

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 27 (2005)
Indiana University

**Speaker-independent Factors Affecting the Perception of
Foreign Accent in a Second Language¹**

Susannah V. Levi, Stephen J. Winters and David B. Pisoni

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

¹ Preparation of this manuscript was supported by grants from the National Institutes of Health to Indiana University (NIH-NIDCD T32 Training Grant DC-00012 and NIH-NIDCD Research Grant R01 DC-00111). We wish to thank Jennifer Karpicke and Chirstina Fonte for help with data collection.

Speaker-independent Factors Affecting the Perception of Foreign Accent in a Second Language

Abstract. Previous research on the perception of foreign accent has largely focused on speaker-dependent factors such as Age of Learning and Length of Residence which are specific to an individual speaker. Factors that are independent of particular speakers and their language learning history have also been shown to affect perception of second language speech. The present study examined two speaker-independent factors—lexical frequency and listening context—that affect the perception of foreign-accented speech. Using a seven-point scale, two groups of listeners rated speakers on how much of a foreign accent they displayed. Listeners in the Auditory-Only listening context heard only the target stimuli, while listeners in the Auditory + Orthography listening context were presented with both the auditory signal and an orthographic display of the target word. The results revealed that lexical frequency affects the perception of the degree of foreign accent; higher frequency words were consistently rated as sounding 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 nonnative speakers. The results suggest that structural and methodological factors independent of the speakers' actual speech articulations or developmental history affect the perception of degree of foreign accent and that such factors should be considered when interpreting the results of studies on the perception of foreign accented speech.

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 degree of foreign accent of a speaker, however, is not based exclusively on the amount of acoustic and articulatory mismatches between nonnative 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 individual. 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, MacKay, & Flege, 2001 for a review).

Additional factors which 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 judgments of perceived degree of foreign accent. They point out that scales with fewer intervals may produce ceiling effects and therefore are not 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. Native controls serve to confirm that listeners are correctly performing the task by testing that they can distinguish native from nonnative speech. Using native controls also ensures that listeners use a wider range of the rating scale. Characteristics of the listener can affect the perceived degree of foreign accent, as well. Several studies have varied whether naïve listeners (e.g., Flege & Fletcher, 1992; Flege, Munro, & MacKay, 1995) or experienced listeners such as linguists (e.g., Fathman, 1975) or ESL teachers (e.g., Piper & Cansin, 1988) serve as raters. Thompson (1991) found that naïve listeners tended to perceive a greater degree of foreign accent than experienced listeners, although Bongaerts, van Summeren, Planken, & Schils (1997) did not find a significant difference. Taken together, these studies show that speaker-independent 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.² 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 & 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 the low frequency words. In a similar study, Savin (1963) examined listeners' response errors. Incorrect responses tended to be words of higher frequency than the target word. In a lexical decision task, Luce & Pisoni (1998) asked listeners to determine whether a target stimulus was a word or a nonword. They 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 and asked them to repeat the words as quickly as possible. The target words were spoken by several different talkers. Goldinger predicted that the 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

² We consider lexical frequency to be a speaker-independent factor because we consider it to be a property of a linguistic community. Though no two speakers have exactly the same frequency for all of their lexical items, we contend that globally, listeners with similar levels of education (in this case, students at Indiana University) will have similar lexical frequencies. In this experiment, we are examining the effects of frequency on the listener's perception of foreign accented speech and are therefore testing a homogeneous population.

than high frequency words, suggesting that subjects were more sensitive to the surface acoustic-phonetic details in the low-frequency words than the 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; Pierrehumbert, 2001, 2002; Kirchner, 1999, 2004). 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 which 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. 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 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.

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"). Knowledge of the intended target in the Auditory + Orthography context should facilitate the perception of degraded speech stimuli (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005). 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 noise vocoded speech, a type of speech stimulus that simulates the signal heard by cochlear implant users.³ 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 learning of severely degraded, noise vocoded speech regardless of the modality in which the original undegraded sentences are presented. As Davis et al. concluded, "pop-out must be at a non-acoustic, phonological level or higher" (pg. 230).

Presenting a word to a listener in orthographic form while he/she hears a nonnative production of the word may induce similar "pop-out" effects, since foreign accented speech can be regarded as a form of degraded speech. The effects of this type of pop-out on the perception of degree of foreign accent are

³ This type of speech is created by filtering the original signal into six logarithmically spaced frequency bands.

unclear, however. One possibility is that simultaneously presenting the auditory and orthographic representations of the target word together will cause nonnative speech samples to be rated as less accented. If a nonnative production of the target word is ambiguous or difficult to understand, presenting the target word with orthography on the screen may promote a type of pop-out effect to occur where the “degraded”, nonnative 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 nonnative 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’ focuses listeners’ attention to the 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 nonnative 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.

Methodology

Materials

Twelve female and ten male German L1/English L2 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 head-mounted 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 and 360 German words was produced by each speaker. Each word was of the form consonant-vowel-consonant (CVC) and was selected from the CELEX English and German databases (Baayen et al. 1995). Speakers read each word 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 which 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. Presentation of the test items was blocked by language, but all within-language items were randomized. Items that were produced incorrectly or too loudly were noted and re-recorded in the same manner following each recording block. The total recording time for each

language block was approximately one hour for each speaker. Speakers were given the option of recording both sets of language items on either the same day or on two separate days, but all speakers elected to record all stimuli in a single recording session.

This process yielded recordings which 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. All edited tokens were then normalized to have a uniform RMS amplitude of 66.4 dB. Only the English words which had been both edited and normalized in this way were presented to the listeners in this study.

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 for 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/hr for their time.

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, Ash, & Boberg, 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.

Words from both languages varied in frequency based on counts from the CELEX database. For the purposes of analysis, the English 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.

Listeners

A total of 87 listeners participated in this experiment; Forty-two were assigned to the Auditory-Only context and 45 were assigned to the Auditory + Orthography context. Twenty-seven listeners were eliminated (polylingual/nonnative 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 listeners for 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). Each listener participated in only one of the two listening contexts. All listeners received partial course credit for their participation.

Procedure

The experiment was implemented on Macintosh G3 computers running a customized SuperCard (version 4.1.1) stack. Listeners sat in front of these computers in a quiet testing room while wearing Beyerdynamic DT-100 headphones. 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 nonnative speakers. All listener ratings were converted to normalized z-scores per listener prior to completing any statistical analyses.

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 eleven monolingual and thirteen 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 listeners had the option of listening to the target words again before making their responses. After rating each token, participants clicked an on-screen button to play the next token. The entire study took approximately one hour for most listeners to complete. Participants in the Auditory + Orthography context 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 only once and to 21.5% of the tokens two or more times. A repeated-measures ANOVA with listening context (Auditory-Only vs. Auditory + Orthography) as a between-subjects factor and with native language of the speaker (L1-English vs. L2 English) as a within-subjects factor revealed main effects of both listening context ($F(1,58)=5.688, p=.020$) and native language ($F(1,58)=5.617, p=.021$), but no interaction. Listeners listened to stimuli more often in the Auditory-Only context than in the Auditory + Orthography context (means: 21.5% vs. 10.5%, respectively). Furthermore, listeners listened more often to the native speakers of English than to the nonnative speakers of English (16.6% vs. 15.3%).

Results

A repeated measures ANOVA with lexical frequency (low, medium, or high) and native language of the speaker (native vs. nonnative) as within-subjects variables and listening context (Auditory-Only or 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 repeated measures ANOVA revealed a significant main effect of lexical frequency ($F(2, 116) = 44.8, p < .001$). Paired-samples t-tests revealed significant pair-wise differences between low vs. medium frequency, between medium vs. high frequency, and between low vs. high frequency (all $p \leq .002$). The direction of this effect indicated that lower frequency words were rated as more accented than higher frequency words. The mean z-scores for the ratings for each frequency group are presented in Table 1. A main effect of native language of the speaker was also found ($F(1, 58) = 1214.8, p < .001$). Native speakers were rated as having less foreign accent overall than nonnative speakers (see Table 2). The main effect for listening context was not significant ($F(1, 58) = 2.29, p = .135$).

Frequency	Mean (SD)
Low	.048 (.093)
Medium	-.049 (.054)
High	-.090 (.071)

Table 1. Overall means and standard deviations of the z-scores of the ratings by three levels of frequency.

Speaker	Mean (SD)
Native English	-.39 (.14)
Nonnative English	.33 (.12)

Table 2. Overall means and standard deviations of the z-scores for native vs. nonnative speakers.

The analysis also revealed significant interactions between lexical frequency and native language of the speaker ($F(2, 116) = 6.51, p = .002$), lexical frequency and listening context ($F(2, 116) = 13.81, p < .001$), and native language of the speaker and listening context ($F(1, 58) = 8.76, p = .004$). The three-way interaction was not significant.

The interaction between lexical frequency and native language of the speaker is shown in Figure 1. Paired-samples t-tests revealed that this interaction was due to a different pattern of ratings for the medium and high frequency words 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 $p \leq .001$). In contrast, for the nonnative speakers, low frequency words were rated as more accented than both medium and high frequency words ($p < .001$), but there was no significant difference between the medium and high frequency words ($p=.213$).

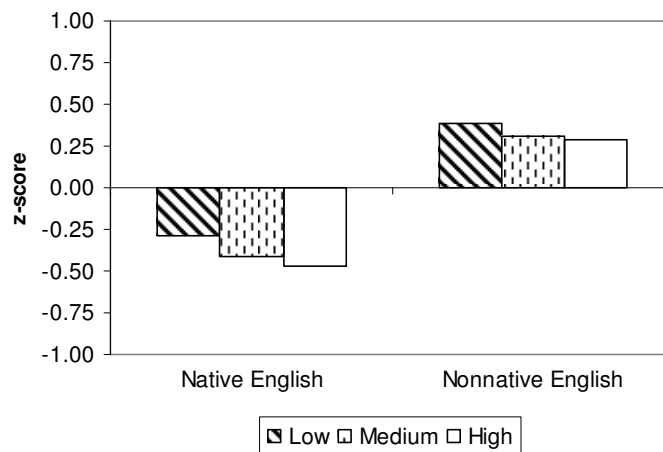


Figure 1. Mean z-score ratings for native and nonnative speakers for each of the three frequency conditions.

Figure 2 shows the interaction between lexical frequency and listening context. Paired samples t-tests revealed that this interaction was also the result of a different pattern of ratings for the medium and high frequency words. In the Auditory-Only context, low frequency words were rated as more accented than medium frequency words, which were in turn rated as more accented than high frequency words (all $p < .001$). In the Auditory + Orthography context, however, only low frequency words were rated as more accented than medium and high frequency words (both $p \leq .005$), while no significant difference was observed between medium and high frequency words ($p = .855$).

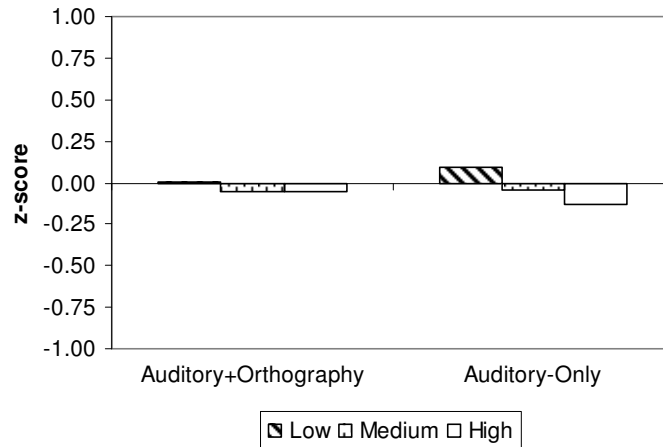


Figure 2. Mean z-score ratings for Auditory + Orthography and Auditory-Only contexts for each of the three levels of frequency.

Post-hoc t-tests on the native language of the speaker by listening context interaction revealed significant differences between the two listening contexts for both speaker groups. The cross-over interaction is illustrated in Figure 3. Native speakers were rated as less accented in the Auditory + Orthography context than in the Auditory-Only context ($p = .004$), whereas nonnative speakers were rated as less accented in the Auditory-Only context than in the Auditory + Orthography context ($p = .004$).

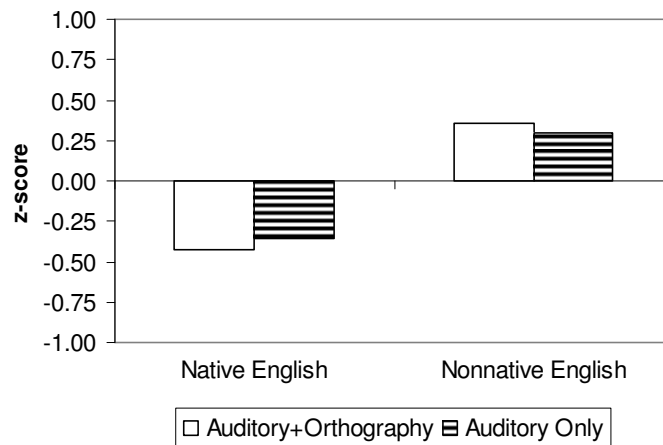


Figure 3. Mean z-score ratings for native and nonnative speakers for each of the two listening contexts.

Discussion

The results of this study demonstrate that two speaker-independent factors, lexical frequency and listening context, affect the perception of foreign accent in spoken words. High frequency words were rated as less accented than low frequency words. This result replicates earlier findings reported by Goldinger (1997) and 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. Highly variable, unusual, nonnative, “accented” productions of a target word will therefore be 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 often and therefore have many fewer exemplars in memory. Thus, a nonnative production of a word must be a closer acoustic match to the few exemplars in memory in order to be rated as a good exemplar of that word type.

The effect of lexical frequency also entered into an interaction with the native language of the speaker. 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 attenuated for the nonnative speech, however. For nonnative tokens, only the low frequency words had significantly higher accent ratings than the medium and high frequency words, which did not significantly differ from one another.

The attenuation of the lexical frequency effect for the nonnative tokens may have been caused by the relationship between incoming acoustic stimuli and their stored exemplars. If 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 which are acoustically similar to only a few exemplars, as was observed for native tokens. Nonnative tokens, however, are likely to have fewer acoustically similar stored exemplars than native tokens, especially for naïve listeners who have little if any experience with nonnative speech. Because the nonnative tokens lie in sparsely populated areas of the exemplar space in memory, the differences between high and medium frequency words may be eliminated. The frequency effect may remain for native speech because native productions of high and medium frequency words are in densely populated portions of the acoustic space where differences in frequency of the exemplars are likely to be apparent. The reason why the frequency effect remains for the low frequency nonnative productions may be the result of a different processing strategy for low frequency words. Very low frequency words may be processed as nonwords for some of the listeners and therefore receive significantly lower ratings. The lack of the expected frequency effects may also 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. The resulting frequency groups were therefore continuous in the level of frequency. Since the difference between high and medium frequency was arbitrary and adjacent, the differences between the two highest levels of frequency may have been too small.

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 the 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. In the Auditory-Only listening context, the perceived degree of foreign accent was significantly different for words of all three levels of frequency. In the Auditory + Orthography context, however, accent

ratings for the high and medium frequency words were not significantly different from one another. 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 and compare it with stored exemplars in memory. In the Auditory + Orthography context, the process of auditory word recognition and lexical access are bypassed because the correct word is displayed visually on the computer screen. The attenuation of frequency effects on the perceived degree of foreign accent in the Auditory + Orthography context is consistent with numerous studies showing that effects of lexical frequency which are observed in open-set word recognition tasks disappear in analogous closed-set word recognition tasks (Pollack, Rubenstein, & Decker, 1959; Sommers, Kirk, & Pisoni, 1997; Clopper, Pisoni, & Tierney, in press). Since the Auditory + Orthography listening context eliminates the process of auditory word recognition from influencing ratings of accentedness, the perceived accentedness of a target word in this listening context must be based solely on its acoustic-phonetic properties, rather than on how familiar or unfamiliar the listener may be with the lexical item itself. In other words, in the Auditory + Orthography context, the nativeness ratings are based exclusively on acoustic-phonetic or phonological differences between the stimulus and existing exemplars and not on knowledge of the lexical properties of the items.

Nonetheless, low frequency words were consistently rated as more accented than both medium and high frequency words in both the Auditory + Orthography context and the Auditory-Only context. Because this effect was observed across both listening contexts, low frequency words may be processed in a fundamentally different way than high and medium frequency words, perhaps because listeners remain unfamiliar with them even after they have been informed of the identity of the word. It is also possible that the higher accent ratings for the low frequency words may reflect differences in the productions of these words. In a study that manipulated lexical frequency and neighborhood density, Wright (2003) found that speakers differed in the degree of vowel reduction/centralization as a result of these two 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). Similarly in our data, low frequency words may exhibit less fluency and may include more hyper-articulated segments, causing them to be consistently perceived as less natural and therefore more accented.

Presentation context also influenced the degree of perceived accentedness by interacting 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 nonnative 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. 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, then, requires listeners to only judge the accentedness of a stimulus based on acoustic-phonetic similarity with existing exemplars of a particular word type. In their classic study of speech intelligibility, Miller, Heise, and Lichten (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. In one experiment, they found that threshold (50 % correct) was reached at -14 dB SNR in a two-word vocabulary task but that a -4 dB SNR was needed for a 256-word vocabulary task.

These findings illustrate that more noise may be added to stimuli when there are fewer response alternatives while maintaining the same amount of intelligibility. Miller et al. argued that speech intelligibility is not determined by the stimulus item alone, but also by its context. 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. The availability of context may account for why the 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.

This explanation does not, however, account for the ratings of the nonnative speakers in the two listening contexts. The nonnative speakers were instead rated as more accented in the Auditory + Orthography context than in the Auditory-Only context. Since the process of word recognition is bypassed in the Auditory + Orthography listening context, accent ratings will be based solely on the acoustic-phonetic or phonological mismatch between a stimulus and stored exemplars. Presenting the target word to listeners orthographically in this context may highlight how poorly a nonnative production of that word matches its stored exemplars. Hence, nonnative 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 the context modulates a listener's judgment. Listeners 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 speaker in this study has an important implication for future nativeness rating studies. Presenting words to listeners in an Auditory + Orthography context makes nonnative speakers sound *more accented* while making native speakers sound *less accented* than in the Auditory-Only context. The Auditory + Orthography listening context makes the accent ratings for the two groups of speakers diverge in the appropriate directions; native speakers are rated as less accented and nonnative speakers as more accented.

Conclusion

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 nonnative speakers of English to be rated as more accented than in the Auditory-Only context.

These findings have several implications for future research on accent perception. First, these results demonstrate that researchers need to consider the role that lexical frequency plays in studies that measure degree of foreign accent. If the effects of frequency are to be avoided, 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 degree of foreign accent; in the Auditory + Orthography context, native speakers were rated as less accented while nonnative 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 from native and nonnative speakers.

The results of this study also have several theoretical implications. Our findings show that an “accent” is not just a feature of a speaker's voice or how well a speaker is able to phonetically

approximate native speech, but also depends on the process by which that voice is perceived. We have shown here that this perceptual process is partially dependent upon non-acoustic properties of the signal such as lexical frequency and listening context. Previous work has shown that the intelligibility of L2 speakers reflects not only the actual acoustic accuracy of nonnative 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 nonnative speakers depended on the language background of the listeners. Nonnative listeners in both a matched and mismatched native language background performed equally well in a sentence intelligibility task with proficient nonnative talkers and with native talkers. Native listeners, on the other hand, found all the nonnative talkers to be less intelligible than native talkers. The process of accent perception is therefore shaped and modified 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 talker, but also prior experience of the listener and factors that affect 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 nonnative speakers more leniently (i.e. as less accented) than naïve listeners do. This result may occur because experienced listeners have more exposure of L2 speech and therefore more L2 speech exemplars in memory than naïve listeners. Therefore, experienced listeners are more likely to find an acoustic match to incoming exemplars of L2 speech and rate these tokens as less accented than naïve listeners do. Increasing the amount of experience that naïve listeners have with L2 speech by presenting them with many tokens of L2 speech should therefore make them more tolerant raters of foreign accent in speech produced by unfamiliar L2 talkers. 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".

References

- Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1995) The CELEX Lexical Database (Release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].
- Bent, T. & Bradlow, A.R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, *114*, 1600-1610.
- Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language, *Studies in Second Language Acquisition*, *19*, 447-465.
- Clopper, C.G., Pisoni, D.B., & Tierney, A.T. (in press). Effects of open-set and closed-set task demands on spoken word recognition. *Journal of the American Academy of Audiology*.
- Davis, M.H., Johnsruide, I.S., Hervais-Adelman, A., Taylor, K. & McGettigan, C. (2005). Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the Comprehension of Noise-Vocoded Sentences. *Journal of Experimental Psychology: General*, *134*, 222-241.

- Fathman, A. (1975). The relationship between age and second language productive ability, *Language Learning*, 25, 245-253.
- Flege, J.E. & Fletcher, K.L. (1992). Talker and listener effects on degree of perceived foreign accent, *Journal of the Acoustical Society of America*, 91, 370-389.
- Flege, J.E., Munro, M.J., & MacKay, I.R.A. (1995) Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125-3134.
- Goldinger, S. (1997) Words and voices: perception and production in an episodic lexicon. In K. Johnson & J. Mullennix (eds.), *Talker variability in speech processing*, pp. 33-66 (San Diego: Academic Press).
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, 28, 267-283.
- Hintzman, D. (1986) Schema abstraction in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- Hintzman, D. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95, 528-551.
- Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America*, 29, 296-305.
- Johnson, K. (1997) Speech perception without speaker normalization: an exemplar model In K. Johnson & J. Mullennix (eds.), *Talker variability in speech processing*, pp. 145-165. San Diego: Academic Press.
- Kirchner, R. (1999). Preliminary thoughts on “phonologisation” within an exemplar-based speech processing system. *UCLA Working Papers in Linguistics*, 1 (*Papers in Phonology* 2), 207-231.
- Kirchner, R. (2004) Exemplar-based phonology and the time problem: a new representational technique. Poster presented at LabPhon 9 Conference, June 28th, 2004.
- Labov, W., Ash, S. & Boberg, C. (2006). *The Atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.
- Luce, P.A. & Pisoni, D.B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear & Hearing*, 19, 1-36.
- Miller, G.A., Heise, G.A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, 41, 329-335.
- Oyama, S. (1976) A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, 5, 261-283.
- Pierrehumbert, J. (2001) Exemplar dynamics: word frequency, lenition, and contrast. In J. Bybee & P. Hopper (eds.), *Frequency effects and the emergence of linguistic structure*, pp. 137-157. Amsterdam: John Benjamins.
- Pierrehumbert, J. (2002) Word-specific phonetics. In C. Gussenhoven & N. Warner, (eds.), *Laboratory Phonology VII*, pp. 101-140. Berlin: Mouton de Gruyter.
- Piper, T. & Cansin, D. (1988). Factors influencing the foreign accent, *The Canadian Modern Language Review*, 44, 334-342.
- Piske, T., MacKay, I.R.A., & Flege, J.E. (2001) Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, 29, 191-215.
- Pollack, I., Rubenstein, H., & Decker, L. (1959) Intelligibility of known and unknown message sets. *Journal of the Acoustical Society of America*, 31, 273-279.
- Savin, H.B. (1963). Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America*, 35, 200-206.
- Sommers, M.S., Kirk, K.I., Pisoni, D.B. (1997) 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 & Hearing*, 18, 89-99.

- Southwood, M.H. and Flege, J.E. (1999) Scaling foreign accent: direct magnitude estimation versus interval scaling. *Clinical Linguistics & Phonetics*, 13. 335-349.
- Thompson, I. (1991) Foreign accents revisited: the English pronunciation of Russian immigrants, *Language Learning*, 41, 177-204.
- Wright, R. (2003). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, & R. Temple (eds.), *Papers in Laboratory Phonology, VI: Phonetic Interpretation*, pp. 75-87. Cambridge, UK: Cambridge University Press.

