# COMPUTATION OF L2 SPEECH RHYTHM BASED ON DURATION AND FUNDAMENTAL FREQUENCY

*Elisa Pellegrino, Lei He and Volker Dellwo*

*University of Zürich*
*elisa.pellegrino@uzh.ch*

**Abstract:** Rhythmic characteristics of speech vary between native and non-native speakers. Studies comparing the rhythmic properties of L1 and L2 speech based on rhythm metrics have shown that this relationship is far from straightforward. It seems evidently the case that the difference between native and non-native speech is a complex interaction of a variety of rhythmic cues (duration, F0 and intensity). In this study we extended the durational domain by F0 and tested whether metrics combining duration and F0 (henceforth combined measures) could better account for the rhythmic differences between L1 and L2 speech. To test this, 5 native Mandarin speakers and 5 Italian learners of Mandarin recorded The North Wind and the Sun in Mandarin. Besides, each Italian speaker also recorded the same text in Italian. Each sentence in L1-L2 Chinese and in L1 Italian was segmented into syllables. We calculated duration- and F0-based metrics (Δsyllable duration/F0; r-PVI syllable duration/F0; Varco syllable duration/F0; nPVI syllable duration/F0), as well as combined metrics (Δcombined, Varco combined, rPVI combined and nPVI combined). Results show that measures taking syllable duration and F0 separately do not differentiate consistently between L1 and L2 Chinese, while combined metrics reflect more systematically the variability pattern between L1 and L2 Chinese speech rhythm. More research is, however, needed to explore the significance of combined metrics as opposed to duration- and F0-based measures for the perceived rhythmic differences between L1 and L2 speech.

## 1 Introduction

Over the past few years language-specific rhythmic characteristics have been associated with the durational properties of consonantal and vocalic intervals, and a wide array of interval-based measures (henceforth rhythm metrics) have been developed to quantify between-language rhythmic differences. [1] proposed to measure the proportion over which speech is vocalic (%V) together with the variability of consonantal and vocalic intervals over an entire utterance (ΔC, ΔV). [2] suggested to calculate the average durational differences between two consecutive vocalic or consonantal intervals (r-PVIC, n-PVI-V). Some of the original metrics were further developed: [3] and [4], for example, designed the normalized variants of ΔC and ΔV (varcoC and varcoV) to control for speech rate variations. (For a detailed overview of the various metrics, see [5]).

The very nature and robustness of these metrics for the rhythmic classification of languages have been recently disputed for a variety of reasons: a) they focus exclusively on durational properties, thus reducing rhythm to timing; b) they disregard the role played by other prominence-bearing acoustic parameters in the creation of speech rhythm, c) they provide rhythmic values to languages that varied considerably across different data collection methods, sentence materials, segmentation procedures and speaking styles [6]. There have subsequently been developed alternative models of speech rhythm which focused on prominence-lending parameters other than duration, such as intensity [7], [8], [9], fundamental frequency [10], loudness [11], [12], and the frequency components of the amplitude envelope [13].

Which model of speech rhythm has been employed in research on second language speech? Despite the reservations about the rhythmic metrics, the approach that has been most widely applied in second language research is the one focusing on the temporal characteristics of consonantal and vocalic intervals. Although there is evidence that native and non-native speech vary considerably in terms of their segmental durational characteristics (e.g. [14]-[15] and [16]), the search for systematic rhythmic differences between L1 and L2 speech, or among learners at different proficiency levels has provided inconclusive results. The findings across studies seem largely to depend on the rhythm metric and on the L1/L2 pair under investigation.

In a review of the relevant literature, [17] found that %V distinguished native speakers of English from French and Spanish learners, but not from Japanese learners [15], [16], [18]. The consonantal and vocalic PVI displayed different values between Korean learners of English and English native speakers [19], but did not separate either Spanish-accented English from L1 English or English-accented Spanish from L1 Spanish [15], [20]. Similar divergences can be derived from the literature on the development of L2 rhythm. For instance, [21] and [22] found that rate normalized metrics distinguished beginner, intermediate and advanced learners of English with German and French as their L1s. Conversely, [19] did not find any significant differences between Korean learners of English at different proficiency levels.

This considerable disagreement across L2 studies further undermines the validity of rhythm metrics as instruments to quantify the rhythmic properties of languages. Given that L2 speakers can sound rhythmically dissimilar from the native speakers even for deviations in intonation, stress, pitch range and intensity [23], [24], [25], a more suitable approach to compare L1 and L2 speech rhythm or to trace the acquisition of L2 rhythm should integrate the contribution of acoustic correlates of prominence other than duration. In light of previous findings showing that duration and F0 are interdependent cues to perceived rhythm [26], in the present study we extended the durational domain by F0 and tested whether metrics taking into duration and F0 (henceforth combined metrics) capture more reliably the rhythmic variability between L2 and L1 speech than metrics based only on either duration or F0.

## 2 The Experiment

### 2.1 Participants, materials and recordings

5 Italian learners of Mandarin and 5 native Mandarin speakers were recorded in the anechoic room of the Language Center of the University of Naples L'Orientale, while performing a reading task. Both groups read The North Wind and the Sun in Mandarin. Each Italian native speaker also recorded the same text in Italian. Native Italian speakers were all university students from the Campania region, ranged in age from 21 to 23, had studied Chinese for three years only in a classroom setting. According to the language competence self-assessment grid (CEFR [27]), they were pre-intermediate learners (B1). The native Mandarin speakers ranged in age between 25 and 36. They graduated in China in humanities and were advanced learners of Italian (C1). Given the gender-related differences in $f_0$ values, both groups of informants were composed of only female speakers.

### 2.2 Segmentation and measurements

In order to compute duration, F0 and combined metrics (duration and F0), each sentence in L1 - L2 Chinese and in L1 Italian was segmented into syllables using Praat [28]. The syllabification criteria were more acoustic than phonological. Syllabification of Mandarin was based on the sequence of phones that correspond to each orthographic character (fig. 1). However, where two stops met or one stop was left adjacent to an affricate, the two sounds were merged into the right syllable; 2) where two nasals were next to each other, they were merged into the right syllable except when a fault-like boundary was discernable between the two nasals. As regards Italian: /sC/ clusters were treated as heterosyllabic clusters given the general tendency to shorten the duration of preceding vowels ([29], [30] and [31]. Long consonants were treated as the onset of the following syllable since no discontinuities occurred in the signal (fig. 2).

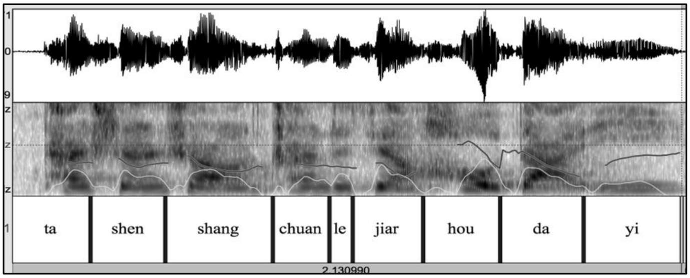


**Figure 1** – An example of speech segmentation in Chinese

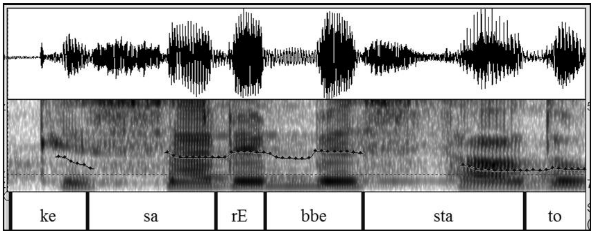


**Figure 2** – An example of speech segmentation in Italian

From the syllable tier, we calculated raw and normalized duration-based metrics:

- ΔS: standard deviation of syllable duration
- r-PVI-S: averaged difference in duration between consecutive syllables
- VarcoS: normalized index of variability of syllable duration
- n-PVI-S: averaged of the mean difference in duration between consecutive syllables.

We also measured the average f0 across each syllable and, based on interval-based metrics, we devised the following F0 metrics: ΔS.F0, VarcoS.F0 and r-PVIS.F0 and n-PVIS-F0. In addition, we developed combined metrics based on both duration and F0. For each syllable we measured the duration in seconds and mean F0 in Hz. We considered syllable duration ($D$) and mean syllabic F0 (f0) as two elements of a single vector (*mF0*). We calculated the standard deviation and the PVI of the norm of such vector ($= \sqrt{D^2 + mF0^2}$) sentence-wise, resulting in the following hybrid measures: Δcombined, Varco combined, rPVI combined and nPVI combined.

## 2.3 Data analysis and results

To test the significance of between-language variability (L1 Chinese, L2 Chinese and L1 Italian) on the three groups of rhythmic measures (duration, f0 and duration and F0 combined), we run a series of one-way ANOVA in SPSS with the values of the metrics as the dependent variables and the languages as the independent variable.

### 2.3.1 Duration-based measures

A significant effect of language was found for all metrics, with the exception of N-PVI-S. ΔS: [$F$ (2,42) = 21.37, $p < 0.0001$]; VarcoS: [$F$ (2,42) = 9.14, $p < 0.001$]; r-PVI-S: [$F$ (2,42) = 11.96, $p < 0.0001$]; n-PVI-S: [$F$ (2,42) = 1.233, $p = 0.302$] (fig. 3).

Post hoc tests (Tukey HSD) conducted on the three metrics where the main effect of language was found indicated that:

- On ΔS: Chin_L2 was significantly higher than Chin_L1 ($p < 0.001$) and Ita_L1 ($p < 0.001$), whereas the difference between the latter two groups was not significant (p = 0.585) (fig. 3a)
- On VarcoS: Chin_L2 was not significantly lower than Chin_L1 ($p = 0.221$), whereas Ita_L1 was significantly higher than both Chin-L1 ($p = 0.037$) and Chin-L2 ($p < 0.001$) (fig. 3b)
- On r-PVI-S: Chin_L2 was significantly higher than Chin_L1 ($p < 0.001$) and Ita_L1 ($p < 0.001$), whereas the difference between the latter two groups was not significant ($p = 0.757$) (fig. 3c).

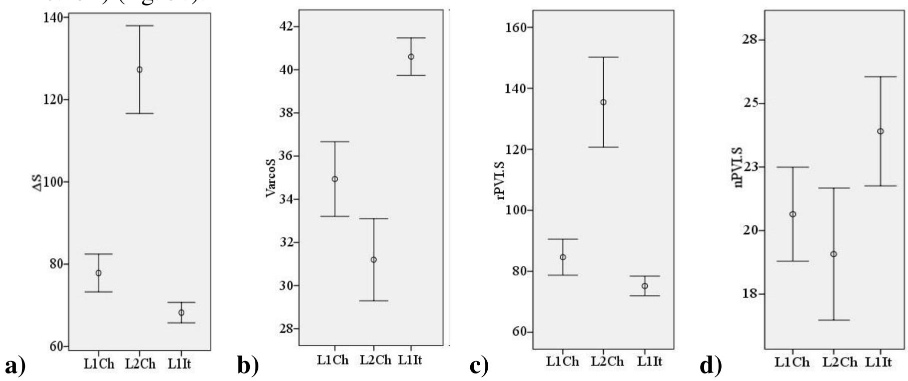


**Figure 3** – Means and standard errors of metrics based on syllable duration per language group (L1Chin, L2Chin and L1It); ΔS (a), VarcoS (b), rPVIS (c), nPVIS (d).

### 2.3.2 F0-based measures

A significant effect of language was found for all metrics, with the exception of VarcoF0. ΔS.F0: [$F$ (2,42) = 3.496, $p = 0.039$]; VarcoS.F0: [$F$ (2,42) = 1.048, $p = 0.36$]; r-PVIS.F0: [$F$ (2,42) = 11.9, $p < 0.0001$]; n-PVIS.F0: [$F$ (2,42) = 6.741, $p = 0.003$] (fig. 4).

Multiple comparisons (Tukey HSD) conducted on the three metrics where the main effect of language was found revealed that:

- On ΔS.F0: Chin_L2 was not significantly different either from Chin_L1 ($p = 0.256$) nor Ita_L1 ($p = 0.567$), whereas the difference between the latter two groups was significant ($p = 0.031$) (fig. 4a).
- On r-PVI-S: Chin_L2 was not significantly different from Chin_L1 ($p = 0.991$) but was significantly higher than Ita_L1 ($p < 0.001$). The difference between the latter two groups also was significant ($p <0.001$) (fig. 4c).
- On n-PVIS.F0: the differences between Chin_L2 and Chin_L1 as well as between Chin_L1 and Ita_L1 were insignificant ($p > 0.05$), whereas Chin_L2 was significantly higher than Ita_L1 (p = 0.002) (fig. 4d)

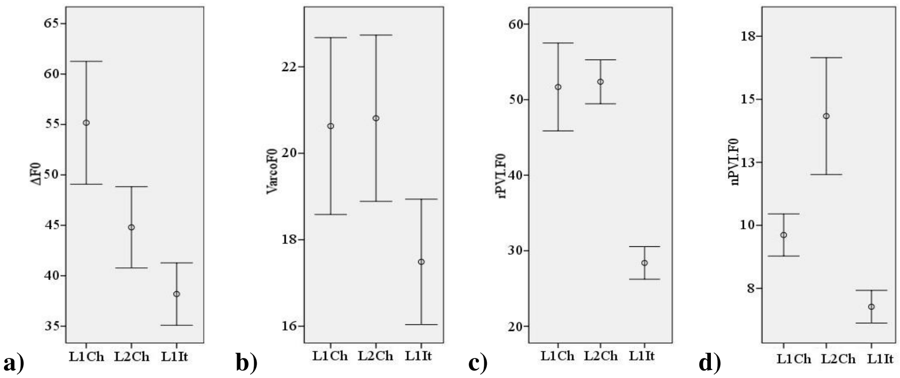


**Figure 4** – Means and standard errors of metrics based on mean F0 per language group (L1Chin, L2Chin and L1It); ΔF0 (a), VarcoF0 (b), rPVIF0 (c), nPVIF0 (d).

### 2.3.3 Combined measures (duration and F0)

A significant effect of language was found for all metrics: Δcombined [$F$ (2,42) = 24.72, $p$ < 0.0001]; Varco combined [$F$ (2,42) = 7.649, $p$ = 0.001]; rPVI combined [$F$ (2,42) = 5.715, $p$ = 0.006]; and nPVI combined [$F$ (2,42) = 5.238, $p$ = 0.009] (fig. 5)

Post hoc multiple comparisons (Tukey HSD) showed that:

- for Δcombined: Chin_L2 was significantly different from Chin_L1 and Ita_L1 ($p$ < 0.01), whereas the differences between the latter two groups were insignificant ($p$ = 0.381) (fig. 5a)
- for Varco combined: Chin_L2 was significantly different from Chin_L1 and Ita_L1 ($p$ < 0.01), while Ita_L1 was not significantly higher than Chin_L1 ($p$ = 0.988) (fig. 5b)
- for rPVI: Chin_L2 was not significantly different from Chin_L1 ($p$ = 0.198), whereas the differences with Ita_L1 were significant ($p$ < 0.01). The differences between Chin_L1 and Ita_L1 were insignificant ($p$ = 0.244) (fig. 5c)
- for nPVI: Chin_L2 was significantly higher than both Chin_L1 ($p$ = 0.016) and Ita_L1 ($p$ = 0.026), while Ita_L1 and Chin_L1 were not significantly different ($p$ = 0.977) (fig. 5d)

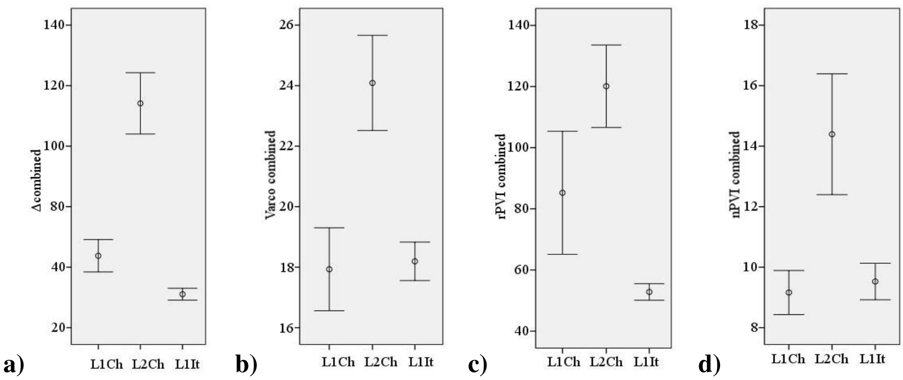


**Figure 5** – Means and standard errors of combined metrics (duration and F0) per language group (L1Chin, L2Chin and L1It); Δcombined (a), Varco combined (b), rPVI combined (c), nPVI combined (d).

## 3 Discussion and conclusion

The overall results of the study show that metrics taking into account the variability of syllable duration and F0 separately do not distinguish consistently between L1 and L2 Chinese. The results largely depend on the acoustic parameter (syllable duration or mean syllabic F0) and on the metric examined (table 1: rows 1-6, columns: 3-4).

Regarding the durational characteristics of syllabic intervals (table 1: light grey columns), L2 Chinese overshoots the values of their native and target language in terms of ΔS and r-PVIS but it approximates the target pattern when the values are normalized for speech rate (Varco and nPVI_syllable duration). Given that the former two metrics have been repeatedly proven to be highly influenced by the speakers' speech rate [3], [4] the higher variability of syllabic intervals in L2 Chinese speech can be thus explained by the slower speech rate of L2 Chinese speakers. These results coupled with studies showing that L2 learners displayed increased consonantal and vocalic variability compared to the native speakers ([14],[15], [19]).

As regards the F0-based metrics (table 1: white columns), the differences between the two groups did not reach the significance level, suggesting that L2 Chinese learners approximate the native group in the production of lexical tones. One possible explanation to the nearly-native like performance of L2 Chinese speakers in F0-based metrics can be found in the extensive tone training they had done while studying Chinese at the University. There is evidence indeed that the L2 learners' ability to perceive Mandarin tones can improve significantly after perceptual tone training and this improvement transfers also to the production domain [32], [33].

**Table 1:** Metrics that significantly distinguish the three language pairs.

| | Row | L2 Chin vs L1Chin | Chin_L2 vs Ita_L1 | Ita_L1 vs Chin_L1 |
|---|---|---|---|---|
| **Duration -based measures** | 1. | **ΔS** | **ΔS** | **-** |
| | 2. | **rPVI-S** | **rPVI-S** | **-** |
| | 3. | **-** | *VarcoS* | *VarcoS* |
| **F0 – based measures** | 4. | **-** | **-** | ΔS.F0 |
| | 5. | **-** | *rPVI.F0* | *rPVI.F0* |
| | 6. | **-** | nPVI.F0 | - |
| **Combined measures** | 7. | **Δcombined** | **Δcombined** | **-** |
| | 8. | **-** | r-PVI combined | **-** |
| | 9. | **Varco combined** | **Varco combined** | **-** |
| | 10. | **nPVI combined** | **nPVI combined** | **-** |

A total different picture is displayed when L1 and L2 Chinese speakers are compared by measures combining duration and F0 (table 1, dark grey columns). Three out of the four combined measures gave values for L2 learners that were significantly higher from both L1 Chinese and L1 Italian (fig. 5). Interpreting these outcomes in the light of interlanguage hypothesis [34], it seems that pre-intermediate learners of Chinese have created 'a new rhythmic system' that is neither L1-like nor L2-like.

Although the results seem to be in favor of an approach that measures L2 speech rhythm combining two acoustic correlates of prominence, more research on different L1/L2 pairs is needed to verify whether these metrics are actually reliable for quantifying rhythmic differences between L1 and L2 speech or among learners at different proficiency levels. Moreover, given that speech rhythm is a perceptual phenomenon, it would be important to verify to what extent combined measures correlate with listeners' ratings of perceived accentedness.

## 4 References


[1] RAMUS, F., M. NESPOR, and J. MEHLER: *Correlates of linguistic rhythm in the speech signal*. In: *Cognition* 73, 265-292, 1999.

[2] GRABE, E., and E. L. LOW: *Durational variability in speech and the Rhythm Class Hypothesis*. In: C. GUSSENHOVEN, N. WARNER (Eds.) *Laboratory Phonology* **7**, 515–545. Berlin, New York, Mouton de Gruyter, 2002.

[3] DELLWO, V.: *Rhythm and Speech Rate: A Variation Coefficient for deltaC*. In: P. KARNOWSKI, P. and SZIGETI, I. (Eds.): Language and language-processing, 231-241. Frankfurt am Main, Peter Lang, 2006.

[4] WHITE, L., and S. L. MATTYS: *Calibrating rhythm: First language and second language studies*. In: *J. Phonetics* 35, 501–522, 2007.

[5] LOUKINA, A., G. KOCHANSKI, B. ROSNER, E. KEANE, and C. SHIH: *Rhythm measures and dimensions of durational variation in speech*. In: *J. Acoust. Soc. Am.* 129, 3258-3270, 2011.

[6] ARVANITI, A.: *The usefulness of metrics in the quantification of speech rhythm*. In: J. *Phonetics* 40, 351–373, 2012.

[7] KOHLER, K. J.: *The perception of prominence patterns*. In *Phonetica* 65(4): 257–269, 2008.

[8] KOHLER, K. J.: *Rhythm in speech and language: a new research paradigm*. In: *Phonetica* 66(1–2), 29-45, 2009.

[9] HE, L.: *Syllabic intensity variations as quantification of speech rhythm: evidence from both L1 and L2*. In *Proceedings of Speech Prosody 2012* 466–469. Shanghai, China, 2012.

[10] CUMMING, R. E.: Perceptually informed quantification of speech rhythm in pairwise variability indices. In *Phonetica* 68(4): 256–277, 2011.

[11] FUCHS, R.: *Integrating variability in loudness and duration in a multidimensional model of speech rhythm: evidence from Indian English and British English*. In *Proceedings of Speech Prosody 2014*, 290–294. Dublin, Ireland, 2014.

[12] GALVES, A., J. GARCIA, D. DUARTE, and C. GALVES: Sonority as a basis for rhythmic class discrimination. In *Proceedings of Speech Prosody 2002*, Aix-en-Provence, 323–326, 2002.

[13] TILSEN S., and A. ARVANITI: *Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages*. In *J. Acoust. Soc. Am.* 134 (1), 628-639, 2013.

[14] GUT, U.: *Non-native speech rhythm in German*. In: *Proceedings of 15th International Congress of Phonetic Sciences 2003*, Barcelona, 2437–2440, 2003.

[15] TORTLE A., and D. HIRST: *Rhythm metrics and the production of L1/L2 English*. In: *Proceedings of Speech Prosody* 2010, Chicago, 2010.

[16] WHITE L., and S. L. MATTYS: Calibrating rhythm: First and second language studies. In: *Journal of Phonetics* 35, 501-522, 2007.

[17] GUT, U. *Rhythm in L2 speech*. In D. GIBBON, D. HIRST and N. CAMPBELL (Eds.): *Rhythm, melody and harmony in speech: Studies in honour of Wiktor Jassem. Special edition of Speech and Language Technology*, 14/15, 83–94. Poznan, Adam Mickiewicz University Press, 2012.

[18] GRENON, I., and L. WHITE. *Acquiring rhythm: A comparison of L1 and L2 speakers of Canadian English and Japanese*. In: H. CHAN, H. JACOB, and E. KAPIA (Eds.)

*Proceedings of the 32nd Annual Boston University Conference on Language Development*, 155-166. Somerville, Cascadilla, 2008.

[19] Jang, T.Y.: Speech Rhythm Metrics for Automatic Scoring of English Speech by Korean EFL Learners. In: *Malsori Speech Sounds. The Korean Society of Phonetic Sciences and Speech Technology* 66, 41–59, 2008.

[20] MOK P.K., and V. DELLWO: *Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English*. In: *Proceedings of the Speech Prosody* 2008, Campinas, 423-426, 2008.

[21] ORDIN, M., and L. POLYANSKA: *Perception of speech rhythm in second language: The case of rhythmically similar L1 and L2*. In: *Frontiers in Psychology*, *6*(MAR), 1–15, 2015.

[22] POLYANSKAYA, L., ORDIN, M., and M.G. BUSÀ: Relative Salience of Speech Rhythm and Speech Rate on Perceived Foreign Accent in a Second Language. In: *Language and Speech*, 1–23, 2016.

[23] KASHIWAGI, A., and M. SNYDER: *American and Japanese listener assessment of Japanese EFL speech: Pronunciation features affecting intelligibility*. In: *The Journal of Asia TEFL*, *5*(4), 27–47, 2008.

[24] MAGEN, H. S.: The perception of foreign-accented speech. In: *Journal of Phonetics*, *26*(4), 381–400, 1998.

[25] KASHIWAGI, A., and M. SNYDER: *Speech characteristics of Japanese speakers affecting American and Japanese listeners evaluations*. In: *Working Paper in TESOL & Applied Linguistics* 10(1), 1-14, 2010.

[26] CUMMING, R. E.: *The interdependence of tonal and durational cues in the perception of rhythmic groups*. In: *Phonetica* 67(4): 219-42, 2010.

[27] Council of Europe. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge University Press, Cambridge, 2001.

[28] BOERSMA P., and D. WEENINK. 2011. *Praat: doing Phonetics by computer*, retrieved from http://www.fon.hum.uva.nl/praat/

[29] NESPOR, M.: *La fonologia*, Bologna, Il Mulino, 1993.

[30] LOPORCARO M.: *On possible onsets*. In: J.R. RENNISON & K. KÜHNHAMMER (Eds.): *Phonologica 1996. Syllables*!? *Proceedings of the Eighth International Phonology Meeting*, Vienna 1996, 133-151. The Hague, Holland Academic Graphics, 1999.

[31] MAROTTA G. 1995. La sibilante preconsonantica in italiano: Questioni teoriche ed analisi sperimentale. In R. AJELLO, and S. SANI (Eds.): *Scritti linguistici e filologici in onore di Tristano Bolelli*, 393-437. Pisa, Pacini.

[32] WANG, Y., M. M. SPENCE, A. JONGMAN, and J. A. SERENO: Training American listeners to perceive Mandarin tones. In: *Journal of the Acoustical Society of America* 106, 3649-3658, 1999.

[33] WANG, Y., JONGMAN, A., and J. A. SERENO: *Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training*. In: *Journal of the Acoustical Society of America* 113, 1033-1044, 2003.

[34] Selinker, L. *Interlanguage.* In IRAL 10, 209–231, 1972.