TEMPORAL VARIABILITY IN STRONG VERSUS WEAK FOREIGN ACCENTED SPEECH

Chris Davis and Jeesun Kim
The MARCS Institute, Western Sydney University
chris.davis;j.kim@westernsydney.edu.au

ABSTRACT

This study examined temporal variability in the production of strongly or weakly accented L2 speech using a novel index of variation, the normalized Multi-Scale coefficient of variation (MSCVnorm) applied to vowel and consonant durations. Recordings of Korean and French talkers reading aloud a 69-word English passage were selected based on L1 English ratings of foreign accent to form two extreme groups; talkers with a strong or weak foreign accent. Recordings of Australian English talkers saying the same passage were used as comparison. The MSCVnorm scores for both Korean and French talkers differed as a function of foreign accent strength; with strongly accented speech more variable than weakly accented speech. MSCVnorm scores for the weakly accented speech did not differ from English native speech. These results suggest that a strong accent may be an indication that the timing of a talker's L2 productions may lack consistency (at least for read speech).

Keywords: Foreign accent; L2 speech; variability.

1. INTRODUCTION

When a person learns a second language (L2) in adulthood, the way that she/he speaks it may consistently differ from that of native speakers. For example, the way that certain vowels and consonants are pronounced may be different, and such differences will typically go beyond the variations that normally occur when speaking an L1. In such cases the person can be said to have a foreign-accent.

Research on foreign-accent has mainly focused on segmental phenomena [1]; although non-segmental contributions have also been investigated (e.g., [2]). For example, in an early study, Munro [2] low-pass filtered English sentences spoken by native English speakers and Mandarin-speaking learners of English to render them unintelligible (i.e., no discernible segmental information was available). Native English listeners were presented with these filtered speech stimuli and asked to rate the likelihood that each was spoken by a native English talker. Munro showed that listeners gave significantly higher ratings to the filtered L1 English utterances; a result that demonstrated that listeners were able to detect some aspect of foreign accent based only on rhythmic differences.

Although the Munro study demonstrated that listeners can pick up on rhythmic differences when given, what Munro described as the "musical" portion of speech, the issue remains whether rhythm affects accent judgments with intact speech. To examine this, Polyanskaya and colleagues [3] used resynthesized sentences in which native English segments were given durations based on the production of learners of English who had different levels of proficiency. It was found that ratings of perceived foreign accent were influenced by the level of L2 English proficiency; this was interpreted as showing that speech rhythm plays a role in the perception of foreign accent.

Data from another recent study, however, goes against the position that speech rhythm has a strong role in producing a foreign accent. That is, Sereno et al [4] conducted a similar study using synthesized speech and employed a fully factorial design, i.e., native segments were given non-native rhythm; non-native segments were given native rhythm, etc. Participants made accent judgments on these sentences and transcribed them to assess intelligibility. The results showed that resynthesizing with non-native rhythm did not influence accent ratings even though it did influence intelligibility.

In studies that have used filtered or resynthesized speech there remains a possibility that some effects may have arisen due to the unnaturalness of the tokens and doubts about whether findings would apply to unaltered speech stimuli. To address such concerns, studies have taken a different approach where natural speech can be used.

In this approach, the idea is to select L1 and L2 languages that have different rhythms and then to determine whether the rhythm of a talker's L2 is like that of their L1 [5]. Note that this approach does not in itself assess whether speech rhythm contributes to foreign accent, however, evidence for this can then be obtained by assessing the extent to which any L1 influence is associated with ratings of foreign accent.

To conduct such a study, it is necessary to use a rhythm metric [6] and to examine language learning where the L2 and L1 languages have different rhythms; however, selecting a metric and language pairs opens a range of issues. Consider the selection of a rhythm metric. Such metrics typically measure

the duration of speech segments and produce indices of variability. This focus on timing fits with the basic intuition that rhythm is a temporal phenomenon (although see [7]) yet the question of which properties to evaluate and how to best characterise the temporal dimension remains fraught.

A set of measures of variability in speech timing was developed by Ramus and colleagues [8]. Ramus et al took vowels and consonants as the units over which timing was considered and focused particularly on vowel duration. This latter focus was motivated by the extent to which vowels are reduced varies markedly across languages. The variability in vowel and consonant durations was indexed by taking the average standard deviation of their durations over a sentence. Dellwo & Wagner [9] noted that standard deviation measures were strongly influenced by speaking rate and proposed a modification that simply normalised the vowel and consonant measures by speaking rate.

The above measures characterise the global characteristics of speech timing, other measures summarise local characteristics, i.e., changes that occur over adjacent intervals, (e.g., Grabe and colleagues [10; 11]). For example, the measure that Low and colleagues developed, the Pairwise Variability Index (PVI), takes local variation into account by measuring the variability of all pairs of successive intervals (of a defined type) in a time series [11]. In the normalized PVI (nPVI), the difference in duration between pairs is calculated and then normalized by the mean duration of the pairs and multiplied by 100. This measure thus captures temporal variability in terms of a single measure that only uses adjacent interval (local) information.

Although the above local and the global indices of speech rhythm differ in the size of the window over which variability is considered, they only consider non-hierarchical relationships. For example, the nPVI provides a zeroth-order distributional statistical measure that does not measure higher-order relationships. However, a key property of the distribution of speech energy is that it fluctuates and correlates across different time-scales. In our view, then, it is important that a measure of variation in speech duration takes account of multiple scales.

In the current study, then we used a recently developed multi-scale index of variability (MSCVnorm) [12] to examine the durational variability of L2 and L1 speech segments (vowels and consonants). Using specially designed time-series, Abney and colleagues [12] have demonstrated that the MSCV measure is sensitive to short-term and long-term correlations. Further, they have shown that an MSCV analysis can differentiate between the durational properties of English and French read speech.

The other issue identified above concerned the selection of L1 and L2 languages that differ in speech rhythm. Recently, there has been a debate as to whether the classically proposed classification of languages based on rhythm is supported by the evidence [13], and even whether such a classification is appropriate [14]. There are issues in the categorical rhythm class hypothesis, whereby all languages must unambiguously fit within one of three rhythm classes (i.e., stress-, syllable- and mora-timed). However, for current purposes, the debate about whether languages can be typed by rhythm class is somewhat peripheral; as what matters is whether there is evidence that the selected languages differ in rhythm. To this end, we selected English as the L2 and both French and Korean as L1 languages. As mentioned above, Abney et al [12] have shown that English and French differ in the multi-scale variability of their segments. Likewise, English and French have been shown to differ using other rhythm metrics [11]. Korean was selected as an additional L1 because it has rhythm properties similar to French [15], and because unlike French and English that share an orthography, Korean and English do not, and differences in native orthography may play a role in the rhythm when reading aloud (as per the stimuli used in the current study, see below).

In summary, in the current study we examined the variability of L2 (English) versus L1 speech (either French or Korean) using the MSCV measure. We also used ratings of L2 foreign accent to select extreme groups (weak versus strong accent) to investigate whether the speech segments from a talker with a strong foreign accent had different durational variability than those of a talker with a weak accent.

2. METHOD

2.1. Speech stimuli

We used recordings of read-speech as this allowed the content to be controlled. Audio files were downloaded from [16]. Each recording consisted of a person reading the same 69-word passage that contained most of the consonants, vowels, and clusters of standard American English (see [16]). Recordings from 35 Korean speakers of L2 English (24 Female; Mean Age = 31.8 years; SD = 12.9) were used. These speakers begun learning English at various ages (Mean = 13 years; SD= 5.9) and had resided for various lengths of time in English speaking countries (Mean = 8.3 years; SD= 8.4). We also used a further set of English L2 recordings that consisted of 27 French speakers of L2 English (13

Female, Mean Age = 30.9 years; SD = 13.6). These speakers begun learning English at various ages (Mean = 11.6 years; SD= 2.7) and had resided for various lengths of time in English speaking countries (Mean = 5.8 years; SD= 11.8). In addition, recording of 32 native English speakers (14 Female; Mean Age = 29.4 years; SD = 10.1) were used for comparison.

2.2. Rating foreign accent

To quantify foreign accent, we had three L1 English raters listen to the L2 speech recordings and judge the extent of foreign accent on a 0-to-9 point scale (0 being no accent, 9 being strong accent). Ratings for the Korean and French recordings were conducted in separate sessions at least one week apart. We did not specify what was meant by foreign accent but left that up to each rater to decide (i.e., we did not specifically mention speech rhythm, etc.).

Using these ratings, we selected two extreme groups for the Korean and French talkers. For Korean, the Strong accent group (N = 12, 10 Females) had a mean rating of 7.2. These talkers had an average age of 44.2 years; had begun speaking English at an average age of 15.8 years and had an average period of residence in an English-speaking country of 11.4 years. The weak/no accent group (N = 10, six females) had a mean rating of 1.6. These talkers had an average age of 23.4 years; had begun speaking English at an average age of 11.1 years and had an average period of residence in an English-speaking country of 7.3 years. For the French talkers, the strong accent group (N = 5, 1 Female) had a mean rating of 6.3. These talkers had an average age of 27.6 years; had begun speaking English at an average age of 11.2 years and had an average period of residence in an English-speaking country of 2.1 years. The weak accent group (N = 7, 3 Females) had a mean rating of 1.4. These talkers had an average age of 30.6 years; had begun speaking English at an average age of 11.4 years and had an average period of residence in an English-speaking country of 3.4 years.

2.3. The MSCV analysis

The MSCV analysis was conducted using the Matlab scripts referenced by [12]. The MSCV provides a measure of the difference between a local coefficient of variation for a specific time window and the overall coefficient of variation for all the time samples. The MSCV is calculated by (1).

$$(1) \qquad MSCV(T) = \frac{\sigma(T)}{\mu(T)}$$

where σ is the standard deviation, and μ is the mean.

The normalised version of the MSCV is divided by the global coefficient of variation (CV) and divided by the number of window sizes (NT), as in (2) below.

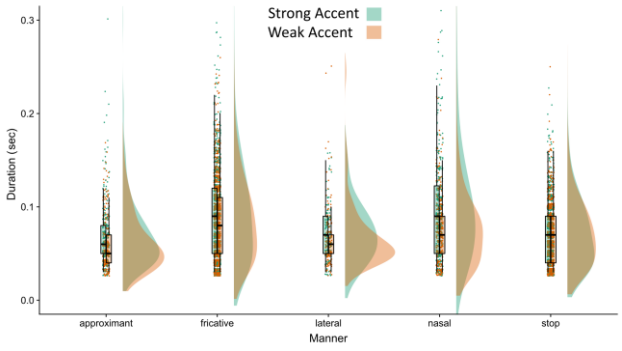
$$(2) \qquad MSCVnorm = \frac{\frac{\sum_{i=2}^{T} MSCV(T)}{CV}}{NT}$$

The MSCVnorm measure reflects the extent to which variation in a time series is heterogeneous over time scale. Series that are homogeneous across time scales tend to have a MSCVnorm value of about 1.0 (as determined in the simulation studies of [12]; values less than 1.0 indicate an increase in multiscale structure. In the current analysis, we report both the nPVI and MSCVnorm. In the MSCV analysis, T was set as a power of 2 and it ranged between 2 and L/2-1, where L was the number of measurements in the time series.

3. RESULTS

We first present descriptive summaries of vowels and consonants duration for the strong and weak foreign accent Korean and French L2 English recordings and the L1 English ones. Mean consonant durations varied considerably across the five accent groups. A generalised linear mixed-effect model (with talker as a random effect) was fitted to the duration values, using the LME4 R package, [17], and p values were obtained by the lmerTest package [18]. There was a significant difference in durations as a function of accent group, $F = 5.927$, $p < 0.001$; and Manner, $F = 60.098$, $p < 0.001$ and an interaction between these variables, $F = 4.557$, $p < 0.001$. Figure 1 presents the median durations as a function of the strength of foreign accent.

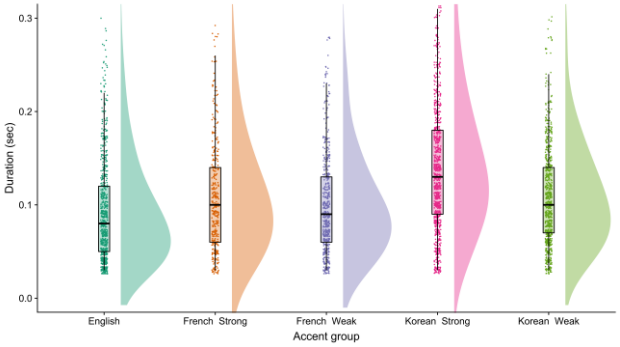
Figure 1. Median duration (sec) of consonants for Strongly accented English L2 (both Korean & French) versus Weakly accented English L2



There was an effect of strong versus weak accent, $F = 10.769$, $p < 0.001$; and effect of manner, $F = 56.702$, $p < 0.001$ and an interaction between these variables, $F = 8.829$, $p < 0.001$.

Figure 2 shows the median duration of the vowels as a function of accent group.

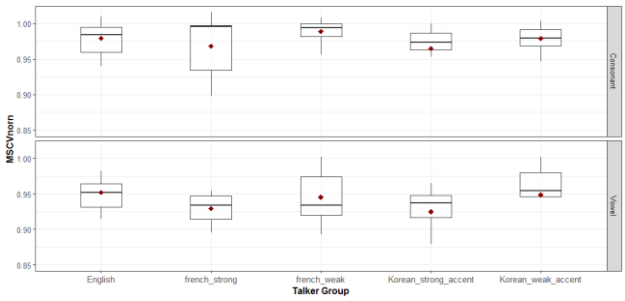
Figure 2. Median vowel duration (sec) for Korean and French L1 Strongly accented English L2; weak/no accented English L2 and English L1 speech**.**



The vowel durations also varied across the accent groups, with mean durations for the strongly accented talkers longer than those of the weak accented ones; this was particularly the case for the strongly accented Korean L2 English talkers. A LMM (with talker as a random factor) was fitted to the duration values. There was a significant effect of accent group, $F = 18.468$, $p < 0.001$ and an effect of accent strength, $F = 16.376$, $p < 0.001$.

Figure 3 shows the mean MSCVnorm scores as a function of accent group. As can be seen, the mean values (red dots) are lower for speech rated as having a strong accent compared to speech having a weak accent or L1 English. This indicates that the strongly accented speech had more heterogeneity of variance across the various window sizes.

Figure 3. Mean (red dot) MSCVnorm values for consonant (top) & vowel segments for each accent group.



An LMM (with talker as a random effect) was fitted to the MSCVnorm scores. The analysis contrasted the scores for consonants and vowels and strong and weak accent (English L1 scores were not included). There was a significant difference for consonants compared to vowels, $F = 25.31$, $p < 0.001$. The difference between the MSCVnorm scores for weak and strong accented L2 speech was significant (strong accent had a lower value), $F = 4.937$, $p = 0.032$. These two variables did not interact, $F = 0.101$, $p = 0.75$. There was no significant difference between the MSCVnorm values for the weakly accented and the L1 English speech, $F = 0.002$, $p = 0.97$.

As a comparison to the MSCV results, Pairwise variability indices (nPVI) were calculated. These scores were generally higher for the strong vs. weak accented speech. An LMM (with talker as a random effect) was fitted to the nPVI scores. The analysis contrasted the scores for consonants and vowels and strong and weak accent (English L1 scores were not included). There was no significant difference in the nPVI scores of consonants and vowels, $F = 0.3101$, $p = 0.581$. There was a significant difference between the scores for the weak and strong accented L2 speech (strong accented speech having a higher value), $F = 5.388$, $p < 0.05$. There was no interaction between these variables, $F = 0.125$, $p = 0.725$ and no difference between the nPVI values for the weakly accented and the L1 English speech, $F = 2.395$, $p = 0.134$.

4. DISCUSSION

Abney et al [12] showed that MSCVnorm scores were lower when calculated over English vowel durations than French ones (when nPVI was taken into account). This finding was interpreted as demonstrating the English (read) speech had more multiscale variability than French; and was thought to be related to English having more complex syllables. Other studies have shown that nPVI scores higher are typically higher for English than French utterances [11].

The current results for L2 speech showed differences between strong and weakly accented English for both the MSCVnorm and nPVI measures. However, the direction of these differences was opposite to that in the above studies of L1 speech.

The duration data from consonants and vowels (Figures 1 and 2) showed that the L2 productions that had high foreign accent ratings were longer (and more variable) than those rated as having a weak accent. This result is consistent with the strongly accented L2 productions having higher across-scale heterogeneity of variance and higher nPVI scores than the weakly accented ones and suggests that a strong foreign accent may be an indication that an L2 talker has not obtained consistency in their productions. In this regard, it may be that the idea of determining whether a foreign-accent is a legacy of a talker's L1 has distracted from viewing L2 speech production in terms of a complex motor task for which the skill of talkers varies.

5. REFERENCES

[1] Riney, T. J., Takada, M., & Ota, M. (2000). Segmentals and global foreign accent: The Japanese flap in EFL. Tesol Quarterly, 34(4), 711-737.

[2] Munro, M. J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. Studies in Second Language Acquisition, 17(1), 17-34.

[3] Polyanskaya, L., Ordin, M., & Busa, M. G. (2017). Relative salience of speech rhythm and speech rate on perceived foreign accent in a second language. Language and speech, 60(3), 333-355.

[4] Sereno, J., Lammers, L., & Jongman, A. (2016). The relative contribution of segments and intonation to the perception of foreign-accented speech. Applied Psycholinguistics, 37(2), 303-322.

[5] Kawase, S., Kim, J., & Davis, C. (2016). The Relative Contributions of Duration and Amplitude to the Perception of Japanese-accented English as a Function of L2 Experience. Proceedings of Speech Prosody, 746-750.

[6] Fuchs, R. (2015). Speech rhythm in varieties of English: Evidence from educated Indian English and British English. Chapter 3. Springer.

[7] Kohler, K. J. (2009). Rhythm in speech and language. Phonetica, 66(1-2), 29-45.

[8] Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. Cognition, 73, 265-292.

[9] Dellwo, V., & Wagner, P. (2003). Relations between language rhythm and speech rate. In Proceedings of the 15th international congress of phonetics sciences (pp. 471–474). Barcelona.

[10] Ling, L. E., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. Language and speech, 43(4), 377-401.

[11] Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. Papers in laboratory phonology, 7(515-546).

[12] Abney, D. H., Kello, C. T., & Balasubramaniam, R. (2017). Introduction and application of the multiscale coefficient of variation analysis. Behavior research methods, 49(5), 1571-1581.

[13] Arvaniti, A., & Rodriquez, T. (2013). The role of rhythm class, speaking rate, and F0 in language discrimination. Laboratory Phonology, 4(1), 7-38.

[14] Nolan, F., & Jeon, H. S. (2014). Speech rhythm: a metaphor?. Phil. Trans. R. Soc. B, 369(1658), 20130396.

[15] Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. Language and speech, 51(4), 343-359.

[16] Weinberger, S. (2015). Speech Accent Archive. George Mason University. Retrieved from http://accent.gmu.edu.

[17] Bates, D., Machler, M., Bolker, B. M., and Walker, S. C. (2015). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. 67(1), 1–48.

[18] Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. R package version, 2(0)-33.