkers with the same foreign accent.

Speech variability has an effect on both speech processing and representation (McLennan and Luce, 2005). Maintaining detailed representations of talker-specific characteristics assists word identification for known talkers (Nygaard *et al.*, 1994). In contrast, both intra-talker and inter-talker variability can cause performance decrements. For example, in word identification and memory tasks, accuracy suffers when stimuli vary in speaking rate or style, or include multiple talkers or tokens (Bradlow *et al.*, 1999; Magnuson and Nusbaum, 2007; Mullennix *et al.*, 1989; Nygaard *et al.*, 1995; Sommers and Barcroft, 2006; Sommers *et al.*, 1994; Uchanski and Braid, 1998). However, changes across certain stimulus dimensions, such as amplitude and fundamental frequency, do not negatively influence perceptual performance (Bradlow *et al.*, 1999; Magnuson and Nusbaum, 2007; Nygaard *et al.*, 1995; Sommers and Barcroft, 2006; Sommers *et al.*, 1994). The phonetic relevance hypothesis (Sommers and Barcroft, 2006) reconciles these seemingly contradictory findings by arguing that not all sources of variability are equal; only acoustic properties that influence phonetic perception impair speech processing. For example, speaking rate and amplitude affect phonetic perception differently. Amplitude variation, in which speech stimuli are presented at higher or lower sound pressure levels, does not change how listeners perceive phonemic categories. In contrast, variations in speaking rate can affect how acoustic information is mapped onto phonemic categories (Miller and Liberman, 1979); the same acoustic stimulus may be perceived as one phoneme at one speaking rate and a different phoneme at a different speaking rate. In other words, only acoustic-phonetic variability relevant to phoneme categorization will hinder spoken word recognition (Sommers and Barcroft, 2006).

^{a)}Author to whom correspondence should be addressed. Electronic mail: tbent@indiana.edu

Here, the effect of variability on word identification was further investigated by testing another source of acoustic-phonetic variability—the presence of a foreign accent. Foreign-accented speech presents a particularly challenging example of variability in the speech signal. In addition to compensating for fine phonetic details that introduce substantial variability in native speech (e.g., idiolect differences, positional effects, coarticulatory effects, and speaking rate differences), listeners also need to be able to accommodate more extreme deviations from native language norms in foreign-accented speech. Foreign-accented speech can differ from native speech in both the segmental domain (e.g., phoneme additions, substitutions, deletions, and distortions) and the suprasegmental domain (e.g., stress, rhythm, and intonation differences).

Although foreign-accented speech can deviate substantially from native norms, the variability introduced is systematic and structured. The systematic relationship between the sound structures in the speaker's first and second languages results in speech production similarities across talkers from the same native language background. In other words, although foreign-accented productions of a given phoneme might differ acoustically from native norms, productions across speakers from the same native language background can be quite similar. For example, when English and Portuguese speakers are learning French, speakers within each language group make similar vowel production errors, whereas speakers across the two groups make different vowel production errors. The reason for this is that English and Portuguese differ from French in the number of high vowels in their inventories. English and Portuguese have only two high vowels (/i/ and /u/), whereas the vowel inventory of French contains three high vowels (/i/, /y/, and /u/). When English and Portuguese speakers are learning French, they tend to show systematic yet different substitution patterns when producing the French vowel they lack in their native inventory (/y/). A native English speaker is likely to produce the French /y/ as /u/, whereas a native Portuguese speaker tends to substitute /i/ for /y/ (Rochet, 1995). Due to these regularities within talkers and across talkers from the same language background, listeners can both perceptually adapt to specific foreign-accented talkers and can generalize learning from several talkers from the same native language background to novel talkers with the same foreign accent (Bradlow and Bent, 2008; Clarke and Garrett, 2004; Sidaras *et al.*, 2009; Weil, 2001). Learning how to better perceive foreign-accented speech requires listeners to change the way they map acoustic information onto phoneme categories (Sumner, 2011). Indeed, listeners have been shown to categorize the same acoustic vowel information in different ways depending on the regional accent of the sentential context (Evans and Iverson, 2004).

The current study investigates whether perceptual compensation for multiple foreign accents results in a word recognition decrement above and beyond that caused by multiple talkers with the same foreign accent. Specifically, native English listeners' word identification in conditions that included talkers with different foreign accents was compared to conditions in which multiple talkers with the same

foreign accent produced the words. In the experiments presented here, one foreign-accented word was presented for identification on each trial. Each word was mixed with speech-shaped noise (no babble was used). Due to the systematic acoustic-phonetic variability in speech production across talkers from the same native language background, we hypothesized that spoken word identification would be more accurate in conditions with multiple foreign-accented talkers from the same language background than conditions in which talkers with different foreign accents were included. Following the phonetic relevance hypothesis, these predicted results would be consistent with the notion that foreign-accent variation causes a reduction in spoken word recognition accuracy because the presence of multiple foreign accents is a source of acoustic-phonetic variation that affects phonetic perception. That is, the same acoustic input may be interpreted differently depending on the accent of the talker. Listening to a group of talkers who deviate from native language norms in similar ways may promote accurate word recognition. In contrast, accommodating trial-to-trial variations from talkers who demonstrate different sets of deviations from native language norms may hinder word recognition accuracy.

II. EXPERIMENT 1

In Experiment 1, we tested whether accommodating trial-to-trial variations in the foreign accent of talkers would result in performance decrements on a word identification task. Talker or speaking rate variation negatively impacts word identification due to the additional processing resources required to perceptually accommodate these differences. Similarly, when there are multiple foreign accents, mapping the acoustic information from foreign-accented speech to stored lexical items may be less accurate than in multiple talker conditions with a single foreign accent. The three conditions in this experiment extend previous work on perceptual accommodation in two ways. First, we assessed whether the robust finding of word identification decrements in multiple-talker conditions compared to single-talker conditions extends to foreign-accented speech. Second, in multiple-talker conditions, we assessed whether the presence of multiple foreign accents compared to a single foreign accent adds an additional processing cost for spoken word identification.

A. Method

1. Participants

Fifty-nine monolingual listeners with normal speech and hearing recruited from Indiana University and the surrounding Bloomington, IN community (18 male, 41 female) with an average age of 22 years (range = 19–28) participated in Experiment 1. Ten additional participants were excluded from final data analysis due to computer error resulting in incomplete data ($n = 9$) or outlier performance ($n = 1$; see details in Sec. II B). Normal hearing was evidenced by passing a pure-tone hearing screening of 20 dB hearing level (HL) at 500, 1000, 2000, and 4000 Hz, as well as 25 dB HL

at 250 Hz. Participants were randomly assigned to one of the three conditions (20 each in the single accent and multiple accent conditions and 19 in the single talker condition). Prior to participation in the experiment, all participants completed an informed consent form and a language background questionnaire. In the language background questionnaire, participants indicated the extent of their exposure to non-native speakers of English. Most participants reported limited exposure to non-native speakers (e.g., interactions at restaurants with non-native speakers, interactions during brief travel abroad or vacation experiences, a friendship with a non-native speaker, or instruction from an international professor or teaching assistant). None of the participants indicated extensive interactions with non-native speakers. Participants were paid for their participation.

2. Stimuli

The stimuli consisted of 200 unique monosyllabic words selected from the Phonetically Balanced-Kindergarten (PB-K) (Haskins, 1949) and Word Identification by Picture Identification (WIPI) lists (Ross and Lerman, 1970). Lexical characteristics of the words were obtained from the Hoosier Mental Lexicon (Nusbaum *et al.*, 1984). Of the 200 words, data were available for 176 of the words. Overall, words were highly familiar (average familiarity = 6.95 on a 7-point scale, standard deviation = 0.14), low in neighborhood density (average = 13.4, standard deviation = 7.6), and varied in frequency from 1 to 10 595 (average = 440, standard deviation = 1444). The recordings of these words were taken from the Hoosier Database of Native and Non-Native Speech for Children (Bent, 2010). This database includes recordings of words, sentences, and paragraphs produced by talkers from seven language backgrounds. All non-native talkers in the database had started studying English at the age of 10 or later and lived in the United States or other English speaking countries for four or fewer years.

Recordings were made in a sound-attenuated booth with a headset microphone (Shure Dynamic WH20XLR, Niles, IL) and a digital recorder (Marantz PDM670, Mahwah, NJ). All words, sentences, and paragraphs were segmented into individual wav files and equated for root-mean-square (RMS) amplitude.

Fifty words from one male native Japanese talker were included in all three conditions. These 50 words are referred to as the “target set.” The talkers producing the other 150 words varied in each condition. For the single talker condition, the male Japanese talker who produced the target set also produced the other 150 words. For the single accent condition, three additional native Japanese talkers (1 male and 2 female) each contributed 50 words, which were added to the target set. The multiple accent condition included the target set as well as 50 words from each of the following talkers: one German female, one Spanish male, and one Korean female. In an effort to minimize differences in intelligibility across conditions, words in each condition were matched as closely as possible in intelligibility based on previous word identification testing in quiet (Bent, 2010). It should be noted here that an inherent difference between the

single talker and multiple talker conditions was that gender varied in the two multiple talker conditions (single accent and multiple accent), but was necessarily constant in the single talker condition.

3. Procedure

Participants were seated individually in front of a computer (Mac mini, Apple, Cupertino, CA) with a 19.5-inch monitor in a sound attenuated booth. Up to three participants were tested at the same time. A custom-written Python script controlled stimulus presentation. Each participant was presented all 200 words in random order binaurally over headphones [Sennheiser (Old Lyme, CT) HD280 Pro] at an average signal level of 68 dB A. The words were embedded in a speech-shaped noise at +5 dB signal-to-noise ratio that was 1 s longer than the word with 500 ms of noise before the word and 500 ms after. On each trial, the section of noise was randomly selected from a 1-min noise file. The noise was added to ensure that no listener performed at ceiling in the task. The listeners’ task was to listen to each word and type in what she/he heard. The task was self-paced and listeners were presented each word only one time. No feedback was given as to the accuracy of the listener’s response. After typing a response for each trial, the listener pressed enter or a “next” button on the monitor, which advanced him/her to the next trial. After responding to 100 words, listeners received a short break.

As noted above, the same set of words by the same talker (i.e., the target set) was included in all three conditions. Thus, differences in intelligibility for the target set of 50 words should be related to the context in which the words were presented. For the single talker condition, listeners heard the target set within the context of other words by the same talker. In the single accent condition, listeners heard the target set in the context of three other Japanese talkers. Therefore, listeners in the single accent condition needed to perceptually compensate for talker differences, but the foreign accent remained constant throughout the experiment. In the multiple accent condition, listeners heard the set of 50 words in the context of three other talkers each with a different foreign accent, none of which was the same as the target talker. Therefore, listeners in the multiple accent condition needed to perceptually compensate for both general talker differences as well as differences in foreign accent across the talkers.

B. Results

The data were scored for word identification accuracy of the target set (i.e., the 50 words produced by the native Japanese talker included in all three conditions). Data from listeners whose scores were more than three standard deviations below or above the mean were assumed to be outliers and were excluded from analysis. Based on this procedure, one outlier in the single talker condition was removed. Responses were evaluated with a strict scoring criterion: words with added or deleted morphemes were counted as incorrect, and homophones and obvious misspelling were counted as correct.

Accuracy data from the target set of words in each condition were subjected to a one-way analysis of variance (ANOVA) with condition (single talker, single accent, multiple accent) as the between-subjects factor. The results revealed a significant main effect of condition, $F(2,58) = 5.28$, $p = 0.008$ (see Fig. 1). *Post hoc* pairwise comparisons (Statistical Package for the Social Sciences Bonferroni adjusted p -values are reported) were conducted to determine significant differences between conditions. Performance in the single talker condition was significantly better than performance in the single accent, $p = 0.014$, and the multiple accent conditions, $p = 0.029$. Performance in the single accent and multiple accent conditions was not significantly different.

C. Discussion

Listeners were able to accurately identify more words in the target set in the single talker condition compared to either of the conditions in which multiple talkers were present. Numerous studies have demonstrated a word identification accuracy advantage for single-talker conditions compared to multiple-talker conditions for native talkers. This study extends the single talker word identification accuracy advantage finding to a new talker population—foreign-accented talkers. Word identification accuracy in single- and multiple-talker conditions has been compared previously with foreign-accented sentences (Bradlow and Bent, 2008). A single-talker advantage for two of the four tested talkers was found, but caution is needed in interpreting these results because the scores were based on substantially different numbers of sentences in the single- and multiple-talker conditions. Thus, the current study extends previous results without the potential methodological confounds noted in the earlier study.

In single-talker conditions, listeners are able to exploit stability in talker characteristics and do not need to perceptually normalize for talker characteristics on each trial because

the initial normalization process can be used throughout the experiment. In contrast, the process of mapping the acoustic input to stored phonemic and lexical categories is more difficult when listeners have to adjust to trial-to-trial variation in the talker because they must normalize each time a talker change occurs. Here, the term “normalization” refers to the process of using talker-specific vocal characteristics to assist in the mapping between the acoustic input and stored representations. We do not assume that the normalization process involves a stripping away of extra-linguistic or indexical information. The process of talker normalization in conditions with multiple talkers appears to result in a similar processing cost with native- or foreign-accented talkers compared to conditions with a single native- or foreign-accented talker. When listeners are presented with foreign-accented talkers in multiple talker conditions, they are simultaneously making perceptual adjustments that are necessary to perceive foreign-accented speech accurately and completing the normalization process for talker-specific vocal characteristics that is required with any talker. The adjustments for foreign-accented speech are necessary because the acoustic realization of phoneme and lexical categories by foreign-accented talkers may differ substantially from the listener’s own phonemic and lexical categories and from the majority of his/her stored phonemic and lexical exemplars.

In contrast to the significant difference between the single- and multiple-talker conditions, there was no word identification advantage for multiple-talker conditions with a single foreign accent compared to those with multiple foreign accents. This finding suggests that the perceptual challenges posed by the inclusion of multiple talkers may be different than those posed by the inclusion of multiple foreign accents. That is, listeners showed more accurate word recognition in the single-talker condition than multiple-talker conditions presumably because they were able to exploit the stability of talker-specific characteristics in the single-talker condition. Listeners showed equivalent performance in the single- and multiple-accent conditions suggesting that listeners are not able to take advantage of the speech production regularities across talkers from the same native language background to promote accurate word recognition. It is possible that this type of accent normalization does not occur because listeners are primarily focused on talker-specific vocal characteristics. Previous work has demonstrated that talker effects are greater when speech processing is slower as is the case with foreign-accented speech (McLennan and Gonzalez, 2012). The resources needed to compensate for talker differences may not allow adequate resources for extracting the similarities across talkers with the same native language. Further, different processes may account for the decrease in word recognition for accented speech in general compared with conditions with multiple native talkers in noise. Assuming a native listener with normal hearing, word recognition will only be impaired in multiple talker conditions when adequate levels of noise are added. There is, therefore, a loss of information (Mattys *et al.*, 2012), which listeners can compensate for by relying on the stability of talker characteristics within a listening period. In contrast, the word recognition difficulties,

FIG. 1. Word identification accuracy on the target set by listeners in the three conditions in Experiment 1. The top and bottom of the box represent the 75th and 25th percentiles, respectively. The middle line in the box represents the median. The top whisker represents the 90th percentile and the bottom whisker represents the 10th percentile. The circles indicate individual participants who fell above the 90th percentile or below the 10th percentile.

which arise from single or multiple accented talkers, result from “acoustic-phonetic deviations from expectations” (Mattys *et al.*, 2012, p. 959). The normalization processes may differ for talker variability versus accent variability as the initial word recognition difficulties stem from different processes.

It remains possible that foreign-accent variability would incur a processing cost if the task or stimuli differed. The methodology used in the current experiment diverged from that used in previous studies investigating the influence of variability on word recognition. In the current experiment, the intelligibility of one specific set of words was compared relative to the characteristics of the other words presented in each condition. In previous studies, the intelligibility of all the words in each condition had been compared. There were several reasons the current methodology was selected. Intelligibility differences across non-native speakers are vastly more variable than across native speakers. Testing the intelligibility of one specific set of words (i.e., the “target set”) allowed for the assessment of the accent and talker context without other confounding factors. However, this methodology may not have been sensitive enough to capture differences across conditions, especially considering the range of individual variability across listeners. Further, this methodology limited the assessment of accents in the single-accent condition to one foreign accent. It is possible that inclusion of other accents or talkers would result in an intelligibility advantage for single-accent listening conditions over multiple-accent conditions. Although there are similarities in the acoustic-phonetic features across talkers from the same language background, there are differences as well due to talker-specific production patterns, language learning history, and proficiency, among other factors. Some sets of non-native talkers with the same first language may have more homogeneous production patterns and thus may more readily invoke a single-accent benefit.

III. EXPERIMENT 2

Experiment 2 was designed to further investigate foreign-accented word recognition in single-accent and multiple-accent conditions. In this experiment, a mixed design was employed to compare intelligibility in single-accent versus multiple-accent conditions. All listeners were exposed to both single-accent and multiple-accent blocks (a within-subjects variable) while the accent presented in the single-accent blocks varied across listeners (a between-subjects variable). There were several advantages to this design compared with the fully between-subjects design of Experiment 1. First, this method allowed for the examination of the accent variability effect in four sets of talkers, each with a different language background, rather than focusing on one foreign accent, as in Experiment 1. Second, the design allowed for greater experimental power for the overall analysis of single- versus multiple-accent conditions because every listener received both single- and multiple-accent blocks. Finally, in both the single- and multiple-accent conditions, multiple talkers from each foreign accent

were included so that there were multiple examples of each accent. In contrast, Experiment 1 had only one talker as a representative for each accent in the multiple-accent condition. Including multiple talkers from each accent should reduce the possibility that idiosyncratic characteristics of any one talker will impose a substantial impact on patterns of performance.

A. Method

1. Participants

Participants in Experiment 2 included 64 monolingual listeners with normal speech and hearing recruited from Indiana University and the surrounding Bloomington, IN community (22 male, 42 female) with an average age of 22.2 years (range = 18–34). None of the participants in Experiment 2 were tested in Experiment 1. Three additional participants were tested but excluded from final data analysis due to experimenter error. Inclusion criteria and participant payment were identical to Experiment 1.

2. Stimuli

The stimuli consisted of 320 unique monosyllabic words selected from the PB-K lists (Haskins, 1949), WIPI lists (Ross and Lerman, 1970), Lexical Neighborhood Test (Kirk *et al.*, 1995), the Northwestern University–Children’s Perception of Speech test (Elliott and Katz, 1980), and the Pediatric Speech Intelligibility test (Jerger and Jerger, 1984). As in Experiment 1, lexical characteristics of the words were obtained from the Hoosier Mental Lexicon (Nusbaum *et al.*, 1984). Of the 320 words, data were available for 274 of the words. Overall, words were highly familiar (average familiarity = 6.97 on a 7-point scale, standard deviation = 0.11), low in neighborhood density (average = 13.9, standard deviation = 7.5), and varied in frequency from 1 to 10 595 (average = 340, standard deviation = 1181). As in Experiment 1, the recordings of the words were taken from the Hoosier Database of Native and Non-Native Speech for Children (Bent, 2010) and were equated for RMS amplitude.

Recordings from 16 non-native talkers were used as stimuli. Of the 16 talkers, there were 4 talkers (2 male and 2 female) from each of the following native language backgrounds: German, Japanese, Korean, and Spanish. These talkers met the same criteria for inclusion as those in Experiment 1. Word intelligibility for the talkers from each language background based on previous testing in quiet (Bent, 2010) was as follows: German, 85, 86, 96, and 97%; Japanese, 69, 74, 77, and 84%; Korean, 74, 83, 84, and 86%; and Spanish, 72, 74, 75, and 80%.

3. Procedure

Experiment 2 employed a mixed design in which each listener was presented with a single-accent word block and a multiple-accent word block. Participants were divided into four groups (with 16 participants assigned to each) based on the language background of the talkers in the single-accent block (i.e., German, Japanese, Korean, or Spanish). That is, each listener only received one single-accent block in which

they were presented with words produced by talkers from one of the four accents tested. In the single-accent block, talkers from one foreign accent produced 160 words (2 male and 2 female talkers produced 40 words each). In the multiple-accent block, listeners were presented with words produced by 16 talkers with 4 different foreign accents (i.e., 4 German, 4 Japanese, 4 Korean, and 4 Spanish talkers). Gender of the talkers was balanced within each language background. Each of the 16 talkers in the multiple-talker block contributed 10 words for a total of 160 words. The order of the single-accent and multiple-accent blocks was counter-balanced across participants.

The words were also divided into two word sets. The words assigned to the single-accent or multiple-accent blocks were counter-balanced across participants. The words within a block were presented in a different randomized order for each participant. Participants received three breaks—one after each set of 80 words. The listener task, equipment, signal-to-noise ratio, stimulus presentation level, stimulus presentation procedures, and scoring procedures were the same as in Experiment 1.

The rationale for the design employed was to allow for the comparison of word recognition in single- versus multiple-accent listening conditions while controlling for inherent intelligibility differences across talkers. Although there were differences in intelligibility across talkers, these differences were controlled for because each talker contributed an equal number of words to the single-accent blocks and multiple-accent blocks. The design was based on previous methodologies used to investigate other dimensions of variability (Sommers, 1998; Sommers and Barcroft, 2006). However, prior investigations have not compared two multiple-talker conditions. The design necessitated the use of different numbers of talkers in the single- and multiple-accent blocks. Selecting a subset of specific talkers for the multiple-accent blocks could have skewed the results depending on which talkers were selected. That is, word recognition in the multiple-accent blocks could have been made artificially lower just by selecting specific talkers from each language background who were lower in intelligibility. Having multiple-accent blocks with the same number of talkers as the single-accent blocks while controlling for inherent intelligibility differences among the talkers would have required an onerous number of conditions to balance the talkers included. Further, previous investigations on how token variability influences speech perception have shown a decrement for the inclusion of four exemplars compared to one, but no further decrement when 16 exemplars were included (Uchanski and Braid, 1998).

B. Results

1. Overall comparison of intelligibility in single-versus multiple-accent listening conditions

The data were scored for word identification accuracy of the single-accent blocks compared to the multiple-accent blocks. The data were analyzed with a repeated-measures ANOVA with accent condition (single accent versus multiple accents) as the within-subjects variable and order (single-accent

FIG. 2. Word identification accuracy for talkers from the four language backgrounds in single-accent (black bars) and multiple-accent (gray bars) blocks in Experiment 2. Scores in the two blocks represent a within-listener comparison. Error bars represent standard error values.

block first versus multiple-accent block first) and language of the single-accent block (German, Japanese, Korean, or Spanish) as between-subjects variables. There were significant main effects of accent condition, $F(1,56) = 19.37$, $p < 0.001$, and language, $F(3,56) = 34.99$, $p < 0.001$. The main effect of accent condition occurred because listeners were more accurate in the single-accent block ($M = 52.2\%$ correct) compared to the multiple-accent block ($M = 49.6\%$ correct). The main effect of language occurred due the differences in intelligibility across the four accents: German was the most intelligible and Japanese, Korean, and Spanish were relatively similar in intelligibility (see Figs. 2 and 3). There was also a significant interaction between accent condition and language, $F(3,56) = 137.34$, $p < 0.001$. This interaction resulted from the finding that listeners in the German condition were more accurate in the single-accent block than the multiple-accent block. Listeners in the other language conditions showed equivalent or better

FIG. 3. Word identification accuracy for talkers from the four language backgrounds in single-accent (black bars) and multiple-accent (gray bars) blocks in Experiment 2. Scores in the two blocks represent a between-listener comparison. Error bars represent standard error values.

performance in the multiple-accent block. Other main effects and interactions were not significant.

2. *Intelligibility of each foreign accent in single-accent versus multiple-accent blocks*

Word recognition scores were compared between the single-accent blocks and the multiple-accent blocks for the talkers from each language background. This analysis was completed in two ways (see Table I). First, a within-subjects comparison was conducted. The listeners' scores on their single-accent block were compared with their word recognition scores for the talkers with the same accent in the multiple-accent block. For example, listeners in the German-accented condition received one score for the 160 words produced by the German-accented talkers in their single-accent block and received a second score for the 40 words produced by the same German-accented talkers in the multiple-accent block. This analysis allowed for a direct comparison of how intelligible the same set of talkers was in a context in which listeners heard only one accent versus a context in which they heard multiple accents. Paired *t*-tests were conducted to compare performance for each language background. The listeners assigned to the German-accented single-accent condition were more accurate on those talkers in the single-accent block compared to the multiple-accent block, $t(15) = 3.83$, $p = 0.002$, (see Fig. 2). In the other language background conditions, performance was not significantly different in the single-accent blocks compared to the multiple-accent blocks.

The second analysis was conducted to control for possible adaptation due to learning in the single-accent blocks. That is, the listeners who received the single-accent block first may have learned about the acoustic-phonetic features of the foreign accent (Bradlow and Bent, 2008; Clarke and Garrett, 2004; Sidaras *et al.*, 2009). Therefore, when the listeners were presented with the stimuli in the multiple-accent block, their performance may have been enhanced compared to the level expected if they would not have been exposed to the words in the single-accent block. In this case, differences between the single-accent and multiple-accent blocks would be attenuated due to adaptation effects. Further, conducting an analysis with only the listeners who received the multiple-accent block first would have reduced the number of listeners to eight per accent group and thus, severely limited experimental power. Therefore, to account for possible adaptation in the single-accent block, a between-subjects

analysis was conducted. In this analysis, the scores for the single-accent block were the same as for the within-subjects analysis described above (i.e., the word recognition accuracy scores on the single-accent block from the 16 listeners who were assigned to that particular accent). In contrast to the within-subjects analysis described above, the scores for each accent in the multiple-accent block were taken from the listeners who did *not* receive that accent in their single-accent block (see Table I). For example, all listeners heard 40 words from the Korean-accented talkers in the multiple-accent block. In this between-subjects analysis, the word recognition scores for these 40 words were analyzed based on the listeners who were assigned to a language background other than Korean for their single-accent block (i.e., German, Japanese, and Spanish). Thus, for Korean, the single-accent score was derived from the 16 listeners who were presented with Korean-accented talkers in the single-accent block and the multiple-accent score was derived from the 48 listeners who were assigned to the other three language background conditions in their single-accent block. Independent-samples *t*-tests were conducted to analyze performance for each language background separately. For three of the talker language backgrounds (German, Japanese, and Korean), intelligibility was higher in the single-accent block compared to the multiple-accent block: German, $t(61.83) = 2.88$, $p = 0.005$; Japanese, $t(44.17) = 2.20$, $p = 0.033$; Korean, $t(62) = 3.91$, $p < 0.001$ (Fig. 3). Intelligibility was not different in the single- versus multiple-accent blocks for the Spanish-accented talkers.

C. Discussion

The results from Experiment 2 suggest that perceptually compensating for multiple foreign accents in a word recognition task may cause a subtle, detrimental effect compared to multiple-talker conditions with a single foreign accent. Intelligibility was higher in the single-accent blocks overall compared with the multiple-accent blocks suggesting that there is a word recognition benefit in multiple-talker conditions when all the talkers have the same foreign accent. This result adds to the findings demonstrating a benefit for word recognition in conditions with single talkers, speaking rates, and speaking styles. The benefit in these conditions arises because listeners can perceptually normalize for one talker, rate, or style, and use that perceptual lens to facilitate recognition of words with similar characteristics. Similarly, because the mapping between the acoustic signal and phonemic categories are

TABLE I. Description of which participants' data were used to calculate the multiple-accent scores in the within-subjects and between-subjects analyses from Experiment 2. The single-accent scores were always derived from the 16 listeners who received the specific accent in their single-accent block.

Single-accent language assignment	Multiple-accent scores							
	German		Japanese		Korean		Spanish	
	Within	Between	Within	Between	Within	Between	Within	Between
German ($n = 16$)	X			X		X		X
Japanese ($n = 16$)		X	X			X		X
Korean ($n = 16$)		X		X	X			X
Spanish ($n = 16$)		X		X		X	X	

more consistent when listening to talkers with a single foreign accent than to talkers with multiple foreign accents, word recognition should be facilitated. For example, if a listener heard [rit], they may interpret it differently depending on the accent context. If they had been previously exposed to a number of different German-accented words, they may realize that German-accented talkers frequently devoice word final stops (e.g., “ground” produced as [graʊnt] and “end” produced as [ɛnt]) and interpret the word as likely being “read.” However, if the listener was exposed to Korean-accented English, they may have noted that Korean-accented talkers frequently substitute /t/ for /θ/ (e.g., “mouth” produces as [maʊt] and “thumb” produced as [tʌm]) and may interpret the acoustic input [rit] as “wreath.” Presentation of consistent deviations from native norms—which occurs for talkers from the same native language background—appeared to provide a small word recognition advantage.

To further explore how the single-accent effect may have differed across the language backgrounds tested in the current experiment, intelligibility of talkers from each foreign accent was examined in the single-accent versus multiple-accent blocks. In these analyses, the single-accent effect differed depending on the language background of the talkers. The variation in the effect across talker language backgrounds may be due to factors such as overall intelligibility level, homogeneity of the talkers within each language background, and listener experience with the accent.

The German-accented talkers showed a consistent single-accent effect. These talkers were more intelligible in the single-accent block than in the multiple-accent block in both analyses (within listeners and across listeners). The intelligibility level of these talkers may account for listeners’ ability to more readily benefit from single-accent listening conditions. Overall, the German-accented talkers were the most intelligible talkers included in the current study (e.g., 70% correct in the single-accent block). It may be easier to extract the accent regularities and determine how they deviate from native language norms with talkers who are more intelligible overall, with fewer and more consistent deviations from native pronunciation norms. In contrast, listeners accurately identified less than half of the words spoken by the talkers from the other language backgrounds. At these intelligibility levels without feedback, listeners may not have been accessing the correct lexical items frequently enough to tap into the consistencies within accents and robustly facilitate word recognition in single-accent listening conditions. However, it should be noted that the listeners assigned to the Korean-accented talkers for the single-accent block demonstrated the largest gain in the between-subjects analysis, suggesting that factors other than intelligibility are at play in determining the presence or extent of a single-accent benefit for word recognition.

In contrast to the findings for the German-accented talkers, the Japanese- and Korean-accented talkers showed the single-accent advantage for the between-subjects analysis, but not the within-subjects analysis in Experiment 2. The Spanish-accented talkers did not show a single-accent advantage in either type of analysis. In addition to the intelligibility factor discussed above, the attenuation or absence of

the single-accent effect in these cases may be related to the homogeneity of acoustic-phonetic patterns within- or across-talkers from these language backgrounds. Although an effort was made to recruit talkers who were similar in their language learning history (e.g., all talkers had been in the United States for four years or less) and dialect (e.g., all of the Spanish speakers were from Colombia), certain sets of talkers in the current experiment may have been more consistent in their production patterns than others. For example, Spanish does not have interdental fricatives and, therefore, speakers from this language background commonly have some difficulty acquiring these consonants. One speaker may consistently substitute /t/ for /θ/, while another substitutes /f/ for /θ/. Word recognition would be more difficult under these conditions than in situations in which all talkers made the same substitution. Similarly, within-talker deviations may vary in consistency. Some talkers may show consistent patterns of deletion, substitution, or distortion for sounds they are having difficulty acquiring, whereas other talkers may vary across words in the strategy they adopt toward difficult sounds. Without controlling specifically for the phonemic constitution of the words included in the current study, it is difficult to quantify homogeneity among talkers from the same language background. Future studies that systematically manipulate the consistency in deviations from native norms across talkers may help to determine how the level of consistency within- and across-talkers from a language background contributes to a single-accent word recognition advantage. Examining word recognition for talkers with varying regional dialects could also help to avoid within- and across-talker inconsistencies in the realization of particular sound contrasts. Native talkers from within one dialect region tend to produce consistent acoustic-phonetic realizations of lexical items (e.g., Wells, 1982) and all talkers would be highly proficient in the language. However, the processing mechanisms may be different for regional dialects compared to foreign accents (Adank *et al.*, 2009; Flocia *et al.*, 2006; Goslin *et al.*, 2012). Thus, perceptual compensation for regional dialect variability may be a distinct process from the foreign-accent compensation process investigated here.

Last, listener experience with a particular accent may interact with the single-accent benefit. Although all listeners in the current study did not report extensive experience with any of the foreign-accent included here, it is likely that listeners were most familiar with Spanish-accented English prior to participation in the experiment because Spanish is the most commonly spoken language in the United States other than English. With high familiarity signals, it is possible that listeners benefit less from low variability conditions. For example, in Sommers and Barcroft (2006), the speaking style with the smallest benefit between the single- and multiple-speaking style conditions was the “normal” speaking style. Similarly, in single- versus multiple-talker conditions, it is possible that listeners would not show the same variability effect with highly familiar voices. The interaction of familiarity and variability effects should be investigated further by providing listeners with explicit exposure to the accent or talker prior to testing listeners’ ability to

accommodate trial-to-trial variation along specific dimensions. Previous methods that have demonstrated perceptual adaptation to regional dialects and foreign accents (Bradlow and Bent, 2008; Clarke and Garrett, 2004; Floccia *et al.*, 2006) could be administered before the completion of a word recognition task comparing single- and multiple-accent conditions. In this type of experimental design, the impact of familiarity and adaptation could be experimentally manipulated. For conditions in which the talkers are lower in intelligibility, a period of training with feedback could assist listeners in mapping the non-canonical pronunciations to stored lexical items. By accurately making this mapping through the use of feedback, listeners could determine how talkers from a particular native language background deviate from native language norms. After a period of exposure with feedback, listeners may be more likely to demonstrate a word recognition advantage in single-accent conditions.

IV. GENERAL DISCUSSION

The presence of multiple talkers compared to a single talker impaired word identification, as seen in Experiment 1. These findings extend previous results demonstrating the deleterious effect of talker variability on speech processing to foreign-accented speech. The results suggest that the talker normalization process is not contingent on the realization of phoneme contrasts that map onto the bulk of exemplars stored in memory. In other words, the process of adjusting to talker-specific characteristics—such as making the necessary perceptual adjustments for vocal tract size—occurs regardless of whether there is a match between the talker's and the listener's native language background (Winters *et al.*, 2008).

This study also demonstrated that under some circumstances there is a word recognition benefit for conditions in which listeners are presented with talkers from one native language background versus multiple language backgrounds. This result adds to the literature demonstrating that variability along several stimulus dimensions—including talker, speaking rate, speaking style, and now foreign accent—results in decrements to word recognition. However, the magnitude of the effect was relatively small compared to the effects seen for other types of variability and appeared to be dependent on the experimental methodology and the characteristics of the talkers included in each accent set. This set of experiments is a first attempt to examine the effects of foreign-accent variability on spoken word recognition. Future investigations will begin teasing out the very complicated, real-world factors that influence word recognition under single- and multiple-accent listening conditions.

Adjusting for foreign-accent differences involves a number of different perceptual processes including adjusting for phoneme categories that are outside the normal range of variability for native talkers (i.e., phoneme distortions) as well as processing phoneme substitutions or “bad maps” (Sumner, 2011, p. 132). In the single-accent and multiple-accent conditions, listeners were making these perceptual adjustments for foreign-accented speech and simultaneously adjusting to trial-to-trial variation in talker-specific charac-

teristics. When adjusting to talker variation from trial to trial, there is a small but significant benefit to listening to talkers from the same native language background compared to those from different language backgrounds. It is possible that normalization for talker variability nearly obscures the adjustment process that occurs with foreign-accent variability because these two factors are confounded in the current experiments. That is, there was not a condition in which foreign-accent variability existed but the talker remained constant (i.e., the same talker producing various words with different foreign accents). A condition of this type would be more analogous to the methodology used with speaking rate and style variation in which the talker remains constant but the rate or style changes from trial to trial. Unlike style or rate variation, production of foreign-accent variation within a single talker is not a typical form of variability and would require the recruitment of someone skilled at producing different foreign accents. Further, the imitation of foreign accents would not necessarily reproduce the type and extent of variability introduced by actual foreign-accented talkers.

The results from the current studies provide some support for the Phonetic Relevance Hypothesis: under specific listening conditions, there was a benefit for single-accent listening conditions over multiple-accent listening conditions. It was hypothesized that word recognition with multiple foreign-accented talkers from a single language background should be more accurate than multiple foreign-accented talkers from different language backgrounds because the language background of the talker affects the mapping of acoustic information onto phoneme categories. When presented with words produced by talkers from the same language background, listeners can interpret acoustic information more consistently from talker to talker than with talkers from multiple language backgrounds. The current results add to the investigations of other stimulus dimensions that change phonetically relevant properties of the speech signal such as talker, speaking rate, or speaking style variability.

V. CONCLUSION

The two experiments described here explored the influence of talker and foreign-accent variability on word identification. The results from these studies extend the findings of talker variability effects to a new talker population—foreign-accented talkers. Further, there appears to be a subtle but significant benefit for word recognition in conditions with multiple talkers from a single language background compared to multiple talkers from multiple language backgrounds. This single-accent benefit differed across the foreign accents included in the investigation. Future studies should investigate how the effect of overall intelligibility level, consistency of acoustic-phonetic patterns across talkers from a language background, and listener experience contribute to the perceptual benefit of single-accent conditions over multiple-accent conditions.

ACKNOWLEDGMENTS

We would like to thank Charles Brandt for programming the experiments, Kierra Villines, Nancy Eastman,

Steven Elmlinger, Eileen Sisk, Jessica Copperman, and Marissa Ganeku for data collection assistance, and the NIH-NIDCD (Grant No. R21-010027) and Indiana University for providing funding to support this research.

- Adank, P., Evans, B. G., Stuart-Smith, J., and Scott, S. K. (2009). "Comprehension of familiar and unfamiliar native accents under adverse listening conditions," *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 520–529.
- Bent, T. (2010). "Native and non-native speech database for children," *J. Acoust. Soc. Am.* **127**, 1905.
- Bradlow, A. R., and Bent, T. (2008). "Perceptual adaptation to non-native speech," *Cognition* **106**, 707–729.
- Bradlow, A. R., Nygaard, L. C., and Pisoni, D. B. (1999). "Effects of talker, rate, and amplitude variation on recognition memory for spoken words," *Percept. Psychophys.* **61**, 206–219.
- Clarke, C. M., and Garrett, M. F. (2004). "Rapid adaptation to foreign-accented English," *J. Acoust. Soc. Am.* **116**, 3647–3658.
- Elliott, L. L., and Katz, D. (1980). *Development of a New Children's Test of Speech Discrimination (Technical Manual)* (Auditec, St. Louis, MO).
- Evans, B. G., and Iverson, P. (2004). "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences," *J. Acoust. Soc. Am.* **115**, 352–361.
- Floccia, C., Goslin, J., Girard, F., and Konopczynski, G. (2006). "Does a regional accent perturb speech processing?," *J. Exp. Psychol. Hum. Percept. Perform.* **32**, 1276–1293.
- Goh, W. D. (2005). "Talker variability and recognition memory: Instance-specific and voice-specific effects," *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 40–53.
- Goldinger, S. D. (1997). "Words and voices: Perception and production in an episodic lexicon," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego), pp. 33–66.
- Goldinger, S. D. (1998). "Echoes of echoes? An episodic theory of lexical access," *Psychol. Rev.* **105**, 251–279.
- Goslin, J., Duffy, H., and Floccia, C. (2012). "An ERP investigation of regional and foreign accent processing," *Brain Lang.* **122**, 92–102.
- Haskins, H. (1949). "A phonetically balanced test of speech discrimination for children," Master's, Northwestern University, Evanston, IL.
- Jerger, S., and Jerger, J. (1984). *Pediatric Speech Intelligibility Test* (Auditec, St. Louis, MO).
- Kirk, K. I., Pisoni, D. B., and Osberger, M. J. (1995). "Lexical effects on spoken word recognition by pediatric cochlear implant users," *Ear Hear.* **16**, 470–481.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). "Perception of the speech code," *Psychol. Rev.* **74**, 431–461.
- Luce, P. A., and Lyons, E. A. (1998). "Specificity of memory representations for spoken words," *Mem. Cognit.* **26**, 708–715.
- Magnuson, J. S., and Nusbaum, H. C. (2007). "Acoustic differences, listener expectations, and the perceptual accommodation of talker variability," *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 391–409.
- Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). "Speech recognition in adverse conditions: A review," *Lang. Cognit. Processes* **27**, 953–978.
- McLennan, C. T., and Gonzalez, J. (2012). "Examining talker effects in the perception of native- and foreign-accented speech," *Attention Percept. Psychophys.* **74**, 824–830.
- McLennan, C. T., and Luce, P. A. (2005). "Examining the time course of indexical specificity effects in spoken word recognition," *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 306–321.
- Miller, J. L., and Liberman, A. M. (1979). "Some effects of later-occurring information on the perception of stop consonant and semivowel," *Percept. Psychophys.* **25**, 457–465.
- Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (1989). "Some effects of talker variability on spoken word recognition," *J. Acoust. Soc. Am.* **85**, 365–378.
- Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (1984). "Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words," *Research on Speech Perception Progress Report No. 10*, 357–376.
- Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (1994). "Speech perception as a talker-contingent process," *Psychol. Sci.* **5**, 42–46.
- Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (1995). "Effects of stimulus variability on perception and representation of spoken words in memory," *Percept. Psychophys.* **57**, 989–1001.
- Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (1993). "Episodic encoding of voice attributes and recognition memory for spoken words," *J. Exp. Psychol. Learn. Mem. Cogn.* **19**, 309–328.
- Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- Rochet, B. L. (1995). "Perception and production of second-language speech sounds by adults," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 379–410.
- Ross, M., and Lerman, J. (1970). "A picture identification test for hearing-impaired children," *J. Speech Hear. Res.* **13**, 44–53.
- Sidas, S. K., Alexander, J. E. D., and Nygaard, L. C. (2009). "Perceptual learning of systematic variation in Spanish-accented speech," *J. Acoust. Soc. Am.* **125**, 3306–3316.
- Sommers, M. S. (1998). "Spoken word recognition in individuals with dementia of the Alzheimer's type: Changes in talker normalization and lexical discrimination," *Psychol. Aging* **13**, 631–646.
- Sommers, M. S., and Barcroft, J. (2006). "Stimulus variability and the phonetic relevance hypothesis: Effects of variability in speaking style, fundamental frequency, and speaking rate on spoken word identification," *J. Acoust. Soc. Am.* **119**, 2406–2416.
- Sommers, M. S., Nygaard, L. C., and Pisoni, D. B. (1994). "Stimulus variability and spoken word recognition. 1. Effects of variability in speaking rate and overall amplitude," *J. Acoust. Soc. Am.* **96**, 1314–1324.
- Sumner, M. (2011). "The role of variation in the perception of accented speech," *Cognition* **119**, 131–136.
- Uchanski, R. M., and Braid, L. D. (1998). "Effects of token variability on our ability to distinguish between vowels," *Percept. Psychophys.* **60**, 533–543.
- Weil, S. A. (2001). "Foreign accented speech: Adaptation and generalization," Master of Arts, The Ohio State University, Columbus, OH.
- Wells, J. C. (1982). *Accents of English: An Introduction* (Cambridge University Press, Cambridge, UK).
- Winters, S. J., Levi, S. V., and Pisoni, D. B. (2008). "Identification and discrimination of bilingual talkers across languages," *J. Acoust. Soc. Am.* **123**, 4524–4538.