Perceived foreign accentedness: Acoustic distances and lexical properties

Vincent Porretta^{1,3} · Aki-Juhani Kyröläinen² · Benjamin V. Tucker¹

Published online: 19 May 2015
© The Psychonomic Society, Inc. 2015

Abstract In this study, we examined speaker-dependent (acoustic) and speaker-independent (lexical) linguistic influences on perceived foreign accentedness. Accentedness ratings assigned to Chinese-accented English words were analyzed, taking accentedness as a continuum. The speaker-dependent variables were included as acoustic distances, measured in relation to typical native-speaker values. The speaker-independent variable measures were related to the properties of individual words, not influenced by the speech signal. To the best of the authors' knowledge, this represents the first attempt to examine speaker-dependent and speaker-independent variables simultaneously. The model indicated that the perception of accentedness is affected by both acoustic goodness of fit and lexical properties. The results are discussed in terms of matching variability in the input to multidimensional representations.

Keywords Decision making · Speech perception · Psycholinguistics

Traditionally, nonnative speakers have been thought to have a foreign accent because they fail to produce native-like speech; that is, they do not produce second language (L2) speech sounds (or sequences of L2 speech sounds) as a native speaker would. The existing literature on foreign-accented speech indicates that a wide variety of variables may influence both the presence and the degree of foreign accent (Piske, MacKay, &

Flege, 2001). Despite previous research, it is unclear which factors may affect the perception of gradient foreign accentedness.

Speaker-dependent variables such as age of L2 learning, length of residence in the target-language country, gender, formal instruction, motivation, language learning aptitude, and amount of continued native language (L1) use have been shown to influence a nonnative speaker's accent. As was indicated by Piske et al. (2001), both age of L2 learning and amount of continued L1 use appear to have the greatest effects on the degree of foreign accentedness. If foreign accent is indeed a failure to achieve native-like productions, factors such as age of L2 learning, instruction, and L1 use may capture other underlying articulatory and phonological properties. Speech sounds are produced by manipulating vocal articulators, and nonnative speakers may lack practice in making the necessary articulatory gestures in terms of both place and timing, thus resulting in a perceived accent (Flege, 1980). The phonology of a speaker's L1 may also influence his/her ability to produce nonnative phones by interfering with the speaker's ability to produce L2-specific phonetic detail. Speech sounds that are similar between the two languages appear to interfere with each other because the speaker may judge them to be acoustically different realizations of the same category (Flege & Hillenbrand, 1984). Additionally, the phonotactic rules of the speaker's L1 have been shown to affect accent in the L2; specifically, L2 sequences that align with the phonotactics of the speaker's L1 are judged to be less accented than those that do not (Park, 2013). Given this, nonnative productions indeed differ from native productions on a variety of temporal and spectral acoustic measures (Baker et al., 2011; Flege & Hillenbrand, 1984; Munro, 1993; Wayland, 1997). Specifically, this research has shown that the word duration, vowel duration, voice onset time, and formant values in nonnative productions deviate from typical native speaker values.

Given the acoustic differences between nonnative and native productions, native listeners appear sensitive to them, and

✉ Vincent Porretta
porretta@ualberta.ca

¹ University of Alberta, Edmonton, Alberta, Canada

² University of Turku, Turku, Finland

³ Department of Linguistics, University of Alberta, 2-40 Assiniboia Hall, Edmonton, Alberta T6G 2E7, Canada

can detect the presence of an accent in as little as 30 ms of a burst release (Flege & Hillenbrand, 1984). Listeners seem to use this fine-grained acoustic information when rating the degree of foreign accentedness. A number of studies have shown that accentedness ratings (for various L1–L2 combinations) can be predicted by many of the acoustic variables discussed above (Munro, 1993; Porretta & Tucker, 2012; Wayland, 1997). Witteman, Weber, and McQueen (2013) stated that vowels drive differences in perceived accentedness because they can vary significantly from standard forms, and Munro (1993) showed this for L1 Arabic productions of English vowels. In that study, five native English-speaking listeners (all linguists) rated the accentedness of vowel productions from native Arabic speakers. It was shown that first-formant (F1) and second-formant (F2) frequency measures were most predictive of those ratings, though the predictors varied by vowel. In a similar study, Wayland (1997) found that ratings (from native Thai speakers) of native English-speaking productions of Thai words were driven by spectral rather than temporal measures; however, the set of predictors varied for each of the different Thai tones. Prior to the study presented here, Porretta and Tucker (2012) found that global accentedness ratings of Chinese-accented talkers could be predicted by interactions between F1 and F2 values, as well as between vowel duration and word duration taken from a separate set of words from the same speaker. Thus, at least in the case of Chinese-accented English, the relationship between measures can affect perceived foreign accentedness.

Because it is often noted that accentedness appears to be a gradient, the concept of perceptual distance has been advanced for both regional- and foreign-accented speech (Clarke & Garrett, 2004; Floccia, Goslin, Girard, & Konopczynski, 2006; Goslin, Duffy, & Floccia, 2012). Perceptual distance acknowledges nonnative deviations and places accentedness on a sliding scale according to a particular accent's acoustic distance from a given native variety. This scale has foreign accents occupying the far end and regional accents somewhere in the middle, closer to the native accent. The underlying assumption of perceptual distance is that a single mechanism handles the processing of this variation (see Floccia et al., 2006). This continuum can be examined through the application of distance measures, calculated in relation to a typical native production. This aligns conceptually with the observation that accent strength varies in magnitude across speakers. Furthermore, some acoustic studies (Munro, 1993; Porretta & Tucker, 2012; Wayland, 1997) have attempted to explore the foreign accentedness continuum by employing acoustic distance measures. These studies have shown that acoustic distances can be used to predict accentedness ratings, suggesting that native listeners match nonnative tokens to the distributional properties of their learned native language. This matching process can be considered an evaluation of the goodness of fit of a token relative to the representation of a

native-like production. As was noted by J.B. Pierrehumbert (2003b), any given language occupies a particular region of phonetic space. It may be that the region occupied by a foreign accent does not match directly with that of the native accent. Thus, this conceptualization of accentedness can be quantified as the distances between the regions along any psychoacoustic dimension. By calculating multiple distance measures, it is possible to begin to understand the type of multidimensional comparison that listeners carry out. Although it may be that perceived accentedness is likely affected by the acoustic properties of the signal, it seems less intuitive that it may also be affected by speaker-independent properties of the words themselves. These properties include lexical frequency—an estimation of how often a word occurs in a language (e.g., Hasher & Zacks, 1984); phonological neighborhood density—an estimation of how many phonologically similar words exist in a language (e.g., Luce & Pisoni, 1998); and phonotactic probability—an estimation of how probable a given phonological sequence is in a language (e.g., Vitevitch & Luce, 1999). From a psycholinguistic perspective, the properties of words and their relationships to each other contribute to the organization of the mental lexicon (see Pierrehumbert, 2003b). These lexical variables have long been shown to affect behavioral measures such as reaction times, especially in native spoken word recognition. However, only a few studies have examined lexical variables as they relate to the recognition of foreign-accented speech (Imai, Walley, & Flege, 2005) or the perception of foreign accentedness (Levi, Winters, & Pisoni, 2007).

Imai et al. (2005) investigated the role of phonological neighborhood density in the recognition of native- and Spanish-accented English words. Although the focus of their study was to investigate the lexicon of L2 learners, the data from their native-English controls are particularly relevant here. Using an offline measure of word recognition (transcription accuracy), Imai et al. showed that lexical frequency and neighborhood density influenced word recognition of Spanish-accented English. Words with higher lexical frequency were recognized better than low-frequency words, whereas recognition was better for words from sparse neighborhoods than for words from dense neighborhoods. Although the task was inherently different from the one presented here, it shows that such variables as lexical frequency and phonological neighborhood density are still likely to be at play when dealing with foreign-accented speech. In a task more similar to the one presented here, Levi et al. (2007) examined the influence of lexical frequency on the foreign accentedness ratings assigned to English words produced by native German speakers. They found that words with higher lexical frequencies resulted in lower accentedness ratings. These two studies, taken together, suggest that not only the properties of the acoustic signal produced by the speaker, but also aspects of the lexical representations in the mental lexicon, can affect the recognition of accented speech and judgments about it. These

lexical properties may mediate the process by which listeners match new input to native-like representations and evaluate goodness of fit.

The existing research on the perception of foreign accentedness has indicated that speaker-dependent factors, particularly those contained within the speech signal, influence the perception of gradient foreign accent. Additionally, this perception may also be affected by speaker-independent factors. However, these speaker-dependent and -independent variables have not been examined together, nor with the same foreign accent. Thus, in the present study we investigated the perception of gradient foreign accentedness and the potential factors that may influence the matching process. An investigation of this sort, examining the perception of gradient foreign accentedness, provides an opportunity to develop an understanding of what contributes to the perception of variability in spoken language. Specifically, it indicates the type of information that must be represented in the lexicon and the relationships among representations.

In this study, we asked: (a) How do speaker-dependent variables (temporal and spectral acoustic distance measures) influence the perceived foreign accentedness of Chinese-accented English? (b) Do speaker-independent variables (lexical measures in addition to frequency) affect the perception of foreign accentedness? (c) If yes, how do these lexical variables influence perception?

That both lexical and acoustic properties might affect the perception of gradient foreign accent bears on the process by which nonnative variation is matched against a typical native-like production. With regard to speaker-dependent variables, we predicted that greater acoustic distance from a native reference point would result in stronger perceived foreign accentedness for all acoustic measures considered, due to the magnitude of the mismatch. As for the speaker-independent variables, we predicted that easing the matching process (by means of increased the activation or probability of the lexical item) would result in weaker perceived foreign accentedness. If speaker-dependent and speaker-independent variables are shown to be factors contributing to the degree of perceived foreign accentedness, this approach would bring forth the dimensions involved in assessing the goodness of fit of nonnative productions. The dimensions affecting the evaluation would then speak to the contents of the representations and the relationships among them.

Methodology

Participants

Thirty participants (24 female, six male) were recruited from the University of Alberta campus area and ranged in age from 18 to 33 years old ($M = 22.1$, $SD = 3.97$). All reported having

normal hearing and being native speakers of North American English. Participants received \$10 for completing the experimental task.

Materials

Recordings of nine male native Chinese speakers and seven male native English speakers were retrieved from the NU Wildcat Corpus of native- and foreign-accented English (Van Engen et al., 2010). The Chinese speakers were all listed as native speakers of Mandarin Chinese. Each recording contained a word list being read three times by a single talker, and a subset of 40 monosyllabic words was selected for use in this study (see Table A1 in the Appendix). Additionally, global accentedness ratings were obtained from the corpus; a detailed description of the ratings task can be found in Van Engen et al. (2010). This global rating was based on each talker's reading of the Stella Passage (Weinberger, 2013) and was judged on a scale of 1 to 9 (1 corresponding to *no foreign accent*, and 9 corresponding to a *very strong foreign accent*). The selected Chinese speakers represented a broad range of accentedness based on these mean global ratings ($M = 5.96$, $SD = 1.35$, $\min. = 3.1$, $\max. = 7.41$).

For the present study, the first repetition of the word list was chosen and extracted to serve as the individual stimuli for this study. Five measurements were taken from each token for each talker: (a) word duration, (b) vowel duration (of both monophthongs and diphthongs), (c) F1 frequency, (d) F2 frequency, and (e) F3 frequency. The word and vowel boundaries of the tokens from each speaker were segmented by hand in Praat (Boersma & Weenink, 2011), inspecting both the waveform and spectrogram. The beginning of the word was marked at consonant burst/onset, and the end of the word was marked after consonant aspiration or frication. For word-initial stops with a negative voice onset time (VOT), the beginning was marked at the onset of voicing. The beginning of each vowel was marked at the beginning of regular glottal pulses and the onset of the voicing bar, and the end of the vowel was marked at the decrease in F2 energy. Vowels adjacent to sonorant consonants presented difficulty, and changes in amplitude were used to identify the boundary, which was then verified auditorily. Although we do not believe that these are the only acoustic properties that a listener may make use of, previous research has indicated that these are likely contributing factors to foreign accentedness.

Using an automatic script in Praat, temporal measurements of word duration and vowel duration were extracted, along with three spectral measurements: F1–F3 measured at the midpoint. Formant values that appeared to be mispredicted values ($n = 17$) were hand inspected in the spectrogram, and new measurements were taken.

Stimuli

The recordings of the nine native Chinese speakers along with one native English speaker were taken as the stimuli for this experiment. The other six native English speakers, which we will refer to as the *native acoustic reference*, were used to calculate the mean values of the acoustic variables from which distance measures could be calculated. As such, they were not included as talkers in the rating task. Individual sound files were created for each token, which were then normalized for amplitude. These stimuli were arranged into ten blocks, each containing all 40 words from the word list (four words from each talker). Within each block, the word order was pseudorandomized. This was done to ensure that raters would hear all talkers in all blocks, thus representing the full range of productions.

Procedures

Participants completed the task seated at a computer in a quiet room. The stimuli were presented in DMDX (Forster & Forster, 2003) with over-the-ear headphones adjusted to a comfortable volume. Participants were instructed that they would see a word on screen and hear an auditory token of that word. Presenting the written word in conjunction with the auditory stimulus has been shown to make perceived differences between native and nonnative speakers more salient (see Levi et al., 2007). Participants were asked to rate how much of a foreign accent each stimulus had on a scale from 1 (*no foreign accent*) to 9 (*very strong foreign accent*). This response was made via computer keyboard. If a response was not made within 6 s, the program automatically proceeded to the next trial. Short, self-paced breaks were provided between blocks.

Analysis and results

Ratings

The goal of this study is to examine the factors that affect the perception of gradient foreign accentedness. Because we were interested in the underlying concept of accentedness as a continuum, we ultimately approximated it by using mean rating values for each item similar to those from other studies (e.g., Balota, Pilotti, & Cortese, 2001; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). Thus, before calculating and modeling these mean values, it was necessary to verify four aspects of the ratings obtained in this study: (a) that raters were using the scale in the appropriate manner (i.e., that they did not reverse the scale); (b) the use of the ratings scale across all participants; (c) the spread of ratings assigned to each talker; and (d) the agreement of raters across items.

Inspection of ratings

In total, 12,000 ratings were collected, 14 of which contained missing values (0.12% of the data). These missing values were approximately equally spread across the ten talkers and were removed before proceeding with the analysis. The ratings were first visually inspected, and it was verified that no participant reversed the rating scale. Second, the use of the scale across participants was inspected. The distribution of the assigned ratings indicated that no single value was over- or underused (see Table 1). Third, the spread of ratings for each talker was visually inspected, and all talkers appeared to be rated across a reasonable spread of the scale. On the basis of these inspections, no further data were removed.

Rater agreement

To assess the agreement among raters, we calculated two types of intraclass correlation coefficients (ICC): ICC2 and ICC2k, implemented in the package psych, version 1.4.5, of the statistical environment R, version 3.1.0 (R Development Core Team, 2014). The ICC is the ratio of the variance in the data explained by the variance between raters (Shrout & Fleiss, 1979). When the ICC is at 1, this indicates perfect agreement. The ICC2 is a measure of the absolute agreement in the ratings when each rater provides a rating for each token, corresponding to a two-way (Token $\times$ Rater) analysis of variance (ANOVA). Additionally, ICC2 treats raters as a random sample, lending itself to generalizability to a larger population. ICC2 is then an index of rater reliability based on agreement. Although this is informative regarding the reliability of a typical single rater, we were not interested in training listeners to rate accented speech. Rather, we were interested in the perception of foreign accentedness on average along this scale. For this reason, ICC2k was preferred, because it reflects the reliability of raters as a group, averaged together. This measure is of particular interest because the goal is to have each token rated by a group of raters and take the final rating of a token to be the mean of the group. The ICC2k indicates a high level of agreement on the mean accentedness value when the raters are viewed as a random group. The summary statistics of the ICC2 and ICC2k are given in Table 2, in which the p value corresponds to the null hypothesis test (i.e., if the intraclass correlation is zero).

Mean ratings

Mean item ratings were then calculated for each token by averaging over raters, henceforth referred to as the mean item rating. Item means ranged from 1.03 to 8.73, with standard deviations from 0.18 to 2.5. When these item means were aggregated by talker, the talker means ranged from 2.4 to 6.6, with standard deviations from 1.8 to 2.33. The distribution of the mean item ratings by talkers is illustrated in Fig. 1.

Table 1 Distribution of the assigned ratings

Rating	Frequency
1	1,356
2	1,234
3	1,497
4	1,611
5	1,451
6	1,531
7	1,497
8	901
9	908

Interestingly, these were highly correlated [$r(8) = .9, p < .001$] with the global accentedness rating (the Stella paragraph rating) available for each talker in the NU Wildcat Corpus (Van Engen et al., 2010). This further verifies the results of the intraclass correlation statistics and indicates that the raters in the present study arrived at judgments similar, on average, to those of the listeners in the Stella ratings from the Wildcat Corpus. This suggests that their judgments of accentedness may have been based on similar information. Thus, we assume that the underlying concept of degree of accentedness is likely to be related across the studies.

Model

The mean item ratings were modeled using generalized additive mixed modeling (GAMM; Wood, 2006), implemented in the mgcv package, version 1.7-29. GAMM does not assume a linear relationship between predictors as ANOVA or linear models do, and it allows for complex interactions between two or more numeric predictors. Because accentedness is thought to be a gradient, GAMM allows us to model the possibly wiggly functional form of the predictors (and interactions). Thus, GAMM is capable of handling possible nonlinearities in the data. This point is particularly important for the present study, due to the working conception of foreign accentedness as a continuum. It may be that predictors affect this continuum differently at different points along it. Additionally, GAMM allows for the inclusion of random effects, similar to linear mixed-effects models (see Baayen, Davidson, & Bates, 2008). A detailed description of the model-fitting process is presented below, following a description of the input variables.

Table 2 Intraclass correlation statistics

Type	ICC	F Value	df1	df2	p Value	Lower Bound	Upper Bound
ICC2	0.46	32.14	386	11194	<.01	0.42	0.50
ICC2k	0.96	32.14	386	11194	<.01	0.96	0.97

With regard to the application of GAMM, it has been previously applied successfully to model a variety of linguistic data. In particular, it has been used for investigating dialectal variation (Wieling, Montemagni, Nerbonne, & Baayen, 2014; Wieling, Nerbonne, & Baayen, 2011), event-related potentials (Kryuchkova, Tucker, Wurm, & Baayen, 2012; Tremblay & Baayen, 2010), prosodic prominence (Arnold, Wagner, & Baayen, 2013), and reaction times (Baayen, 2010a, 2010b). Outside of language, it has been used extensively in the field of ecology (see Zuur, Ieno, Walker, Saveliev, & Smith, 2009).

Input variables

Acoustic variables Formant values (F1–F3) were log-normalized for comparison across speaker-specific vowel spaces. The vowel-to-word ratio was calculated by dividing the vowel duration of a word by the total duration of that word. If accentedness is a result of nonnative productions approaching native-like acoustic targets (to varying degrees), quantifying the distance from native speaker norms allows for the examination of how variation along different variables affects perceived accentedness. Native speaker norms are presumed to be the acoustic values of a typical speaker. Here, as in other studies (Munro, 1993; Wayland, 1997), this was operationalized as a native acoustic reference, obtained by averaging across multiple native speakers for a particular variable. The six native English speakers subjected to acoustic analysis for calculation of the native acoustic reference were not included as talkers in the rating task. For each numeric variable, the token value of each talker was subtracted from the native acoustic reference. The absolute value of that difference yielded a positive number representing the magnitude of the distance between the production in question and the native acoustic reference. The acoustic variables are summarized in Table 3. An example of the log F1 and log F2 for the vowel /ɛ/ in *bet* is presented in Fig. 2.

Lexical variables The lexical frequencies for each word were retrieved from COCA (Davies, 2008), and the number of phonological neighbors for each word was extracted from the English Lexicon Project (Balota et al., 2007). As in other studies, neighbors were defined as words with a one-phoneme difference (e.g., /bæt/ is neighbors with /bæʔ/, /bæg/, and /set/). Furthermore, the phonotactic probability of each word was calculated using the Phonotactic Probability Calculator provided online by Vitevitch and Luce (2004). Both of these have been argued to capture the process of

Fig. 1 Boxplots for mean item ratings by talker. The black squares indicate the grand averages over items for each talker

matching words to phonological templates when searching the mental lexicon (Pierrehumbert, 2003b). The lexical variables are also summarized in Table 3.

Standardization and residualization The pairwise correlations of the input variables were computed to check for possible collinearity among them. In the case of collinearity, it is possible to assess the effect of a collinear input variable, though it is difficult to reliably assess the direction of the effect (Dormann et al., 2013). All continuous input variables were standardized in order to reduce spurious correlations between them by subtracting the mean and dividing by one standard deviation. However, even after standardization, phonotactic probability remained correlated with neighborhood density [$r(398) = -.44, p < .001$]. As in Jaeger (2010), residualization

was carried out to remove collinearity between these variables by fitting a linear model to the standardized phonotactic probability as a function of the standardized neighborhood density. We took phonotactic probability as the response variable because it requires an accumulation of phonemic sequences over words in the lexicon, whereas neighborhood density only requires phonological words and the relationships between them. Thus, it seems that, at least conceptually, neighborhood density is a simpler variable than phonotactic probability. The residuals from this model were extracted and, as expected, after residualization the correlation between phonotactic probability and neighborhood density was removed [$r(398) = 0, p = 1$]. Importantly, the residualized phonotactic probability was highly correlated with the original measure [$r(398) = .9, p < .001$]; thus, the residualized variable can be interpreted in the same manner as the original.

Table 3 Summary of input variables

Variable	Range	<i>M</i>	<i>SD</i>
Log F1 distance	0–1.13	0.16	0.16
Log F2 distance	0–1.04	0.15	0.14
Log F3 distance	0–0.5	0.09	0.07
Vowel-to-word ratio distance	0–0.32	0.06	0.05
Log COCA frequency	5.93–14.55	9.39	1.75
Neighborhood density	1–58	27.50	15.34
Phonotactic probability	0.05–0.33	0.17	0.06

Model fitting and evaluation

The input variables described above were fit to the response variable (the mean item rating) with by-word and by-talker random intercepts. By-word random intercepts allow for the possibility that some words may be more likely than others to sound accented. By-talker random intercepts allow for the possibility that also different talkers may be more likely than others to sound accented. For each predictor, a nonlinear functional relation with the response variable was allowed for by using a smooth function. For interactions, tensor product

Fig. 2 Log F1 and log F2 distances for the vowel /ε/ in “bet” are shown as dotted lines in relation to the native acoustic reference (square symbol) for the most accented Chinese talker (dot symbol) and the English talker (triangle symbol)

smooths were used, allowing for a wiggly surface (Baayen, 2010b; Wood, 2006).

The model was fit using the backward step-wise elimination procedure described in Zuur et al. (2009). All input predictors were first included as main effects, and the number of smoothing parameters was optimized using maximum likelihood. Second, their contributions to the model were evaluated using two criteria. The first criterion was the estimated p value of the smoothing parameter. The estimated p value of the smoothing parameter indicates whether or not the functional form of the predictor is different from zero. If greater than the conventional alpha level of .05, the predictor was considered for removal. The second criterion was the Akaike information criterion (AIC; Akaike, 1998). The use of AIC is an information-theoretic approach that supplies information on the strength of evidence for a particular model, given the data (Burnham & Anderson, 2002). Lower AIC values indicate a better model, because the loss of information is minimized. This concept of information loss is then applied to model selection by comparing the AIC values between different models. If the removal of a predictor led to an increase of less than 2 it was eliminated, because it did not lead to a substantial loss of information. This approach is particularly advantageous because it is not affected by the order of variables entered into the model.

Through this model-fitting process, log F3 distance and log COCA frequency were eliminated. Finally, motivated interactions among the acoustic and lexical predictors were included and examined using the same criteria. Doing so, two interactions emerged—one between log F1 distance and log F2 distance, and another between residualized phonotactic probability and neighborhood density. Thus, the model consisted of the following predictors: random intercepts for word and talker, the main effect of vowel-to-word ratio distance, the interaction between log F1 distance and log F2 distance, and the interaction between residualized phonotactic probability and neighborhood density. Visual inspection showed that the residuals of this model were approximately normally distributed, thus meeting the assumptions of the model and indicating a good fit to the data.

To evaluate the best model, we calculated Δ AIC values for the main effects and interactions in the model (see Table 4). This allowed us to rank the predictors in terms of the strength of their evidence. This was done by subtracting the AIC of the model including the predictor (i.e., the best model) from the AIC of the model without the predictor (i.e., the simpler model). As a rule of thumb, a $\Delta < 2$ suggests substantial evidence for the simpler model; Δ values between 3 and 7 indicate considerably less support for the simpler model; and $\Delta > 10$ indicates that the simpler model is very unlikely (Burnham &

Table 4 Model AIC, Δ AIC, and AIC weight for the rating model

Model	<i>df</i>	AIC	Δ AIC	AIC Weight
Best model	34.66	1,214.96	N/A	0.91
w/o tensor: Phonotactic prob., Neighborhood dens.	41.67	1,219.61	4.66	0.09
w/o tensor: Log F1 distance, Log F2 distance	33.62	1,241.75	26.79	0.00
w/o smooth: Vowel-to-word ratio distance	34.70	1,243.46	28.50	0.00

Anderson, 2002). Additionally, we calculated AIC weight values for the model variants (see Table 4). AIC weights are computed on the set of models and give a proportion indicating how often a particular model would be chosen as the most likely, given the data. These two measures give an indication of the importance of the predictor within the model and the likelihood of a particular model.

Looking at the Δ AIC values, it can be seen that the smooth function for vowel-to-word ratio distance produces the largest change, and thus indicates that it plays a large role in explaining the data. The interaction of log F1 distance and log F2 distance also plays an important role, similar in impact to vowel-to-word ratio distance. Although the interaction of phonotactic probability and neighborhood density plays a smaller role, it still contributes to the model likelihood. As can be seen in Table 4, on the basis of the AIC weights, the best model would be chosen as being most likely 91% of the time.

After evaluating the best model, it was refitted using restricted maximum likelihood in order to optimize the number of smoothing parameters, as is recommended in Zuur et al. (2009). This final model, reported below, explains 62.9% of the deviance, showing that the model is able to capture important facets of gradient of foreign accentedness.

Results

The results of the final model are reported along with a visualizations of the effects. Table 5 presents the statistics for the parametric and smooth terms in the model. The column labeled “edf” indicates the estimated degrees of freedom of the smooth functions. When they are equal to 1, as in the case of the vowel-to-word ratio distance measure, the effect is approximately linear. The effects are visualized in Fig. 3. Importantly, zero on all axes represents the mean.

The interaction of log F1 distance and log F2 distance was included because F1 and F2 are established variables involved in vowel categorization. The result of this interaction is presented as a regression surface in Fig. 3 (top left panel). The contour lines represent the estimated mean item rating values fitted by the model. As both log F1 and log F2 distances increase, mean item ratings also increase. This indicates that increased spectral deviations from the native acoustic

reference result in higher mean item ratings. Additionally, because there are more contour lines per scale unit along log F1 distance, the increase in accent strength appears to be more precipitous than with log F2 distance.

A linear main effect of vowel-to-word ratio distance can be seen in Fig. 3 (top right panel). The vowel-to-word ratio represents the amount of the word duration that is subsumed by the duration of the vowel. The vowel-to-word ratio distance represents how far that ratio is from the native acoustic reference for a given word. Increased deviation from the native acoustic reference is correlated with higher mean item ratings.

The interaction of neighborhood density and residualized phonotactic probability was included because both variables are involved in word identification and phonological matching (Imai et al., 2005; Luce & Pisoni, 1998; Pierrehumbert, 2003b). The result of this interaction is presented as a regression surface in Fig. 3, bottom left. Again, the contour lines represent the estimated mean item rating values fitted by the model. As neighborhood density and residualized phonotactic probability both increase, mean item ratings decrease. In particular, mean item ratings are high across neighborhood densities when residualized phonotactic probability is low (to the left in this panel). Likewise, mean item ratings are high across residualized phonotactic probabilities when neighborhood density is low (to the bottom in this panel). However, when both neighborhood density and residualized phonotactic probability are high, accentedness ratings steadily decrease. This indicates that when the phonemic sequence is probable and the word has many phonological neighbors, mean item ratings are lower (i.e., less foreign accent).

Discussion and conclusions

In the present study, we aimed to examine, in a single model, the acoustic and lexical variables that influence the accentedness ratings assigned to 40 English words. These words were spoken by ten different talkers: one native English-speaking talker and nine native Chinese-speaking talkers. On the basis of previous studies, it was predicted that temporal and spectral acoustic distance measures (i.e., speaker-dependent factors) would influence the perception of gradient foreign accentedness. Specifically, increased distance from the native acoustic reference (i.e., the mean native speaker value) would fit less well to a native-like

Table 5 Generalized additive mixed model reporting parametric coefficients (Part A), along with estimated degrees of freedom (Edf), reference degrees of freedom (Ref. *df*), *F* and *p* values for the tensor products, and random effects (Part B) for the rating model

A. Parametric Coefficients	Estimate	Std. Error	<i>t</i> Value	<i>p</i> Value
Intercept	4.76	0.33	14.40	<.0001
B. Smooth Terms	Edf	Ref. <i>df</i>	<i>F</i> Value	<i>p</i> Value
Tensor: Res. phonotactic prob., Neighborhood dens.	3.53	3.73	7.47	<.0001
Smooth: Vowel-to-word ratio distance	1.00	1.00	31.90	<.0001
Tensor: Log F1 distance, Log F2 distance	3.25	3.46	10.95	<.0001
Random smooth: Word	17.92	36.00	1.01	<.00001
Random smooth: Talker	8.75	9.00	34.87	<.0001

representation and would thus result in stronger perceived foreign accentedness. In addition, it was predicted that lexical variables (speaker-independent factors) would also influence perception along the continuum of accentedness. Previous research using ratings had suggested that higher lexical frequency would result in lower perceived foreign accentedness. It was predicted that increased probability and activation of a particular item within the mental lexicon might ease the matching process and lead to lower perceived foreign accentedness.

The analysis of the individual ratings indicates that when raters are viewed as a randomly sampled group, the mean item ratings offer a highly reliable measure of the accentedness continuum. This was further corroborated by the high correlation between the mean talker ratings (based on word-specific

ratings collected in this study) and the global accentedness ratings for each talker available in the NU Wildcat Corpus.

The mean item ratings were modeled using GAMM. The results indicate that, indeed, both acoustic and lexical variables affect the perception of gradient foreign accentedness. This suggests that both are involved in the matching of a particular token to the representation of what constitutes a native-like production. The content of the acoustic signal (temporal and spectral properties) is what must be decoded and matched. This signal is dependent on the abilities of the talker to approximate a native-like production. However, lexical properties such as frequency, phonotactic probability, and the number of phonological neighbors are properties of the lexicon and are not dependent on a talker's production. These may also affect the matching process and, subsequently, the perception of foreign

Fig. 3 Effects of the final model. Top left panel: Contour plot of the interaction of log F1 distance and log F2 distance. Top right panel: Partial effect of vowel-to-word ratio distance, with 95% confidence bands. Bottom left panel: Contour plot of the interaction of residualized

phonotactic probability and neighborhood density. For both contour plots, light gray represents higher mean item ratings (stronger accent), and black represents lower mean item ratings (weaker accent)

accentedness. By including both acoustic and lexical predictors in the model, we controlled for the effects of one on the other.

The results of this study show that multiple acoustic measures predict the variability of foreign accentedness ratings when they are considered as distances from a native acoustic reference. The present study replicates the results of Munro (1993), who showed that distance variables predict the accentedness ratings assigned to English vowels spoken by native Arabic speakers. Here, we demonstrated the same with Chinese-accented English words, showing that the distances of both F1 and F2 values from typical native productions positively correlate with the strength of perceived accentedness. Also, as was suggested by Witteman et al. (2013), deviations in vowel quality appear to be a driving force in the perception of foreign accentedness. On the basis of the Δ AIC results, it appears that the interaction between the log F1 and log F2 distances is quite important in the assessment of foreign accentedness. This stands to reason, particularly for monosyllabic words, because F1 is a cue to vowel height and F2 a cue to vowel frontness. Jointly, these measures relate to vowel categorization, and variability along these frequencies may lead to possible miscategorization, which may particularly influence the perception of monosyllabic words.

Additionally, the present results add to those of previous studies by showing the role of temporal properties such as word and vowel durations—more specifically, their relationship to each other. Here, the temporal relationship of vowel-to-word ratio indicates the amount of the word subsumed by the vowel, and the distance of this from a typical native value positively correlates with the strength of perceived accentedness. It appears that speakers who are more able to approximate this durational pattern are perceived as being less accented. This, taken together with the spectral properties, supports the idea that listeners assess acoustic features including the word durations, vowel durations, and formant values produced by the speaker and compare them against native speaker values. It should be noted, however, that we do not believe these to be the only acoustic properties that listeners use to evaluate the degree of foreign accent. It is likely that other properties (e.g., VOT and formant transitions/trajectories) also influence judgments of accentedness.

These acoustic results support the concept of perceptual distance (Floccia et al., 2006) as a framework for understanding the phenomenon of foreign accent. As we have seen here, perceptual distance can be quantified along any number of acoustic parameters, and the magnitudes of these deviations generally lead to stronger perceived accentedness. Listeners may maintain fine phonetic and distributional information about what constitutes a native-like production. This could be represented either as a set of exemplars (Johnson, 1997, 2006; Pierrehumbert, 2001, 2003a; Walsh, Möbius, Wade, & Schütze, 2010) or as a

single prototype (Iverson & Kuhl, 1995; Samuel, 1982) that encodes multiple dimensions. In either case, this information is then available for matching new input produced by different speakers. Listeners then seem to be able to judge the goodness of fit of a given token to the representation of a native production along multiple acoustic dimensions. Non-native speakers who are more successful at producing speech sounds closer to typical native values are indeed perceived as being more native-like. It is also interesting to note that not all of the words spoken by the native talker were rated by listeners as being perfectly native. This indicates that even among native tokens, perceptual distance may be taken into consideration by listeners for handling variation in native speech, suggesting the employment of similar mechanisms for both native and nonnative speech. However, given the results of the model, perceptual distance is not the only influence on the degree of perceived foreign accentedness.

The lexical properties of particular words also appear to exert influences on the perception of gradient foreign accentedness. This is particularly important because it shows that factors independent of the speaker's ability to phonetically approximate native productions influence this perception. The present model tests these lexical predictors when all other variables are held constant. That said, we failed to replicate the results of Levi et al. (2007), who showed that lexical frequency influences ratings of foreign accentedness. Specifically, their results indicate that perceived foreign accentedness decreases as lexical frequency increases. However, lexical frequency was not a significant predictor in our model. One reason for this may be that there were not enough words to provide sufficient power for detecting the effect. Our 40 monosyllabic words were extracted from the NU Wildcat Corpus, and therefore we could not control the distribution across lexical frequencies. Levi et al. specifically controlled for frequency at three levels (i.e., low, mid, and high), had a greater number of words, and also showed that the frequency effect was reduced when both orthographic and auditory forms of information were available.

The model does indicate the influence of both neighborhood density and phonotactic probability, and in particular, their interaction. Specifically, as both neighborhood density and phonotactic probability increase, perceived accentedness decreases. According to Pierrehumbert (2003b), lexical neighborhoods and phonotactics provide general information about the lexicon and involve matching an input to some phonological template. The present result of phonotactic probability is similar to work on well-formedness (Hay, Pierrehumbert, & Beckman, 2004; Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997). Vitevitch et al. (1997) found that ratings of well-formedness of English nonce words reflect the probability of phonotactic sequences. Subsequently, Hay et al. (2004) found similar results indicating that well-formedness is gradient with

this probability. Here it seems that when the phonemic sequence is more probable, perceived foreign accentedness is reduced. This can be thought of as a frequency-like effect that facilitates the matching process and the judgment of well-formedness with respect to nativeness.

With regard to neighborhood density, as neighborhoods become more dense, perceived accentedness decreases. The result suggests that denser neighborhoods may provide greater activation within the lexicon, thus easing the matching process. Processing studies have shown that effects associated with neighborhood density are modality-specific. In visual word recognition, greater neighborhood density facilitates processing in both lexical decision (Andrews, 1992; Yates, Locker, & Simpson, 2004) and sentential context (Yates, Friend, & Ploetz, 2008). This facilitative effect has been explained in terms of the extent of activation—namely, more neighbors trigger more overall activation, and result in easier word identification. Conversely, in spoken word recognition, greater neighborhood density inhibits processing (Vitevitch & Luce, 1998). This inhibitory effect is explained in terms of lexical competition. This modality-specific effect of neighborhood density is likely also related to how and when information becomes available for processing; visual information becomes available rather quickly, whereas auditory information unfolds sequentially over time. With regard to rating spoken stimuli, one might expect that dense neighborhoods would provide less “wobble room” in actual production, due to the competition among neighbors. Indeed, in a study of transcription accuracy, Imai et al. (2005) showed that their native control group correctly identified fewer dense-neighborhood accented words. Similarly, in a ratings task, one might expect that raters would judge tokens from dense neighborhoods more harshly than tokens from sparse neighborhoods. For example, *bit*, which has many neighbors, would be rated as more accented than *through*, which has few neighbors, because raters would require the acoustics of the tokens to be more precise in order to avoid confusion with neighbors. However, this was not the case in the results obtained in the present study. It is important to note that the task of rating tokens is rather different than word recognition, particularly with the present design, in which participants saw the written form of the word on screen while the auditory stimulus was played. This presentation context was chosen because Levi et al. (2007) found that an auditory + orthography presentation increased the perceived differences between native and nonnative speakers. However, we do not believe that the effect reported here is simply a matter of presentation context, but that of the matching process. Participants were asked to judge how native the token was as compared to their knowledge of native productions of that word. The visual information may have assisted participants so that, rather than needing to first identify the

auditory word form before making the judgment, they could evaluate the auditory token relative to the intended lexical item. In this way, the judgment of the match could be given without solely relying on the auditory recognition process to ensure proper lexical selection. It is conceivable that highly accented words might be misinterpreted or not interpreted as items existing in the listener’s lexicon.

We believe that the immediacy of the visual information leads to an overall increase of activation in the neighborhood, thus facilitating the comparison between the auditory token and the reference representation. Similar to the effect of phonotactic probability, this increase in activation is then reflected in the ratings. However, given the disparity in the results for online processing and that presentation context can modulate lexical effects, as was indicated by Levi et al. (2007), further studies will be required in order to tease apart the potential influences of the task on the perception of accentedness with regard to lexical variables such as neighborhood density. Whatever is the case, it is important to highlight that lexical variables are expected to influence the ratings of foreign accentedness, regardless of modality. In a strictly auditory task, in which listeners can rely only on the spoken input, it might be that the effect of neighborhood density on perceived accentedness would change to be more in line with the results from spoken word recognition tasks. In this case, however, it would be necessary to carefully control intelligibility, so as to ensure that listeners understood the intended lexical item. Although intelligibility and accentedness are related and covary, they have been shown to be at least partially independent (Derwing & Munro, 1997).

Another interesting result is the relative strengths of evidence for given predictors in the final statistical model. In this study, ΔAIC values were used to rank the predictors. This ranking indicates that the acoustic (i.e., speaker-dependent) variables have the strongest influence on perceived foreign accentedness. The lexical (i.e., speaker-independent) variables were also found to play a role, but to a lesser extent. Importantly, the contribution of the lexical variables, though a smaller effect, was above and beyond that of the acoustic variables, because the model held the them statistically constant. This analysis does not rule out the possibility that lexical variables may influence acoustic–phonetic productions, such as higher-frequency tokens being produced more accurately (Pierrehumbert, 2001; Wright, 1979). However, as we discussed above in the Standardization and Residualization section, pairwise Pearson’s correlations between the acoustic and lexical variables were low (ranging from $-.15$ to $.22$). This aligns with the results reported by both Levi et al. (2007) and Imai et al. (2005). So, it seems that properties of the lexicon can influence perceived foreign accentedness beyond the effect of perceptual distance.

The goodness of fit of a given token is not solely dependent on the relationship between the acoustic signal and a native-like representation; rather, this fit is affected by probabilistic information about a particular lexical item, as well as by its

relation to other items. This acoustic and lexical matching process appears to be multifaceted in nature, such that the shorter the acoustic distances and the better the fit to probable lexical properties, the less accented a token is judged to be. Whether or not the native-like representation is based on a single prototype or a set of exemplars that give rise to a prototype, the results point to a lexical representation that contains multidimensional probabilistic and distributional information about what constitutes a native-like accent as well as various properties of the lexical item. Listeners (in this case, raters) likely use all of this learned information when evaluating a token's goodness of fit within their native language. Matching processes occur across multiple dimensions, and perceived foreign accentedness may reflect the ease with which a match occurs across both acoustic and lexical properties, rather than strict perceptual distance. It thus seems that the perception of variation (at least at the word level) is affected by the acoustic distance from native-like representations as well as by properties specific to the lexical items themselves. Therefore, an L2 talker's ability to approximate typical native-speaker values on acoustic measures is only part of what affects the strength of perceived foreign accentedness, with probabilistic properties of the lexicon modulating the perception of accentedness. So, perceived foreign accentedness may be better described as an index of the ease of the matching process.

Moving forward from these results, it would be interesting to examine the roles of linguistic context and listener experience with foreign-accented speech. In a more typical communicative setting, speakers make their own lexical choices. Thus, it is possible that listeners have developed inferences about the types of words that nonnative speakers choose to use, since there is evidence that nonnative speakers tend to opt for high-frequency words in their productions (see Crossley & McNamara, 2009). Listeners might then use this information to infer something about the speaker's proficiency, which in turn could affect ratings of foreign accentedness. This might be tested by placing words in sentence contexts and having listeners make ratings to specific words. With regard to experience, it may be that the distributional information against which a nonnative token is compared changes over time, through experience (see Piske et al., 2001). If this is the case, one could expect that raters with differing levels of experience might find nonnative variability more or less acceptable, due to changes in the distribution subsequently affecting the matching process. This would indicate that language experience influences the development of the representation of what constitutes a native production for individual listeners.

Author note The data contained in this article were previously included as part of the first author's dissertation. Also, we thank the associate editor and the anonymous reviewers for their constructive comments on the manuscript.

Appendix: Experimental items

Table 6 Experimental monosyllabic words selected from the NU Wildcat Corpus word list

bait	booze	peep	sheets
bat	boss	poop	shoes
bees	bought	pop	shoots
beet	box	pot	shop
bet	but	put	shots
bit	cheese	risks	soup
bite	chirp	rubbed	splash
block	chop	screen	spring
boat	girl	searched	stripe
boot	legs	sheep	through

References

- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199–213). New York, NY: Springer. doi:10.1007/978-1-4612-1694-0_15
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 234–254. doi:10.1037/0278-7393.18.2.234
- Arnold, D., Wagner, P., & Baayen, R. H. (2013). Using generalized additive models and random forests to model prosodic prominence in German. In F. Bimbot et al. (Eds.), *Proceedings of Interspeech 2013* (pp. 272–276). Lyon, France: International Speech Communication Association. Retrieved from www.isca-speech.org/archive/interspeech_2013/i13_0272.html
- Baayen, R. H. (2010a). Demythologizing the word frequency effect: A discriminative learning perspective. *Mental Lexicon*, 5, 436–561.
- Baayen, R. H. (2010b). The directed compound graph of English: An exploration of lexical connectivity and its processing consequences. In S. Olson (Ed.), *New impulses in word-formation [Linguistische Berichte Sonderheft 17]* (pp. 383–402). Hamburg, Germany: Buske.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. doi:10.1016/j.jml.2007.12.005
- Baker, R. E., Baese-Berk, M., Bonnassee-Gahot, L., Kim, M., Van Engen, K. J., & Bradlow, A. R. (2011). Word durations in non-native English. *Journal of Phonetics*, 39, 1–17. doi:10.1016/j.wocn.2010.10.006
- Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, 29, 639–647. doi:10.3758/BF03200465
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445–459. doi:10.3758/BF03193014
- Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer (Version 5.3.61). Retrieved January 12, 2012, from www.praat.org
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer.

- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, 116, 3647–3658. doi:10.1121/1.1815131
- Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, 18, 119–135.
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 400+ million words, 1990–present*. Salt Lake City, UT: Brigham Young University. Retrieved from www.american corpus.org
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, 20, 1–16.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27–46.
- Flege, J. E. (1980). Phonetic approximation in second language acquisition. *Language Learning*, 30, 117–134. doi:10.1111/j.1467-1770.1980.tb00154.x
- Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America*, 76, 708–721. doi:10.1121/1.391257
- Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? A lexical decision study in French listeners. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1276–1293. doi:10.1037/0096-1523.32.5.1276
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35, 116–124. doi:10.3758/BF03195503
- Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, 122, 92–102. doi:10.1016/j.bandl.2012.04.017
- Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, 39, 1372–1388. doi:10.1037/0003-066X.39.12.1372
- Hay, J., Pierrehumbert, J., & Beckman, M. E. (2004). Speech perception, well-formedness and the statistics of the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation: Papers in laboratory phonology 6* (pp. 58–74). Cambridge, UK: Cambridge University Press.
- Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *Journal of the Acoustical Society of America*, 117, 896–907. doi:10.1121/1.1823291
- Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, 97, 553–562.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61, 23–62. doi:10.1016/j.cogpsych.2010.02.002
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic Press.
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34, 485–499. doi:10.1016/j.wocn.2005.08.004
- Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, 122, 81–91. doi:10.1016/j.bandl.2012.05.005
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44, 978–990. doi:10.3758/s13428-012-0210-4
- Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America*, 121, 2327–2338.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
- Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, 36, 59–66.
- Park, H. (2013). Detecting foreign accent in monosyllables: The role of L1 phonotactics. *Journal of Phonetics*, 41, 78–87. doi:10.1016/j.wocn.2012.11.001
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam, The Netherlands: Benjamins.
- Pierrehumbert, J. B. (2003a). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46, 115–154. doi:10.1177/00238309030460020501
- Pierrehumbert, J. B. (2003b). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probability theory in linguistics* (pp. 177–228). Cambridge, MA: MIT Press.
- Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191–215. doi:10.1006/jpho.2001.0134
- Porretta, V., & Tucker, B. V. (2012). Predicting accentedness: Acoustic measurements of Chinese-accented English. *Canadian Acoustics*, 40, 34–35. Retrieved from <http://jcaa.caa-aca.ca/index.php/jcaa/article/view/2526>
- R Development Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org/
- Samuel, A. G. (1982). Phonetic prototypes. *Attention, Perception, & Psychophysics*, 31, 307–314. doi:10.3758/BF03202653
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428. doi:10.1037/0033-2909.86.2.420
- Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London, UK: Continuum.
- Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat corpus of native and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, 53, 510–540.
- Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9, 325–329. doi:10.1111/1467-9280.00064
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374–408. doi:10.1006/jmla.1998.2618
- Vitevitch, M. S., & Luce, P. A. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, 36, 481–487. doi:10.3758/BF03195594
- Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, 40, 47–62.
- Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel exemplar theory. *Cognitive Science*, 34, 537–582. doi:10.1111/j.1551-6709.2010.01099.x

- Wayland, R. (1997). Non-native production of Thai: Acoustic measurements and accentedness ratings. *Applied Linguistics*, 18, 345–373.
- Weinberger, S. H. (2013). *Speech accent archive*. Fairfax, VA: George Mason University. Retrieved from <http://accent.gmu.edu>
- Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R. H. (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language*, 90, 669–692. doi:10.1353/lan.2014.0064
- Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, 6(e23613), 1–14. doi:10.1371/journal.pone.0023613
- Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, 75, 537–556. doi:10.3758/s13414-012-0404-y
- Wood, S. N. (2006). *Generalized additive models: An introduction with R* (Vol. 66). Boca Raton, FL: Chapman & Hall/CRC Press.
- Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition*, 7, 411–419. doi:10.3758/BF03198257
- Yates, M., Friend, J., & Ploetz, D. M. (2008). The effect of phonological neighborhood density on eye movements during reading. *Cognition*, 107, 685–692.
- Yates, M., Locker, L., Jr., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*, 11, 452–457. doi:10.3758/BF03196594
- Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. (2009). *Mixed effects models and extensions in ecology with R*. New York, NY: Springer.