Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 2aSC: Linking Perception and Production (Poster Session)

2aSC45. Acoustic Analysis of Perceived Accentedness in Mandarin Speakers' Second Language production of Japanese

Peipei Wei* and Kaori Idemaru

***Corresponding author's address: University of Oregon, Eugene, Oregon 97405, peipei@uoregon.edu**

Many second language (L2) learners, particularly adult learners, retain foreign accent on their L2 production. What acoustic sources give rise to the perception of foreign accent? This study examines beginning and intermediate Chinese learners' productions of Japanese in terms of segmental and suprasegmental features and investigates relationship between acoustic characteristics of the L2 production and accentedness ratings provided by native Japanese listeners. Results of acoustic examination indicated that learners' production varied considerably from that of native speakers in terms of durational features of stops, spectral features of some vowels, pitch (F0) peak alignment and F0 contour. Multiple regression analysis identified that the second formant of /u/, F0 peak alignment and contour to be the strong predictors of perceived accent, accounting for nearly 90% of variance. These findings confirmed Flege's hypothesis of Speech Learning Model- L2 sounds that are similar to L1 sounds, while subphonemically distinct, seem to pose greater difficulty for acquisition than dissimilar sounds. Moreover, longer classroom experience was found to show limited effects in reducing perceived accent, with slightly greater effects on segmental than suprasegmental variables.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4800815]
Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

The phenomenon of foreign accent in second language (L2) learners' production has long attracted researchers' interests (Piske, MacKay & Flege 2001 for review). A broad range of speaker characteristics has been carefully examined, and many studies converge in demonstrating that the onset age of learning and length of residency in the community where the language is spoken exert a crucial influence on the development of a perceived foreign accent (e.g., Flege, 1988, 1995), whereas studies are inconclusive about the influence of other factors, such as gender, formal instruction, motivation and so forth (Suter, 1976; Thompson, 1991; Elliott, 1995; Flege et al. 1995, 1996; Moyer, 1999).

In addition to speaker characteristics, it is also important to understand what acoustic properties of non-native speech give rise to the perception of a foreign accent. A few published studies that have examined acoustic correlates of foreign accents (e.g., Wayland 1997 on American-accented Thai; Trofimovich & Baker 2006 for Korean accented English) reveal that there are certain salient acoustic features that seem to affect perception of a foreign accent more than others. Wayland's (1997) examined L2 Thai production by English speakers. The relationship between the acoustics of the L2 productions and accentedness rating by native Thai listeners revealed that suprasegmental features such as the peak, valley and range of fundamental frequency (F0) influenced perceived accent than segmental features including vowel formants. Trofimovich and Baker (2006) examined suprasegmental features in L2 English production by Korean learners and native English speakers' accentedness rating on the L2 productions. The findings indicated that duration of pause and speech rate have more influence on perceived accentedness than stress timing and peak alignment. These studies have renewed interest in the question of whether inaccurate pronunciation of consonants and vowels (i.e., segmentals) contributes to the sense of an accent (or perception of greater fluency) more than inaccurate execution of rhythm and intonation (i.e., suprasegmental), or vice versa.

An important consideration for research on this topic is the question of the generality or specificity of accent perception across different languages. It is not difficult to imagine that what comprises a sense of accent may be different from language to language, while also depending on the native language of the non-native speaker. At the same time, it is also conceivable that, given the common human cognitive auditory faculty, we all detect a foreign accent in similar ways cross-linguistically. We cannot begin to address this question until we have a database of studies examining multiple languages and language pairings between the target language and native language.

Thus the broad objective of the present study is to contribute to an understanding of the acoustic correlates of a foreign accent. To attain this goal, the present study examines the effect of both segmental features (i.e., consonants and vowels) and prosodic features (i.e., F0 alignment and contour) on perceived foreign accent, investigating the non-native production of Japanese sentences by Chinese learners (Production Study) and relating the L2 acoustics with the accentedness rating by native Japanese listeners (Rating Study).

PRODUCTION STUDY

Methods

Participants

Seventeen Chinese learners of Japanese, recruited at an American university, participated in the production study (11 female, 6 male, mean age=22.3). All were native Mandarin speakers and used it on a daily basis. These learners were divided into two groups based on the level of Japanese language course they were assigned to at the university. The first group of learners (2Y: n=10) was in the second year Japanese language course. Among the second group (4Y: n=7), two had taken and 5 were currently taking the fourth year Japanese language course. In addition, 10 native Japanese speakers (NS: 7 female, 3 male, mean age=21), recruited at the same university, participated as the control group. All Japanese native speakers came from areas where the standard Tokyo dialect is spoken.

Speech Materials and Procedure

Six short Japanese sentences were selected as stimulus sentences (Table 1) and embedded in a delayed repetition task (Flege et al, 1995; Trofimovich and Baker, 2006). To create stimuli for the task, two native Japanese speakers produced six question-response-repeated question sequences (See (1) below for an example). In this delayed repetition task, the participant heard a question spoken by a make speaker followed by a response spoken by a female speaker, and then the question repeated by the male speaker.

For example, the following sequence (1) was used to elicit the first stimulus sentence.

(1) Question (male): *Nihongo no kurasu wa dou desu ka? 'How is your Japanese class?'*
Response (female): *Tanoshii desu yo. 'It is fun.'*
Question (male): *Nihongo no kurasu wa dou desu ka? 'How is your Japanese class?'*

Each participant was presented with the six recorded sequences in a random order, and each sequence was presented 3 times consecutively. All the recordings were conducted individually in a sound booth using a flash digital recorder (Marantz PMD 670) and a standing microphone (SHURE Beta 87) at a sampling rate of 44 kHz and 16-bit quantization. The experimenter (the first author) was present during the recording to present the recorded sequences using E-Prime experiment software (Psychology Software Tools, Inc.) and to ensure target productions.

TABLE 1. List of target answers in Japanese with mora count and English translation.

Japanese sentence	Mora count	English translation
Tanoshii desu yo.	7	[It] is really fun.
Kuruma ga kaitai desu.	10	[I] want to buy a car.
Nihongo no jisho ga hoshii desu ne.	14	[I] really want to buy a Japanese dictionary.
Daigaku no tonari ni arimasu yo.	14	[It] is next to the university.
Ashita wa eiga o mitai desu ne.	14	[I] want to see a movie tomorrow.
Rokuji ni okimasu.	8	[I] get up at 6.

Acoustic measurements

The third production of each stimulus sentence from the speakers was analyzed. If there was significant disfluency or errors, the second production was used. For segmental variables, the duration of stop closures and VOTs as well as F1 and F2 values of all vowels including long vowels (/a/, /i/, /ii/, /u/, /e/, /ee/, /o/) were measured using Praat 5.2.18 (Boersma & Weenink, 2005). The standard method of segmentation was used to measure segmental durations (e.g., Idemaru & Guion, 2008) and the vowel formants were measured at the vowel mid-point and then Lobanov-normalized to correct for gender-related variation (Thomas & Kendall, 2007).

For suprasegmental variables, F0-peak hit, F0-peak distance and F0 contour were examined. If the F0 peak of the stimulus sentence in L2 productions occurred in the same syllable as any of native speakers' (NS) production for that sentence, it was considered as 1 hit, otherwise 0 hit. For each participant, the total possible number of hit was 6 for 6 stimulus sentences. F0-peak distance was measured in milliseconds as the distance from the vowel where the F0 peak actually occurred in L2's production to the vowel where F0 peak occurred in NS productions. F0-peak hit and distance were examined as pitch peak alignment. The contour score was intended to capture the tendency of pitch fluctuation in a sentence. The pitch contours of native speaker's production of each sentence were divided into contour segments and described by either "flat", "rising", "falling" or the combination of these three labels, yielding 27 segments in six sentences. If the description of the contour of L2 production matched any description of the contours of native speakers, "1" was scored; otherwise "0" was scored.

Results

The mean values of stop duration measurements for three groups are reported in Figure 1. A series of one-way ANOVAs with the group as a factor found group differences in all measures except for the closure of /t/ [d-closure: $F(2, 23)=6.40$, $p<.05$; d-VOT: $F(2, 24)=4.76$, $p<.05$; t-VOT: $F(2, 24)=3.97$, $p<.05$; g-closure: $F(2, 23)=4.47$, $p<.05$; g-VOT: $F(2, 23)=6.30$ $p<.05$; k-closure: $F(2, 24)=5.20$, $p<.05$; k-VOT: $F(2, 24)=20.48$, $p<.05$]. Tukey post hoc tests further revealed that 2Y group differed from NS in the durations of d-closure, d-VOT, t-VOT, k-closure, and k-VOT; whereas 4Y group differed from NS in d-VOT, g-closure, g-VOT, and k-VOT [p .$<.05$ for all]. 2Y and 4Y did not differ from each other in any of the durational measurements. These results suggest tendencies that L2 learners produced stops with longer VOTs than native speakers, while learners produced longer closure for voiced stops and shorter closure for voiceless stops than native speakers.

The mean values of normalized F1 and F2 formants for vowels are reported in Figure 2. A series of one-way ANOVAs with the group as a factor found group differences only in F1 of long vowel /ee/ and F2 of /u/ [/ee/ F1: $F(2, 24)=5.43$, $p<.05$; /u/ F2: $F(2, 24)=29.96$, $p<.05$]. Post hoc tests indicated that for F1 of /ee/, 2Y, but not 4Y, was different from NS; For F2 of /u/, both 2Y and 4Y were different from NS [$p<.05$ for all], while the two learner groups were not different from each other. These results indicate that the second year students' /ee/ was produced with more open jaw position than native speakers' /ee/, and both second and fourth year students produced /u/ further back than native speakers did.

The mean values of three suprasegmental variables are reported in Figure 3. One-way ANOVAs with the group as a factor indicated that F0-peak hits and F0-contour score differed among the groups [F0 peak hits: $F(2, 24)=19.66$, $p < .01$; F0 contour score: $F(2, 24)=27.68$, $p < .01$]. Post hoc tests showed that for both variables 2Y and 4Y were different from NS [$p < .05$ for both], while there was no difference between the two learner groups. These results indicate that learners placed the F0 peak locations and contours different from native speakers.

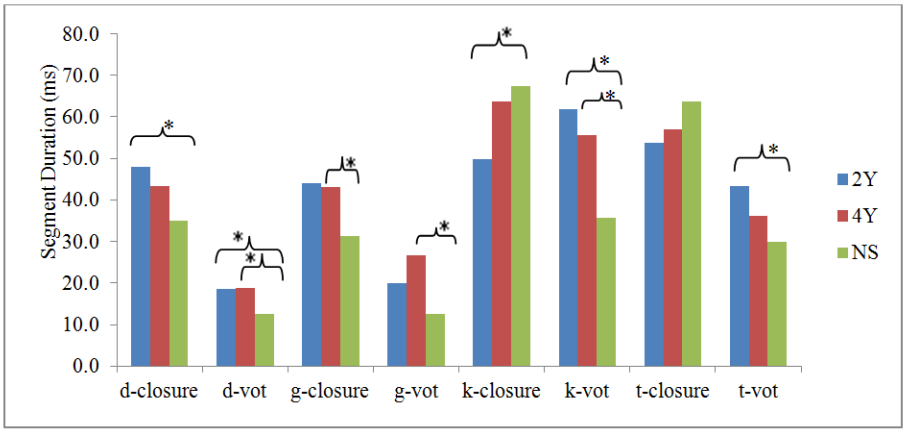

FIGURE 1. The mean durations of stop closure and VOT were calculated based on three groupings: 2nd year, 4th year and Native Speaker groups and the comparison of these three groups comparison was indicated in the graph below.

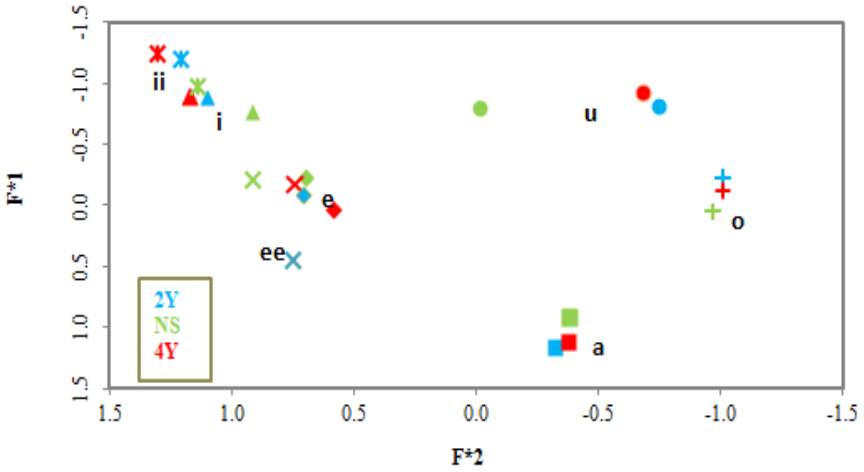

FIGURE 2. Mean value of Lobanov normalized F1&F2 of all vowels

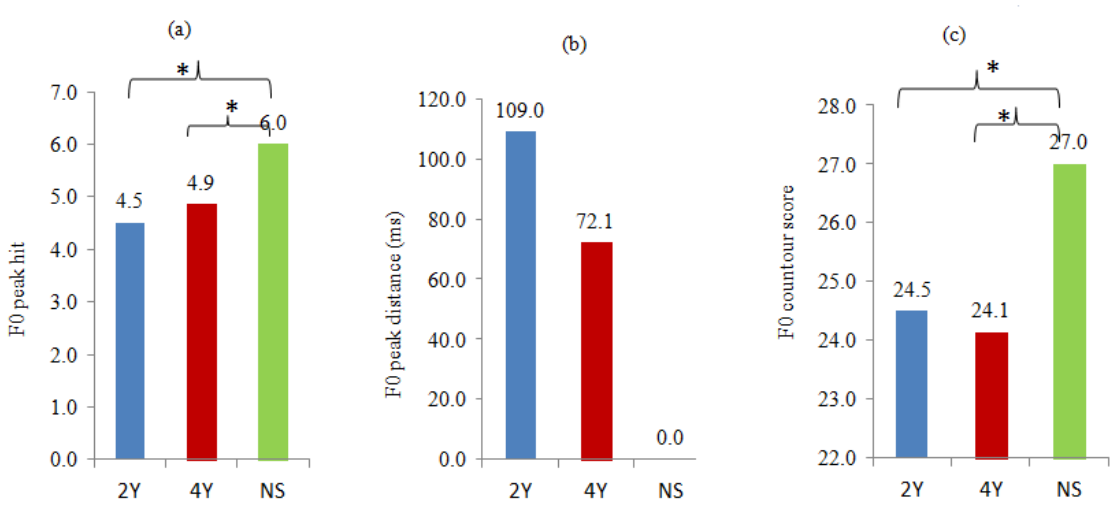

FIGURE 3. Mean value of three suprasegmental variables (F0 peak hit, F0 peak distance, F0 contour score) across three groups.

RATING STUDY

Methods

Participants

Ten native Japanese speakers were recruited at the same university where the production study was conducted (5 female, 6 male, Mean age = 23.3). They were all speakers of the standard Tokyo dialect, and none participated in the production study.

Stimuli

All utterances (the third or the second trial of each stimulus sentence) from the three groups (2Y, 4Y and NS) analyzed in the production study were used as stimulus. These stimuli were band-pass filtered at 700-1300 Hz and amplitude normalized to 75 dB to eliminate segmental information but retaining prosodic information (Trofimovich and Baker, 2006). This modification created two sets of stimuli, unfiltered and filtered. The unfiltered stimuli (the original utterances) were used to obtain accentedness rating for speech that includes all acoustic information, including segmental and prosodic. The filtered stimuli were used to obtain accentedness rating for speech that only retains prosodic information to focus on the influence of prosody on perceived accent. There were 324 stimuli in total (27 speakers × 6 sentences × 2 sets (filtered and unfiltered).

Procedures

The 10 native Japanese raters listened to utterances and rated each sentence on degree of foreign accent. Participants were tested individually in front of a computer monitor in a sound-attenuated booth wearing headphones. The perception experiment was delivered through E-Prime (Psychology Software Tools, Inc.). The experiment was broken up into two equal blocks of 162 trials: the first block presented randomized filtered stimuli, and the second block presented the randomized unfiltered stimuli. Before beginning the experiment, participants practiced rating four practice trials (2 filtered recordings and 2 unfiltered recordings), which were not included in analysis.

Participants rated the accentedness of each utterance using a horizontal visual analog scale displayed on a computer monitor (Urberg-Carlson, K., B. Munson, et al., 2009). For each utterance, the participant was prompted to rate each utterance for degree of foreign accent by sliding the bar in the middle of the scale using a computer mouse. The leftmost point on the bar indicated no accent, or speech that sounded like a native Japanese speaker. The rightmost point on the bar indicated the strongest accent possible. Participants were told they could drag the bar anywhere between those points. Participants also had the choice of listening to the utterance as many times as they wished, and could play the sound again by clicking anywhere on the screen.

Analyses

The mean rating scores were examined to compare accent ratings of 2Y, 4Y and NS speakers' productions. Furthermore, using logistic regression analysis, the accent rating scores were related to the acoustic measurements obtained in the production study to explore the acoustic source of perceived foreign accent in the learners' productions.

Results

First, the mean ratings for each speaker were averaged across six sentences and submitted to separate one-way ANOVAs for unfiltered and filtered conditions with group as a factor. These tests indicated a main effect of group for both unfiltered and filtered conditions [unfiltered productions: $F(2, 24)=121.91$, $p<.05$; filtered productions: $F(2, 24)=358.91$, $p<.05$]. Tukey post hoc tests revealed that both 2Y and 4Y received higher accentedness rating scores than NS for both filtered and unfiltered conditions. While the ratings of unfiltered sentences did not differ between the two learner groups, the ratings of filtered sentences were higher for 2Y than 4Y [$p < .01$ for all]. These results suggest that whereas native listeners detected foreign accent on the learners compared to the native speakers, they detected more accent in the second year learners than in the fourth year learners only when the utterances had prosodic information.

Next, in order to examine which acoustic properties contribute to the perception of foreign accent, the acoustic measurements were submitted to stepwise multiple regression. One analysis was run with both segmental measurements (VOT duration, closure duration, normalized vowel F1 and F2) and suprasegmental measurements (F0-peak hits, F0-peak distance, F0 contour scores) as predictors and accent rating scores as the dependent variable for unfiltered stimuli (Table 2). The best prediction model retained F2 in /u/ and F0-peak hits as significant predictors and explained 88% of the rating data. The beta coefficients of the model indicate that while both F2 in /u/ and F0-peak hits (-.643 and -.438 respectively) affected the perception of accent, F2 in /u/ did so more than the pitch factor.

Another analysis was run with the three suprasegmental measurements as predictors for the accent rating scores on filtered productions. This analysis allows us to focus on the influence of prosodic factors in the absence of segmental information. The results showed that the best model was the one with F0-peak hits and F0 contour scores as predictors (Table 3), accounting for 67% of the data. The beta coefficients of the model indicate that while both F0-peak hits and F0 contour scores (-.469 and -.409 respectively) affected the perception of accent, F0-peak hit did so more than the pitch factor.

TABLE 2. Multiple regression models on ratings of unfiltered productions

				Change Statistics				
Model	**R**	**R Square**	**Std. Error of the Estimate**	**R Square Change**	**F Change**	**df1**	**df2**	**Sig. F Change**
1	.855	.731	13.948	.731	65.243	1	24	.000*
2	.937	.878	9.589	.147	27.780	1	23	.000*

Model 1: u. F2
Model 2: u. F2 and F0-peak hits

TABLE 3. Multiple regression models on ratings of filtered productions

				Change Statistics				
Model	**R**	**R Square**	**Std. Error of the Estimate**	**R Square Change**	**F Change**	**df1**	**df2**	**Sig. F Change**
1	.790	.625	12.478	.625	41.613	1	25	.000*
2	.830	.689	11.593	.064	4.964	1	24	.000*

Model 1: F0-peak hits
Model 2: F0-peak hits and F0 contour scores

A comparison of the two analyses indicates that alignment of the F0 peak (F0-peak hit) is an important factor. When the segmental information is absent, F0-peak hit alone explains the 63% of the data (Model 1 in Table 3). However, when both segmental and suprasegmental information is present, a segmental property, the second formant of /u/ production, affects the perception of accent, followed by F0 peak alignment. The current results

suggest that both segmental and suprasegmental properties are important for describing perceived accent in the second language production of Japanese by Chinese speakers.

DISCUSSION AND CONCLUSION

The production study showed that the Chinese learners of Japanese produced some acoustic features differently from the native speakers. The learners' stops in terms of VOT and closure durations as well as some vowel formants (F1 in /ee/ and F2 in /u/) differed from those of native speakers in terms of measurement values. Interestingly, however, only F2 in /u/ among all vowel and consonant measurements was identified as a factor contributing to the perception of accent. As we saw in Figure 2, the vowel /u/ produced by native speakers is almost centralized. In comparison, learners' /u/ is produced further back. Mandarin Chinese, the native language of our Chinese learners, has the vowel /u/, and the Chinese /u/ is a back vowel unlike Japanese /u/. Therefore, it appears that these Chinese learners used the Chinese vowel /u/ for the production of Japanese /u/. This observation seems to provide evidence for Speech Learning Model—similar yet distinct sounds present in both L1 and L2 systems pose greater difficulty for acquisition than dissimilar sounds (Flege, 1995). In our case, the subtle subphonemic difference in Chinese /u/ and Japanese /u/ was perceived as a source of the learners' foreign accent on Japanese productions.

For suprasegmentals, the two properties, pitch peak alignment and pitch contour, distinguished the learner productions and native speaker productions, and both affected perceived foreign accent in Chinese learners production of Japanese. Thus, the more learners displaced pitch peak and the more leaners used wrong pitch contour, the heavier their accent was perceived. Although both Wayland (1997) and Trofimovich and Baker (2006) suggested primacy of prosodic factors influencing perceived foreign accent, our results suggest that both segmental and suprasegmental factors are important in describing Chinese learners' accent on their Japanese production. This finding emphasize the importance of examining various native-target language pairs to develop broad understand of the nature of foreign accent.

Furthermore, our study found little difference between learners in the second year and those in the forth year Japanese courses in terms of their acoustics and accent ratings. The only place we observed the difference between the two learner groups was accent ratings of utterances that only retained prosodic information (filtered utterances). Although further research is needed to explore this issue, the finding may suggest a possibility that suprasegmentals are more impervious to L2 experiences than segmental, a view consistent with Wayland (1997).

REFERENCE

Boersma, P., & Weenink, D. (2005). Praat: doing phonetics by computer (version 4.3. 14).[Computer program]. Retrieved May 26, 2005.

Elliott, A. R. (1995). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *The Modern Language Journal, 79*(4), 530-542.

Flege, J. E. (1988). The production and perception of foreign language speech sounds. *Human communication and its disorders: A review, 2*, 224-401.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 233-277.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *J. Acoust. Soc. Am, 97*(5), 3125-3134.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of/ɹ/and/l. *The Journal of the Acoustical Society of America, 99*, 1161.

Idemaru, K., & Guion, S. G. (2008). Acoustic covariants of length contrast in Japanese stops. *Journal of the International Phonetic Association, 38*(02), 167-186.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of phonetics, 29*(2), 191-215.

Suter, R. W. (1976). PREDICTORS OF PRONUNCIATION ACCURACY IN SECOND LANGUAGE LEARNING1. *Language learning, 26*(2), 233-253.

Thomas, E. R., & Kendall, T. (2007). NORM: The vowel normalization and plotting suite. http://ncslaap.lib.ncsu.edu/tools/norm/index.php

Thompson, I. (1991). Foreign Accents Revisited: The English Pronunciation of Russian Immigrants*. *Language learning, 41*(2), 177-204.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28*(01), 1-30.

Urberg-Carlson, K., Munson, B., & Kaiser, E. (2009). Gradient measures of children's speech production: Visual analog scale and equal appearing interval scale measures of fricative goodness. *The Journal of the Acoustical Society of America, 125*, 2529.

Wayland, R. (1997). Non-native production of Thai: Acoustic measurements and accentedness ratings. *Applied linguistics, 18*(3), 345-373.