sciendo

Research in Language, 2018, vol. 16:2 DOI: 10.2478/rela-2018-0012

WHICH PHONETIC FEATURES SHOULD PRONUNCIATION INSTRUCTIONS FOCUS ON? AN EVALUATION ON THE ACCENTEDNESS OF SEGMENTAL/SYLLABLE ERRORS IN L2 SPEECH

ZHIYAN GAO
George Mason University
zgao@gmu.edu

STEVEN WEINBERGER
George Mason University
weinberg@gmu.edu

Abstract
Many English language instructors are reluctant to incorporate pronunciation instruction into their teaching curriculum (Thomson 2014). One reason for such reluctance is that L2 pronunciation errors are numerous, and there is not enough time for teachers to address all of them (Munro and Derwing 2006; Thomson 2014). The current study aims to help language teachers set priorities for their instruction by identifying the segmental and structural aspects of pronunciation that are most foreign-accented to native speakers of American English. The current study employed a perception experiment. 100 speech samples selected from the Speech Accent Archive (Weinberger 2016) were presented to 110 native American English listeners who listened to and rated the foreign accentedness of each sample on a 9-point rating scale. 20 of these samples portray no segmental or syllable structure L2 errors. The other 80 samples contain a single consonant, vowel, or syllable structure L2 error. The backgrounds of the speakers of these samples came from 52 different native languages. Global prosody of each sample was controlled for by comparing its F0 contour and duration to a native English sample using the Dynamic Time Warping method (Giorgino 2009). The results show that 1) L2 consonant errors in general are judged to be more accented than vowel or syllable structure errors; 2) phonological environment affects accent perception, 3) occurrences of non-English consonants always lead to higher accentedness ratings; 4) among L2 syllable errors, vowel epenthesis is judged to be as accented as consonant substitutions, while deletion is judged to be less accented or not accented at all. The current study, therefore, recommends that language instructors attend to consonant errors in L2 speech while taking into consideration their respective phonological environments.

Keywords: accentedness, speech perception, pronunciation instruction

1. Introduction

Pronunciation, rather than vocabulary or grammar, has been found to be a major factor that impairs communication (Grant and Brinton 2014), and non-native

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

accent carries certain stigma that often leads to negative social and workplace outcomes (Gluszek and Dovidio 2010). English learners, therefore, place great importance on the correctness of their pronunciation. For example, most Polish students surveyed in (Waniek-Klimczak, Rojczyk and Porzuczek 2015) do not want non-English phonetic features to be present in their L2 English speech, and 94% of them desire to speak like a native English speaker. In other words, the learning objective for some learners is not only about achieving intelligibility, but also about speaking with a native-like accent. Although achieving a native-like accent is often conditioned by various extra-linguistic elements, recent research on pedagogy does show that accent can be reduced via explicit instruction of phonological forms (Couper 2006; Nair, Krishnasamy and De Mello 2006). However, instructional time is usually limited in conventional ESL classrooms, which makes it impossible for instructors to attend to individual difficulties of every student. Therefore, it could be more efficient to set priorities on correcting L2 phonetic features that are most "foreign" to native speakers. Previous studies on foreign accent indeed show that the degree of perceived foreign accent is affected by various phonetic cues (Munro and Derwing 2001). However, the relative importance of different segmental and syllable structural cues is not readily clear. The current study investigates whether different types of segmental and syllable structural cues differ in their relative impacts on accentedness perception. 110 native English speakers are recruited to provide accentedness ratings on 100 non-native English speech samples. Each speech sample is a short unsynthesized two-word audio snippet, containing only one segmental or syllable structure L2 pronunciation error. The results provide evidence showing that native English speakers do judge L2 errors differentially. The findings could potentially enable a more efficient curriculum for pronunciation instruction.

2. Segmental correlates of foreign accent

Foreign-accented speech displays a variety of phonetic characteristics that differentiate it from native speech. Indeed, "foreign accent" is usually considered an issue of perception, rather than production. Only those perceivable phonetic deviations in non-native speech are considered features of "foreign accent". As Munro and Derwing (1998) defined it, foreign accent is "the extent to which an L2 learner's speech is *perceived* to differ from native speaker norms". Therefore, research on foreign accents often relies on perception experiments to investigate the phonetic characteristics that might correlate with foreign accent. Among investigations on the segmental correlates of foreign accent, consonant errors are often found to be of vital importance. Several studies have found that VOT duration associates with perceptual accentedness in L2 English speech (Flege and Eefting 1987; McCullough 2013). Liquid errors might also associate with foreign accent. For example, the substitution of Japanese flap (i.e. [ɾ]) for English liquids

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

[ɹ] and [l] were considered accented by native English speakers (Riney Takada and Ota 2000).

Findings on the impact of vowel quality change on accentedness perception are not conclusive. McCullough (2013) finds that foreign accentedness associated with vowel formant changes. Greater formant frequency deviations (i.e. mean formant frequency) from native speaker norms leads to higher ratings of accentedness (i.e. more foreign accented). This finding is consistent with several other studies, which also show independent effects of both static F1 and dynamic/static F2 values on accentedness ratings (Munro 1993; Wayland 1997). However, conflicting findings have been reported elsewhere. Major (1987) found that foreign accentedness might associate with some vowels but not with others. Chan, Hall, and Assgari (2016) argue that it is vowel space, rather than deviations of format frequency, that correlates with foreign-accentedness.

3. Syllable Structure Correlates of Foreign Accent

Most research on foreign accent perception has focused on the impact of vowel, consonant and prosody. Fewer studies investigated the impact of syllable structure change on accentedness perception. L2 syllable production often involves some form of a simplification strategy, namely, segment epenthesis or segment deletion (Sato 1984; Hansen 2001). Some suggest that segment epenthesis is more important than consonant errors in L2 speech in signaling foreign accentedness. Epenthetic schwa, for example, was perceived as more accented than consonant feature changes (e.g. [tʃ] to [ʃ]) (Magen 1998). However, evidence is lacking on whether segment deletion could also be indicative of foreign-accentedness. After all, strategies such as obstruent coda deletion is also a prominent feature in native speech (Labov 1997; Tagliamonte and Temple 2005; Demuth, Culbertson and Alter 2006). Take t/d-deletion in English as an example. Native English speakers are more likely to delete /t/ or /d/ when they are past tense morphemes (e.g. /d/ in "called") than when they are part of the stem (e.g. /d/ in "hold") (Guy 1991). Non-native speakers' t/d-deletion strategy, however, does not seem to be bound by the grammatical conditions of /t,d/ (Hansen 2001; Edwards 2011). Although there are indeed differences between deletion strategies in native and non-native speech production, there is a paucity of evidence on whether the differences affect foreign accent perception.

4. Accentedness Rankings of L2 Errors

While most studies investigated only a few phonetic deviations, Magen (1998) and van den Doel (2006) compiled a list of L2 phonetic variants and directly compared their perceptual accentedness or severity. In Magen (1998), two Spanish speakers each recorded 96 sentences in English, from which 56 phrases

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

were selected for acoustic manipulation. For each phrase, Magen (1998) acoustically edited out one L2 error (e.g. editing out epenthetic schwa, or lengthening VOT duration on [pʰ]), which would ideally make the altered phrases less accented than the original ones. Ten native English speakers provided their accentedness judgment on the synthesized phrases and their unaltered counterparts. By comparing judgment ratings on the altered and unaltered phrases, Magen (1998) found that epenthetic schwa, vowel quality change (e.g. [ʃɪp] becomes [ʃip]), consonant manner change (e.g. [tʃ] becomes [ʃ]) significantly affect accentedness perception, while stop voicing (e.g. VOT shortening) does not. While Magen (1998) mainly focused on Spanish speakers' L2 English production, van den Doel (2006) focused on Dutch speakers' L2 English production. To provide natural sounding stimuli, van den Doel (2006) asked native English speakers to mimic L2 errors that are common among Dutch speakers (e.g. "bed" becomes "bet"). He then placed these stimuli in carrier phrases (e.g. she lay in bed/bet for most of the day.) and asked native English speakers to first identify the "error" presented in each phrase, and then provide a "severity" rating on each "error". The results showed that lexical stress shift and the uvularization of English [ɹ] are the most severe among all errors. Although various consonant errors (e.g. VOT shortening) and vowel errors (e.g. [æ] becomes [e]) were considered severe to native English speakers, consonant and vowel errors in general did not show any apparent difference in severity.

Magen (1998) and van den Doel (2006) both studied a specific group of L2 English speakers, and both provided an accentedness or severity ranking of different types of L2 errors. They both found that lexical stress shift and vowel epenthesis are indicative of accentedness, but they seemed to disagree on whether stop voicing changes (i.e. VOT changes) affects accentedness perception. The two studies also applied different approaches to achieve experimental control. Magen (1998) resorted to acoustic manipulations, while van den Doel (2006) had native English speakers mimic L2 errors. Both strategies have advantages and shortcomings. Acoustic manipulation could be quite precise in altering specific signals, but one might question the "naturalness" of the altered sound. Native speakers' mimicry of L2 errors might indeed achieve "naturalness". It however raises questions about whether the mimicry is truly representative of L2 speech. Both Magen (1998) and van den Doel (2006) placed stimuli in carrier phrases. However, the phonological environment of each target stimulus was not well controlled.

The current study aims to address the potential problems in previous research by obtaining stimuli that are both natural and representative of L2 speakers with various language backgrounds. Instead of acoustic manipulation, the current study employs a Dynamic Time Warping method (Giorgino 2009) to control for prosody in the least intrusive manner. The term "error" was used in previous studies (van den Doel 2006) to refer to any types of differences between L2 speech and its target. As these studies often show, some so-called "errors" were not considered

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

accented by native speakers. The term "error", therefore, does not necessarily imply "mistake". The current study adopts the term "L2 errors" used in previous studies to refer to differences between L2 speech and its target, while fully acknowledging that some "errors" could indeed be native-like.

5. The present study

The current study aims to investigate the relative importance of different segmental and syllable structure errors in foreign accent perception. 11 types of consonant errors, five types of vowel errors and two types of syllable structure errors from a large-scale speech archive are assembled to enable a more detailed comparison between different types of errors. This study will further explore whether consonant errors in general are more foreign accented than vowel errors or syllable structure errors. Human transcribers (i.e. professionally trained phoneticians) are recruited to identify the errors. Short audio snippets are used as stimuli without acoustic manipulation. That is, we have left nonnative prosody intact. Prosodic information is controlled for by calculating the DTW distance between nonnative F0 contours and native ones. Lexical stress is also controlled for by excluding any speech sample that involves the misplacement of lexical stress. The control stimuli in this study consisted of nonnative speech samples that have no segmental errors but may exhibit nonnative-like prosodic features. Native English speakers are recruited from Amazon Mechanical Turk to provide accentedness judgements on the stimuli. The results provide direct comparisons between consonant, vowel and syllable errors.

5.1. The experiment

5.1.1. Stimuli

The stimuli are audio speech samples extracted from the Speech Accent Archive (Weinberger 2016), which currently consists of 2,608 paragraph readings by speakers of various language backgrounds. All speakers were recorded reading the "Stella" passage at a university laboratory or their own residence (See Appendix A for the paragraph). 5 phrases were selected from the "Stella" passage for this experiment (Table 1). We opted to use General American English (GA) as the benchmark (See the "correct" condition in Table 1). Deviations from GA were considered "errors". 20 tokens of each phrase were chosen from the archive, five of which have only one consonant error, five of which have only one vowel error, five of which have only one syllable error, and another five were labeled as "correct", because they are representations of GA, yielding 100 audio snippets in total. The errors are all phonemic alternations. Sub-phonemic changes such as vowel lengthening are not included. The intensity of the 100 audio snippets were normalized to 75dB using PRAAT (Boersma and Weenink 2015).

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

The determination of errors was based on the IPA transcriptions available on the Speech Accent Archive. The transcriptions are relatively reliable because the transcribers were phonetically trained transcribers. The transcriptions were vetted by at least three transcribers before being uploaded online. A recent study recruited an additional 67 phonetically trained people to transcribe a selection of audio clips from the Speech Accent Archive (Weinberger et al. 2017). The results show that 72% of the 67 participants' transcriptions matched the vetted ones, which lend further support to the validity of the vetted transcriptions.

Table 1. Illustration of stimuli conditions

	consonant error	**vowel error**	**syllable error**	**correct**
please call	[bliz kʰɑl]	[pʰliz kʰol]	[pʰəliz kʰɑl]	[pʰliz kʰɑl]
ask her	[æsk hər]	[ɑsk həɹ]	[æs həɹ]	[æsk (h)əɹ]
six spoons	[sɪks spunʃ]	[siks spunz]	[sɪks əspunz]	[sɪks spunz]
five thick	[faɪv tɪk]	[fav θɪk]	[faɪvə θɪk]	[faɪv θɪk]
small plastic	[smɔl̩ pʰlæstɪk]	[smɔl pʰlæstik]	[smɔl pʰlæsɪk]	[smɔl pʰlæstɪk]

The vowel error condition consists of five types of vowel problems, namely vowel raising, vowel backing, vowel fronting, vowel lowering and vowel shortening. The vowel shortening error in this study refers specifically to the shortening from [aɪ] to [a] in word "five". There are two types of syllable errors, namely consonant deletion and vowel insertion. Consonant deletion refers to the deletion of a consonant at coda position (e.g. [kʰɑl] to [kʰɑ]) or within a consonant cluster (e.g. [pʰlæstɪk] to [pʰlæsɪk]). The deletion of /h/ in "ask her" was not treated as an error, because this type of /h/-dropping is also common in native speech (Milroy 1983). Vowel insertion involves prothesis (e.g. [spunz] to [əspunz]), anaptyxis (e.g. [pʰliz/ to /pʰəliz]) and paragoge (e.g. [æsk] to [æskə]). 11 types of consonant errors were included in this experiment, ranging from feature changing (e.g. [s] to [ʃ]) to consonant replacement (e.g. [θ] to [t]).

100 audio snippets were collected from 93 different non-native speakers. 52 different L1s were represented in the stimuli. To ensure that the stimuli are produced by nonnative English speakers, only late learners' speech samples are selected. 91 of the speakers started to learn English after age six. One speaker started to learn English at age five; one started at age four. The last two speakers were considered nonnative English speakers because they reported to have acquired English in academic settings, and their speech samples do show nonnative-like patterns. Previous studies show that native speakers are generally able to tell the native language of a nonnative speaker from his/her L2 speech (Kunath and Weinberger 2010). It is possible that one's bias for or against a certain language might affect one's accentedness judgement on speech samples produced by people of that language group. This potential confound is not accounted for in the current study. However, raters of this study might not be able

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

to successfully classify the L1 backgrounds of the speakers, because each stimulus is considerably short and contains only one segmental error.

Prosodic cues have been found to be important in identifying foreign accent (Munro and Derwing 2001; Kang Rubin and Pickering 2010; Morrill and Gao 2016). However, given the stimuli in this study are very short, it is unlikely that prosodic characteristics will be of much importance. Nevertheless, the current study controlled for prosody of the stimuli with a Dynamic Warping Method (DTW). The DTW is a non-linear algorithm that looks for the dissimilarity between two temporal sequences of data and calculates the costs to align one with the other (Rilliard Allauzen and de Mareüil 2011). It generates a DTW score that represents the dissimilarity between the two sets of data. The larger the DTW score, the more dissimilar the two sets of data are. In the current study, the DTW algorithm takes F0 values of a native speech sample [1]as the reference, and F0 values of a nonnative speech sample as the input. A DTW score, thus, represents the intonational dissimilarity between the native and nonnative speech samples. A native English speech sample was chosen from the Speech Accent Archive. The same five phrases as listed in Table 1 were extracted from the native speech sample as references. For each phrase, the F0 value at each millisecond was extracted in Praat with the auto-correlation algorithm (Boersma and Weenink 2015). Artifacts were removed by smoothing with a bandwidth of 5Hz. The F0 values were then converted to semitones relative to 1 Hz. The same process was carried out for all the 100 snippets produced by nonnative speakers. To allow for cross-speaker comparison, the semitones were then normalized for each speaker. The DTW function was then implemented in R with the DTW package (Giorgino 2009) to calculate the warping costs between each of the 100 snippets and its corresponding native speech sample. The DTW scores were then used in the analysis to account for prosodic information of the snippets.

5.1.2. Procedure

Participants (i.e. raters) listened to each of the 100 audio snippets and were then asked to judge the degree of the accent exhibited in the snippet on a 9-point Likert-like scale. Following the practice of similar studies (McCullough 2013; Huang and Jun 2015), only the endpoints of the scale were marked. A rating of one means the speaker has no foreign accent at all. A rating of nine means the speaker has a very strong foreign accent. To reduce the order-effect, the presentation of the stimuli was randomized. The 100 audio snippets were first divided into five blocks, each of which contains one token per condition per phrase, yielding 20 stimuli per block (five phrases x four conditions). The interface of the experiment provides a button and a 9-point rating scale. The stimulus is played once the participants hit the button, after which the rating scale will appear. Participants provide their accentedness judgement by choosing a number from one to nine on the rating scale, and then move on to the next trial.

[1] The native speech sample was provided by a 42-year-old male native GA English speaker from Pittsburgh, PA.

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

Unlike van den Doel (2006), raters in the current study were not required to identify or locate the error in each stimulus, because the error had already been pointed out by the vetted transcriptions. There are 100 trials in total. At the end of the experiment, the raters were asked to take a demographic survey, which collected information on the raters' age, gender, L1/L2, occupation, current residence and birth place. The maximum time allowed for completing this experiment was 30 minutes. Raters on average spent 12.34 minutes (SD=3.20) on the experiment. The experiment was programmed with HTML. Trial randomization was achieved via JavaScript.

5.1.3. Participants

Participants (i.e. raters) were 110 adult native English speakers recruited via Amazon Mechanical Turk (MTurk), a web-application that allows researchers to conduct survey-based experiments. Previous literature has shown that results of behavioral experiments conducted on MTurk were comparable to results of similar experiments conducted in lab settings (Sprouse 2010; Enochson and Culbertson 2015). Difallah, Filatova and Ipeirotis (2018) recently showed that there are about 2,000 participants being active on MTurk at any given time. 51% of them are female, 49% of them are male. About 75% of the participants are from the United States. Indian participants represent 16% of the population. The rest are from Canada, Great Britain, Philippines and Germany. Since the current study aims to investigate accentedness judgement of American English speakers. The experiment was made accessible only to people with a U.S. IP address. To increase the reliability of responses, the experiment required participants to have an approval rating of at least 95%. That is, the participants' previous work on MTurk has been approved at least 95% of the time. Of the 110 recruited participants, 62 were female, 46 were male, and another 2 participants did not report their gender. All of them reported to be born and currently residing in the United States. All of them reported that that they were native speakers of English. We therefore assume that the participants are native speakers of American English. All participants were paid $0.50 upon completion of the experiment. two of the participants reported having speech or hearing related disorders. Responses from these two participants were thus removed, yielding 108 participants in total. The age of participants ranged from 20 to 66. The mean age was 33.50 (SD=12.51).

5.2. Results

5.2.1. Segmental influences

The mean ratings across all 4 conditions (where each audio snippet was rated on a scale from 1 to 9) was 4.81 (SD =2.21). The larger the number, the more accented a snippet was judged. As expected, the participants assigned higher ratings for snippets with segmental and syllable structure errors (M=5.10, SD=2.15) than for snippets without segmental or syllable structure errors

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

(M=3.94, SD=2.16). Ratings for consonant errors (M=5.66, SD=2.04) are on average higher than ratings for syllable structure errors (M=4.96, SD=2.17), which is on average higher than vowel errors (M=4.69, SD=2.13). Figure 1 demonstrates the mean ratings of each condition, where the error bars represent the 95% confidence intervals.

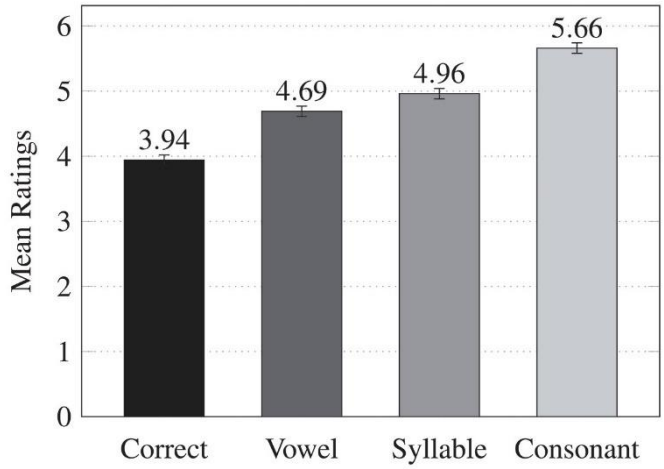

Figure 1. Mean ratings by error type on the scale from 1 to 9.

Linear mixed effects models were employed with the lme4 package in R (Bates et al., 2014) to investigate the segmental and syllable structure influences on foreign accent perception. The conditions were contrast coded to examine the effects of segmental errors (i.e. consonant vs. syllable, syllable vs. vowel, vowel vs. correct). To investigate accentedness ratings across the 100 trials, trial numbers were included as a fixed effect. The three condition contrasts and the interactions between trial numbers and each contrast were included as fixed effects. To control for prosody, the logarithmic DTW score (DTW henceforth) of each stimulus was included as another fixed effect. The two-way interactions between DTW and the contrasts, the two-way interaction between DTW and trial numbers, and the three-way interactions between DTW, the contrasts, and the trial numbers were also included as fixed effects. Raters were included as a random effect with the five phrases as its random slope. Stimuli were included as another random effect.

Model comparisons showed that the contribution of DTW to model fit is not significant (χ^2 = 2.64, p =0.104) and none of the interactions involving DTW achieved significant contribution to model fit, suggesting that the intonation of the audio snippets might not have affected accentedness ratings. The contrast between consonant and syllable errors significantly contributes to model fit (χ^2 = 18.83, p < .001), showing that stimuli with consonant errors were perceived as more accented than stimuli with syllable errors in general. The contrast between syllable

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

and vowel errors also significantly contributes to model fit ($\chi^2 = 17.26$, $p < .001$), showing that stimuli with syllable errors were perceived as more accented than stimuli with vowel errors. In addition, the contrast between stimuli with vowel errors and stimuli without segmental errors also contributes significantly to model fit ($\chi^2 = 13.32$, $p < .001$), showing that stimuli with vowel errors were perceived as more foreign-accented than stimuli without segmental errors.

These results suggest that all the 3 types of errors contributed to the perception of foreign-accent. However, stimuli with consonant errors were perceived as being more accented than the other two. Among the 3 types of errors, stimuli with vowel errors were perceived to be the least accented.

5.2.2. Phonological Environment

The analysis above showed that consonant errors are judged to be more accented in general. It might be too hasty to draw the conclusion that all consonant errors are more accented than the other two types of errors. As mentioned in previous sections, individual errors were placed in different phonological context, which might be of some importance in identifying errors and consequently influencing accentedness ratings. Vowel reduction, for example, might be considered an error if the vowel belongs to a stressed syllable. It might not be an error if the vowel is not stressed. In the case of monosyllabic function words or pronouns (e.g. "to", "her"), vowel reduction is often obligatory (Selkirk 2011); using full vowels could instead be nonnative-like. The analyses on individual segmental errors were thus carried out for each of the 5 contexts.

Liner mixed effects models were implemented to compare accentedness ratings on each individual error within a given context. For example, 6 types of stimuli were represented in context "please call", namely, vowel insertion (i.e. [pʰliz] to [pʰəliz]), VOT shortening (i.e. [pʰ,kʰ] to [p,k]), final devoicing (i.e. [pʰliz] to [pʰlis]), vowel raising (i.e. [kʰɑl] to [kʰol]), coda deletion (i.e. [kʰɑl] to [kʰɑ]) and stimuli with no segmental or syllable structure errors. The models took "stimuli type" as a fixed effect. Trial number, and the interaction between trial number and stimuli type were also included as fixed effects. Participants were included as a random effect with condition as its random slope. Stimuli were entered as a second random effect. To enable the comparison between ratings on any two given types of stimuli (e.g. VOT shortening vs. vowel raising), the condition variable was contrast coded (e.g. VOT shortening vs. vowel raising, VOT shortening vs. no error, etc.). The results of model comparisons for each context are summarized in the following tables, where ">>" shows the direction of significant differences. The types of errors on the left of ">>" received significant higher ratings than types on the right of ">>". Ratings for the types of errors on the same side of ">>" did not differ significantly from one another.

Table 2. Hierarchy of relative impacts of individual errors in the phrase "please call"

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

1.	Vowel insertion, VOT shortening >>Vowel raising >> no error
2.	Vowel insertion, VOT shortening, Final devoicing >> Coda deletion
3.	Final obstruent coda devoicing, Vowel raising >> Coda deletion, no error

Table 3. Hierarchy of relative impacts of individual errors in the phrase "ask her"

1.	Vowel insertion, vowel backing, r-trilling >> vowel raising, vowel fronting, vowel lowering, no error
2.	Vowel insertion, vowel backing, r-trilling, Coda /k/ deletion >> no error

Table 4. Hierarchy of relative impacts of errors in phrase "small plastic"

1.	[ɬ] retroflexing, [ɬ] to [r] >> [s] voicing, [t] deletion, VOT shortening, vowel lowering, vowel tensing, vowel raising, no errors,

Table 5. Hierarchy of relative impacts of errors in phrase "six spoons"

1.	[z] to [ʃ] >> [sp] to [sp^{h}], [n] deletion>> vowel laxing, no errors
2.	[z] to [ʃ] >> [n] deletion, vowel insertion, vowel tensing, vowel fronting,
3.	[sp] to [sp^{h}] >> vowel insertion, vowel tensing, vowel fronting, vowel laxing, no errors

Table 6. Types of stimuli for phrase "five thick"

1.	[θ] to [st] >> [θ] to [t], [θ] to [f], coda [v] deletion, no errors
2.	[θ] to [st] >> vowel shortening, vowel insertion, vowel tensing>>[θ] to [t] >> [θ] to [f]
3.	Vowel shortening, vowel insertion, vowel tensing >> [θ] to [f], coda [v] deletion, no errors

Several notable generalizations can be drawn from the observations on individual types of errors. First, stimuli with consonant errors do seem to be perceptually more accented than stimuli with vowel errors in all five contexts, but phonological environment affects the accentedness of some consonant errors. For example, accentedness of VOT shortening might be affected by the phonological context where the shortening happens as illustrated in Figure 2, where * marks the statistically significant difference between accentedness ratings of a given stimulus (e.g. [pl]) and its target form (e.g. [p^{h}l]).

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

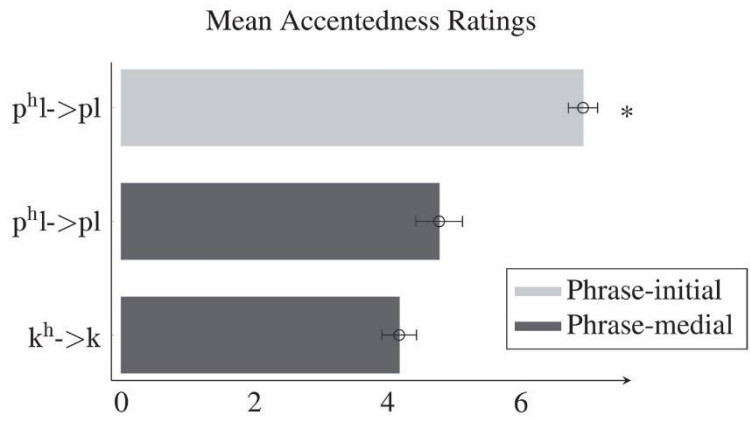

Figure 2. Mean Accentedness Ratings on VOT shortening

VOT shortening in "please call" (i.e. [pʰl] to [pl]) was assigned higher ratings than stimuli without segmental or structural errors (i.e. control stimuli). However, ratings on VOT shortening in "call" (i.e. [kʰ] to [k]) and in "small plastic" (i.e. [pʰl] to [pl]) was not significantly higher than stimuli without segmental or syllable structure errors. The reason might be that [pʰl] in "please call" is the initial segment of the whole utterance, which usually carries a longer VOT in native speech, as a result of prosodic domain-initial strengthening (Keating et al., 2004). Therefore, the shortening of [pʰl] in "please call" is not only a consonant alternation, but also defies rules in the prosodic domain, which might have led to higher accentedness ratings.

The effect of phonological environment was also observed on the accentedness of vowel errors. Figure 3 shows the accentedness ratings of vowel tensing (i.e. [ɪ] to [i]) in 3 environments. Only vowel tensing in "thick" was perceived as more accented than the control stimuli. The reason could be that sound sequence [θik] is not as common as [stik] or [sik] in English. In other words, English phonotactics could have affected accentedness perception. According to Vitevitch and Luce (2004)'s calculation, sound sequence [θik] has a 9% probability to occur in English context, while the probabilities for [tik] and [sik] to occur are 21% and 27%. The low phonotactic probability of [θik] could have given rise to the impression of foreignness.

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

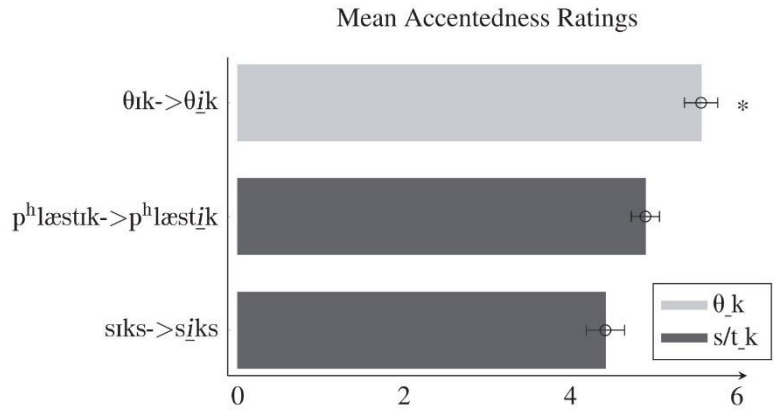

Figure 3: Mean Accentedness Ratings on Vowel Tensing

Similar effects of phonological environments have been found on syllable structure errors. Coda deletion is often allowed in native speech. In most contexts, coda deletion was indeed perceived to be less accented than other errors. Vowel insertion, on the other hand, is not normally allowed in native speech. Vowel insertion was indeed perceived as more accented than other types of errors. Interestingly, obstruent coda deletion in "ask her" (i.e. [æsk] to [æs]) was rated as accented, showing that native speakers of English are sensitive to the environment where coda deletion could happen.

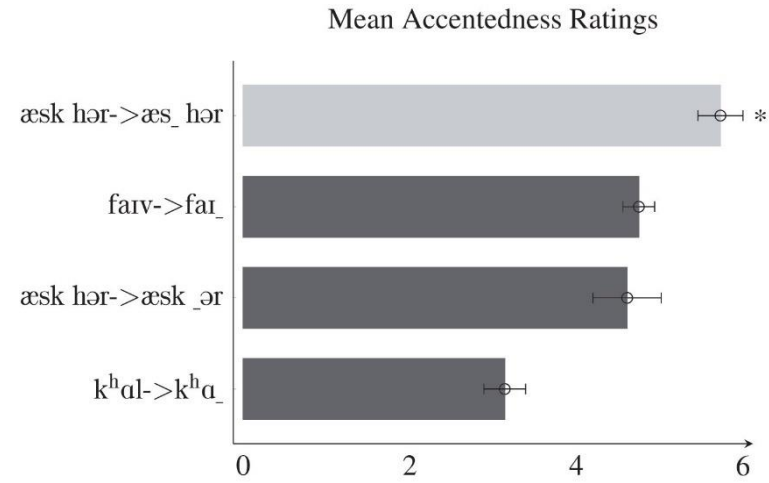

Figure 4: Accentedness Ratings on Coda Deletion

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

Stimuli with cluster internal epenthesis (i.e. [pʰl] to [pʰəl]) and stimuli with coda epenthesis (i.e. [æsk] to [æskə]) are more accented than stimuli with consonant errors in their respective contexts. Prothesis of s-clusters was not as accented as the other 2 types of epenthesis. The reason can be attributed to the perceptual similarities between the s-clusters with prothesis (i.e. [əsp]) and the unaffected one (i.e. [sp]) since [əsp] preserves the falling sonority profile of [sp] (Gouskova 2001). Word final epenthesis in "ask", on the other hand, changed the falling sonority (i.e. [sk]) to a rising one (i.e. [kə]). These results show that the effect of syllable errors on accentedness concerns both the specific type of errors and the environment the errors are in.

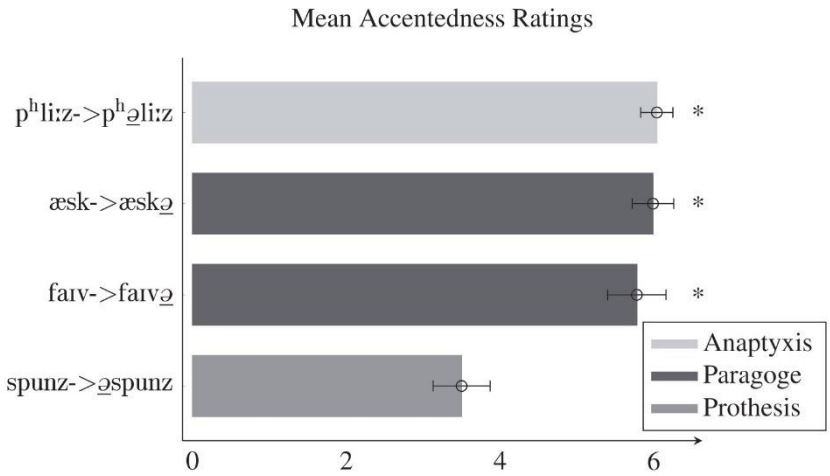

Figure 5: Accentedness Ratings on Vowel Epenthesis

5.3. Summary

The first part of the analyses focused on ratings on four types of stimuli, namely stimuli with consonant errors, vowel errors, syllable structure errors, and stimuli without errors. The results show that stimuli with consonant errors were rated as more accented than stimuli with vowel and syllable structure errors, which in turn were rated as more accented than stimuli without errors. Consonant errors were always rated higher than other types of error. Stimuli with no segmental or syllable structure errors always received lower accentedness ratings. Syllable structure and vowel errors were always rated lower than consonant errors and higher than stimuli without segmental errors. However, further analysis showed that accentedness ratings on the same type of errors may vary depending on the phonological context of the errors.

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

6. General Discussion and Conclusion

This study finds that General American English stimuli with consonant errors are, in general, judged to be more accented than stimuli with vowel or syllable structure errors. However, different consonant errors do not carry equal weight in foreign accent perception. As shown in the discussion above, the most accented stimuli were the ones with a nonnative sound (e.g. retroflex [ɻ], and trill [r]). In comparison, the alternations between native consonant phonemes were rated as relatively less accented (i.e. [θ] to [f]). The degree of acoustic distinctions between a substitute and its target sound might also attribute to the degree of foreign accent. For example, [θ] to [t] was rated as more accented than [θ] to [f]. The current study also shows that the effect of VOT shortening on foreign accent perception is much more prominent phrase-initially than phrase-medially. The reason for such phenomenon was attributed to native speakers' sensitivity to the existence/absence of the domain-initial strengthening effect on domain-initial aspirated plosives, which might also account for conflicting findings on the accentedness of VOT shortening/lengthening in previous literature (Gonzalez-Bueno 1997; Magen 1998; Riney and Takagi 1999). The effect of vowel errors on accentedness perception is not as clear as that of consonant errors. Several reasons might account for the mixed findings presented here. First, accentedness of some vowel errors was also affected by phonological environment. Second, vowel quality change might often be perceived as dialectal rather than foreign accented. Depending on the raters' own dialects and their exposure to other varieties of English, many types of "errors" could be native-like.

Although stimuli with syllable errors were in general less accented than stimuli with consonant errors, and more accented than stimuli without segmental errors, different types of syllable errors seem to affect accentedness rating differently. For example, stimuli with cluster internal epenthesis (i.e. [pʰl] to [pʰəl]) and stimuli with coda epenthesis (i.e. [æsk] to [æskə]) are more accented than stimuli with consonant errors in their respective contexts. Prothesis of s-clusters was not as accented as the other 2 types of epenthesis. The reason can be attributed to the perceptual similarities between the protheized s-cluster (i.e. [əsp]) and the original one (i.e. [sp]). Consonant deletion also exhibited different degree of impact on accentedness perception. Coda [v] deletion in "five thick" and coda [ɬ] deletion in "please call" did not contribute much to accentedness ratings. However, coda [k] deletion in "ask her" was considered accented. These results show that the accentedness of syllable errors associates with both the specific type of errors and the environment the errors are in.

Given these findings, it might be beneficial for pronunciation instructions to set priorities on correcting consonant errors and vowel epentheses, while taking into account phonological environment. As shown in the current study and some previous research (Munro and Derwing 2006; Wilson and Davidson 2013), phonological environment and English phonotactics do have an impact on accentedness perception. The reason for such observation could be further pursued

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

along two lines of research. First, the distribution of the substituted segment and its substitution (Munro and Derwing, 2006). For example, substituting /n/ for /l/ was found to be perceptually more accented than substituting /ð/ for /d/, because /n/ and /l/ participate more frequently in minimal pairs in word initial and final positions, while the /ð/-/d/ contrast distinguishes relatively few minimal pairs. A detailed survey on the distribution of the substituted segment and its substitutions is needed to further investigate which substitution is perceptually more accented. Second, the finding that consonant errors are more accented coincides with early research on speech perception, which often finds that consonant perception is categorical and vowel perception is relatively continuous. Sensitivity peaks were found at boundaries of consonant phonemes, but not always at boundaries of vowel phonemes (Fry et al. 1962; Pisoni 1973), which might imply that listeners are more sensitive to consonantal alternations than to vowel alternations. The claim that vowel perception is continuous was often disputed by later studies, which showed that listeners are sensitive to vowel boundaries (Repp and Crowder 1990; Iverson and Kuhl 2000). Without disputing the categorical nature of vowel perception, several recent studies have provided empirical evidence showing that vowel perception is relatively continuous, in comparison to consonant perception (Kronrod, Coppess and Feldman 2012; Altmann et al. 2014), lending support to Fry et. al (1962) and Pisoni's (1973) early findings. Results from the current study might potentially support the latter claim.

The current study focused on phonetic features of L2 speech. However, sociolinguistic elements such as one's own dialect and familiarity with L2 speech could potentially affect accentedness judgements. As shown in van den Doel (2006), British English speakers and American English speakers do not always agree on which L2 errors are accented. Raters of the current study are from 33 states within the continental United States. Some of them are from regions where the local dialects are quite different from GA (e.g. Texas, Georgia, New York etc.). Native speakers of southern American English or people who are familiar with southern American English might be more tolerant to monophthongizations such as [aɪ] to [a], because such sound change is similar to the phenomenon of off-glide deletion in many varieties of southern American English (Labov, Ash and Boberg 2005). Due to a large presence of Hispanic population in California, Arizona and Texas (Ennis, Ríos-Vargas and Albert 2011), raters from these regions are very likely to have been exposed to Spanish accented English, and thus could be more familiar with Spanish speakers' L2 English speech errors (e.g. s-cluster prothesis). Future research is needed to further investigate how one's familiarity with certain L2 errors affects accentedness perception. Due to the limited access to raters' personal information, the current study cannot warrant a detailed investigation on these extra-linguistic factors. However, since raters of the current study are spread out across the U.S., the current study is likely to have achieved its goal of drawing a general picture of American English speakers' perception of accented speech.

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

References

Altmann, Christian, Uesaki, Maiko, Ono, Matsuhashi, Masao, Mima, Tatsuya and Hidenao Fukuyama. 2014. Categorical speech perception during active discrimination of consonants and vowels. *Neuropsychologia*, 64, 13–23.

Boersma, Paul and David Weenink. 2015. Praat: Doing phonetics by computer [Computer program]. Version 5.3. 23. Available from: http://www.fon.hum.uva.nl/praat/ [Accessed: 24th January 2015]

Chan, Kit Ying, Hall,Michael, and Ashley Assgari. 2016. The role of vowel formant frequencies and duration in the perception of foreign accent. *Journal of Cognitive Psychology*, 29 (1), 1–12.

Couper, Graeme. 2006. The short and long-term effects of pronunciation instruction. *Prospect*, 21 (1), 46–66.

Demuth, Katherine, Culbertson, Jennifer and Jennifer Alter. 2006. Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, 49 (2), 137–173.

Difallah, Djellel, Filatova, Elena and Panos Ipeirotis. 2018. Demographics and Dynamics of Mechanical Turk Workers. In *Proceedings of the 18th ACM International Conference on Web Search and Data Mining (WSDM)*. 135–143.

van den Doel, Rias. 2006. *How friendly are the natives? An evaluation of native speaker judgements of foreign-accented British and American English*. PhD Dissertation. University of Utrecht, Utrecht: LOT.

Edwards, Jette G Hansen. 2011. Deletion of /t, d/ and the Acquisition of Linguistic Variation by Second Language Learners of English. *Language Learning*, 61 (4), 1256–1301.

Ennis, Sharon, Ríos-Vargas, Merarys and Nora G Albert. 2011. *The Hispanic population: 2010*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Enochson, Kelly and Jennifer Culbertson. 2015. Collecting psycholinguistic response time data using Amazon Mechanical Turk. *PloS one*, 10 (3), e0116946.

Flege, James and Wieke Eefting. 1987. Production and perception of English stops by native Spanish speakers. *Journal of phonetics*, 15, 67–83.

Fry, Dennis, Abramson, Arthur, Eimas, Peter and Alvin M Liberman. 1962. The identification and discrimination of synthetic vowels. *Language and speech*, 5 (4), 171–189.

Giorgino, Toni. 2009. Computing and visualizing dynamic time warping alignments in R: the dtw package. *Journal of statistical Software*, 31 (7), 1–24.

Gluszek, Agata and John Dovidio. 2010. Speaking with a nonnative accent: Perceptions of bias, communication difficulties, and belonging in the United States. *Journal of Language and Social Psychology*, 29 (2), 224–234.

Gonzalez-Bueno, Manuela. 1997. Voice-onset-time in the perception of foreign accent by native listeners of Spanish. *IRAL-International Review of Applied Linguistics in Language Teaching*, 35 (4), 251–268.

Gouskova, Maria. 2001. Falling sonority onsets, loanwords, and syllable contact. *CLS*, 37 (1), 175–185.

Grant, Linda and Donna Brinton. 2014. *Pronunciation myths: Applying second language research to classroom teaching*. Ann Arbor: University of Michigan Press.

Guy, Gregory R. 1991. Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change*, 3 (1), 1–22.

Hansen, Jette G. 2001. Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics*, 22 (3), 338–365.

Huang, Becky H and Sun-Ah Jun. 2015. Age matters, and so may raters. *Studies in Second Language Acquisition*, 37 (04), 623–650.

Iverson, Paul and Patricia K Kuhl. 2000. Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception & Psychophysics*, 62 (4), 874–886.

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

Kang, Okim, Rubin, Don and Lucy Pickering. 2010. Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, 94 (4), 554–566.

Kronrod, Yakov, Coppess, Emily and Naomi H Feldman. 2012. A unified model of categorical effects in consonant and vowel perception. In *Proceedings of the 34th annual conference of the cognitive science society*. 629–634.

Kunath, Stephen and Steven Weinberger. 2010. The wisdom of the crowd's ear: speech accent rating and annotation with Amazon Mechanical Turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*. Association for Computational Linguistics, 168–171.

Labov, William. 1997. Resyllabification. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 145–180.

Labov, William, Ash, Sharon and Charles Boberg. 2005. *The atlas of North American English: Phonetics, phonology and sound change*. New York and Berlin: Walter de Gruyter.

Magen, Harriet. 1998. The perception of foreign-accented speech. *Journal of phonetics*, 26 (4), 381–400.

Major, Roy. 1987. Phonological similarity, markedness, and rate of L2 acquisition. *Studies in Second Language Acquisition*, 9 (01), 63–82.

McCullough, Elizabeth. 2013. *Acoustic correlates of perceived foreign accent in non-native English*. PhD Dissertation. The Ohio State University, Ohio: Columbus.

Milroy, Jim. 1983. On the Sociolinguistic History of H-dropping in English. In Davenport, Michael, Hansen, Erik, and Hans Frede Nielsen (eds.), *Current topics in English historical linguistics*. Odense University Press, 37–53.

Morrill, Tuuli and Zhiyan Gao. 2016. Discriminability of non-native tonal contours in low-pass filtered speech. *The Journal of the Acoustical Society of America*, 139 (4), 2162–2163.

Munro, Murray J. 1993. Productions of English Vowels by Native Speakers of Arabic: Acoustic Measurements and Accentedness Ratings. *Language and Speech*, 36 (1), 39–66.

Munro, Murray and Tracey Derwing. 1998. The Effects of Speaking Rate on Listener Evaluations of Native and Foreign-Accented Speech. *Language Learning*, 48 (2), 159–182.

Munro, Murray and Tracey Derwing. 2001. Modeling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Studies in second language acquisition*, 23 (04), 451–468.

Munro, Murray and Tracey Derwing. 2006. The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34 (4), 520–531.

Nair, Ramesh, Krishnasamy, Rajasegaran and Geraldine De Mello. 2006. Rethinking the teaching of pronunciation in the ESL classroom. *The English Teacher*, (35), 27–40.

Pisoni, David. 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13 (2), 253–260.

Repp, Bruno and Robert Crowder. 1990. Stimulus order effects in vowel discrimination. *The Journal of the Acoustical Society of America*, 88 (5), 2080–2090.

Rilliard, Albert. Alexandre Allauzen. and Philippe Boula de Mareüil. 2011. Using Dynamic Time Warping to Compute Prosodic Similarity Measures. In *INTERSPEECH*. 2021–2024.

Riney, Timothy, Takada, Mari and Mitsuhiko Ota. 2000. Segmentals and global foreign accent: The Japanese flap in EFL. *Tesol Quarterly*, 34 (4), 711–737.

Riney, Timothy and Naoyuki Takagi. 1999. Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning*, 49 (2), 275–302.

Sato, Charlene. 1984. Phonological processes in second language acquisition: Another look at interlanguage syllable structure. *Language Learning*, 34 (4), 43–58.

Selkirk, Elisabeth. 2011. The syntax-phonology interface. In Goldsmith, John, Riggle, Jason, and Alan Yu (eds.), *The handbook of phonological theory*. Oxford: Blackwell, 435–483.

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

Sprouse, Jon. 2010. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods*, 43 (1), 155–167.

Tagliamonte, Sali and Rosalind Temple. 2005. New perspectives on an ol'variable:(t, d) in British English. *Language Variation and Change*, 17 (03), 281–302.

Thomson, Ron. 2014. Myth 6: Accent reduction and pronunciation instruction are the same thing. In Grant, Linda and Donna Brinton (eds.), *Pronunciation myths: Applying second language research to classroom teaching*. Ann Arbor: University of Michigan Press, 160–187.

Vitevitch, Michael and Paul Luce. 2004. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, 36 (3), 481–487.

Waniek-Klimczak, Ewa, Rojczyk, Arkadiusz and Andrzej Porzuczek. 2015. 'Polglish'in Polish Eyes: What English Studies Majors Think About Their Pronunciation in English. In *Teaching and Researching the Pronunciation of English*. Springer, 23–34.

Wayland, Ratree. 1997. Non-native Production of Thai: Acoustic Measurements and Accentedness Ratings. *Applied Linguistics*, 18 (3), 345–373.

Weinberger, Steven. 2016. *Speech accent archive* [online]. Geroge Mason University. Available from: http://accent.gmu.edu. [Accessed: 6th May 2016]

Weinberger, Steven, Nelson, Jill, Kunath, Stephen, Gao, Zhiyan, Luu, Vu and Thao vy Vo. 2017. *Transcribing non-native speech: the development of a crowdsourcing tool to evaluate perceptions of accented speech*. Presented at the 11th International Conference on Native and Non-native Accents of English, Łódź, Poland.

Wilson, Colin and Lisa Davidson. 2013. Bayesian analysis of non-native cluster production. In Kan, Seda, Moore-Cantwell, Claire, and Robert Staubs (eds.), *Proceedings of the Northeast linguistics society 40*. 265–276.

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC

Appendix

The “stella” passage:

Please call Stella, ask her to bring these things with her from the store: six spoons of fresh snow peas, five thick slabs of blue cheese and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

Bereitgestellt von George Mason University | Heruntergeladen 03.12.19 22:10 UTC