5-2019

The effects of pitch and speaking rate on foreign accented speech perception.

Rebecca F. Davis
University of Louisville

The Effects of Pitch and Speaking Rate on Foreign Accented Speech Perception

By

Rebecca Davis

Submitted in partial fulfillment of the requirements for Graduation summa cum laude

and

for graduation with Honors from the Department of Psychology

University of Louisville

March, 2019

Abstract

Perception of foreign accent is typically studied using an accentedness rating task. For example, native English listeners rate the degree of accentedness in sentences produced by non-native English speakers. However, in past studies, it has been unclear what criteria participants used to judge accentedness. Here, native English speakers rated the accentedness of Korean-accented English sentences on a scale from 1 (strong accent) to 9 (little to no accent). Participants rated sentences that were unmodified or had one acoustic property removed. In one block, pitch contours of sentences were flattened and set to their mean values. In another block, speaking rates were set to the grand mean of all speaking rates (3.8 syllables/second). This way, changes in accentedness ratings across unmodified and modified sentences were attributable to the acoustic property that was removed. Accentedness ratings were not systematically influenced by manipulations of pitch contours, but were influenced by speaking rate manipulations. Increasing speaking rate (to 3.8 syllables/second) made sentences sound less accented than their unmodified versions; decreasing speaking rate made sentences sound more accented than their unmodified versions. Results suggest that speaking rate directly contributes to ratings of foreign accentedness.

The Effects of Pitch and Speaking Rate on Foreign-Accented Speech Perception

Introduction

In today's time, people are becoming exposed to more individuals who speak with a foreign accent than in the past. It has become nearly impossible to have not encountered at least one individual speaking with an accent. Some speakers require more effort on the part of the listener in order to understand the accented speech (Baese-Berk et. al. 2013). But what exactly is a foreign accent comprised of? An accent is when a person is speaking in a non-native language while at the same time using speech characteristics from the individual's own native language (O'Brien, 2016), or, to an untrained listener, a foreign accent is simply sounded unlike the native language (Haynes-Harb & Hacking 2015). This can be measured using "accentedness," which is subjectively how strong the listeners perceive a foreign-accent to be. Listeners can usually identify when someone is speaking with an accent, however, the specific components of speech that an accent is comprised of are still unclear.

There have been many studies dealing with listeners' perception on accentedness (e.g., Derwing & Munro, 1997; Haynes-Harb & Hacking, 2015; O'Brien, 2016). Several studies have focused on differentiating accentedness (strength of accent), intelligibility (how well the listeners understand what is being spoken), and comprehensibility (perceived intelligibility – how well the listeners think they understand). In these studies, in order to judge accentedness, researchers asked listeners to rate on a scale from 1 to 9 (1 being very heavily accented and 9 being little to no accent) without giving them clear criteria on how to rate accentedness (Derwing & Munro, 1997). This methodology is problematic because, for an untrained listener not given any criteria

in how to rate the speaker's accentedness, intelligibility and comprehensibility might have been used to judge accentedness.

Haynes-Harb and Hacking (2015) viewed this method as being very subjective and not producing enough information to scientifically judge what accentedness is. After completing an accentedness rating task, participants completed a survey asking them what specific criteria they used in order to judge accentedness. In fact, 23% of participants said that they had used prosodic features (auditory qualities of speech such as pitch, rhythm, tempo, etc.) to judge accentedness and 15% said that they used speaking rate (Haynes-Harb & Hacking, 2015). Other criteria participants mentioned in the survey are segmentals – vowel and consonant pronunciation accuracy – (92%), grammar and syntax (46%), enunciation and mumbling (38%), and fluency and vocabulary (8%). This survey was beneficial in shedding some light on past results using this methodology, but still fails to delve deeper into exactly which prosodic features listeners are using.

Pitch is commonly used to tell individuals' voices apart from one another. Winters (2013) argued that accentedness ratings are based in part on the pitch contours of sentences (how pitch changes over time throughout the sentence), which differ for native and non-native speakers. Speaking rate is also used as an indicator of how adept second-language learners are at their second language (Winters & O'Brien 2013). After removing features that contributes to higher accentedness ratings, I hypothesized that participants would report the speakers as being less accented (i.e. The manipulated blocks would be less accented than the control block).

The following research project took pitch and speaking rate, two of the prosodic features mentioned from the Haynes-Harb & Hacking survey (2015), and directly manipulated them in order to judge their relative contributions to perceived accentedness. The other criterion the

participants mentioned were controlled for, in this study, by the selecting of the stimuli so that each speaker was enunciating enough to be understood and so that each speaker was not misspeaking words. The experiment asked listeners to rate accentedness in three different blocks. One block was the control block in which no acoustic manipulations were done, and it was just the Korean-accented English speakers saying the sentences as they were recorded. In another block, the pitch of the speakers was flattened to each speaker's individual pitch average. And lastly, a block set each speaker's rate to the overall average of all of the speakers' talking rate. These manipulations removed the uncertainty that other studies have over which aspect was truly used to judge accentedness rating because we are controlling for just one criterion in each block. In the case of this study, any differences from the control to the manipulated data would be due to the manipulations.

The findings from this study are important because identifying specific ways to lessen an accent can make a big difference with how non-native speakers interact with the native speakers' world. If, for instance, a person knew how to modify their own pitch to sound less accented (i.e., by controlling the voice to neither be high nor low like what was done with the stimuli in this study), then that would make it easier for them to be understood. There are whole realms of possibilities dependent upon this and future research done on the subject.

Methodology

Participants

There were 30 participants that self-reported being native English speakers and having normal hearing. The participants were students from the University of Louisville. Participants

were recruited through the Department of Psychological and Brain Sciences Research SONA system. Each participant received course credit for participating in the study.

Stimuli

Sentences used in this experiment were found from the Wildcat Corpus of Native- and Foreign-accented English (Van Engen et al., 2010). There were 28 native-Korean talkers with each reading a different sentence, with two female talkers saying two sentences each (30 sentences total). Having 30 speakers was decided upon due to wanting to be as generalizable to how people interact in the general population. The sentences chosen were decided upon based on recording quality and clarity of the sentence. Sentences in which the speaker misspoke words or sentences that had background noise were not chosen in case those reasons would impact the participants' ratings on accentedness (those criteria actually belong to intelligibility and comprehensibility, not accentedness). Also, there were an even number of male and female speaker sentences. To do so, two of the female speakers had to be used twice with 2 different sentences (i.e. 28 unique speakers and 30 unique sentences). Half of the sentences were chosen from the high predictability section (where the last word of the sentence was guessable) and the other half from the lower predictability section (where the last word of the sentence was not guessable as it contained unusual word combinations).

The regularly accented block acted as a control. This allowed for a baseline to be set up in order to detect changes in accentedness from this block to the manipulated blocks.

Pitch Contour Manipulation. The pitch contour of each sentence was edited in Praat (Boersma & Weenink, 2018). Any pitch points that were clearly not part of the sentence were manually removed. After doing so, the average pitch was calculated for each sentence. Then,

each sentence was set to that average pitch by replacing all of the original pitch points with the average pitch. The new sentences had the flattened pitch contours.

Speaking Rate Manipulation. The duration (how long it took the speaker to recite the sentence) and number of syllables were calculated. The number of syllables was divided by duration, and this yielded the speaking rate for each sentence. These numbers were averaged, producing the overall average speaking rate across the speakers of 3.8 syllables per second. Each sentence was then scaled to have the average speaking rate. This was done by using Pitch Synchronous Overlap and Add (PSOLA) in Praat. Each sentence's scaling factor (the rate at which the sentence was sped up or slowed down in order so it would have the mean speaking rate) divided by the mean rate across all sentences. For example, one sentence had a rate of 5.12 syllables per second. 5.12 was divided by the overall average of 3.8 and that equaled a scaling factor of 1.34. That individual sentence was slowed down by a factor of 1.34 so that it would have a new speaking rate of 3.8 syllables/second. If the scaling factor was over 1, the sentence was slowed down; if the scaling factor was under 1, the sentence was sped up. Overall, 16 sentences were sped up and 14 were slowed down in order to meet the target rate of the 3.8 syllables per second.

Procedure

Participants read and signed a consent form. After doing so, participants were led to a sound isolated booth and wore headphones. The participants were given oral and written instructions about the task. On each trial, participants heard a sentence and were asked to rate accentedness on a scale of 1 to 9 (1 being very accented and 9 being little to no accent). Participants responded by clicking the mouse on buttons labeled 1 through 9. One block tested the original sentences with no manipulation, one block tested sentences with flattened pitch

contours, and another block tested sentences with averaged speaking rate. The three blocks were presented in counterbalanced orders, controlling for the chance that order effects will occur between the three blocks. Each block contained 30 trials which were tested in random orders, as well. The listeners were allowed to take brief breaks between blocks as needed. The whole experiment took participants no more than 30 minutes.

Analysis

The data that was analyzed for this study was the change from accentedness ratings of the control block to the accentedness ratings of the manipulated pitch and speaking rate blocks. As this experiment was analyzing ordinal data, it was not possible to use a paired-samples t-test for comparing two measures. Rather, in this study, the nonparametric Wilcoxon signed-rank test was used, which allowed for the comparison of two groups of ordinal data.

Results

Each of the participants completed the experiment by rating their perceived accentedness of the speakers on a scale of 1 to 9 (1 being very accented and 9 being little to no accent). For each sentence in a given block, the mean rating was calculated across all participants. The following are the histograms of each sentence's average accentedness rating in each block:

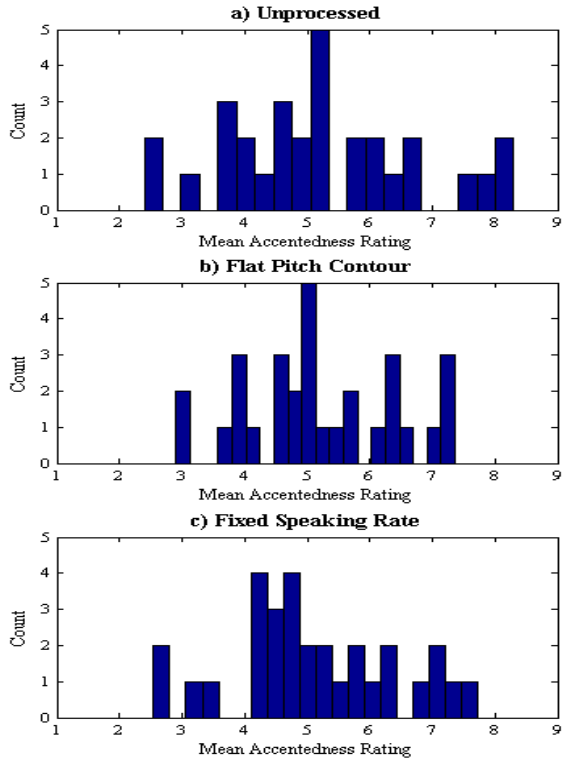

Figure 1: Histograms of the mean ratings of each sentence. The y-axis depicts the frequency of responses and the x-axis shows mean accentedness ratings (1= very accented, 9=little to no accent). The top figure is the control block, the middle figure is the flattened pitch block, and the bottom figure is the fixed speaking rate block.

As shown in the Figure 1, each of the ratings broadly follow a normal distribution. There did not look to be a significant change in ratings based off this data (i.e., a rightward shift of distributions in Figures 1b or 1c to higher numbers, or, lower accentedness ratings). This inference was reached by nonsignificant Wilcoxon signed rank tests (between unprocessed and flat pitch contour ratings *[Z = -0.7463, p = .4555]*, and between unprocessed and fixed speaking rate rating *[z = 0.6479, p = .5170]*). Given the (normal) shapes of the distributions in Figure 1 and the transformation of ratings into mean ratings, it is possible that paired-samples t-tests would be the more appropriate statistical test here, therefore a t-test was also performed between

unprocessed and flat pitch contour ratings [t(29) = -0.4149, p = .6813] and between unprocessed and fixed speaking rate rating [t(29) = 0.853, p = .4007]. The same conclusion that there was no significant change was reached from both analyses.

The manipulations done in this study did not yield significant effects on the change in rating. The statistical analyses above were expecting group-level shifts in accentedness, but it is possible that not all sentences were equally resilient/susceptible to our manipulations-see the wide distributions in Figure 1a. Also, it is important to note that there was a difference in how the pitch block was manipulated versus the speaking rate stimuli. When manipulating pitch, each speaker was set to their own individual average pitch. This differed from the speaking rate manipulation, in which speakers were set to the average speaking rate of all 30 talkers. Given that these manipulations were relative to a mean speaking rate, this meant that some talkers were sped up and others were slowed down depending on their relation to the mean. In looking at the data by item, there is actually a correlation present between whether the speakers were slowed down or sped up, and the change in accentedness rating relative to the unprocessed sentences. The speakers' sentences that were sped up were rated as sounding less accented than the unprocessed versions of those sentences. Also, the sentences that were slowed down were rated as more accented than the unprocessed versions of those sentences. As shown below in Figure 2, there is a negative correlation present between original speaking rate and the change in rating due to rate standardization *[r: -0.78, $p < .001$]*. Also, each plot point is labeled with the corresponding control group sentence rating in order to see if the magnitude of the initial accentedness rating influenced how much the rating was changed (i.e. would all speakers who were rated as highly accented be grouped around the same area on the graph?) and, as the graph

shows, the initial accentedness rating had no bearing on the relationship.

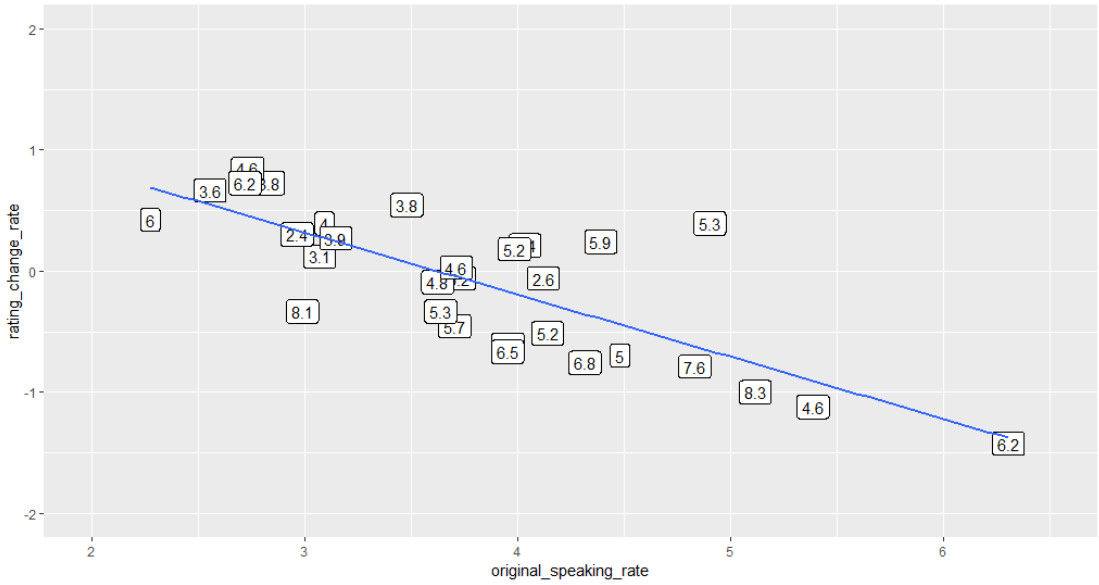

Figure 2: The x-axis depicts the original speaking rate for each sentence in syllables per second (Mean: 3.8 syllables/second). The y-axis shows the change in ratings, calculated as fixed speaking rate sentences minus unprocessed speaking rate sentence ratings. The positive values mean that the fixed speaking rate sentences were rated as less accented than the unprocessed sentences while the negative values mean the fixed speaking rate sentences were rated as more accented than the unprocessed sentences. Each point on the graph represents a sentence and the number is its mean rating in the unprocessed condition. Slower sentences that were sped up to 3.8 syllables per second were rated as less accented (positive rating change) and faster sentences that were slowed down to 3.8 syllables per second were rated as more accented (negative rating change).

There was a systematic relationship uncovered when analyzing speaking rate by item. Since that was the case, this encouraged reanalysis of stimuli in the pitch contour flattened condition by item as well. This new analysis compared whether the speakers' average pitches correlated with the change in accentedness ratings relative to the unprocessed version of these stimuli. As shown in figure 3, this was not the case *[r=-0.08, p = 0.6813]*. It was also considered whether rating changes might have corresponded to how much pitch contours varied or not before flattening them, but again, the data did not yield a significant correlation *[r=-0.12, p =*

0.5242].

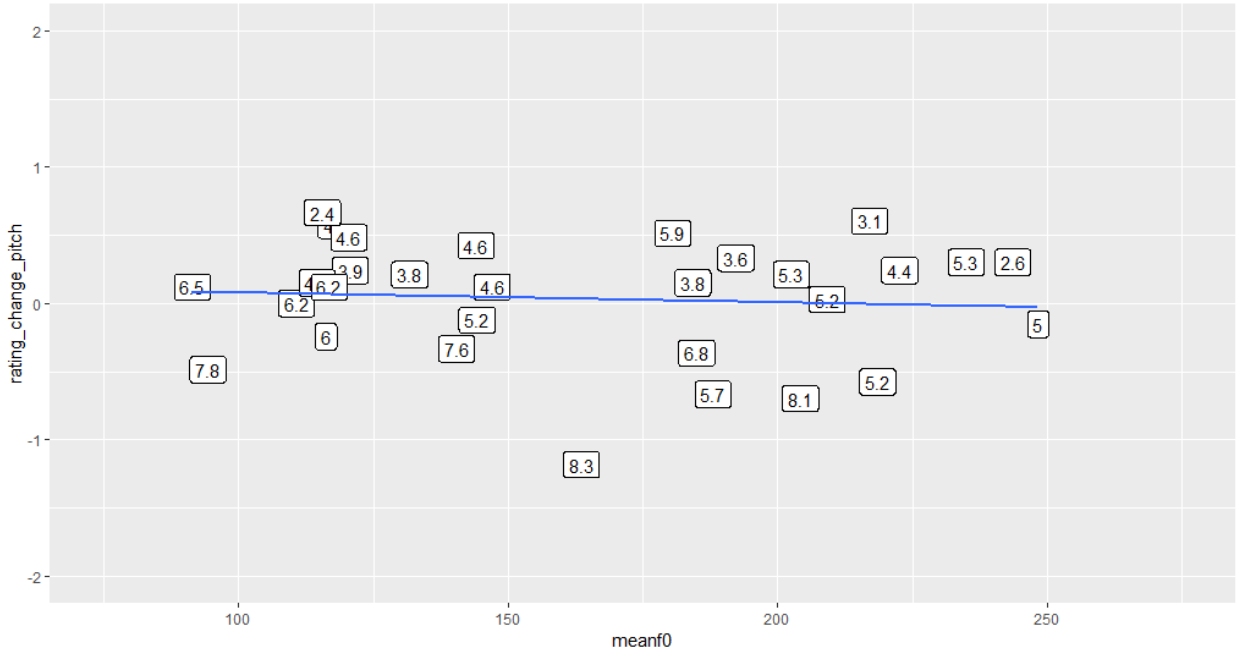

Figure 3: The x-axis depicts the mean f0 for each sentence. The y-axis shows the change in ratings, calculated as flattened pitch sentence ratings minus unmanipulated pitch sentence ratings. The positive values mean that the manipulated pitch sentences were rated as less accented than the unmanipulated pitch sentences while the negative values mean that the manipulated pitch sentences were rated as more accented than the unmanipulated sentences. Each point on the graph represents a sentence and the number is its mean rating in the unmanipulated condition.

Discussion

In previous studies, researchers have differentiated accentedness from intelligibility and comprehensibility; those are two aspects of non-native speech that, in the past, have been mistaken for accentedness. In knowing this confusion, it is unclear if the criterion that untrained listeners used in rating accentedness belonged to accentedness or intelligibility and comprehensibility. Due to this, in this research study, pitch and speaking rate were controlled for to make certain that any changes that would have been found would be from the manipulation of pitch or speaking rate respectively.

Participants were asked to listen to three distinct blocks and rate how accented they thought the speaker was. One block consisted of the original 30 sentences with no manipulation, similar to how other experiments of this type have been run. This was the baseline which we compared to the other two blocks. Of the other two blocks, one had flattened the pitch of the sentences to each individual speaker's average pitch. The other block set each speaker's rate to an overall average. Any changes between the control block and the manipulated blocks would be due to these specific manipulations and not by any other outside influence. Neither the manipulation of pitch nor the manipulation of speaking rate had a statistically significant change in accentedness from the initial control ratings of accentedness.

There could be two reasons as to why the pitch manipulation did not yield significant changes in accentedness ratings. The first is that flattening pitch makes the sentences sound unnatural and robotic. In a study done by Tamagawa et, al. (2011), participants rated synthetic voices as more accented than human speech. It could be that participants misinterpreted that with what accentedness is. Explained further in the limitations section, just because the experiment was controlled so that pitch was the only manipulation for that block, it is still not certain whether participants considered pitch in their ratings. The participants were not told the definition of accentedness, or what properties make up accentedness, so that they could make their ratings accordingly. Like the study by Derwing and Munro (1997) and the majority of studies done on this subject, it is unclear what criteria participants used to rate accentedness. There is no way of knowing if the participants were consistent between each other, and that could have skewed the data.

For the speaking rate block, there was no overall shift in accentedness ratings as per the histogram and the two analyses performed (see Figure 1c). A reason for this is that we failed to

take into account individual speaker differences when making the manipulation. With pitch, we set each speaker to their own individual average pitch, while, with speaking rate, we set each speaker to the average speaking rate of all of the speakers, which was about 3.8 syllables per second. In this block, we averaged together the speaking rate of all 30 speakers and set them all to the same average rate. By doing this, some speakers were slowed down in rate and others were sped up. Not doing the same manipulation to every individual (like what was done to pitch, matching the speakers to their own f0) could have impacted the study by trying to make the speakers all sound universally the same instead of letting the speakers sound like their own individual selves.

Even though the speaking rate manipulated sentences were not systematically different in overall changes in ratings, there were systematic differences in rating by item. As mentioned in the previous paragraph, depending on if they were above or below the speaking rate mean (3.8 syllables/second) some talkers' speaking rates were sped up or slowed down. For the speakers whose sentences were sped up, they were rated as less accented than their control sentences. Conversely, for the speakers whose sentences were slowed down, those sentences were rated as more accented than the control. This suggests an important point, that a key part of what makes people perceive accentedness is how slow or fast the speaker is. This finding could be impactful in helping second language learners be more perceived as being less accented, thus, making them more likely to be understood. When a speaker is at a point in their second language learning where they are ready to have natural conversations, it is helpful to know that speaking faster would contribute to native speakers understanding the non-native speaker better. This is not just with accentedness, but intelligibility and comprehensibility as well (Derwing & Munro, 1997).

In combating the limitations from the previous paragraphs, there are a couple of things that could be done to this study to improve it for the future. For starters, there is a high variability of the stimuli with having 28 different talkers and 30 different sentences. It could also be beneficial to repeat this study using a single sentence spoken by 30 speakers to see if sentence difference created too much variability, making the results not significant. The manipulation of speaking rate could also be changed to make everyone faster by a fixed amount instead of the speakers' speed going in either direction from the global mean. This study could also be contrasted with one in which the participants were told of what criteria to rate accentedness with. That might make the results more reliable because there would be a higher probability that participants are measuring what accentedness is, instead of confusing accentedness with intelligibility and comprehensibility.

References

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America*, *133*(3), EL174-EL180.

Boersma, Paul & Weenink, David (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.42, retrieved 15 August 2018 from http://www.praat.org/

Bradlow, A.R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 16* (2), 707-729

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19(1), 1-16.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125-3134.

Grosjean, F. (1985). The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual & Multicultural Development*, 6(6), 467-477.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31-38.

Hayes-Harb, R., & Hacking, J. F. (2015). Beyond rating data: What do listeners believe underlies

their accentedness judgments?. *Journal of Second Language Pronunciation*, 1(1), 43-64.

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, *10*(2), 135-159.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and speech*, *38*(3), 289-306.

Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Studies in Second Language Acquisition*, *23*(4), 451-468.

O'Brien, M. G. (2016). Methodological choices in rating speech samples. *Studies in Second Language Acquisition*, 38(3), 587-605.

Tamagawa, R., Watson, C. I., Kuo, I. H., MacDonald, B. A., & Broadbent, E. (2011). The effects of synthesized voice accents on user perceptions of robots. *International Journal of Social Robotics*, *3*(3), 253-262.

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 6 across conversational dyads with varying language alignment profiles. *Language and Speech*, 53(4), 510-540.

Winters, S., & O'Brien, M. G. (2013). Perceived accentedness and intelligibility: The relative

contributions of F0 and duration. *Speech Communication*, *55*(3), 486-507.