Intelligibility of foreign-accented words: Acoustic distances and gradient foreign accentedness

Vincent Porretta and Benjamin V. Tucker

University of Alberta
porretta@ualberta.ca, bvtucker@ualberta.ca

ABSTRACT

Intelligibility and degree of accentedness are interrelated aspects of non-native speech. Previous research suggests that foreign accentedness is influenced by phonetic distance measures [7]. These distance measures may also influence the intelligibility of individual words. In the present study we further investigate the relationship between the intelligibility of native- and Mandarin-accented English words and acoustic distance measures (both spectral and temporal). We also examine the functional relationship between intelligibility and ratings of foreign accentedness assigned to the same words. Intelligibility was based on transcription accuracy scores and acoustic distances were based on formant and duration measurements in relation to mean values from a set of native talkers. The results indicate that temporal and spectral distances influence word intelligibility and that the functional relationship between intelligibility and accentedness is non-linear.

Keywords: Foreign-accented speech, intelligibility, gradient accentedness, acoustic distance.

1. INTRODUCTION

Foreign-accented speech can present significant challenges for listeners. It can vary along numerous dimensions simultaneously resulting in a perceived accent and possibly disrupting speech intelligibility. Intelligibility and accentedness are known dimensions of non-native speech and have been shown to be related and partially correlated to one another [5]. Despite their interrelated nature, Munro and Derwing [7] show that highly accented speech does not necessarily preclude full intelligibility of utterances, underlining the partially independent nature of the dimensions.

Acoustic distances (from typical native speaker values) have been implicated as factors affecting the perception of gradient foreign accentedness [7, 9]. Given this, along with the interrelationship between intelligibility and accentedness, it would seem reasonable to expect that this type of distance measure might also influence the intelligibility of non-native speech.

Acoustic distance metrics have previously been employed in a study of non-native speech intelligibility [6]. Examining the mean intelligibility of two American English vowels produced by native Japanese speakers it was shown that formant distance affected intelligibility (though differentially for each vowel). While this has been shown for the intelligibility of single vowel productions, the effect of distances has not been demonstrated for the intelligibility of words.

Duration and its distance from typical native values may also play a role in intelligibility. Word duration has been examined as an aspect of non-native speech which varies from native distributions [2]. Also, vowel duration has been shown to influence identification of vowels produced by non-native speakers [6]. In addition to raw durations, temporal (durational) distance can also be calculated similar to spectral distance measures and has been shown to influence ratings of degree of foreign accentedness [9]. Therefore, a distance metric applied to duration may also prove fruitful in the study of intelligibility.

If intelligibility and accentedness are indeed related to one another, one would expect both temporal and spectral distances to influence intelligibility at the word level. Also, while intelligibility and accentedness have been shown to be partially correlated, the nature of their relationship is not clear. Thus, the present study investigates the following questions: 1) Do temporal and spectral distance measures affect single word intelligibility? 2) What is the functional relationship between intelligibility and degree of foreign accentedness? We address these questions through statistical models of intelligibility data making use of acoustic measurement and previously-collected ratings of foreign accentedness [9].

2. METHODOLOGY

2.1. Participants

One hundred and twenty participants (90 Female, 30 Male) were recruited for participation in the experiment. All reported normal hearing and that they are native speakers of North American English.

2.2. Materials

Word list recordings of nine male native Mandarin Chinese speakers and seven male native English speakers were retrieved from the NU Wildcat Corpus of Native- and Foreign-Accented English [11]. A subset of 40 monosyllabic words was selected from the list for use in this study and was extracted from each speaker's recording. The word and vowel boundaries of the tokens were marked by hand in Praat [3]. Five measurements were taken from each token: 1) word duration; 2) vowel duration; 3) midpoint F1 frequency; 4) midpoint F2 frequency; and 5) midpoint F3 frequency. Individual sound files were created for each token which were then normalized for amplitude.

The tokens of the nine native Chinese speakers along with one native English speaker were taken as the stimuli for this experiment. The stimuli were arranged into 10 lists, each containing all 40 words from the word list (4 words from each talker). This ensured that any participant would only hear each word once. The other six native English speakers were used to calculate mean values of the acoustic variables from which distance measures could be calculated. As such, they were not included as talkers in the transcription task.

2.3. Accentedness

As part of a previous study [9], accentedness ratings were collected for the stimuli presented in this experiment. Thirty native English-speaking participants rated each item on a scale of 1 (no foreign accent) to 9 (very strong foreign accent). The ratings were subsequently averaged to produce a mean rating for each item. The items presented in the current study are distributed over a broad range of accentedness. Item ratings ranged from 1.03 to 8.73 suggesting a reasonable amount of pronunciation variability among the talkers presented in the intelligibility task.

2.4. Distance calculation

To quantify the distance (for a given dimension) between a particular production and average native production, it was necessary to calculate a Native Acoustic Reference value for each dimension [cf., 7]. This was obtained by averaging across the values extracted from the six native English speakers not included as talkers in the experiment.

Formant values (F1–F3) were log-normalized and a vowel-to-word ratio was calculated by dividing the vowel duration of a word by the total duration of that word. For each of these numeric variables, the value of a given token 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.

2.5. Procedure

Participants were randomly assigned to one of the 10 lists. They completed the task seated at a computer in a sound-attenuated booth. Stimuli were presented binaurally in the clear using E-Prime 2.0 (Psychology Software Tools, Inc.). Participants were instructed to listen to the auditory stimulus and type the word in its standard spelling. Ten practice items (one from each talker) were provided prior to the experimental items so that participants could familiarize themselves with the task.

3. ANALYSIS AND RESULTS

Transcription accuracy (i.e., correct/incorrect) was scored automatically in E-Prime; however, prior to analysis, all responses marked as incorrect were checked by hand. Homophones, like 'but' and 'butt' were considered interchangeable, and minor spelling errors that did not change the phonemic form were re-scored as correct. This represented 3% of the data (146 of 4800 data points).

Modeling was carried out in R [10], version 3.1.0 utilizing generalized additive mixed modeling (GAMM), implemented in the package mgcv [12], version 1.7-29. GAMM, which has previously been applied to linguistic data [e.g., 1, 9], does not assume a linear relationship and possible non-linear functional form is estimated based on the data. If the estimated degrees of freedom (EDF) is greater than 1, then the effect is not linear.

3.1. Overall Intelligibility

Overall item intelligibility was checked by means of proportion correct across participants. Mean intelligibility ranged from 0 to 1 (M = 0.73, SD = 0.32), indicating that some items were always transcribed correctly while others not. Participant performance on the task was checked by means of proportion correct across items. Mean participant performance in transcription accuracy ranged from 0.58 to 0.88 (M = 0.72, SD = 0.07).

3.2. Intelligibility and Acoustic Distances

To investigate the effect of acoustic distance on intelligibility, transcription accuracy was modeled using binomial GAMM. Each distance variable was included as a smooth function of the response

variable along with a smooth for trial and by-item and by-subject random intercepts. Using an information theoretic backwards step-wise elimination procedure [13], log F2 distance, log F3 distance, and trial were removed from the model.

Figure 1: Partial effect of vowel-to-word ratio distance on intelligibility with 95% confidence intervals.

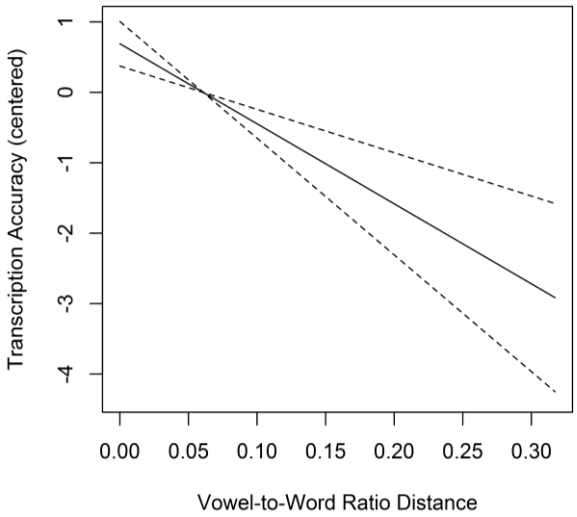
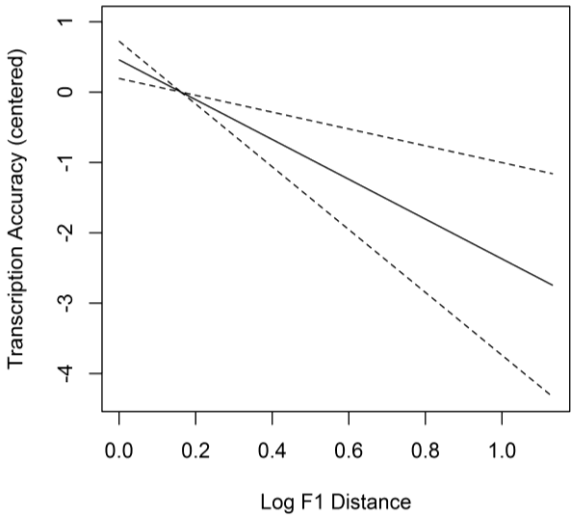


Figure 2: Partial effect of log F1 distance on intelligibility with 95% confidence intervals.

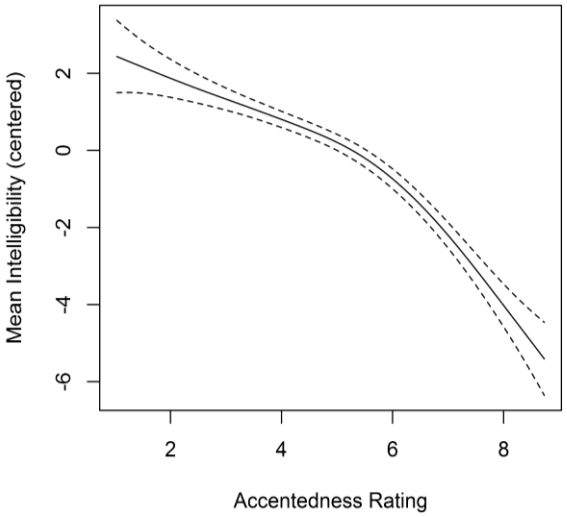


Both vowel-to-word ratio distance and log F1 distance produced significant linear effects (vowel-to-word ratio distance: EDF = 1, F = 19.02, p < 0.0001; log F1 distance: EDF = 1, F = 12, p < 0.001). As seen in Figure 1, as vowel-to-word ratio distance increases, intelligibility decreases. This indicates that as the overall durational relationships of the monosyllables deviates from a typical native production, the less intelligible the word is. The effect of log F1 distance is similar, as illustrated in Figure 2. As F1 moves farther from an average native value, the less likely the word will be transcribed correctly.

3.3. Intelligibility and Gradient Accentedness

To investigate the effect of accentedness rating on intelligibility, mean item intelligibility was modeled as a function of mean item accentedness using GAMM. Mean item intelligibility was calculated as the average transcription accuracy. This proportion was then logit transformed, producing an unbounded response variable. In this model, accentedness rating was included as a smooth function of the response variable along with by-word and by-talker random intercepts. During the backwards step-wise elimination procedure no variable was removed.

Figure 3: Partial effect of accentedness rating on intelligibility with 95% confidence intervals.



We find that when accentedness increases, intelligibility decreases, as illustrated in Figure 3. Interestingly, this effect is not linear (EDF = 3.6, F = 60.86, p < 0.0001). Ratings in the upper third of the continuum appear to be more detrimental to the intelligibility of the word.

4. DISCUSSION AND CONCLUSIONS

This study demonstrates that in single word transcription both temporal and spectral measures influence intelligibility. Specifically, log F1 distance and vowel-to-word ratio distance hold a negative relationship with intelligibility. The greater the deviation of F1, the more likely participants are to provide an incorrect response. This is most likely due to misrecognition of the vowel in the monosyllables. For vowel-to-word duration ratio, greater deviation also leads to increased erroneous responses. Failing to approximate a native-like relationship may change the perception of one or

more phonemes in the monosyllable, suggesting that durational measures of consonants may also play a role.

Given these results, listeners may indeed be sensitive to highly probable values for multiple phonetic dimensions. This provides support for the idea that listeners are sensitive to and have expectations about what constitutes a native-like production against which an incoming signal can be compared. This sensitivity and expectation is likely built up through experience with different talkers. Increased distance from likely native values is then detrimental to intelligibility of non-native speech.

Additionally, we observed the functional relationship between intelligibility and foreign-accentedness of individual words. Specifically we confirmed the previously reported interrelationship between intelligibility and accentedness [8]. We extend these previous findings by demonstrating that this negative relationship appears to be non-linear. This result indicates that while increasing accentedness generally decreases intelligibility, strong accent ratings particularly disrupt intelligibility, perhaps due to the confluence of multiple acoustic factors. Because accentedness rating is a more holistic measure of acoustic appropriateness it may capture more dimensions than can be reasonably measured for a given set of words.

Munro and Derwing [8] suggest a distinction be drawn between the two constructs given that some highly accented productions are nonetheless intelligible. In the present study we have evidence indicating that a one-to-one correspondence indeed does not apply. However, Munro and Derwing used sentences as stimuli while the present study makes use of monosyllabic words. This sentential context likely influences the intelligibility of individual words through contextual expectation. The relationship shown in the current study is based on single monosyllabic words in isolation. Single word intelligibility tasks such as this remove the possible influence of sentential context and focus on the processing of the specific individual words to which the acoustic measurements (or accentedness ratings) relate. Therefore, a more direct comparison can be drawn between the two dimensions based on the properties of the speech itself.

It would be particularly interesting to examine in greater depth listener experience with the accent in question. One might expect a reduced effect of distance as experience increases. Previous work has indicated that listeners can indeed achieve talker-independent generalization for sentence level intelligibility through specific, lab-based training [4]. Additionally, given the functional relationship between intelligibility and accentedness, generalization (or the speed of generalization) may be affected by degree of foreign accentedness.

In summary, the present study has demonstrated that intelligibility of words in isolation is affected by acoustic distance and that the relationship between intelligibility and foreign accentedness is non-linear. Altogether this presents complementary and additional evidence for understanding the influences of both acoustic distance measures and gradient foreign accentedness on second language intelligibility.

5. REFERENCES

[1] Baayen, R. H. 2010. Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, 5(3), 436–561.

[2] Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K. J. & Bradlow, A. R. 2011. Word durations in non-native English. *J. Phon.*, 39(1), 1–17.

[3] Boersma, P. & Weenink, D. 2014. Praat: Doing phonetics by computer. Version 5.3.61.

[4] Bradlow, A. R. & Bent, T. 2008. Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707–729.

[5] Derwing, T. M. & Munro, M. J. 1997. Accent, Intelligibility, and Comprehensibility. *Stud. Second Lang. Acquis.*, 20(1), 1–16.

[6] Kewley-Port, D., Akahane-Yamada, R., & Aikawa, K. 1996. Intelligibility and Acoustic Correlates of Japanese Accented English Vowels. *Proc. of ICSLP 96*, 450–453.

[7] Munro, M. J. 1993. Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Lang. and Speech*, 36(1), 39–66.

[8] Munro, M. J. & Derwing, T. M. 1995. Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Lang. Learning* 45, 73–97.

[9] Porretta, V., Kyröläinen, A.-J., Tucker, B. V. In press. Perceived foreign accentedness: Acoustic distances and lexical properties. *Atten Percept Psychophys*. 1–36.

[10] R Development Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Version 3.1.0.

[11] 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. *Lang. and Speech*, 53(4), 510–540.

[12] Wood, S. N. 2014. mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation. R package version 1.7-29.

[13] Zuur, A., Ieno, E.,Walker, N., Saveliev, A., & Smith, G. 2009. *Mixed effects models and extensions in ecology with R*. New York: Springer.