CONTRIBUTION OF TIMING PATTERNS INTO PERCEIVED FOREIGN ACCENT

Leona Polyanskaya[1], Mikhail Ordin[1], Christiane Ulbrich[2]

Universität Bielefeld, Fakultät für Linguistik und Literaturwissenschaft[1],

Phillips Universität, Marburg[2]

l.polyanskaya@uni-bielefeld.de

Abstract: We studied the influence of speech rate and speech rhythm into perceived foreign accent (FA). Although the effect of speech rate in foreign accent has been investigated quite extensively, much less is known about the contribution of speech rhythm into perceived accentedness. Speech rate and speech rhythm are interdependent, and thus they make an overlapping effect, however, speech rate makes a unique impact on FA. We were interested in separate, unique effect of speech rhythm and also in the relative contribution of rhythm and tempo into FA. Our study has shown that speech rate and speech rhythm both contribute into perception of accentedness. As the combined contribution of speech rate and rhythm is substantially greater than that of speech rhythm alone, we conclude that speech tempo and speech rhythm also make unique, separate and independent contributions into perceived FA. We also show that deviations in timing patterns are overridden by accented phonemic realizations on lower levels of proficiency in second language (L2), but become perceptually more salient for native speakers of the target language as the mastery in L2 increases.

1 Introduction

1.1 Foreign Accent

Linguistic community is still lacking universally accepted definition of foreign accent (FA). Gut defines FA as a set of differences on pronunciation from what is considered to a norm by native speakers of the target language [1]. However, the distinction between a native and a non-native speaker is not always clear. Besides, there are many varieties – social, regional, professional – within each widely used language, and what is considered to be a norm by one speaker will not always belong to a norm by another speaker from a different social or regional background [2]. That is why we will speak about the differences between first language (L1) and second language (L2) speakers. By the L1 speakers we will mean those who have acquired the target language in infancy and early childhood and were raised in the linguistic environment of the target language. L2 speakers have acquired the target language later in life, after their L1 was already learnt. Usually L1 speakers can easily detect L2 speakers whose speech exhibit influences from their L1, and these peculiarities of L1 pronunciation in L2 speech are perceived by monolingual L1 speakers of the target language and create the effect of FA.

Incorporation of elements of one language (L1) into another (L2) is the phenomenon known as linguistic interference. Phonetic interference can happen on segmental level (i.e., in realisation of sounds) and on prosodic level. An example of segmental interference are differences in voice onset time (VOT) between L1 and L2. A lot of studies have been done and the general result is that although L2 speakersare able to produce language-specific differences in voicing onset, but their phonetic implementation of VOT in L2 was nevertheless different from that exhibited by L1 speakers of the target language, the overview of the studies regading VOT and other segmental and prosodic differences between L1 and L2 can be found in [3]. An example of prosodic interference is the use of L1 intonation patterns in L2 speech, as documented in [4], [5] and many other studies.

1.2 Timing patterns in FA perception

When we speak about prosody, we speak about tonal characteristics (fluctuations of fundamental frequency (f0) and intonation, timing patterns (durations, speech rate and speech rhythm), intensity variations in spectrum and overall prominence, stress. In some studies researchers tried to separate the effect of prosody and segmental realizations on FA [6], [7]. In our study we will concentrate on investigating the effect of timing patterns on FA. In particular we will focus on speech rate (measured in syllables per second) and in speech rhythm (defined as durational variability of speech intervals, e.g. syllables, vocalic or consonantal intervals in speech stream, feet, etc.)

There have been some attempts to tackle the unique contribution of speech rate into FA. For example, L2 speech was proved to be generally slower than L1 speech. Moreover, speech rate correlates with L2 mastery [8]. Speech rate of L1 English speakers is 5 syllables per second, mean rate for advanced L2 English learners is 4,4 syllables per second, and mean rate for intermediate learners is 3,3 syllables per second [9].

However, too fast tempo also negatively influences intelligibility of L2 speech and enhances perceived accentedness. The optimal speech rate for L2 speech was estimated at 4,76 syllables per second, which is just slightly slower than L1 speech. Faster tempo (over 5 syllables per second) in L2 is perceived as more accented and less intelligible by L1 speakers. Very slow tempo was also rated lower [10].

The contribution of speech rhythm into foreign accent has received substantially less attention in empirical research. One of the studies focused on variations in durations of segments in L1 English and L2 English as spoken by Chinese learners [11]. Durations in L1 English were equalled to those in L2 English and durations in L2 English were equalled to those of L1 English. Then the sentences with warped durations were presented to L1 English listeners for accent and intelligibility rating. It was found that intelligibility in L1 English was deteriorated when L2 segmental durations were transferred from Chinese-accented sentences. L2 speech, on the other hand, improved in intelligibility when L1 durations were imposed. This shows that durational variability in segments does contribute to the perceived FA indeed.

1.3 Speech rhythm and rhythm metrics

Durational variability is captured by so-called rhythm metrics. The values of some of these metrics depend on speech tempo [12], [13]. For example, standard deviation of syllable durations depends on the overall duration of the syllables. That means, the higher the speech tempo, the shorter the syllables. The shorter the syllables, the smaller the standard deviation of syllabic duration. Other metrics are resistant to variations in speech tempo, for example, %V. Even when speech tempo increases and consequently the durations of vocalic and consonantal intervals decreases, the proportion of vocalic to consonantal durations remain relatively stable. nPVI and Varco metrics include normalization components, which supposedly makes the values of these metrics less dependent on the overall durations of the segments.

The Control and Compensation index (CCI) suggested in [14] measures a language-specific degree of segmental lengthening and shortening. Some languages allow lower degree of compression and maintain segmental durations in unstressed position more strictly than other languages. To estimate the measure of control over segmental durations regardless of stressed/unstressed position is calculated by the following formula:

$$CCI = \frac{100}{m-1} \sum_{k-1}^{m-1} \left| \frac{d_k}{n_k} - \frac{d_{k+1}}{n_{k+1}} \right|$$

where m is the number of intervals in an analysed utterance, d is the (in ms) of k^{th} interval (vocalic or consonantal), and n is the number of segments within the relevant interval. This metric is also robust to the variations in speech rate and can also be calculated on vocalic and consonantal intervals and on syllables.

People are sensitive to timing patterns captured by the rhythm metrics [15]. Besides, there are language-specific differences in these rhythmic patterns, especially between the languages belonging to different rhythm groups, e.g. between French and English, between Spanish and German or between Italian and Russian. Many studies investigated the differences between first and second languages in timing patterns that can be captured by rhythm metrics [16], [17], [18], [19], [20]. The general finding is that individuals cannot easily imitate language specific timing patterns in L2 speech production. Hence L2 speech with different rhythmic properties compared to an L1 might lead to FA in speech production which in turn would be perceivable by native speakers of the L2.

Metric	Description
%V	Percentage of vocalic intervals
ΔV	Standard deviation of vocalic intervals duration
ΔC	Standard deviation of consonantal intervals duration
ΔS	Standard deviation of syllables duration
nPVI-V	Normalized pairwise variability index for vocalic intervals, i.e. mean of differences between successive vocalic intervals divided by their sum and multiplied by 100.
nPVI-C	Normalized pairwise variability index for consonantal intervals divided by their sum and multiplied by 100.
nPVI-S	Normalized pairwise variability index for syllables intervals divided by their sum and multiplied by 100.
rPVI-V	Pairwise variability index for vocalic intervals, i.e. mean of differences between successive vocalic intervals.
rPVI-C	Pairwise variability index for consonantal intervals, i.e. mean of differences between successive consonantal intervals.
rPVI-S	Pairwise variability index for syllables, i.e. mean of differences between successive syllables.
VarcoV	Coefficient of variation of vocalic intervals, i.e. standard deviation divided by the mean.
VarcoC	Coefficient of variation of consonantal intervals, i.e. standard deviation divided by the mean.
VarcoS	Coefficient of variation of syllables durations, i.e. standard deviation divided by the mean.

Tabelle 1 – Most widely-used rhythm metrics

The aim of our study is to estimate the unique contribution of speech tempo and rhythmic patterns to perceived FA, that is why we need to separate the effect of speech rate and the effect of speech tempo. We wanted to separate speech rate and other durational cues which are captured by the rhythm metrics. We also had to eliminate the effect of other prosodic cues and phonemic realizations that influence FA by separating these two timing phenomena using acoustic manipulations.

2 Method

We recorded German learners of English on three different proficiency levels in their L2 mastery. All speakers grew up in or near the city of Bielefeld which is a region in North-Rhine Westfalia in Germany, where inhabitants' speech sounds very close to what is perceived as German standard pronunciation.

We ran a sentence elicitation task following the procedure described in [21]. The participants saw pictures, each picture came with a descriptive sentence. The participants were instructed to remember the sentence that accompanied each picture. After that the participants saw the picture without the sentence and had to remember and produce the sentence which came with the picture. This procedure helped us to avoid reading mode and made the speech a bit similar to spontaneous, however, it was highly controlled and allowed us to elicit the same sentence from different speakers, which is crucial for the research using rhythm metrics. The recordings were made in WAV PCM format at 44 kHz, 16 bit in mono.

We selected 15 sentences from out database we collected in Bielefeld. Each sentence was produced by 3 German learners of L2 English who differed in the proficiency level. The proficiency level was determined by 3 teachers of English of a foreign language at the language school, native speakers of English, and also by two phoneticians (they did not assess the sentences we used for the experiments, they assessed the interviews with the speakers). Thus we had 3 versions of 15 sentences, which made 45 sentences in total.

We segmented the sentences in Praat, and measured the duration of the phonemes. Then these phonemic durations were fed into speech synthesizer to produce 45 sentences (3 versions of 15 sentences). We also normalized intonation. We recorded a native English speaker producing the sentences, then we lifted the intonation contours for the sentences, and imposed this contour onto the stimuli during the speech synthesis process. In this way we got the sentences which did not differ in segmental realizations, nor in intonation or accentuation. Thus, the only things which differed between the produced stimuli were speech rhythm and speech tempo.

These stimuli were input into Praat. Each 3 versions for each of the sentence were equalized in duration. Stretching or compressing the sentence keeps the relative durations between the vowels and consonants intact (thus the speech rhythm, i.e. variability in duration of consonantal and vocalic intervals and syllables, does not change), but the tempo is affected by this manipulation. Consequently, we have the stimuli which differ in speech rhythm only, but not in speech rate.

We asked native English speakers to listen to the sentences and to evaluate the degree of foreign accent on each sentence on a 6-point scale, from 6 (native or native-like pronunciation) to 1 (strongest accent). Other FA degrees on the scale were verbalized as 5(mild accent), 4 (moderate accent), 3(rather strong accent), 2 (strong accent). The listener heard the sentence and then rated the FA degree for this sentence. If desired, he could listen to the sentence again (maximum three times), after that the rating was to be given anyway. After the sentence was evaluated, the following sentence was played for rating.

The listeners had to come for the experiment three times, with the interval of 3 weeks between the sessions. One time they listened to the original recorded sentences, then they listened to the stimuli which differed in speech rhythm AND tempo, and finally they listened to the stimuli which differed in speech rhythm only. The order of sessions was randomized for participants.

During each session every rater listened to the stimuli three times, in three blocks of 45 stimuli each. The stimuli were randomized within the blocks. The first 45 stimuli (first presentations of each sentence) were excluded from further analysis. During the first presentations the listeners got familiar with the range of accents they will have to evaluate and consequently formed the baseline, or the reference according to which each stimulus was further evaluated as more or less accented.

Analysis showed very high consistency in ratings between raters (estimated as Chronbach's alpha for all the raters) and within raters (estimated as Spearman correlations between the ratings given to the same stimulus by the same rater when the stimulus was presented second time and the third time, i.e. in the second and the third blocks). The averaged correlation (ρ) for the original sentences was .858 (ranging between .787 and .966); averaged ρ for the modified sentences in which only rhythmic patterns and speech tempo characteristics were preserved and segmental and other prosodic talker-specific characteristics were neutralized was .722 (ranging between .558 and .904); averaged ρ was .575 (ranging between .345 and . 893) for the sentences in which only rhythmic patterns were preserved. Cronbach's alpha was .974 for the original sentences, .937 for sentences with preserved timing patterns only, inclusive of both speech rhythm and speech rate; .904 for sentences which preserved rhythmic patterns only. These valued indicate high consistency between and high agreement within the raters, therefore we averaged the ratings given to each stimulus across all raters and across second and third presentations. Therefore we obtained one average rating for each stimulus in each session, i.e. 45 ratings per session.

3 Analysis and Discussion

A series of one-way ANOVAs was carried out for each of the metrics in order to see which ones differentiate between the proficiency levels of the L2 German speakers of English. In each ANOVA *proficiency level* was introduced as an independent factor with three levels (lower-intermediate learners, upper-intermediate learners and advanced learners), and one of the *rhythm metrics* was introduced in each of the ANOVAs as a dependent variable. All assumptions for ANOVA analysis have been met, including homogeneity of variance (checked by Levene's test), normality and equality of samples.

We have detected significant differences between the proficiency levels for the following rhythm metrics: nPVI-v, $F(2,42) = 6.317$, $p = .004$; Varco-V, $F(2,42) = 5.928$, $p = .005$; Varco-S, $F(2,42) = 3.38$, $p = .002$; CCI-V, $F(2,42) = 3.277$, $p = .048$; and n-PVI-S, $F(2,42) = 2.759$, $p = .03$. (the letter S means that the metric has been calculated on the durations of the syllables, the letter V stands for the durations of vocalic intervals and the letter C denoted the durations of consonantal intervals). Figure 1 displays the changes in the values of the rhythm metrics as a function of L2 mastery growth and the mean values of these metrics for each proficiency level. Pairwise comparisons revealed that these metrics significantly differ between sentences produced by lower-intermediate and intermediate learners, but do not differ between intermediate and advanced learners. The other metrics did not exhibit significant differences between the proficiency levels.

The overall direction of the rhythm metrics changes shows that variability in duration of vocalic intervals and syllables increases as L2 acquisition progresses. This change might indicate that the acquisition progresses in the direction from more syllable-timing towards more stress-timing.

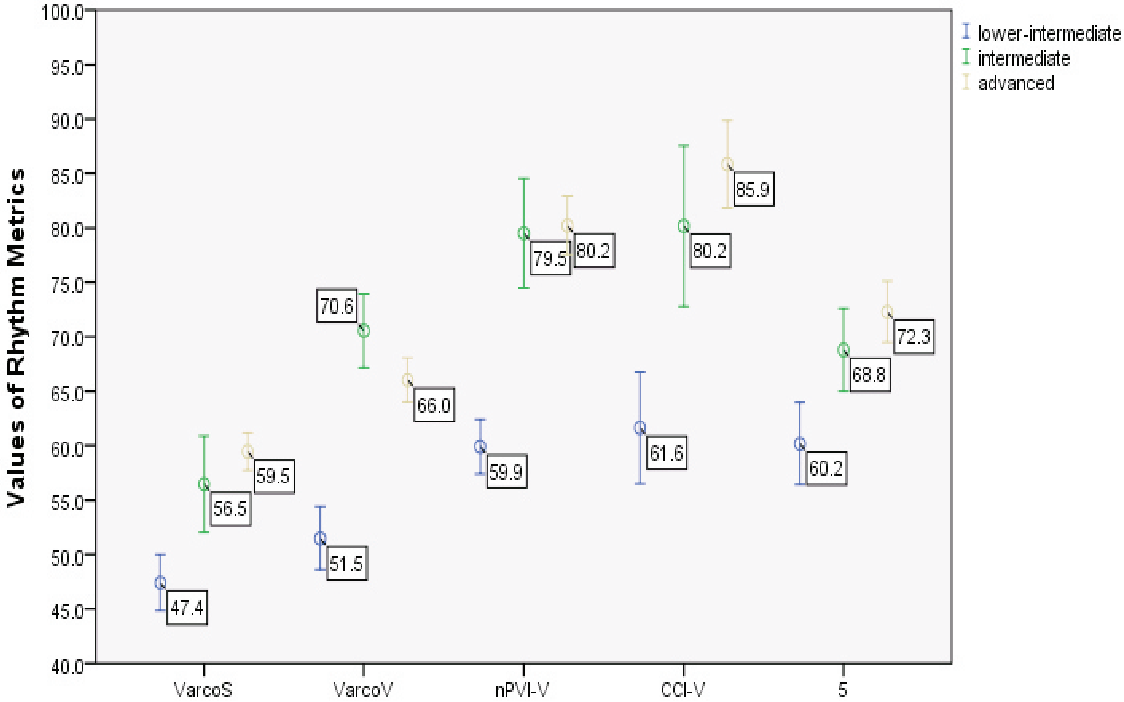

Figure 1. Changes in the values of the rhythm metrics as a function of L2 proficiency grows.

In order to investigate the influence of the L2 learners' proficiency level on FA ratings we carried out ANOVA analyses separately for three different types of stimuli: 1) for the original sentences in which three versions of the same sentence as produced by L2 learners on different proficiency levels differed in segmental realizations and in prosody including intonation, timing, stress, etc. 2) for the set of stimuli in which three versions of the same sentence differed only in speech rte and speech tempo; 3) for the set of stimuli in which three different versions of the same sentence differed only in rhythmic patterns as captured by tempo-normalized rhythm metrics.

The results showed that the proficiency level impacts on the FA ratings given to the original sentences, λ=.086, F(2,28)=148.21, p<.0005. Sentences produced by higher-level L2 speakers are evaluated as less accented than those produced by lower0level learners. The differences between ratings given to lower-intermediate and upper-intermediate and between upper-intermediate and advanced learners are significant, all contrasts at p<.0005 (Figure 2).

For the modified sentences which preserved rhythm and tempo characteristics but neutralized all idiosyncrasies in phonemic realizations and in intonation, proficiency level also influences FA ratings, λ=.218, F(2,28)=50.202, p<.0005. The contrast between ratings given to the sentences produced by lower-intermediate and intermediate L2 learners was significant at p<.0005, and the contrast in ratings given to the sentences by intermediate and advanced L2 learners was significant at p=.007 (Figure 2). As with the case with the original sentences, stimuli from L2 speakers on higher proficiency levels were rated as less accented compared the stimuli from lower-proficiency learners. This reveals that the information contained in the timing patterns only (tempo and rhythm together) is sufficient to make reliable judgement regarding the degree of FA accent.

For the modified sentences that preserved rhythmic patterns but neutralized differences in tempo, proficiency level also determines the FA ratings. However, the contrast between the ratings given to the stimuli from lower-intermediate and upper-intermediate speakers was significant at p=.001, whereas the contrast in ratings of intermediate and advanced L2 learners' productions was not significant, p=.602 (Figure 2). This can be explained by the fact that the rhythm metrics significantly differ between the lower-intermediate and upper-intermediate speakers' productions, but not between upper-intermediate and advanced learners' productions (Figure 1).

Figure 2 also shows that the sentences produced by lower-level L2 learners received lower scores in the original sentences compared to the modified sentences. However, modified sentences produced by advanced L2 speakers received lower scores. This might indicate that on lower level the influence of phonemic realizations, i.e. segmental characteristics on the raters' judgement is greater than on the higher levels. In other words, timing patterns are overridden by foreign accented segmental characteristics in L2 speech if the speakers is on lower proficiency level, and with higher L2 mastery, timing patterns in L2 productions become more perceptually salient for the L1 speakers of the target language.

Based on our results we also conclude that timing patterns do make a unique contribution in to the perceived foreign accent. Contribution of speech rate together with speech rhythm is much higher than that of speech rhythm only. Thus we can conclude that although tempo and rhythm make overlapping contribution into FA, they also make unique, separate contributions.

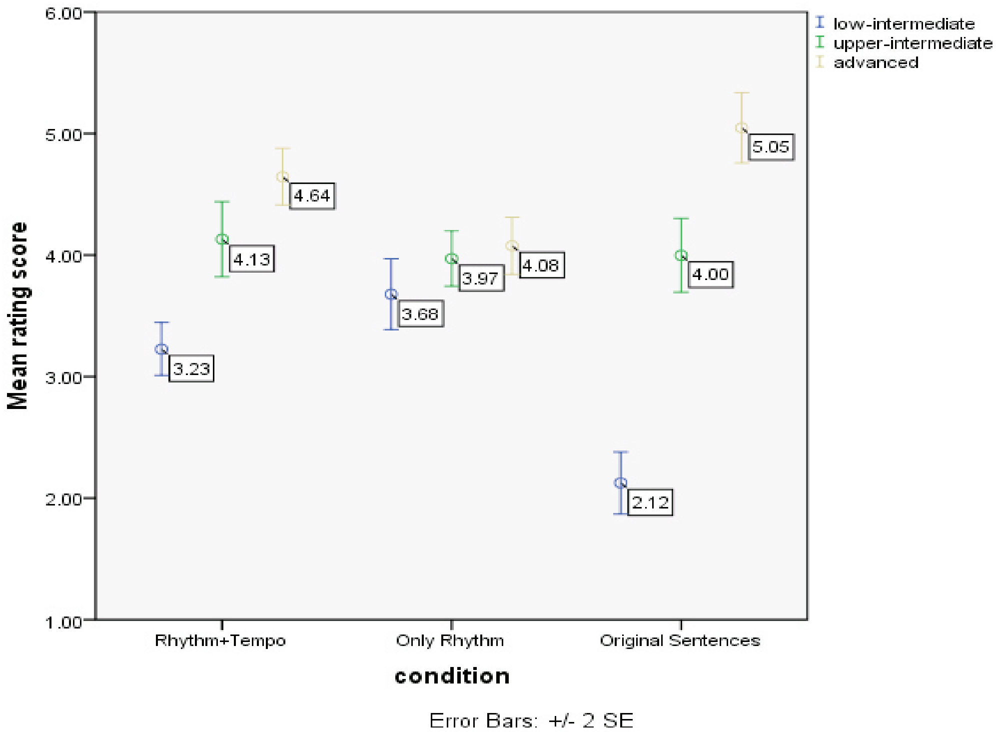

Figure 2. Averaged FA rating scores on the original and modified sentences.

References

[1] Gut, U. (2007). Foreign Accent. In C. Muller (Ed.) Speaker Classification I: Fundamentals, Features, and Methods. Lecture Notes in Computer Science 4343, 75-87.

[2] Labov, W. (1973). Sociolinguistic Patterns (Conduct and Communication). University of Pennsylvania Press.

[3] Mennen, I. (2011). Speech production in simultaneous and sequential bilinguals. In P. Howell and J. van Borsel (Eds.) *Multilingual aspects of fluency disorders*. Bristol: Multilingual Matters, pp. 24-42.

[4] Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. Journal of Phonetics, 32, 543-563.

[5] Ulbrich, C. (in press). German pitches in English: Production and perception of cross-varietal differences in L2. Bilingualism: Language and Cognition
[6] Jilka, M. (2000). The Contribution of Intonation to the Perception of foreign Accent. Doctoral Dissertation, Arbeiten des Instituts für Maschinelle Sprachverarbeitung (AIMS) 6(3), University of Stuttgart.
[7] Boula de Mareüil, P., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent, Phonetica, 63, 247–267.
[8] Munro, M., Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. Language Learning 48, 159–182.
[9] Anderson-Hsieh, J. & Venkatagiri, H. (1994). Syllable duration and pausing in the speech of intermediate and high proficiency Chinese ESL speakers. TESOL Quarterly 28, 807-812.
[10] Munro, M.J., Derwing, T.M. (2001). Modeling Perceptions of the accentedness and comprehensibility of L2 speech. The role of speaking rate. Studies in Second Language Acquisition 23, 451-468.
[11] Tajima, K., Port, R., Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. Journal of Phonetics 25, 1-24.
[12] Dellwo, V. (2006) Rhythm and speech rate: a variation coefficient for deltaC. In Karnowski P. & Szigeti, I. (ed.) *Language and language processing*. Frankfurt am Main: Peter Lang, 231-241.
[13] Dellwo, V. & Wagner, P. (2003) Relations between language rhythm and speech rate. *Proc. of the International Congress of Phonetic Science*, Barcelona, 471-474.
[14] Bertinetto, P. M. & Bertini, C. (2008) On modeling the rhythm of natural languages. Proc. of the 4th International Conference on Speech Prosody, Campinas 2008, 427-430.
[15] Ramus, F., Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. Journal of the Acoustical Society of America 105 (1), 512-521.
[16] Diez, F., Dellwo, V., & Gavalda, N. (2008) The development of measurable speech rhythm during second language acquisition. Journal of Acoustical Society of America 123(5), 3886-3886.
[17] Grenon, I., White, L. (2008). Acquiring rhythm: A comparison of L1 and L2 speakers of Canadian English and Japanese. In Proceedings of the 32nd Boston University Conference on Language Development, 155-166.
[18] Shport, I. (2008). Acquisition of rhythm: Evidence from spontaneous L2 speech. Journal of Acoustical Society of America 123(5), 3328-3328.
[19] Mok, P. (2011). The acquisition of speech rhythm by three-year-old bilingual and monolingual children: Cantonese and English. Bilingualism: Language and Cognition, 14(4), 458-472.
[20] White, L. & Mattys, S. (2007). Calibrating rhythm: First language and second language studies. Journal of Phonetics, 35(4), 501-522.
[21] Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual Spanish- and English-speaking four-and five-year-old children. Journal of Speech, Language, and Hearing Research 50, 999-1014.