Interspeech 2018
2-6 September 2018, Hyderabad

# Characterizing rhythm differences between strong and weak accented L2 speech

*Chris Davis*[1], *Jeesun Kim*[1]

[1]The MARCS Institute, Western Sydney University, Australia

chris.davis@westernsydney.edu.au, j.kim@westernsydney.edu.au

## Abstract

This study examined the rhythmic characteristics of accented L2 speech by using two relatively novel measures of prosodic rhythm: The S-AMPH measure, an index of the degree of synchrony between the stress and syllable amplitude modulation rates; and the Allan Factor measure, that determines the nested clustering of temporal events (in this case peaks in the amplitude envelope) over different timescales. An extreme-group design was used to select strong versus weak foreign accent recordings from a group of Korean and French L2 English talkers saying the same 69-word English passage. For the Korean talkers, both the S-AMPH and the Allan Factor measures differed as a function of the strength of foreign accent. This was not the case for the French talkers, where neither measure differed as a function of foreign accent strength. The difference in outcome between the Korean and French talkers suggests that the measures are not indexing some general property of L2 accent (e.g., production fluency) but rather that picking up some property specific to the strongly accented Korean talkers. We consider several options.

**Index Terms**: foreign accent, speech rhythm, second language, speech production

## 1. Introduction

When someone acquires a second language (L2) in adulthood, aspects of her/his speech may regularly deviate from that of native speakers. That is, these deviations go beyond the variations that normally occur when speaking an L1 and in such cases the person can be said to have a foreign-accent.

In the main, research on foreign-accent has focused on segmental phenomena [1]. However, although less well studied research has also investigated non-segmental contributions to foreign accent (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 segmental information was available). Native English listeners were asked to judge whether these filtered sentences were spoken by a foreign-accented talker or a native English one. Munro showed that listeners could correctly identify foreign-accented speech; a result that demonstrated that listeners are able to detect foreign accent based only on rhythmic differences.

In a more recent study, Polyanskaya and colleagues [3] examined the role of rhythm on the perception of foreign accent by using resynthesized sentences that used native English segments and the segment timing of English learners 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.

Interestingly, Sereno et al [4] conducted a similar study using synthesized speech but 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. So, based on these resynthesized speech studies, it would appear then that the issue of whether the degree of foreign accent relates to differences in speech rhythm is an open one.

Other studies of foreign accent have used natural speech rather than filtered or resynthesized versions. In this research, the idea is to determine whether the rhythm of a talker's L2 is different from that of a native talker, and if the L2 rhythm is like that of the L2 talker's L1 [5]. 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 (if they had the same rhythm it would not be possible to observe a difference). However, there is some evidence that an L2 learner's rhythm will differ from the target language rhythm even if the leaner's L1 and the L2 have similar rhythms [7]. One reason for this may be that the standard rhythm metrics (that typically focus on the durational properties of speech segments, e.g., the timing of vowels and consonants) can be influenced by a range of properties [8] and so do not unambiguously index rhythm.

In the current research we used two novel measures to investigate the rhythmic properties of L2 speech. Here, our interest was in investigating whether the rhythm of L2 talkers who had strong foreign accented speech was different from that of weak accented talkers. The measures we selected are defined at the level of the speech signal rather than at a more abstract level such as the timing of vowel and/or consonant segments (as used in the traditional metrics). We regard these non-abstract measures as having twin advantages, first, they avoid issues such as whether vowel or consonant timing is more important (and whether consonant sonority should be considered), and second because considerable effort is saved in not having to segment and label the speech signal.

The first measure (S-AMPH) indexes the grouping and timing of speech energy across the spectral domain [9]. The second measure (the Allan Factor) indexes the clustering of energy peaks within the temporal domain [10]. In what follows, we briefly outline these measures and review evidence that they are sensitive to prosodic speech properties.

The S-AMPH measure. Leong [9] developed the S-AMPH model as amplitude-based account of prosodic rhythm. The measure indexes the extent of amplitude modulation (AM) synchronization of different frequency bands. Evidence that this measure may be sensitive to speech rhythm comes from studies run by Leong and colleagues that examined the



10.21437/Interspeech.2018-1798

perception of iambic and trochaic rhythms and showed that the pattern of listener judgements was consistent with the view that speech rhythm associated with hierarchical AM patterns in the amplitude envelope [11].

Additionally, Leong and colleagues have used this model to demonstrate that the measure can differentiate between speech styles that use different rhythms [12]. That is, they tested mothers talking to their infants (infant directed speech, IDS) or to other adults (Adult directed speech, ADS). Their approach consisted of analysing the modulation spectrum of IDS and ADS based on three modulation rates and then they determined the degree of synchrony between pairs of these modulation rates. The rates chosen, approximate those of different types of speech cue. It was assumed that a rate of 12-40 Hz, captures properties at the phonemic scale; a rate of 2.5-12 Hz captures syllabic information and that the slowest rate, 0.9-2.5Hz, captures stress and phrasal groupings that contribute to the intonation contour of an utterance.

Consistent with the idea that IDS has different rhythmic properties from ADS, it was found that the phase synchrony of the syllable rate band (2.5 – 12 Hz) and the stress rate one (0.9-2.5 Hz) was greater for IDS than ADS for acoustic frequencies below 700 Hz.

The Allan Factor (AF) presents a different way of measuring rhythm. The Allan Factor is a measure of the coefficient of variation in timing of a point event across multiple timescales. That is, the AF is a statistical method that can distinguish a Poisson process (events occur unpredictably over time) and a process in which events occur non-randomly. Somewhat coincidentally, Falk and Kello [10] used this proposed measure of rhythm to examine differences in IDS and ADS speech styles. That is, they use the AF to examine the temporal distribution of peaks (events) in the speech amplitude envelope (as filtered into four frequency bands) for IDS and ADS. Like Leong and colleagues [11] they found that these speech styles differed, but only in Falk and Kello's case it was found that at multiple time-scales, IDS showed greater event clustering than ADS. Falk and Kello [10] suggested that this increased temporal clustering for IDS was due to increased durational variability and contrast of hierarchically nested linguistic units. Further, they proposed that such variability and contrast in IDS functions to make the speech more interesting and to enhance the infant's level of arousal.

So, what might we predict using these measures about L2 speech production? We presume that a strong accent indicates a talker is not able to consistently obtain the rhythm of the L2, thus would show larger durational variability in his/her L2 production than a talker who was perceived as having very little accent. If this were that case, we would predict higher AF values for the strong compared to the less accented L2 talker.

We are uncertain as to what to predict in terms of the S-AMPH measure. Taking the IDS results of Falk and Kello [10] along with those of Leong and colleagues [12], it can be suggested that higher durational variability in speech goes together with increased phase synchrony between the stress and syllable modulations. Thus, if we are predicting greater durational variability for strongly accented speech, then we should also predict it show greater phase synchrony between stress and syllable modulation rates. However, it may be that the increase in durational variability shown for IDs was due to some systematic underlying production difference which gave rise to increased phase synchrony, whereas any duration variability due to foreign accent may be due to more idiosyncratic factors that do not lead to any increase in phrase synchrony.

# 2. Method

### 2.1. Speech recordings

We used recordings of read-speech as this allowed the content to be controlled, and we contrasted participants whose L1 shared the L2 orthography (French and English) versus those where this was not the case (Korean and English) since L1/L2 script differences may play a role in the rhythm of reading aloud.

Audio files were downloaded from [13]. Each recording consisted of the person reading the same 69-word passage that contained most of the consonants, vowels, and clusters of standard American English (see [13]). 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 raters listen to the L2 speech recordings and judge the extent of foreign accent on a 0 - 9 point scale (0 being no accent, 9 being strong accent). 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 = 6) had a mean rating of 8.39 with no rater giving a score below 7. The weak/no accent group (N = 9) had a mean rating of 2 with no rater giving a score above 4. For the French talkers, the strong accent group (N = 5) had a mean rating of 7, with no rater giving a score below 6; and the weak accent group (N = 6) had a mean rating of 1 with no rater giving a score above 3.

### 2.3. Allan Factor analysis

The Allan Factor Analysis was conducted using a set of Matlab scripts available from [14]. The method detailed in [10] was followed. In brief, the speech recordings were filtered into four frequency bands and peaks in the Hilbert amplitude envelope of each band were calculated to produce a single time series (i.e., collapsed over the frequency bands). The distribution of these events over various time-scales was indexed by using Allan Factor (AF) functions to determine the degree of nested clustering of these peaks. For a formal description of the AF see [10]. In brief, AF variance is calculated by composing a signal into windows of size (T) and totaling the number of events (N) in each window (i). AF variance for timescale T, (AT), is simply the average squared difference of counts between adjacent windows divided by two times the mean count, as in (1).

$$A(T) = \frac{\left\langle (N_i - N_{i+1})^2 \right\rangle}{2\langle N \rangle} \quad (1)$$

Note that due to the shorter duration of the current recordings (~ 30s) only timescales under a few seconds could be calculated, since the largest AF timescale is 1/16th each recording's length.

### 2.4. S-AMPH analysis

The synchrony index between amplitude modulation in the three rate bands used by [12] was calculated using the S-AMPH model [9]. This index represents how in-phase are the modulation envelopes of the selected speech frequency rates (0 = no synchrony, 1 = perfect synchrony). Details of the signal processing steps involved are given in [12]. In brief, using adjacent finite impulse response filters, a waveform is band-pass filtered into five frequency bands and for each band three AM rates extracted from each down-sampled Hilbert envelope. A phase synchrony index (PSI) between pairs of AM rates is calculated according to (2)

$$PSI = \left| \left\langle e^{i(n\theta_1 - m\theta_2)} \right\rangle \right| \quad (2)$$

where $(n\theta_1 - m\theta_2)$ is the phase difference between the two AMs calculated by taking the distance between phase angles using circular distance (modulus $2\pi$).

## 3. Results

Figure 1 shows the AF functions for the Korean talkers who had strong or weak foreign accents and the English native talkers.

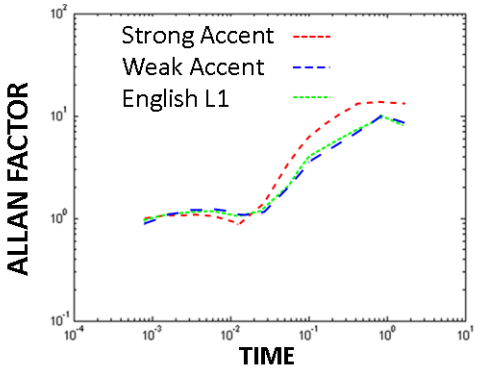


Figure 1: *Mean AF functions for Korean L1 Strong accented English L2; weak/no accent English L2 and English L1 speech. The timescale is in seconds and an Alan Factor of 1 indicates events occurred randomly.*

As can be seen, the strong accent curve started to diverge from the others at about 15 ms and continued to diverge from there. For the key contrast, the difference between the two accent types was tested by repeated measures ANOVA run on the factors accent type (Strong accent; Weak accent) and Time. There was a significant overall effect of accent type (Strong accent vs. weak accent), $F(1,14) = 12.89$, $p < 0.01$ and an interaction of this variable with Time, $F(11,154) = 6.43$, $p < 0.001$. There was a difference between the English L1 values and those of the strong accent, $F(1, 38) = 11.051$, $p < .001$ and no difference between the English L1 values and the weak accent, $F < 1$.

Figure 2 Shows the AF functions for the French talkers who had strong or weak foreign accents and the English native talkers as a comparison group.

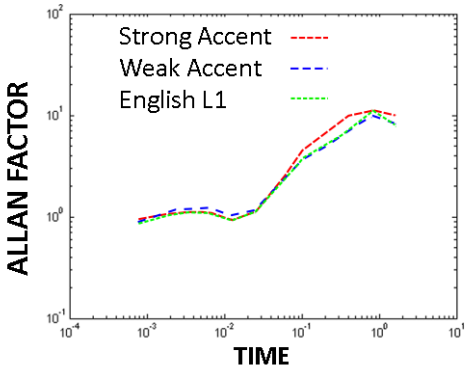


Figure 2: *Mean AF functions for French L1 Strong accented English L2; weak/no accent English L2 and English L1 speech.*

Unlike for the Korean talkers, where the AF differed between strong and weak accent, for the French talkers there was no difference. The ANOVA for strong accent vs. weak accent contrast was not significant, $F(1,8) = 2.24$, $p = 0.165$ and there was no interaction with accent type and Time, $F(11,88) = 1.56$, $p = 0.327$. Also, the omnibus comparison between all three language groups (English L1, French strong and French weak accent) was not significant, $F < 1$.

The S-AMPH PSI results for the Korean talkers are shown in Figure 3.

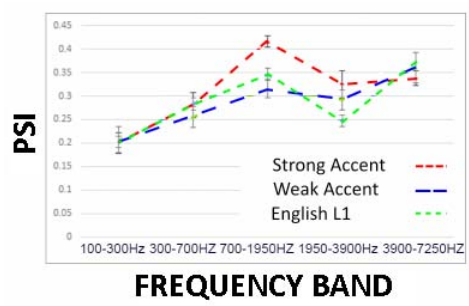


Figure 3: *Mean PSI (Phase Synchrony Index) values computed for the Stress-Syllable AM rates for the Korean talkers who had a Strong accent; weak/no accent as well as the English L1 talkers. A PSI of 0 = no synchrony; 1 = perfect synchrony.*

As can be seen in the figure, the PSI for the strong accent was greater than the weak accent for the intermediate frequency band, 700-1950 Hz. For the key contrast between the PSI values of the strong and weak accented English, a repeated measures ANOVA for this frequency band was used. This analysis indicated that the strong accent had a higher PSI value than the weak accent, $F(1,16) = 14.73$, $p = 0.002$.

Figure 4 shows the PSI scores for the French talkers. Here, the strong and weak accented talkers had very similar PSI scores across all the acoustic frequency bands.

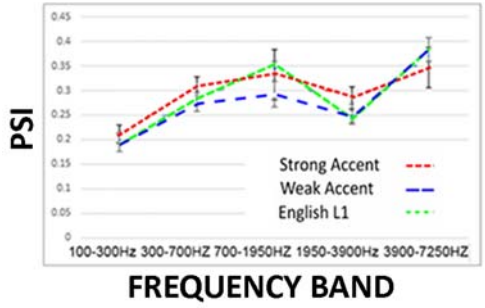


Figure 4: *Mean PSI (Phase Synchrony Index) values computed for the Stress-Syllable AM rates for the French talkers who had a Strong accent; weak/no accent as well as the English L1 talkers.*

The repeated measures ANOVA between the strong and weak accent PSI scores produced an F value that was < 1 (and the other contrast were also not significant).

## 4. Discussion

The results were mixed. For the Korean L2 talkers there was a difference between the strong and weak foreign accents for both the Allan Factor and S-AMPH measures. However, for the French L2 talkers, there was no difference between strong and weak foreign accented speech for either of these measures.

This difference between the results for the Korean and French talkers suggests that whatever the two measures are sensitive to, it is not likely to be some general property associated with having a strong foreign accent (e.g., a general lack of speech fluency) since the French strong accent group should have also produced differences.

What then is the difference between the Korean and French L2 talkers that produced these differences? The difference is unlikely to be linked to a putative rhythm class difference, since (as standardly conceived) both French and Korean are syllable-timed languages [15]. Moreover, the difference would need to set Korean apart from both French and English, since neither the Allan Factor or S-AMPH scores differed between them. One possibility might be an influence from the intonation system of Korean, as this differs from French and English in that for Korean, phrase beginnings are emphasised, whereas for French and English, phrase final lengthening occurs [16; 17].

Another difference between the French and Korean talkers is that for the Korean talkers, reading the English passage aloud required processing an L2 orthography. If it is assumed that the talkers with the most pronounced foreign accents were also those who had less fluency in reading, then this may help explain why these talkers produced a different speech rhythm. This effect may be quite subtle and could occur even though the talker may have a perfect declarative knowledge of English orthography. That is, the real-time pressure of reading aloud may have resulted in these participants adopting a speech style that was more fluent for high frequency shorter words, and slightly more laboured for longer lower frequency ones. This mixed speaking style may have led to an increase in the distributional variation of the timing of peaks in the amplitude envelope (leading to a greater AF), and possibly increased the synchrony between syllables and stress due to more easily read parts being given more prominence. This hypothesis could be tested by conducting the same measurements as above with talkers of another language that does not use the same orthography as English/French.

The null result for the strong versus weak accented French talkers demonstrates that the AF and S-AMPH measures are not sensitive to differences in the degree of foreign accented speech. Of course, the perception of a foreign accent could arise not due to a sense of an anomalous speech rhythm, but due to perception of segments deviating from language norms [4]. In such cases, it perhaps should not be expected that a measure of rhythm would be sensitive. In our view however, it seems unlikely that segmental and rhythmic differences would be dissociated so completely.

In summary then, both the AF and S-AMPH measures show promise in indexing aspects of speech production associated with a strong versus weak foreign accent. However, because this only occurred for the Korean talkers, at this stage the basis of the effect is not clear.

## 5. Acknowledgements

The authors acknowledge support from two ARC Discovery grants, DP130104447 and DP50104600.

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

[7] Ordin, M., & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. The Journal of the Acoustical Society of America, 138(2), 533–544.

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

[9] Leong, V. (2012). Prosodic rhythm in the speech amplitude envelope: Amplitude modulation phase hierarchies (AMPHs) and AMPH models. Doctoral dissertation, University of Cambridge.

[10] Falk, S., & Kello, C. T. (2017). Hierarchical organization in the temporal structure of infant-direct speech and song. *Cognition*, 163, 80-86.

[11] Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: converging perspectives? *Laboratory Phonology*, 4(1), 67-92.

[12] Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2017). The Temporal Modulation Structure of Infant-Directed Speech. *Open Mind*, 1(2), 79-90.

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

[14] http://cogmech.ucmerced.edu/downloads.html

[15] Kim, J., Davis, C., & Cutler, A. (2008). Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and speech*, 51(4), 343-359.
[16] Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, 29(2), 155-190.
[17] Jun, S. A. (2005). Prosody in sentence processing: Korean vs. English. *UCLA Working Papers in Phonetics*, 104, 26-45.