# Syllabic Intensity Variations as Quantification of Speech Rhythm: Evidence from Both L1 and L2

*Lei He*

Unaffiliated Researcher, Hohhot, China

helei@rocketmail.com

## Abstract

In this study, three intensity metrics (ΔS.dB, VarcoS.dB and nPVI.dB) devised on the basis of the well-known durational metrics were tested on both L1 (English and Mandarin) and L2 (L2 English). The results suggested that they were effective in distinguishing "stress-timed" English from perceptually "syllable-timed" Mandarin and L2 English (by Mandarin speakers). These metrics break the impasse that although the L1 and L2 varieties of English were similar on durational measurements, they were perceptually different in rhythmicity [22, 23]. Therefore, it is advised that intensity metrics be applied in future rhythm research.

**Index Terms**: rhythm, intensity, English, Mandarin, L1, L2

## 1. Introduction

Languages have been traditionally classified into three rhythmic categories: "stress-", "syllable-" and "mora-timed" languages. Languages in each category were supposed to have isochronous feet, syllables and moræ respectively [1, 2, 3, 4]. However, later empirical studies failed to find true isochrony in both "stress-" and "syllable-timed" languages [5, 6, 7, 8, 9]. For this reason, Nespor [10] even rejected such rhythmic typology, claiming that the perceptual rhythmic differences between the two types of languages resulted from the non-rhythmic rules idiosyncratic to different languages.

Albeit absolutely isochronous feet or syllables were not instrumentally justified, perceptual experiments among neonates showed that they were able to differentiate languages typologically different from their L1s whereas unable to discern a language with the same rhythmicity as their L1s [11, 12, 13]. Perceptual experiments among adults using low-pass filtered speech also yielded similar results that rhythmically different languages were distinguishable [14, 15]. For this reason, metalinguistic terms like "stress-" and "syllable-timing" have been retained by many linguists, and are placed in quotes here to emphasise their metaphorical usage. In order to corroborate the perceived disparities, various durational metrics have been put forward, among which ΔC, ΔV, %V, rPVI, nPVI, VarcoC and VarcoV [15, 16, 17, 18] were the most influential ones. These metrics have quantified the structural characteristics of the two categories proposed by Dauer [7] and have successfully differentiated canonical "stress-" and "syllable-timed" languages.

Nevertheless, rhythm research within the paradigm of durational metrics has been subject to many criticisms (see [19] for a review). Taking durational metrics as the litmus test for rhythmicity would be problematic since it would simplify the issue: essentially rhythm involves alternating prominent units, and prominence can be signalled by $f_0$ and intensity apart from duration. Hence, Cumming [20] incorporated $f_0$ in her study. Moreover, research on L2 rhythm using durational metrics met an impasse. For example, in Mok and Dellwo [21] as well as He [22, 23] (available upon request), durational metrical results failed to discriminate L1 English from L2 English by native speakers of Mandarin (abbreviated as Eng$_{\mathrm{Man}}$ henceforth), even if Eng$_{\mathrm{Man}}$ was perceptually to be different in rhythmicity. In order to delve into the perceived and measured disparity, I compared the intensity contours of English, Mandarin and Eng$_{\mathrm{Man}}$ in [22, 23] and found that the envelope of the English intensity graph was wavier than those of Mandarin and Eng$_{\mathrm{Man}}$ (see Figure 1 for an illustration).

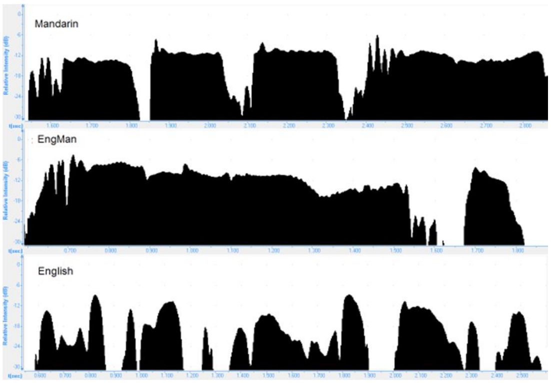

Figure 1: *An illustration of intensity graphs of Mandarin (top), Eng$_{Man}$ (middle) and English (bottom). The time axes are calibrated by intervals of 100ms each.*

Visual inspections of the intensity contours of the three languages suggest that the intensity of "stress-timed" English may be more variable than that of "syllable-timed" Mandarin across the utterance, and the intensity variation of Eng$_{\mathrm{Man}}$ may resemble that of Mandarin. Thus, although English and Eng$_{\mathrm{Man}}$ are similar measured by durational metrics, their perceived rhythmic difference may be due to intensity variations. Therefore, I propose that intensity be incorporated in rhythmic measurements. Based on durational metrics, I devised the following intensity metrics: ΔS.dB, VarcoS.dB and nPVI.S.dB. They are calculated based on the formulae below:

$$\Delta S.dB = \sqrt{\frac{\sum_{i=1}^{n-2} I_i^2 - \left(\sum_{i=1}^{n-2} I_i\right)^2 / (n-2)}{(n-2)-1}} \qquad (1)$$

$$VarcoS.dB = \frac{\Delta S.dB}{\bar{I}} \times 100 \qquad (2)$$

$$nPVI.S.dB = \frac{\sum_{i=1}^{n-2} \left| \frac{I_i - I_{i+1}}{(I_i + I_{i+1})/2} \right|}{n-2} \times 100 \qquad (3)$$

where $n$ = the number of syllables in an utterance (sentence in the present study), $I_i$ = the average intensity of the $i$th syllable, and $\bar{I}$ = the mean intensity of the utterance (i.e., sentence in the present study). The formulae, (1) and (3) in particular, reveal that two syllables (the first and last syllables of the sentences) are excluded because they may not be steadily articulated.

The present study hypothesises that “stress-timed” English has higher standard deviation of syllabic intensities (ΔS.dB), normalised (or *var*iation *co*efficient of) standard deviation of syllabic intensities (VarcoS.dB) and normalised pairwise syllabic intensity variability (nPVI.S.dB), whereas “syllable-timed” Mandarin is lower on all the above mentioned metrics. The interlanguage $\text{Eng}_{\text{Man}}$ and Mandarin are similar on the three metrics.

# 2. Method

## 2.1. Informants, materials and recording

The speech data in [22, 23] were further analysed in this study. The informants were five native speakers of American English and Mandarin each. The speech data of American English were originally part of the pathology-free data set in [24], and was made accessible to the author who had filed a data sharing request and completed an online course “Protecting Human Research Participants” by the National Institute of Health (US). The Mandarin speakers (all Beijing natives) were post-intermediate or advanced L2 English learners. Both American and Chinese informants recorded five English sentences each. Besides, each Mandarin speaker also recorded five Mandarin sentences. The annex lists all the sentences. The recordings were sampled at 44/48 kHz and quantised at 16 bits.

## 2.2. Segmentation and measurements

Each sentence was syllabified based on the spectrogram and waveform using Praat [25]. The syllabification criteria were more acoustic than phonological: 1) where two stops meet or one stop is left adjacent to an affricate, the two sounds were merged into the right syllable; 2) where two nasals are next to each other, they were merged into the right syllable except when a fault-like boundary was discernable between the two nasals; 3) where a consonant is before a vowel, the consonant and the vowel were segmented into the same syllable, albeit the consonant may phonologically be the coda of the previous syllable. The average intensity across each syllable was measured and the intensity metrics were calculated on the Excel spreadsheet before the statistical analysis using R [26].

# 3. Data analysis and results

## 3.1. Data normality

The Kolmogorov-Smirnov tests were conducted to assess data distribution. The scores of ΔS.dB ($Z$ = .706, $p$ = .702, 2-sided), VarcoS.dB ($Z$ = .638, $p$ = .811, 2-sided) and nPVI.S.dB ($Z$ = .884, $p$ = .416, 2-sided) were all normally distributed, meeting the normality assumption of parametric statistics.

## 3.2. ΔS.dB

One-way ANOVA was conducted to analyse all the ΔS.dB scores with languages as the factor. A significant effect of the languages was found (Table 1).

Table 1. *Summary of ANOVA (ΔS.dB ~ languages).*

| | Df | Sum Sq | Mean Sq | F | $\eta^2$ |
|---|---|---|---|---|---|
| Languages | 2 | 9.092 | 4.5458 | 7.7748*** | .1776 |
| Residuals | 72 | 42.097 | .5847 | | |
| Total | 74 | 51.189 | | | |

*** $p$ < .0001

Post hoc tests (Tukey HSD) indicated that English ($\bar{x}$ = 3.622204) was significantly higher than both $\text{Eng}_{\text{Man}}$ ($\bar{x}$ = 3.057624, $p$ < .05) and Mandarin ($\bar{x}$ = 2.786349, $p$ < .001), but $\text{Eng}_{\text{Man}}$ and Mandarin were not significantly different ($p$> .05).

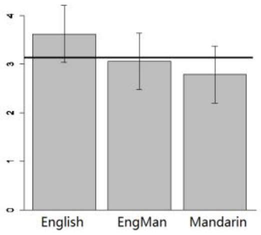

Figure 2: *Means and standard errors of ΔS.dB (the horizontal bar indicates the pooled mean = 3.155392).*

## 3.3. VarcoS.dB

One-way ANOVA was run on all the VarcoS.dB scores with languages as the independent variable. A significant effect of the languages was found (Table 2).

Table 2. *Summary of ANOVA (VarcoS.dB ~ languages).*

| | Df | Sum Sq | Mean Sq | F | $\eta^2$ |
|---|---|---|---|---|---|
| Languages | 2 | 19.112 | 9.5560 | 7.6716*** | .1757 |
| Residuals | 72 | 89.686 | 1.2456 | | |
| Total | 74 | 108.798 | | | |

*** $p$ < .0001

Multiple comparisons (Tukey HSD) revealed that English ($\bar{x}$ = 5.068312) was significantly higher than $\text{Eng}_{\text{Man}}$ ($\bar{x}$ = 4.213533, $p$ < .05) and Mandarin ($\bar{x}$ = 3.867139, $p$ < .001) on VarcoS.dB whereas the difference between the latter two was not significant ($p$ > .05).

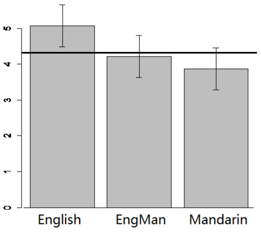

Figure 3: *Means and standard errors of VarcoS.dB (the horizontal bar indicates the pooled mean = 4.382995).*

## 3.4. nPVI.S.dB

One-way ANOVA was run on all the nPVI.S.dB scores with languages as the independent variable. A significant effect of the languages was found (Table 3).

Table 3. *Summary of ANOVA (nPVI.S.dB ~ languages).*

| | Df | Sum Sq | Mean Sq | F | $\eta^2$ |
|---|---|---|---|---|---|
| Languages | 2 | 21.763 | 10.8816 | 6.1966** | .1468 |
| Residuals | 72 | 126.437 | 1.7561 | | |
| Total | 74 | 148.200 | | | |

** $p$ < .001

Post hoc multiple comparisons (Tukey HSD) indicated that English ($\bar{x}$ = 4.921288) was significantly higher than both $\text{Eng}_{\text{Man}}$ ($\bar{x}$ = 3.980015, $p$ < .05) and Mandarin ($\bar{x}$ = 3.649844, $p$ < .005). Nevertheless, the difference between $\text{Eng}_{\text{Man}}$ and Mandarin was insignificant ($p$ > .05).

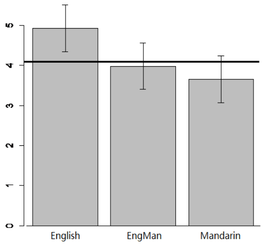

Figure 4: *Means and standard errors of nPVI.S.dB (the horizontal bar indicates the pooled mean = 4.183716).*

In general, the three ANOVA models were significant in distinguishing English from either $\text{Eng}_{\text{Man}}$ or Mandarin. Nonetheless, $\text{Eng}_{\text{Man}}$ and Mandarin were not significantly different from each other on the metrics of ΔS.dB, VarcoS.dB and nPVI.S.dB. The effect sizes ($\eta^2$) of the three models were not satisfactorily high; no more than 20% of the variances were explained. This may be due to small sample sizes (5 sentences × 3 languages × 5 speakers per language = 75 data points). Finally, the models were checked to assess their reliability.

### 3.5. Model checking

Model checking was further performed on the three ANOVA models. Graphically, the plots of constant leverage (residuals vs. languages) were created for "ΔS.dB ~ languages", "VarcoS.dB ~ languages" and "nPVI.S.dB ~ languages" (see Figure 5).

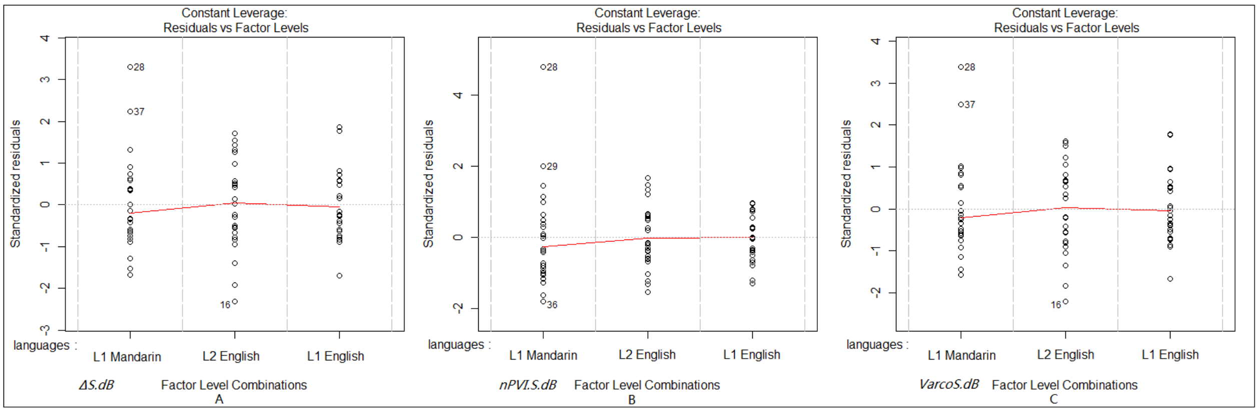

Figure 5: *Plots of constant leverage (residuals vs. languages) of the three ANOVA models.*

It is obvious that from Figure 5A, data points #28, #37 and #16 were potentially influential in the model of "ΔS.dB ~ languages". I tested their influence by repeating the ANOVA without these data points (see the upper portion of Table 4). The interpretation was not affected much except that the significance level changed from α = .0001 to α = .001.

Similarly, the data points #28, #37 and #16 were potentially influential in the model of "VarcoS.dB ~ languages" (see Figure 5C) and the data points #28, #29 and #36 were potentially influential in the model of "nPVI.S.dB ~ languages" (see Figure 5B). Likewise, the ANOVA models were updated by the removal of these points and the results (middle and lower portions of Table 4) indicated that the interpretations were not affected. Therefore, the one-way ANOVAs in §§3.2-3.4 were reliable.

Table 4. *Updated ANOVA tables for model checking.*

| ΔS.dB~langs | Df | Sum Sq | Mean Sq | F |
|---|---|---|---|---|
| Languages | 2 | 5.1588 | 2.57938 | 6.0758** |
| Residuals | 69 | 29.2927 | .42453 | |
| Total | 71 | 34.4515 | | |
| VarcoS.dB~langs | Df | Sum Sq | Mean Sq | F |
| Languages | 2 | 10.177 | 5.0886 | 5.779** |
| Residuals | 69 | 60.756 | .8805 | |
| Total | 71 | 70.933 | | |
| nPVI.S.dB~langs | Df | Sum Sq | Mean Sq | F |
| Languages | 2 | 11.418 | 5.7088 | 5.3416** |
| Residuals | 69 | 73.743 | 1.0687 | |
| Total | 71 | 85.161 | | |

** $p < .001$

### 3.6. Summary

One-way ANOVAs were applied to test whether ΔS.dB, VarcoS.dB and nPVI.S.dB were effective in distinguishing English from Mandarin and $\text{Eng}_{\text{Man}}$. All of the metrics succeeded in discriminating English from Mandarin and $\text{Eng}_{\text{Man}}$. However, the effect sizes were not satisfactory, and could have been improved with more informants.

## 4. Discussion

All the three intensity metrics have fair success in distinguishing "stress-timed" English from "syllable-timed" Mandarin: English was significantly higher than Mandarin on ΔS.dB, VarcoS.dB and nPVI.S.dB. Such result is in line with the hypothesis. "Stress-timed" languages like English may have more fluctuated intensities across the whole utterance where stressed syllables have higher amplitude levels compared with unstressed syllables. However, "syllable-timed" languages like Mandarin may have more levelled intensities across the utterance.

Moreover, the findings of this research coupled with those of [22, 23] (English and $\text{Eng}_{\text{Man}}$ are similar on durational metrics whereas significantly different on intensity metrics) suggested a scale in L2 prosodic acquisition. In the case of L2 English learning among Mandarin speakers, vowel reductions and syllable structures may be more easily learnt than intensity. The acquisition of vowel reduction and syllable structures enables the interlanguage to exhibit similar durational metrics scores to those of native English, but may not suffice to make $\text{Eng}_{\text{Man}}$ a perceptually “stress-timed” language. Thus, the disparity between perceived rhythm and measured rhythm of $\text{Eng}_{\text{Man}}$ indicated the inadequacy of durational metrics in rhythm research, and such inadequacy could be rectified by taking intensity variations into consideration. Mandarin speakers’ insensitivity towards intensity is also supported by the perceptual experiment [27] that only $f_0$ had a decisive role on Mandarin speakers’ judgments of stress, compared with native English speakers who were perceptive of all the three cues of prominence including intensity, $f_0$ and duration.

Two suggestions for further research are made. Firstly, intensity metrics could be applied to a wider range of languages to see if they are effective in distinguishing languages of different perceptual rhythmicity. Also, perceptual experiments could be conducted to explore if the intensity variations have main effects in differentiating rhythmically different languages, or interact with other prosodic elements (e.g. duration and $f_0$) to signal perceived rhythm.

## 5. Conclusion

Three intensity metrics (ΔS.dB, VarcoS.dB and nPVI.S.dB) devised on the basis of durational metrics were adopted among English, Mandarin and $\text{Eng}_{\text{Man}}$ to investigate the effectiveness of these metrics as quantification of speech rhythm. The results showed that all the intensity metrics succeeded in differentiating between canonical “stress-timed” English and “syllable-timed” Mandarin. Moreover, perceptually similar Mandarin and $\text{Eng}_{\text{Man}}$ were similar on all the intensity metrics, although $\text{Eng}_{\text{Man}}$ clustered with English instead of Mandarin on all the durational metrics in [22, 23]. In short, the idea of including intensity variations in quantifying speech rhythm is well motivated.

## 6. Annex

English sentences: 1) The supermarket chain shut down because of poor management. 2) Much more money must be donated to make this department succeed. 3) In this famous coffee shop they serve the best doughnuts in town. 4) The chairman decided to pave over the shopping centre garden. 5) The standards committee met this afternoon in an open meeting.

Mandarin sentences in phonetic symbols (tones omitted): 1) ta tɕiə tɕin tʰian tsau ʈʂʰən kən mamə tɕʰy ʈʂɤ tsia ʈʂʰau ʂʅ mai tɕiao tsɿ. 2) tʰa xau ɕiaŋ tʰiŋ ta tɕia ʈʂʰaŋ na pu tian ʂʅ tsy tə ʈʂu tʰi tɕy. 3) fu tɕin ʈʂɤ tɕia kʰa fei tʰiŋ mai tɕʰyɛn ʂʅ tsui xau tə ʈʂʅ ʂʅ tan kau. 4) ɕiao ʈʂaŋ tsyɛ tiŋ tɕiaŋ ɕyɛ ɕiau tsu tɕʰiəu ʈʂʰaŋ ʈʂʰuŋ ɕin fan ɕiu. 5) tʰa kən tʰuŋ ɕyɛ ʂuo xau tɕin tʰian tsau ʈʂən tsai kʰən tɤ tɕi mən kʰou tɕiɛn mian.

The non-IPA symbols [ʅ] and [ɿ] represent the rhotacised and non-rhotacised non-open central unrounded apical vowels in Mandarin [28].

## 7. References

[1] Classé, A., *The Rhythm of English Prose*. Oxford: Blackwell, 1939.

[2] Lloyd James, A., *Speech Signals in Telephony*. London: Sir Isaac Pitman & Sons, 1940.

[3] Pike, K., *The Intonation of American English*. Ann Arbor: Univ. of Michigan Press, 1945.

[4] Abercrombie, D., *Elements of General Phonetics*. Edinburgh: Edinburgh Univ. Press, 1967.

[5] Pointon, G. E., “Is Spanish really syllable-timed?”, *J. Phonet.*, 8: 293-304, 1980.

[6] Roach, P., “On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages”, in D. Crystal [Ed], *Linguistic Controversies*, pp. 73-79. London: Edwards Arnold, 1982.

[7] Dauer, R., “Stress-timing and syllable-timing reanalyzed”, *J. Phonet.*, 11: 51-62, 1983.

[8] de Manrique, B. A. M. and Signorini, A., “Segmental durations and the rhythm in Spanish”, *J. Phonet.*, 11: 117-132, 1983.

[9] Bertrán, A. P., “Prosodic typology: On the dichotomy between stress-timed and syllable-timed languages”, *Lang. Design*, 2: 103-131, 1999.

[10] Nespor, I., “On the rhythm parameter in phonology”, in I. Roca [Ed], *Logical Issues in Language Acquisition*, pp. 157-195. Dordrecht: Foris, 1990.

[11] Nazzi, T., Bertoncini, J. and Mehler, J., “Language discrimination by newborns: Towards an understanding of the role of rhythm”, *J. exp. Psychol. hum. Percept. Perform.*, 24: 756-766, 1998.

[12] Nazzi, T., Jusczyk, P. W. and Johnson, E. K., “Language discrimination by English-learning 5-month-olds: Effect of rhythm and familiarity”, *J. Mem. Lang.*, 43: 1-19, 2000.

[13] Bosch, L. and Sebastián-Gallés, N., “The role of prosody in infants’native language discrimination abilities: The case of two phonologically close language”, in *EURODPEECH-1997*, pp. 231-234, 1997.

[14] Ramus, F. and Mehler, J., “Language identification with suprasegmental cues: A study based on speech resynthesis”, *J. acoust. Soc. Am.*, 105: 512-521, 1999.

[15] Ramus, F., Nespor, M. and Mehler, J., “Correlates of linguistic rhythm in the speech signal”, *Cognition*, 73: 265-292, 1999.

[16] Grabe, E. and Low, E. L., “Durational variability in speech and rhythm class hypothesis”, in N. Warner and C. Gussenhoven [Eds], *Papers in Laboratory Phonology 7*, pp. 515-543, Berlin: Mouton de Gruyter, 2002.

[17] Low, E. L., Grabe, E. and Nolan, F., “Quantitative characterization of speech rhythm: Syllable-timing in Singapore English”, *Lang. Speech*, 43: 377-401, 2000.

[18] Dellwo, V., “Rhythm and speech rate: A variation coefficient for deltaC”, in P. Karnowski and I. Szigeti [Eds], *Language and Language Processing*, pp. 231-241, Frankfurt: Peter Lang, 2006.

[19] Kohler, K. J., “Editorial”, *Phonetica*, 66: 5-14, 2009.

[20] Cumming, R., “Speech rhythm: The language-specific integration of pitch and duration”, Ph.D. thesis, Univ. of Cambridge, 2010.

[21] Mok, P. P. K. and Dellwo, V., “Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English”, in *Speech Prosody 2008*, Campinas, Brazil, pp. 423-426, 2008.

[22] He, L., “Interlanguage rhythm: A durational metrics study among native speakers of Mandarin and Cantonese learning English”, MSc dissertation in developmental linguistics, Univ. of Edinburgh, 2010.

[23] He, L., “Interlanguage rhythm: A durational metrics study amongst native speakers of Mandarin and Cantonese learning English”, Paper presented at the 16th World Congress of Applied Linguistics (AILA), Beijing, 2011.

[24] Liss, J. M., White, L., Mattys, S., Lansford, K., Lotto, A. J., Spitzer, S. M. and Caviness, J. N., “Quantifying speech rhythm: Abnormalities in the dysarthrias”, *J. Speech Lang. Hear. R.*, 52: 1334-1352, 2009.

[25] Boersma, P. and Weenink, D., “Praat, a system for doing phonetics by computer”, *Glot Int.*, 5: 341-345, 2001.

[26] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/, 2011.

[27] Wang, Q., “L2 stress perception: The reliance on different acoustic cues”, in *Speech Prosody 2008*, Campinas, Brazil, pp. 635-638, 2008.

[28] Pullum, G. and Ladusaw, W. A., *Phonetic Symbol Guide*. Chicago: The Chicago Univ. Press, 1996.