Published in final edited form as:

J Acoust Soc Am. 2007 April ; 121(4): 2327–2338.

Speaker-independent factors affecting the perception of foreign accent in a second language^{a)}

Susannah V. Levi^{b)}, Stephen J. Winters^{c)}, and David B. Pisoni

Speech Research Laboratory, Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47405

Abstract

Previous research on foreign accent perception has largely focused on speaker-dependent factors such as age of learning and length of residence. Factors that are independent of a speaker's language learning history have also been shown to affect perception of second language speech. The present study examined the effects of two such factors—listening context and lexical frequency—on the perception of foreign-accented speech. Listeners rated foreign accent in two listening contexts: auditory-only, where listeners only heard the target stimuli, and auditory + orthography, where listeners were presented with both an auditory signal and an orthographic display of the target word. Results revealed that higher frequency words were consistently rated as less accented than lower frequency words. The effect of the listening context emerged in two interactions: the auditory + orthography context reduced the effects of lexical frequency, but increased the perceived differences between native and non-native speakers. Acoustic measurements revealed some production differences for words of different levels of lexical frequency, though these differences could not account for all of the observed interactions from the perceptual experiment. These results suggest that factors independent of the speakers' actual speech articulations can influence the perception of degree of foreign accent.

I. INTRODUCTION

The ability to speak a second language fluently depends in large part on how well a speaker has been able to acquire the second language (L2) phonology and to accurately realize the intended phonetic targets. The perceived degree of foreign accent of a speaker, however, is not based exclusively on the amount of acoustic and articulatory mismatches between non-native and native productions. Degree of foreign accent also reflects a *listener's* perception of the L2 speech. Many of the factors known to affect the perception of foreign-accented speech are speaker-specific factors that are inherent to a particular speaker. We will refer to these factors as *speaker dependent* since they are dependent upon a particular speaker's language learning history and cannot be directly changed or manipulated by an experimenter. Speaker-dependent factors have received considerable attention in the L2 literature. They include age of learning (the age at which a speaker begins learning a second language), length of residence in an L2 environment, the first language of the speaker, and his/her motivation to attain unaccented or less-accented speech [see Piske *et al.* (2001) for a review].

^{a)}Portions of this work were presented at the 1st ASA workshop on Second Language Speech Learning held on 14–15 May 2005 in Vancouver, B.C., Canada.

© 2007 Acoustical Society of America.

^{b)}svlevi@indiana.edu.

^{c)}Now at the Department of Linguistics, University of Illinois, Urbana, Illinois 61801.

Additional factors that are not inherent to a particular speaker and are not part of the speaker's language learning history can also affect the perception of degree of foreign accent. These factors can be manipulated or controlled by the researcher and often reflect the specific methodology involved in obtaining measures of degree of foreign accent. We will refer to these as *speaker-independent* factors. For example, Southwood and Flege (1999) suggest that different rating scales may affect participants' judgments of perceived degree of foreign accent. They point out that scales with fewer intervals may produce ceiling effects and therefore not be sensitive enough to differentiate L2 speakers.

Different types of elicitation techniques can also affect the degree of perceived foreign accent. Studies investigating the perception of foreign accent have used a variety of techniques to produce their stimulus materials; these techniques vary in whether the L2 speakers spontaneously generate speech, read printed text (words, sentences, or paragraphs), or repeat samples of speech after hearing the intended target produced by a native speaker. Oyama (1976) and Thompson (1991) have found that read speech is judged as more accented than spontaneous speech.

Studies also differ in whether native speaker controls are included in the rating task. Native controls serve to confirm that listeners are correctly performing the task by testing that they can distinguish native from non-native speech. Using native controls also ensures that listeners use a wider range of the rating scale (Flege and Fletcher, 1992). The proportion of native speakers included in a rating set also affects the perception of foreign accent. Flege and Fletcher (1992) found that increasing the proportion of native speakers in the stimulus set caused non-native speakers to be rated as more accented. Characteristics of the listener can affect the perceived degree of foreign accent, as well. Studies have varied whether naive listeners (e.g., Flege and Fletcher, 1992; Flege *et al.*, 1995) or experienced listeners such as linguists (e.g., Fathman, 1975) and ESL teachers (e.g., Piper and Cansin, 1988) serve as raters. Thompson (1991) found that naive listeners tended to perceive a greater degree of foreign accent than experienced listeners, although Bongaerts *et al.* (1997) did not find a significant difference. Flege and Fletcher (1992) also found that if listeners are familiarized with the target sentences, then non-native speakers are rated as more accented. An additional speaker-independent factor can be speaking rate, when the change in rate is caused by experimental manipulation. Munro and Derwing (2001) used speech compression-expansion software to increase and decrease speaking rate by 10%. By using this software they were able to ensure that other properties of the speech (e.g., number of segmental substitutions) remained unchanged. They found that fast stimuli were rated as less accented than stimuli presented at normal and slowed rates. Taken together, these studies show that speaker-independent factors, in addition to speaker-dependent factors, can also affect the perceived degree of foreign accent.

The current study investigated the effects of two additional speaker-independent factors—lexical frequency and listening context—on the perception of degree of foreign accent using an accent rating task. These two factors were chosen because they have been shown to affect speech perception and language processing of native speech. This study extends these two factors to the perception of foreign-accented speech.

Lexical frequency has been found to play an integral role in language processing and may therefore be expected to affect the perception of degree of foreign accent. Lexical frequency affects spoken word recognition (Howes, 1957; Savin, 1963; Luce and Pisoni, 1998), the recognition of words in a gating paradigm (Grosjean, 1980), and word shadowing (Goldinger, 1997). In a word identification task, Howes (1957) mixed words of varying frequency with multiple signal-to-noise ratios. High frequency words exhibited greater intelligibility by being perceived at less favorable signal-to-noise ratios than were low

frequency words. In a similar study, Savin (1963) examined listeners' response errors and found that incorrect responses tended to be words of higher frequency than the target word. In a lexical decision task, Luce and Pisoni (1998) found that listeners responded more quickly and more accurately to high frequency words than to low frequency words.

Goldinger (1997) showed that listeners rely more heavily on the acoustic-phonetic information in the speech signal when they perceive low frequency words than when they perceive high frequency words. Using a word shadowing task, Goldinger presented listeners with both high and low frequency words from a variety of speakers and asked them to repeat the words as quickly as possible. Goldinger predicted that subjects would change their productions to match the different speakers using "spontaneous vocal imitation." The amount of vocal imitation was quantified by comparing how well the response utterances matched the stimulus in fundamental frequency and duration. Goldinger found that low frequency words resulted in higher rates of spontaneous imitation than high frequency words, suggesting that subjects were more sensitive to surface acoustic-phonetic details in low frequency words than in high frequency words.

Goldinger explained these findings within the framework of Hintzman's (1986, 1988) MINERVA2 model, an exemplar-based model of memory (see also Johnson, 1997; Kirchner, 1999, 2004; Pierrehumbert, 2001, 2002). The MINERVA2 model, like other exemplar models, assumes that every exposure to a stimulus creates a memory trace that includes all perceptual details. When a new token (the probe) is heard, it activates an aggregate of all traces in memory, called the *echo*. This *echo* forms the listener's percept. The intensity of the echo depends upon both the similarity of the traces to the probe and the number of these traces. Thus, for speech and language processing, high frequency words induce "generic" echoes because they have many existing traces in memory and are therefore less influenced by any particular probe that enters the perceptual system. Low frequency words, on the other hand, have many fewer existing traces in memory. Any incoming probe will therefore have a greater influence on the subsequent percept. For the low frequency words in Goldinger's word shadowing task, speakers based their repetitions more heavily on the incoming instance-specific information than on traces in memory for low frequency words. Their subsequent productions of low frequency words were therefore affected more by specific properties of the stimulus than were high frequency words.

Working within the framework of exemplar models of speech perception, we hypothesized that the degree to which a speaker is perceived to have a foreign accent will be directly related to the amount of acoustic-phonetic mismatch between the signal and its resulting echo. In a nativeness rating task, we expected listeners' perception of L2 speech to rely more heavily on the acoustic-phonetic features of an incoming speech token for low frequency words. Listeners have fewer exemplars of low frequency words in memory and will thus generate less generic echoes in response to productions of those words. Potential acoustic-phonetic mismatches between productions of those words and their corresponding exemplars in memory should therefore be larger for low frequency words, which should in turn be rated as more accented than high frequency words. An alternative hypothesis is that the lexical frequency of the target word will have no effect on the perception of foreign accent because accent rating does not require accessing the lexicon and therefore can be based solely on the phonetic and phonological properties of the stimulus.

The second speaker-independent factor investigated in this study was the listening context. Spoken words were either presented to participants in the auditory modality alone (auditory-only) or with the addition of a simultaneous orthographic display (auditory+orthography). Previous work has shown that knowledge of the intended target, as in the auditory +orthography context, facilitates the perception of degraded speech stimuli (Davis *et al.*,

2005). Since non-native speech can be considered a type of degraded stimuli, the same type of facilitation should be found. Davis *et al.* use the term “pop-out” to refer to a phenomenon where a degraded speech stimulus immediately becomes comprehensible after it is played to listeners in its original, undegraded form. Davis *et al.* tested the effects of pop-out on a type of noise-vocoded speech that simulates the signal heard by cochlear implant users.¹ In one experiment, they found that listeners were able to correctly report more words from a noise-vocoded target sentence after hearing the sentence in the clear. In another experiment, they found that listeners showed the same advantage, or pop-out effect, after seeing the written version of a noise-vocoded sentence presented on a computer screen. This combination of effects demonstrates that top-down processing can influence the perception of severely degraded, noise-vocoded speech regardless of the modality in which the original sentences are presented. As Davis *et al.* concluded, “pop-out must be at a non-acoustic, phonological level or higher” (p. 230).

The effects of this type of pop-out on the perception of degree of foreign accent are unclear, however. One possibility is that simultaneously presenting the auditory and orthographic representations of the target word together will cause non-native speech samples to be rated as less accented. If a non-native production of a target word is ambiguous or difficult to understand, presenting the target word in orthography on the screen may promote a type of pop-out effect to occur where the “degraded,” non-native production immediately becomes more intelligible. Once the listener knows the intended utterance, possible ambiguities or confusions about which lexical item the listener should retrieve are lost. In this case, the perception of a high degree of foreign accent may also be significantly attenuated.

A second possibility is that simultaneously presenting auditory and orthographic representations of the target word will cause non-native speech samples to be rated as more accented. This outcome might occur because knowledge of the target word may serve as a perceptual benchmark and therefore highlight the amount of mismatch between the target and its corresponding exemplars in memory. An actual example from our data serves to illustrate this point. Several of the L2 speakers in the current study consistently produced word final target/s/as [z]. For these speakers, the target word “noose” [nus] was produced as [nuz] (identical to “news,” which was not one of the target words). Hearing [nuz] while seeing “noose” could focus listeners’ attention on mismatches between the expected and observed productions. It might be expected that listeners would rate these speakers as having more of a foreign accent when they hear the word [nuz] in conjunction with seeing “noose” on the screen than when they simply hear [nuz] alone and could freely conclude that they had heard an accurate production of “news.”

To summarize, the current study examined the effects of lexical frequency and listening context on the perceived degree of foreign accent of native and non-native speakers of English. We predicted that higher frequency words would be rated as less accented than lower frequency words. In terms of the listening context, two competing hypotheses were assessed. The addition of orthographic displays may induce pop-out effects, making the stimuli more intelligible, resulting in their being rated as less accented. Alternatively, the presentation of the target word may cause listeners to focus their attention on mismatches between the target utterance and the actual stimuli, resulting in the stimuli being rated as more accented. Positive results with these two factors will provide additional evidence for the contribution of speaker-independent factors in the perception of foreign accent.

¹This type of speech is created by filtering the original signal into six logarithmically spaced frequency bands.

II. EXPERIMENT: PERCEPTION OF FOREIGN ACCENT

A. Methodology

1. Materials—All speakers were recorded in a sound-attenuated IAC booth in the Speech Research Laboratory at Indiana University. Speech samples were recorded using a SHURE SM98 headmounted unidirectional (cardioid) condenser microphone with a flat frequency response from 40 to 20 000 Hz. Utterances were digitized into 16-bit stereo recordings via Tucker-Davis Technologies System II hardware at 22 050 Hz and saved directly to a PC. A single repetition of 360 English words was produced by each speaker. Each word was of the form consonant-vowel-consonant (CVC) and was selected from the CELEX database (Baayen *et al.*, 1995). Speakers read each word in random order as it was presented to them on a computer monitor in the recording booth. Before each presentation, an asterisk appeared on the screen for 500 ms, signaling to the speaker that the next trial was about to begin. This was followed by a blank screen for 500 ms. After this delay, a recording period began that lasted for 2000 ms. The target word was presented on the screen for the first 1500 ms of this recording period. After the conclusion of the recording period, the screen went blank for 1500 ms, and then another asterisk appeared to signal the beginning of the next recording cycle. Items that were produced incorrectly or too loudly were noted and re-recorded in the same manner following the recording session. The total recording time was approximately one hour.

This process yielded recordings that were uniformly 2000 ms long. Since the actual productions of the stimulus words were always shorter than 2000 ms, the silent portions in the recording before and after each production were manually removed using Praat sound editing software (Boersma and Weenink, 2004). All edited tokens were then normalized to have a uniform RMS amplitude of 66.4 dB.

Words were selected to represent a range of lexical frequencies based on counts from the CELEX database. For the purposes of analysis, words were divided into three equal groups of varying frequency. The 120 lowest frequency words all had a CELEX frequency count of less than or equal to 96, while the 120 highest frequency words all had a frequency of greater than or equal to 586. The remaining 120 words thus all had frequency counts between 96 and 586. The frequency count of homophones (e.g., rite, write, right) was taken to be the frequency count of the most frequent homophone; this homophone was also the word that was presented orthographically to the speakers during the recording sessions.

2. Speakers—Twelve female and ten male German L1/English L2 speakers were recorded. Of the 22 speakers, nine speakers were eliminated due to dialect differences (Austrian German: $N=3$, Southern German: $N=2$, Romanian-German: $N=1$), reported speech or hearing disorders ($N=2$), or only completing part of the recordings ($N=1$). Recordings from the remaining seven female and six male speakers were used in this study. All speakers were paid \$10/h for their time. Demographic variables for the remaining bilingual speakers are given in Table I.

Thirteen native speakers (six male, seven female) of American English were also recorded producing only the list of English words under the same conditions as the bilingual speakers. These speakers were from various dialect areas of American English (Midland: $N=7$, West: $N=1$, South: $N=1$, North: $N=1$, more than one dialect area: $N=3$). [See Labov *et al.* (2006) for descriptions of these dialect labels.] Productions from two of the female speakers were not included in the study due to problems these speakers had with completing the task accurately. Productions from the remaining six male and five female native speakers were included in the study. All of these speakers received partial course credit for their participation.

3. Listeners—A total of 87 listeners between the ages of 18 and 25 participated in this experiment; 42 were assigned to the auditory-only context and 45 to the auditory+orthography context. Twenty-seven listeners were eliminated (polylingual/non-native speakers of English: $N=6$, L2 German: $N=8$, machine malfunction: $N=9$, non-American English dialect: $N=1$, speech/hearing disorder: $N=2$, not completing: $N=1$), resulting in 30 native English listeners in each listening context. None of the remaining listeners had studied German, and only 6 reported having German acquaintances (friend: $N=3$, teaching assistant: $N=2$, professor: $N=1$). All remaining listeners reported no history of a speech or hearing disorder at the time of testing. Each listener participated in only one of the two listening contexts. All listeners were students enrolled in introductory psychology courses at Indiana University and received partial course credit for their participation.

4. Procedure—The experiment was implemented on Macintosh G3 computers running a customized SuperCard (version 4.1.1) stack. Listeners were seated in front of these computers in a quiet testing room while wearing Beyerdynamic DT-100 headphones. Stimuli were presented at a comfortable listening level (approximately 65 dB SPL) to all subjects. The SuperCard stack played productions of individual words to listeners and then presented them with the on-screen question, “How much of a foreign accent did that speaker have?” Participants answered this question by clicking the appropriate button in a seven-point rating scale ranging from 0 (“no foreign accent—native speaker of English”) to 6 (“most foreign accent”) presented on-screen. All listeners were informed that some of the speakers they would hear were native speakers of English and some were non-native speakers. All listener ratings were converted to normalized z -scores per listener prior to statistical analysis.

The auditory tokens of each word were presented to listeners in one of two different ways. Listeners in the auditory-only context heard each word prior to making a judgment of how accented the spoken stimulus was. Listeners in the auditory+orthography context, however, saw the orthographic representation of each word on the computer screen for 500 ms before hearing an auditory production of that word. The orthographic representation of the word remained on screen until the conclusion of the auditory stimulus, after which the listener rated its accentedness.

The experiment was divided into two blocks. In each block, 12 words were randomly selected for presentation from each of the 11 monolingual and 13 bilingual speakers, yielding a total of 288 tokens per block. Listeners thus heard a total of 576 words over the duration of the entire experiment. Each block of words was rated by two different listeners. The experiment was self-paced and took approximately one hour for most listeners to complete.

Participants had the option of listening to stimuli again after the initial presentation. In the auditory+orthography context, participants listened to 89.5% of the tokens only once and to 10.5% more than once prior to making their responses. Participants in the auditory-only context listened to 78.5% of the tokens once and to 21.5% two or more times. An ANOVA of the percentage of trials that listeners elected to hear more than once was conducted with listening context (auditory-only versus auditory+orthography) as a between-subjects factor and with native language of the speaker (L1 English versus L2 English) as a within-subjects factor. The ANOVA revealed main effects of both listening context [$F(1,58) = 5.688$, $p=0.020$] and native language [$F(1,58) = 5.617$, $p=0.021$], but no interaction. Listeners replayed stimuli more often in the auditory-only context than in the auditory+orthography context (means: 21.5% versus 10.5%, respectively). Furthermore, listeners replayed words from native speakers of English more often than from non-native speakers of English (16.6% versus 15.3%).

To assess the overall consistency and reliability of raters, the intraclass correlation coefficient was calculated for raters in each of the two listening contexts (McGraw and Wong, 1996). The intraclass correlation coefficient was high for both groups (auditory-only, 0.734; auditory+orthography, 0.798), indicating a strong degree of agreement among the raters. The reliability for all judges averaged together, sometimes referred to as the interrater reliability coefficient (MacLennan, 1993) was 0.988 for the auditory-only listeners and 0.991 for the auditory+orthography listeners.

B. Results

Before analysis, responses were pooled across speakers and across stimulus items, yielding an average measure of degree of foreign accent for each of the three levels of lexical frequency and the two levels of speakers (native versus non-native). An ANOVA with lexical frequency (low, medium, high) and native language of the speaker (native, non-native) as within-subjects variables and listening context (auditory-only, auditory+orthography) as a between-subjects variable was conducted on the *z*-scores of the nativeness ratings for all listeners. In the presentation of the results, larger *z*-score ratings indicate a greater degree of foreign accent.

The ANOVA revealed a significant main effect of lexical frequency [$F(2,116) = 44.8$, $p < 0.001$]. Paired-samples *t* tests revealed significant pairwise differences between all three levels of lexical frequency (all $p \leq 0.002$). The direction of this effect indicated that lower frequency words were rated as more accented than words of higher frequency. A main effect of native language of the speaker was also found [$F(1,58) = 1214.8$, $p < 0.001$] where native speakers were rated as having less foreign accent overall than non-native speakers. The main effect for listening context was not significant.

The analysis also revealed three significant interactions. The interaction between lexical frequency and native language of the speaker [$F(2,116) = 6.51$, $p = 0.002$] is shown in Fig. 1. Paired-samples *t* tests revealed that ratings for the medium and high frequency words differed between the two groups of speakers. For native speakers of English, low frequency words were rated as having more of a foreign accent than medium frequency words, which were in turn rated as more accented than high frequency words (all $p \leq 0.001$). For non-native speakers, low frequency words were rated as more accented than both medium and high frequency words ($p < 0.001$). No significant difference between the medium and high frequency words ($p = 0.213$) was found for the non-native speakers, although the trend was in the same direction as native English speakers with medium frequency words rated as more accented than high frequency words.

To further investigate this interaction, we examined the differences between the perception of native and non-native words at each level of lexical frequency. The mean difference between native and non-native productions for low frequency words was 0.679, for medium frequency words 0.722, and for high frequency 0.761. These means indicated that the difference between natives and non-natives was enhanced for high frequency words (i.e., a greater difference) relative to lower frequency words. Paired samples *t* tests revealed that the difference between these two groups of speakers was indeed increased in high frequency words when compared to low frequency words ($p = 0.003$) but not when compared to medium frequency words. The difference between native and non-native speakers was smaller for low frequency words than medium frequency words ($p = 0.021$). The interaction between lexical frequency and native language of the speaker, then, is primarily due to the increase in differences between native and non-native speakers for words of high frequency and a decrease in differences for low frequency words.

Figure 2 shows the interaction between lexical frequency and listening context [$F(2,116) = 13.81, p < 0.001$]. Visual inspection of Fig. 2 reveals a decrease of the overall differences between the three levels of lexical frequency in the auditory+orthography context. A one-way ANOVA of difference scores (high frequency-low frequency) revealed a significant difference between the two listening contexts [$F(1,59) = 25.95, p < 0.001$], where differences in perceptual ratings between high and low frequency words were smaller in the auditory+orthography context (mean difference = 0.094) than in the auditory-only context (mean difference = 0.235). Thus, the interaction between lexical frequency and listening context reflects an attenuation of frequency effects in the auditory+orthography context.

The cross-over interaction between native language of the speaker and listening context [$F(91,58) = 8.76, p = 0.004$] is presented in Fig. 3. *Posthoc t* tests of this interaction revealed significant differences between the two listening contexts for both speaker groups. Native speakers were rated as less accented in the auditory+orthography context than in the auditory-only context ($p = 0.004$), whereas non-native speakers were rated as *more* accented in the auditory+orthography context ($p = 0.004$).

C. Discussion

The results of this perceptual ratings study demonstrate that two speaker-independent factors—lexical frequency and listening context—affect the perception of foreign accent in spoken words. First, high frequency words were rated as less accented than low frequency words. This result is consistent with predictions based on exemplar models of speech perception. The more frequently a word occurs in the language, the more often a listener will hear it being spoken, which will in turn lead to encoding more exemplars of the word in memory. Thus, high frequency words are more likely to match (or approximate) an exemplar of a high frequency word in memory and therefore sound comparatively less accented to a native listener of English. Low frequency words, on the other hand, will be experienced less, have many fewer exemplars in memory, and therefore sound more accented.

The effect of lexical frequency also interacted with the native language of the speaker. One analysis of this interaction revealed that differences between native and non-native speakers are increased for high frequency words. In other words, perceptual differences between the two groups of speakers were increased for words of higher frequency. This has implications for future studies of foreign accent perception. If the goal is to enhance differences between native and non-native speakers, words of higher frequency should be used. If, on the other hand, differences between speakers are to be minimized, words of lower frequency should be used. Acoustic measures in the following section indicate that these differences are not due to production differences, at least as measured by vowel formant frequencies. As will be shown, production differences between native and non-native speakers do not vary based on lexical frequency. Thus, the differences found in the ratings are due to listener effects.

The other analysis of the native speaker by lexical frequency interaction revealed that lexical frequency had a stepwise effect on accent ratings for natively produced tokens: high frequency words were rated as less accented than medium frequency words, which were in turn rated as less accented than low frequency words. The effect of frequency was slightly attenuated, however, for the non-native speech where medium and high frequency words did not significantly differ from one another.

The attenuation of the lexical frequency effect for the non-native tokens may have been caused by the relationship between incoming acoustic stimuli and their stored exemplars. If the degree of perceived foreign accent is dependent upon the number of exemplars in memory that are acoustically similar to the input signals, then stimulus tokens that are acoustically similar to many exemplars in memory will be rated as less accented than those

that are acoustically similar to only a few exemplars, as was observed for native tokens. Non-native tokens, however, are likely to have fewer acoustically similar stored exemplars than native tokens, especially for naive listeners who have little if any experience with non-native speech. Because the non-native tokens lie in sparsely populated areas of the exemplar space, differences between high and medium frequency words may be eliminated. Indeed, acoustic analyses presented in the next section show that high frequency non-native words are not better approximations of their native targets than are medium or low frequency words, implying that all non-native tokens, regardless of lexical frequency, are equally deviant from their native targets.

The lack of the expected frequency effects for non-natives may, however, be due to the way the three levels of frequency were created. No *a priori* notion of high, medium, or low frequency was assumed. Instead, the 360 lexical items were simply ranked by lexical frequency and then divided into three equal groups. Since the difference between high and medium frequency was arbitrary and continuous between the levels, differences between the two highest levels of frequency may have been too small.

The production of non-native speech may also be a cause of differences in the perception of native and non-native speech. It is possible that productions of words at these three levels of frequency are different for native speakers, but that productions for non-native speakers are not. Previous work has shown that lexical factors can affect speech production. In a study that explicitly manipulated a combination of lexical frequency and neighborhood density, Wright (2003) found that speakers differed in the degree of vowel reduction/centralization as a result of changes in these lexical factors. In particular, he found that vowels in lexically “easy” words (i.e., high frequency words from sparse lexical neighborhoods) exhibited greater centralization than lexically “hard” words (i.e., low frequency words from dense lexical neighborhoods).

Though differences in the perception of foreign accent differed slightly for native and non-native speakers, globally the results were the same; higher frequency words were rated as less accented than low frequency words. An important point emerges from this observation; lexical effects are found in a task that does not, on the surface, require accessing lexical information. Rating the degree of foreign accent could be done by simply accessing knowledge of the native language phonology and does not inherently require interacting with lexical information. The presence of lexical frequency effects demonstrates that listeners do not bypass the lexicon, but instead use lexical information when making a judgment of foreign accent. It is also important to point out that these effects were found for highly proficient non-native speakers and may not be observed for less fluent bilinguals.²

Although the main effect of listening context did not reach significance, it did have an effect on perceived degree of foreign accent through interactions with both lexical frequency and native language of the speaker. The interaction between listening context and lexical frequency demonstrated that presenting a visual display of the target word on the screen attenuated the effect of lexical frequency. The range of ratings in the auditory+orthography context (between the two extreme values of lexical frequency) was greatly reduced when compared to the auditory-only context. The difference between the two listening contexts with respect to frequency is most likely the result of different processing requirements in the two contexts. In the auditory-only context, listeners must perform both a word recognition task and a nativeness rating task after hearing a stimulus. Listeners must evaluate the stimulus, compare it with stored exemplars in memory, determine the identity of the word, and then evaluate its degree of foreign accent. In the auditory+orthography context, the

²We wish to thank an anonymous reviewer for bringing this point to our attention.

processes of auditory word recognition and lexical access can be bypassed because the correct word is displayed visually on the computer screen.

The attenuation of frequency effects on the perception of foreign accent in the auditory+orthography context is consistent with numerous studies showing that effects of lexical frequency that are observed in open-set word recognition tasks disappear in analogous closed-set tasks (Pollack *et al.*, 1959; Sommers *et al.*, 1997; Clopper *et al.*, 2006b). Since the auditory+orthography listening context eliminates the need for auditory word recognition, the perceived accentedness of a target word in this listening context could be based solely on acoustic-phonetic or phonological differences between the incoming stimulus and existing exemplars, and not on knowledge of the lexical properties of the items. However, some lexical effects, though attenuated, still persist in this context, suggesting that properties of the lexicon influence speech perception even in tasks that do not require direct contact with the lexicon.

Presentation context also influenced the degree of perceived accentedness through an interaction with the native language of the speaker. Native speakers were rated as less accented in the auditory+orthography context than in the auditory-only context. The pattern of results was reversed, however, for non-native speakers who were rated as more accented in the auditory+orthography context than in the auditory-only context. This crossover interaction may reflect differences in the relevant task demands placed on the listener. As discussed above, the auditory+orthography context allows listeners to bypass word recognition because the orthographic presentation serves to limit the possible response alternatives. The auditory+orthography context requires listeners to only judge the accentedness of a stimulus based on acoustic-phonetic similarity with existing exemplars of a particular word type, effectively ignoring phonetically similar words.

In their classic study of speech intelligibility, Miller *et al.* (1951) showed that fewer response alternatives in a word-recognition task leads to higher levels of speech intelligibility at the same signal-to-noise ratio. These findings illustrate that more noise may be added to stimuli when there are fewer response alternatives while maintaining the same level of intelligibility. Miller *et al.* argued that speech intelligibility is not determined by the stimulus item alone, but also by the context in which it is perceived. Likewise, in the present study, the intelligibility of a particular stimulus is increased in the auditory+orthography context because there is essentially only a single response alternative. Though intelligibility and accentedness are not equivalent dimensions, they are correlated (Derwing and Munro, 1997). The availability of context may account for why native speakers are judged as less accented in the auditory+orthography context than in the auditory-only context. In other words, the reduction of response alternatives increases the intelligibility of the individual stimuli and thus the decrease in perceived foreign accent.

This explanation does not, however, account for the ratings of non-native speakers in the two listening contexts. Non-native speakers were instead rated as *more* accented in the auditory+orthography context than in the auditory-only context. Since the process of word recognition can be bypassed in the auditory+orthography listening context, accent ratings will be based solely on the acoustic-phonetic or phonological mismatch between a stimulus item and stored exemplars. Presenting the target word to listeners orthographically in this context may highlight how poorly a non-native production of that word matches its stored exemplars. Hence, non-native productions of words may sound more accented when listeners are informed of the word's identity. In some cases, the auditory percept may even conflict with the orthographic target (e.g., [nuz] with "noose") and therefore result in a significantly higher rating of perceived foreign accent than if the auditory stimulus were presented without its orthographic representation. Data from the number of times listeners

chose to repeat stimuli provide converging evidence that context modulates a listener's judgment. Participants in the auditory-only context listened to stimuli more often than in the auditory+orthography context.

The observed interaction of listening context and native language of the speaker in this study has an important implication for future nativeness rating studies. Presenting words to listeners in an auditory+orthography context results in non-native speakers sounding *more accented* and native speakers *less accented* than in the auditory-only context. The auditory+orthography listening context thus makes the accent ratings for the two groups of speakers diverge in the appropriate directions: Native speakers are rated as less accented and non-native speakers as more accented.

While the listening context effects must be rooted in perceptual processes unrelated to speakers' productions, an account of the results surrounding lexical frequency as rooted entirely in perceptual processes is less plausible. We have argued in this section that differences in accent ratings for high, medium, and low frequency words reflect differences in the perceptions of these words based on the number of stored exemplars and their similarity to an input utterance. As was alluded to above, however, these results may be confounded by concomitant differences in the productions of high, medium, and low frequency words. This hypothesis is further investigated in the next section where acoustic measures of vowels at the three levels of lexical frequency are compared.

III. ACOUSTIC ANALYSIS

A. Methodology

To test the hypothesis that differences found in the perception of high, medium, and low frequency words reflect differences in perception and not production, we examined one measure of speech production, namely, vowel formant frequencies. Vowel formants were selected as a possible locus of production differences between words with different lexical characteristics because other lexical factors have been shown to affect the production of vowels [e.g., Wright (2003) for "easy" versus "hard" words]. Similarly, languages with the "same" vowels have been shown to have different acoustic targets [e.g., Bradlow (1995) for English, Spanish, and Greek vowels]. Furthermore, the exact location of vowel targets in American English is a salient sociolinguistic marker of dialect affiliation (e.g., Labov *et al.* 2006; Clopper *et al.*, 2006a). Thus, listeners in the perceptual rating task are likely to have had some experience discriminating differences in vowel formant frequencies. If production differences in vowel formants exist, they could be used by listeners when performing the accent rating task.

Each target word was segmented into three parts: C1 (onset consonant), V (vocalic nucleus), and C2 (coda consonant). First and second formant measures of the vowel portion were made using Praat sound editing software (Boersma and Weenick, 2004). The formant values reported here come from measures taken at the temporal midpoint of the vowel. The window length was set to 50 ms with a step size of 10 ms. Aberrant formant measures were checked by the first author using a wideband spectrogram and corrected as needed. A total of 91 out of 4080 (2.2%) tokens were hand-corrected in this manner.

Only 170 of the original 360 English words were used in the formant analysis. All words with sonorant coda consonants were eliminated from analysis due to difficulty with determining the boundary between the vowel and the final consonant. All words containing the diphthongs [aj], [oj], and [aw] were also eliminated. Only vowel categories that contained at least three examples at each of the three frequency levels were included in the analysis,

eliminating [a], [o], [ɔ], and [ʊ] from analysis. The remaining vowels [æ], [ʌ], [ɛ], [e], [ɪ], [i], and [u] were used in the analysis.

All formant measures were transformed to z -scores for each person for each formant separately using eq. (1) in (Adank *et al.* 2004; Clopper *et al.* 2006a). This procedure centers each speaker's vowel space at the origin and reduces gender differences while maintaining phonemic differences. In (1), f refers to the formant frequency in hertz (F1 or F2), μ to the overall mean frequency of that formant for a specific speaker, and σ to the standard deviation of that formant for a specific speaker. A plot of the transformed z -scores for native and non-native speakers at each of the three levels of lexical frequency is given in Fig. 4.

$$z = (f - \mu) / \sigma. \quad (1)$$

B. Results and Discussion

Two ANOVAs with vowel (æ, ʌ, ɛ, e, I, i, u) and lexical frequency (high, medium, low) as within-subjects factors and gender (male, female) and native language (native, non-native speakers) as between-subjects factors were calculated for the first formant and second formant z -scores separately. For both F1 and F2, a main effect of vowel was found [$F(6,120) = 914.8, p < 0.001$ for F1 and $F(6,120) = 298.7, p < 0.001$ for F2]. These results merely indicate that vowels have different F1 and F2 targets and will not be further analyzed. A main effect of lexical frequency was found only for F2 [$F(2,40) = 14.2, p < 0.001$]. *Posthoc* paired samples t tests revealed that words with high lexical frequency were more fronted (higher F2) than words of medium and low lexical frequency [$t(167) = 4.58, p < 0.001$ and $t(167) = 4.37, p < 0.001$, respectively]. Both F1 and F2 analyses also revealed significant vowel by lexical frequency interactions [$F(12,240) = 8.02, p < 0.001$ for F1, $F(12,240) = 7.51, p < 0.001$ for F2] and vowel by native language interactions [$F(6,120) = 6.56, p < 0.001$ for F1, $F(6,120) = 3.54, p = 0.003$ for F2]. In addition, the gender by vowel interaction was significant for F1 [$F(6,120) = 4.32, p = 0.001$]. No other main effects or interactions were significant.

To further investigate these interactions, *posthoc* analyses were conducted for F1 and F2 separately. The α level was set to 0.01 due to the large number of *posthoc* analyses. Paired-samples t tests on the vowel by lexical frequency interaction revealed several significant differences between vowels in words of different levels of lexical frequency, as shown in Table II. Five of the seven vowels show some difference between high and low frequency words. This interaction, as well as the overall main effect of lexical frequency found for F2, indicates that phonetic differences between words of different levels of lexical frequency exist. Nonetheless, later analyses will reveal that phonetic differences in formant measures cannot explain all the details of the perceptual interactions involving lexical frequency.

The vowel by native language interaction indicates, not surprisingly, that natives and non-natives produce vowels differently. *Posthoc* one-way ANOVAs of this interaction revealed that non-native productions of [ʌ] were more fronted (higher F2) than native productions [$F(1,23) = 7.987, p = 0.010$]. Additionally, productions of [ɛ] and [u] exhibited differences in F2 that approached significance ($p = 0.033$ and $p = 0.041$). As is apparent in Fig. 4, native productions of [u] exhibited more fronting while non-native productions of [ɛ] exhibited fronting. None of the *posthoc* analyses of this interaction for F1 reached significance, although non-native productions of [æ] were marginally higher than native productions ($p = 0.016$), but lower for [ʌ] and [e] ($p = 0.013$ and $p = 0.048$). The observed differences in formant frequencies between native and non-native productions showed that the non-native speakers, while globally producing target American English vowels correctly, were not perfect in their productions of these vowels. Differences found for [ʌ] and [æ] may be a

result of the absence of these vowels in the native system for the L1-German bilingual speakers. Finally, *posthoc* analyses of the gender by vowel interaction revealed that male productions of [i] were higher (lower F1) than female productions [$F(1,23) = 9.538$, $p = 0.005$].

To further investigate the role that lexical frequency plays in vowel productions of native and non-native speakers, we computed the Euclidean distance in the F1×F2 plane between different levels of lexical frequency. For each speaker, the distance between high and medium lexical frequency, between medium and low lexical frequency, and between high and low lexical frequency for each vowel category was computed. This measure was designed to explore the interaction between native language and lexical frequency that was found in the perceptual ratings data. In the accent rating task, a perceptual difference between all three levels of lexical frequency was found for the native speakers: For non-native speakers the difference between high and medium frequency words was attenuated. To rule out production differences as the source of these perceptual differences, we examined the distance between vowels at different levels of lexical frequency. In particular, we were interested in whether native speakers produced a greater difference between vowels of high and medium frequency words than did non-native speakers. If such a production difference were to exist for native speakers but not for non-native speakers, it could explain the lack of a perceptual difference between high and medium frequency words that was found for the non-native speakers.

ANOVAs were conducted on the distance between high and medium frequency productions, between medium and low frequency productions, and between high and low frequency productions with vowel as the within-subjects factor and native language as the between-subjects factor. Critically, none of these ANOVAs revealed significant differences between the two groups of speakers or in the vowel by native language interaction (all $p > 0.19$). The lack of a main effect of native language, especially in the comparison of high to medium frequency words, suggests that native speakers do not produce larger differences between high and medium frequency words than do non-native speakers. If such a production difference had been found, it could have been the root of the perceptual differences. These acoustic data can therefore partially rule out production (here measured as differences in vowel formants) as the cause of the perceptual difference between the two groups of speakers since native and non-native speakers did not differ in the amount of difference between words of different frequencies.

One additional analysis was conducted on the acoustic data. For each vowel category, Euclidean distances were computed between each non-native speaker's average high frequency production and the overall average native high frequency production. Corresponding distances were calculated for medium and low frequency words. This measure was designed to examine two effects found in the perceptual ratings data: (1) the general trend for high frequency non-native tokens to be rated as less accented than lower frequency tokens and (2) the native language by lexical frequency interaction, which revealed that perceptual differences between native and non-native speakers were increased in words of higher frequency when compared to words of lower frequency. If the former effect were rooted in production differences, then non-native speakers should produce high frequency words *better* (i.e., closer to the native target) than words of lower frequency and therefore be rated as less accented. In other words, the Euclidean distances between native and non-native productions should be smaller for high frequency words and greater for low frequency words. The second perceptual effect mentioned above showed native and non-native speakers diverging in terms of perception for high frequency words and converging for low frequency words. If this perceptual effect arose from production differences, then non-native speakers should produce high frequency words *worse* (i.e., further from the

native target) than low frequency words. In other words, the Euclidean distances between native and non-native productions should be greater for high frequency words and smaller for low frequency words.

Paired-samples *t* tests comparing high to medium, high to low, and medium to low productions revealed no significant differences, indicating that high frequency non-native productions were neither closer to nor further from native targets than medium or low frequency words. Thus, the acoustic analysis of non-native vowel productions from words in the different lexical frequency groups allows us to at least partially rule out the possibility that non-native tokens of high frequency words are acoustically better (or worse) productions. Despite broad differences found in the productions of vowels in words of different levels of lexical frequency, vowel formant production differences do not account for the specific interactions found in the perceptual data. Of course, measurements of other phonetic characteristics (e.g., VOT, degree of obstruent-final devoicing, or vowel formant trajectories) could reveal systematic production differences between levels of lexical frequency. Nonetheless, the lack of consistent effects attributable to an acoustic measure known to be affected by frequency lends support to the claim that the frequency results found in the perception experiment are due primarily to perceptual factors.

IV. GENERAL DISCUSSION AND CONCLUSIONS

The results of the present study demonstrate that two speaker-independent factors—lexical frequency and listening context—affect the perception of degree of foreign accent in isolated spoken words. Listeners consistently perceived high frequency words as less accented than low frequency words. Simultaneously presenting a target word to listeners both auditorily and orthographically attenuated the effect of frequency, however. Furthermore, the addition of orthographic information in the auditory+orthography context caused native speakers of English to be rated as less accented and non-native speakers of English to be rated as more accented than in the auditory-only context. Evidence from acoustic analyses of vowel formant frequencies for native and non-native productions demonstrated that production differences do exist between native and non-native tokens. These analyses further showed that production differences between words at different levels of lexical frequency are found for some, but not all, vowels. However, production differences do not seem to account for some of the perceptual interactions, especially the native language by frequency interaction, thus confirming the role of perception in the ratings of words of different levels of lexical frequency.

The explanation of the effects of lexical frequency and listening context was framed in an exemplar model of speech perception. In exemplar models, words with higher lexical frequency have more stored exemplars in memory. We hypothesized that words of higher frequency would be rated as less accented because the greater number of stored exemplars makes it more likely that an incoming stimulus will be similar to existing targets, resulting in their being rated as less accented. This prediction was supported by our data.

A second effect that can be accounted for within this framework was the attenuation of lexical frequency effects in the auditory+orthography condition. In this listening context, the process of auditory word recognition can be bypassed. A pop-out effect was induced for native tokens presented in the auditory+orthography context. Tokens become more intelligible and are rated as less accented than in the auditory-only context. This effect can be accounted for in an exemplar framework by assuming that in the auditory-only context, some words of the wrong lexical category (e.g., competing “bet” for target “bat”) may cause a decrease in the goodness-of-fit between an incoming token and the target lexical item, increasing the amount of perceived accent. In the auditory+orthography context, on the other

hand, all competing targets and their exemplars are eliminated, causing the incoming token to be rated as less accented. This effect of context on perceived accentedness was reversed for non-natives. Non-native tokens may sometimes be so deviant as to completely switch lexical categories, something less likely to occur among native tokens. In these cases, tokens will be perceived as extremely aberrant and receive very high ratings of accentedness.

Our results may be consistent with other models of speech perception and spoken word recognition. The TRACE model (McClelland and Elman, 1986; McClelland *et al.*, 2006) allows for both bottom-up and top-down processing of speech. Because top-down information can affect processing of ambiguous or degraded sounds, TRACE correctly predicts a decrease in perceived accent in the auditory+orthography listening context. This model alone may not be able to account for the variation of these effects, however. Norris *et al.* (2000) point out that top-down processing in TRACE acts to correct erroneous information in the input and therefore prevents a listener from reliably detecting and perceiving mispronunciations. One might expect that mispronunciations in non-native input would be similarly overlooked, resulting in non-native speakers also being rated as less accented in the auditory+orthography context. However, results from the perceptual ratings indicate that this is not the case. A full examination of our data within the TRACE model is beyond the scope of the current article, though some aspects appear promising and merit further examination.

The findings of the current study have several implications for future research on accent perception. First, the present results demonstrate that researchers should consider the role that lexical frequency plays in perception of degree of foreign accent. If the effects of frequency are to be avoided or minimized, an orthographic representation of the target word can be used to attenuate these effects. Second, presenting target words to listeners both auditorily and orthographically yields different measures of perceived foreign accent; in the auditory+orthography context, native speakers were rated as less accented while non-native speakers were rated as more accented. The auditory+orthography context thus mitigates the effects of lexical frequency on accent ratings and also helps listeners better distinguish speech samples of native and non-native speakers.

The results of this study also have several broader theoretical implications. Our findings show that perceived degree of foreign accent is not just related to a speaker's language learning history or to how well a speaker is able to phonetically approximate native speech, but also depends on the context in which that voice is perceived. We have shown here that perception of foreign accent is partially dependent upon non-acoustic properties of the signal such as listening context and possibly lexical frequency. Previous work has shown that the intelligibility of L2 speakers reflects not only the actual acoustic realization of non-native productions but also the prior experience and history of the listener. For example, Bent and Bradlow (2003) found that the intelligibility of several groups of non-native speakers depended on the language background of the listeners. Non-native listeners in both a matched and mismatched native language background performed equally well in a sentence intelligibility task with proficient non-native and native speakers. Native listeners, on the other hand, found all non-native speakers to be less intelligible than native speakers. Accent perception is therefore shaped to a large extent by a *listener's* past experiences and developmental history. A speaker may therefore only have an "accent" within a specific perceptual framework and listening context. The perception of a foreign accent thus reflects not only properties of the speaker, but also prior experience of the listener and factors that reflect the attunement between speaker and listener.

The influences of two speaker-independent factors—lexical frequency and listening context—on the perception of foreign accent in this study can be accounted for in large part by

casting the process of accent perception more broadly within the framework of exemplar models of speech perception and spoken word recognition. We have assumed that the perception of foreign accent reflects the degree to which there is an acoustic-phonetic mismatch between a stimulus token and the stored exemplars in the listener's memory. The validity and robustness of this theoretical framework can be tested in future research. One possible way to test this framework is to manipulate the amount of experience that listeners have in listening to L2 speech. It has been noted above that highly experienced listeners (e.g., linguists and ESL teachers) sometimes rate the accents of non-native speakers more leniently (i.e., as less accented) than do naive listeners. This result may occur because experienced listeners have more exposure to L2 speech and therefore more L2 speech exemplars in memory than naive listeners. Thus, experienced listeners are more likely to find an acoustic match to incoming tokens of L2 speech and rate these tokens as less accented than naive listeners. Increasing the amount of experience that naive listeners have with L2 speech by presenting them with many tokens of L2 speech should therefore make them more tolerant perceivers of foreign accent in speech produced by unfamiliar L2 speakers. Experimental studies such as these should help increase our understanding of the process by which foreign accents are perceived and also provide us with a more complete picture of what it means for a speaker to "have a foreign accent."

Acknowledgments

Preparation of this manuscript was supported by grants from the National Institutes of Health to Indiana University (NIH-NIDCD T32 Training Grant No. DC-00012 and NIH-NIDCD Research Grant No. R01 DC-00111). We wish to thank Jennifer Karpicke and Christina Fonte for help with data collection and Adam Buchwald and three anonymous reviewers for useful comments on this and previous versions of this manuscript.

References

- Adank P, Smits R, van Hout R. A comparison of vowel normalization procedures for language variation research. *J Acoust Soc Am*. 2004; 116:3099–3107. [PubMed: 15603155]
- Baayen, RH.; Piepenbrock, R.; Gulikers, L. The CELEX Lexical Database (Release 2) [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania [Distributor]; Philadelphia, PA: 1995.
- Bent T, Bradlow AR. The interlanguage speech intelligibility benefit. *J Acoust Soc Am*. 2003; 114:1600–1610. [PubMed: 14514213]
- Boersma, P.; Weenink, D. [accessed 3/19/07] Praat: Doing phonetics by computer (Version 4.2.12). 2004. [Computer program], <http://www.praat.org/>
- Bongaerts T, van Summeren C, Planken B, Schils E. Age and ultimate attainment in the pronunciation of a foreign language. *Stud Second Lang Acquis*. 1997; 19:447–465.
- Bradlow AR. A comparative acoustic study of English and Spanish vowels. *J Acoust Soc Am*. 1995; 97:1916–1924. [PubMed: 7699173]
- Clopper CG, Pisoni DB, de Jong K. Acoustic characteristics of the vowel system of six regional varieties of American English. *J Acoust Soc Am*. 2006a; 118:1661–1676. [PubMed: 16240825]
- Clopper CG, Pisoni DB, Tierney AT. Effects of open-set and closed-set task demands on spoken word recognition. *J Am Acad Audiol*. 2006b; 17:331–349. [PubMed: 16796300]
- Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C. Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence from the comprehension of noise-vocoded sentences. *J Exp Psychol Gen*. 2005; 134:222–241. [PubMed: 15869347]
- Derwing TM, Munro MJ. Accent, intelligibility, and comprehensibility: Evidence for four L1s. *Stud Second Lang Acquis*. 1997; 20:1–16.
- Fathman A. The relationship between age and second language productive ability. *Lang Learn*. 1975; 25:245–253.
- Flege JE, Fletcher KL. Talker and listener effects on degree of perceived foreign accent. *J Acoust Soc Am*. 1992; 91:370–389. [PubMed: 1737886]

- Flege JE, Munro MJ, MacKay IRA. Factors affecting strength of perceived foreign accent in a second language. *J Acoust Soc Am*. 1995; 97:3125–3134. [PubMed: 7759652]
- Goldinger, S. Words and voices: Perception and production in an episodic lexicon. In: Johnson, K.; Mullennix, J., editors. *Talker Variability in Speech Processing*. Academic; San Diego: 1997. p. 33-66.
- Grosjean F. Spoken word recognition processes and the gating paradigm. *Percept Psychophys*. 1980; 28:267–283. [PubMed: 7465310]
- Hintzman D. Schema abstraction in a multiple-trace memory model. *Psychol Rev*. 1986; 93:411–428.
- Hintzman D. Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychol Rev*. 1988; 95:528–551.
- Howes D. On the relation between the intelligibility and frequency of occurrence of English words. *J Acoust Soc Am*. 1957; 29:296–305.
- Johnson, K. Speech perception without speaker normalization: An exemplar model. In: Johnson, K.; Mullennix, J., editors. *Talker Variability in Speech Processing*. Academic; San Diego: 1997. p. 145-165.
- Kirchner, R. UCLA Work Papers Ling. Vol. 1. 1999. Preliminary thoughts on “phonologisation” within an exemplar-based speech processing system. *Papers Phonol*. 2:207–231.
- Kirchner, R. Exemplar-based phonology and the time problem: A new representational technique. poster presented at LabPhon 9 Conference; 28 June; 2004.
- Labov, W.; Ash, S.; Boberg, C. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Mouton de Gruyter; Berlin: 2006.
- www.linguistics.uiuc.edu/labphon9/Abstract_PDF/kirchner.pdf Viewed 3/27/07
- Luce PA, Pisoni DB. Recognizing Spoken Words: The neighborhood activation model. *Ear Hear*. 1998; 19:1–36. [PubMed: 9504270]
- MacLennan RN. Interrater reliability with SPSS for Windows 5.0. *Am Stat*. 1993; 47:292–296.
- McClelland JL, Elman JL. The TRACE model of speech perception. *Cogn Psychol*. 1986; 18:1–86. [PubMed: 3753912]
- McClelland JL, Mirman D, Holt LL. Are there interactive processes in speech perception? *Trends Cogn Sci*. 2006; 10:363–369. [PubMed: 16843037]
- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996; 1:30–46.
- Miller GA, Heise GA, Lichten W. The intelligibility of speech as a function of the context of the test materials. *J Exp Psychol*. 1951; 41:329–335. [PubMed: 14861384]
- Munro MJ, Derwing TM. Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Stud Second Lang Acquis*. 2001; 23:451–468.
- Norris D, McQueen JM, Cutler A. Merging information in speech recognition: Feedback is never necessary. *Behav Brain Sci*. 2000; 23:299–370. [PubMed: 11301575]
- Oyama S. A sensitive period for the acquisition of a nonnative phonological system. *J Psycholinguist Res*. 1976; 5:261–283.
- Pierrehumbert, J. Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J.; Hopper, P., editors. *Frequency Effects and the Emergence of Linguistic Structure*. John Benjamins; Amsterdam: 2001. p. 137-157.
- Pierrehumbert, J. Word-specific phonetics. In: Gussenhoven, C.; Warner, N., editors. *Laboratory Phonology VII*. Mouton de Gruyter; Berlin: 2002. p. 101-140.
- Piper T, Cansin D. Factors influencing the foreign accent. *Can Mod Lang Rev*. 1988; 44:334–342.
- Piske T, MacKay IRA, Flege JE. Factors affecting degree of foreign accent in an L2: A review. *J Phonetics*. 2001; 29:191–215.
- Pollack I, Rubenstein H, Decker L. Intelligibility of known and unknown message sets. *J Acoust Soc Am*. 1959; 31:273–279.
- Savin HB. Word-frequency effect and errors in the perception of speech. *J Acoust Soc Am*. 1963; 35:200–206.

- Sommers MS, Kirk KI, Pisoni DB. Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. *Ear Hear.* 1997; 18:89–99. [PubMed: 9099558]
- Southwood MH, Flege JE. Scaling foreign accent: direct magnitude estimation versus interval scaling. *Clin Linguist Phonetics.* 1999; 13:335–349.
- Thompson I. Foreign accents revisited: The English pronunciation of Russian immigrants. *Lang Learn.* 1991; 41:177–204.
- Wright, R. Factors of lexical competition in vowel articulation. In: Local, J.; Ogden, R.; Temple, R., editors. *Papers in Laboratory Phonology, VI: Phonetic Interpretation.* Cambridge U. P, Cambridge, UK: 2003. p. 75-87.

FIG. 1.

Mean z -score ratings with standard errors for native and non-native speakers for each of the three levels of lexical frequency. Positive values indicate more foreign accent while negative values indicate less foreign accent.

FIG. 2.
Mean z -score ratings with standard errors for auditory +orthography and auditory-only contexts for each of the three levels of lexical frequency.

FIG. 3.

Mean z -score ratings with standard errors for native and non-native speakers for each of the two listening contexts.

FIG. 4.

Mean z-transformed formant values for native and non-native speakers at each level of lexical frequency.

TABLE I

Demographic variables for the bilingual speakers. “Years of English” refers to the number of years speakers have been learning/using English (current age-age of acquisition). “Fluency” is a self-reported measure of English proficiency (1=poor, 5=fluent). The final column provides each speaker’s mean z-score accent rating. Larger z-scores reflect a higher degree of foreign accent.

Speaker	Age of acquisition	Years of English	Length of residence	Fluency	Accent rating
f2	12	9	1	4	0.00
f3	10	14	1	5	0.22
f4	13	13	3	4.5	0.02
f7	9	12	1	5	0.33
f8	13	16	4	5	-0.27
f9	9	16	2	4	0.13
f11	...	...	2	4.5	0.94
m2	12	18	3	5	0.02
m3	10	13	1	4	0.31
m4	11	18	1	4	0.39
m6	13	13	1	5	0.69
m9	12	20	2	4	0.95
m10	10	14	1	5	0.57
Mean	11.2	14.7	1.77	4.54	0.33
SD	1.53	3.06	1.01	0.48	0.37

TABLE II

The *p* values of *posthoc* analyses of the vowel by lexical frequency interaction, collapsed across native and non-native speakers. Comparisons that are significant for either F1 or F2 are bold.

Vowel	Lexical frequency comparisons	F1	F2
æ	high vs. low	0.006	0.005
	medium vs. low		
	high vs. medium		
ʌ	high vs. low		
	medium vs. low	0.003	0.005
	high vs. medium	0.009	0.004
ɛ	high vs. low	0.003	0.000
	medium vs. low		
	high vs. medium	0.000	0.000
e	high vs. low		
	medium vs. low		
	high vs. medium		
ɪ	high vs. low	0.001	
	medium vs. low	0.000	0.000
	high vs. medium		
i	high vs. low	0.003	0.000
	medium vs. low		0.000
	high vs. medium		
u	high vs. low		0.002
	medium vs. low	0.003	
	high vs. medium	0.008	