Acoustic Cues to Tone and Register in Bai: Adult Baseline Data

Allison Benner, John H. Esling

Department of Linguistics, University of Victoria, Canada

abenner@uvic.ca, esling@uvic.ca

Abstract

This paper presents the results of a study of the acoustic cues associated with the tense/lax distinction in Bai, a Tibeto-Burman register tone language spoken in Yunnan, China. The purpose of the paper is to provide baseline adult data for comparison with infant speech in an acoustic study of infants' acquisition of Bai register tones in the second and third years of life. The results show that among adults, F0, F1, and spectral tilt combine to create the tense/lax contrast in Bai. While these three cues tend to be correlated, individual speakers differ in their use, particularly spectral tilt. The patterns in this study suggest that as Bai infants acquire tones in the second and third years of life, their utterances are likely to become structured around these three acoustic cues in previously unattested ways that exemplify the complex interaction between universal physiological and developmental tendencies and the ambient phonological tone system of Bai.

Index Terms: Bai, tone, acoustic cues, laryngeal constriction, first language acquisition

1. Introduction

Bai is a Tibeto-Burman language spoken by approximately 1.6 million speakers in the three dialect regions of Dali, Jianchuan, and Bijiang in Yunnan, China [1]. As shown in Table 1 below, Bai includes eight tones that are phonologically classified as lax or tense, based on pitch and/or phonation type. All these tones have contrastive nasal variants, with the exception of the rising tone. Tense tones are produced with more laryngeal constriction than lax tones, though tones in both registers may be produced with differing degrees of laryngeal constriction. For example, the tense register includes differing degrees of harshness. Tense tones 55+ and 33+ are often produced with tight or 'pressed' phonation, 31+ with harsh voice, and 21 with aryepiglottic trilling. Lax tones 55 and 33 are produced with modal voice, and lax tone 31 with breathy voice. Tone 35 varies between harsh and modal phases.

Table 1. *Bai register tone system.*

	Lax	Tense
high level	55	55+
mid level	33	33+
mid falling	31	31+
low falling		21
Rising		35

While the literature includes a description and classification of the Bai tone system [2] [3] [4], and while detailed studies of the laryngeal articulatory phonetic realization of these tones are available [1] [5], no acoustic studies on Bai tones exist. This study aims to find general trends in the acoustic cues to these tones. As such, the study contributes to the growing recent literature on acoustic cues to register tones [6] [7] [8] [9] and will be used for baseline comparison in our ongoing study of tone acquisition among infants acquiring Bai [10] [11] [12] [13].

2. Methodology

2.1. Data

To examine the acoustic cues to the tense/lax distinction, data were selected from two pre-existing sources: (1) the audio files that accompany the Phonetic Database article for Bai [14], which comprise recordings of a male native speaker of Bai; and (2) field recordings of six native speakers of Bai (5 female, 1 male) made in Yunnan in 2008. The latter recordings were made to find examples of vowels illustrating all eight tones in the paradigm on a single monosyllabic word, and to verify whether speakers made these contrasts in the same manner observed in the speaker in the Phonetic Database. There are some gaps in the data, because some speakers did not recognize words in the paradigm.

The 291 tokens included in this analysis are monosyllabic words containing the vowels /æ/, /a/, /i/, /e/, /u/, and /o/ spanning, where possible, all the non-nasalized tones in the register tone paradigm. In all, the data include 111 tokens of tones in the lax register, and 180 in the tense register. Table 2 shows the distribution of the data among the eight tones. Within the available data, it was not possible to balance the number of tokens by speaker or tone. Cells in Table 2 with at least 7 tokens include recordings of all 7 speakers in the study. In cells with large numbers of tokens, most utterances represent the single male speaker in the Phonetic Database, whose utterances comprise approximately one-third of the total data (102 tokens).

Table 2. *Data included in the analysis, by vowel and tone.*

Tone	/æ/	/a/	/i/	/e/	/u/	/o/	Total
55	7	8	11	3	7	7	43
55+	7	5	9	7	7	1	36
33	7	7	9	1	6	2	32
33+	7	23	10	9	7	8	64
31	7	7	9	1	7	5	36
31+	7	9	3	7	1	5	32
35	1	--	7	1	3	1	13
21	7	10	3	--	7	8	35
Total	50	69	61	29	45	37	291

2.2. Acoustic Analysis

Three acoustic cues were measured for each of the tokens: F0, F1, and spectral tilt. F0 was measured at the beginning, middle, and end of each token. F1 was measured in the middle of each token. Finally, spectral tilt measures, including H1-H2,

10.21437/SpeechProsody.2014-120

H1-A1*, H1-A2*, and H1-A3*, were taken from the middle of each token. All measurements were taken using Praat [15]. Average values for F0, F1, H1-H2, H1-A1*, H1-A2*, and H1-A3* were calculated for individual speakers and for the group as a whole for each of the eight tones, and for the lax and tense registers. Because of the uneven distribution of the data samples between individuals and tones, no statistical tests were performed.

3. Results

3.1. Fundamental Frequency

Tables 3 and 4 below show the average F0 at the beginning and end of each tone for male and female speakers, respectively. Figure 1 illustrates the average F0 values and direction of F0 movement for each tone across speakers. As shown, the eight contrasting tones in the paradigm are distinguished by pitch. Tense tones tend to be higher in pitch than their lax counterparts. For example, 55+, 33+, and 31+ are higher in pitch than 55, 33, and 31, respectively. The pitch range of the lax 31 tone and the tense 21 tone closely overlap (though they differ maximally in phonation type). All speakers follow this pattern. While lax level tones 33 and 55 are produced in different ways by individual speakers (as slightly rising, slightly falling, or small rise-falls), they are consistently produced with more level pitch than corresponding tense tones 33+ and 55+, which all speakers produce with noticeable falls.

Table 3. *Average F0 of tones (Hz), male speakers.*

	55	55+	33	33+	31	31+	35	21
Start	230	268	179	208	166	224	138	160
Mid	236	269	177	196	138	174	191	135
End	237	248	177	193	107	116	222	110

Table 4. *Average F0 of tones (Hz), female speakers.*

	55	55+	33	33+	31	31+	35	21
Start	270	291	249	272	237	288	245	228
Mid	275	276	248	260	205	247	260	206
End	267	227	231	249	160	179	296	169

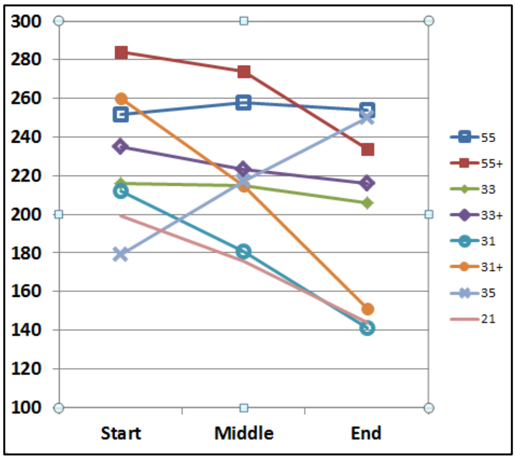

Figure 1: *Average F0 of tones, male and female speakers.*

3.2. F1 Frequencies

The F1 frequency is, on average, higher for tones produced in the tense register (55+, 33+, 31+, 35, and 21) than for tones produced in the lax register (55, 33, and 31) for all vowels studied, reflecting the greater laryngeal constriction in tense tones. Average F1 frequencies for vowels in the tense and lax registers are shown in Figure 2 below.

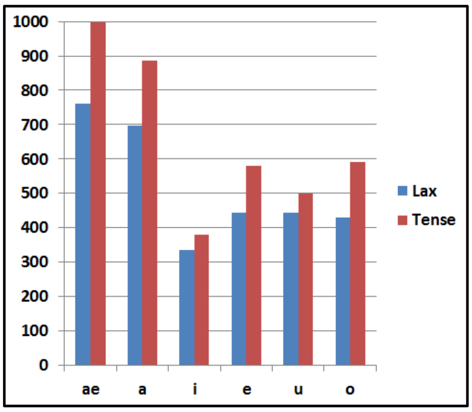

Figure 2: *Average F1 frequencies, all speakers.*

Within this pattern, F1 frequencies differ considerably across tones. Table 5 shows the average F1 for each vowel in the study for each tone. As shown, 55+, 33+, and 31+ have a higher average F1 than 55, 33, and 31, respectively, with the exception of the 33+/33 contrast for /i/. These contrasts in F1 help to preserve distinctions between pairs of tones that are phonologically contrastive and that are potentially confusable on the basis of pitch. Among the tones, across all vowels except /u/, the average F1 for 21 is the highest, and the average F1 for 31 is the lowest, in keeping with the fact that these tones differ maximally in laryngeal constriction: as noted earlier, 21 is produced with harsh voice and aryepiglottic trilling, while 31 is produced with breathy voice.

Table 5. *Average F1, by vowel and tone, all speakers.*

	/æ/	/a/	/i/	/e/	/u/	/o/
55	807	733	357	--	437	457
55+	958	824	405	616	569	593
33	782	681	337	399	400	402
33+	985	867	326	558	442	505
31	694	669	308	392	408	403
31+	1000	868	361	582	493	579
35	875	--	415	490	676	473
21	1056	964	422	--	474	700

F1 frequencies also differ between speakers. Figure 2 depicts the average F1 of /æ/ for speakers 1 to 7 for all tones, except 35 (there was only one recording of this tone in the data). This vowel is chosen as an illustration because the data are balanced across speakers, because tense/lax F1 values occur within a wider range for low vowels, and because the pattern depicted reflects the general trend found for most other vowels in the study. As shown, with the exception of speaker 6, all speakers produce higher F1 values for tense tones than for their lax counterparts (55+, 33+ and 31+ have higher F1 frequencies than 55, 33, and 31, respectively). For 4 of the 7 speakers, 31 has the lowest F1, and 21 the highest. For some speakers, lax tones 33 or 55 have a lower F1 than lax 31, and/or tense 31+ has a higher F1 than tense 21. Across all speakers, with the exception of speaker 6, the F1 frequencies of tense tones 21 and 31+ are both higher than the F1 of lax (breathy) 31.

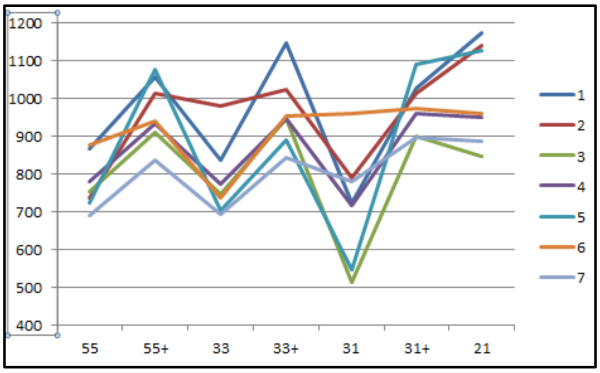

Figure 2: *Average F1 for /æ/, speakers 1-7*.

3.3. Spectral Tilt

As shown in Table 6, average spectral tilt values for the lax register are lower than for the tense register across all six vowels studied and across all four measures of spectral tilt employed (H1-H2, H1-A1*, H1-A2*, and H1-A3*), with the exception of H1-H2 for the vowels /a/ and /i/. These differences reflect the greater laryngeal constriction in the tense tones compared to the lax tones.

Table 6. *Spectral tilt by vowel, tense and lax registers*.

	H1-H2		H1-A1*		H1-A2*		H1-A3*	
	T	L	T	L	T	L	T	L
/æ/	2	7	-10	3	-8	5	4	19
/a/	4	3	-9	-2	-8	3	9	21
/i/	8	7	-1	0	12	18	12	18
/e/	-6	3	-8	3	1	13	6	18
/u/	-2	-1	-3	-2	8	9	27	29
/o/	-3	0	-7	-2	1	15	24	31

As with F1 values, spectral tilt measures varied among tones and among individual speakers. As an illustration of the general trend, Table 7 below shows the average H1-A3* values for each vowel and tone for the seven speakers. Across vowels, H1-A3* is higher for 55, 33, and 31 than for 55+, 33+, and 31+, respectively, with the exception of the 33/33+ contrast for /u/. Among the eight tones, H1-A3* is highest for lax tone 31 for most vowels (/æ/, /i/, /e/, and /u/) and lowest for tense tone 31+ for most vowels (/æ/, /i/, /e/, and /o/). The contrasts in H1-A3* values are most distinct between tense/lax pairs across the tonal paradigm for low vowels /a/ and /æ/, and less so for high vowels /i/ and /u/. These differences likely reflect the degrees of laryngeal constriction to which the production of each vowel is inherently susceptible. More specifically, there may be less latitude to produce tense/lax contrasts with high vowels, which are inherently produced with a more raised tongue and, consequently, a greater degree of pharyngeal expansion than low vowels, which are inherently more susceptible to tongue retraction, larynx raising, and laryngeal constriction [16] [5] [17].

Table 7. *H1-A3*, by vowel and tone, all speakers*.

	/æ/	/a/	/i/	/e/	/u/	/o/
55	16	17	16	17	29	28
55+	5	9	11	8	24	24
33	20	26	16	16	26	29
33+	6	11	15	5	27	28
31	22	22	22	22	33	36
31+	2	8	5	5	27	21
35	11	--	14	11	21	26
21	4	7	8	--	30	20

For all measures of spectral tilt, values range widely across individuals, likely reflecting the different envelope within which speakers produce phonatory contrasts. As an illustration of this point, Figure 3 below shows average H1-A3* values for speakers 1-7 for the vowel /æ/. As can be seen, with the exception of speaker 1, all speakers make a consistent contrast between tense and lax pairs of tones, though some speakers appear to make those contrasts within a more or less constricted setting, and within a narrower or wider range of constriction. Speaker 3, for example, consistently distinguishes between the tense/lax pairs, but does so within a vocal setting that is relatively unconstricted, compared to speakers 4 and 7, who produce these same contrasts within a relatively constricted setting. Of the seven speakers, speaker 5 produces the contrasts within a very wide range, while speaker 1 (with the exception of the 55/55+ contrast) produces most of the contrasts within a very narrow range.

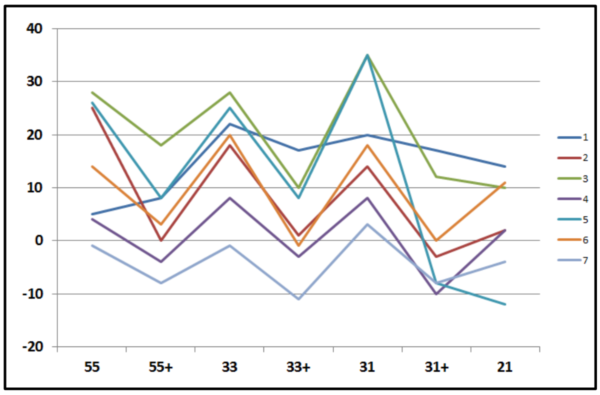

Figure 3: *H1-A3* for /æ/, speakers 1-7*.

On the whole, spectral tilt values are inversely correlated with F1 values, reflecting the relationship of these measures to laryngeal constriction (high spectral tilt values occur in less constricted sounds, which tend to have low F1 values; low spectral tilt values occur in more constricted sounds, which tend to have higher F1 values). In line with this tendency, for all seven speakers, the highest spectral tilt values correspond to lax tones, and the lowest to tense. Similarly, for all speakers, the highest F1 values correspond to tense tones, and the lowest to lax tones. However, there are individual differences in the ways that spectral tilt and F1 values pattern with specific tones. For example, for speaker 5, the highest and lowest F1 values for /æ/ are for tones 21 and 31, respectively, and the lowest and highest spectral tilt values are for those same tones. However, for speaker 1, whose formant values pattern with speaker 5's (highest for 21, lowest for 31), the lowest and highest spectral tilt values are for tones 55+ and 33, respectively. Thus, speakers differ in the ways that they combine the different cues to the tense/lax contrast, though the relationships between the cues are far from arbitrary.

4. Discussion and Conclusion

In summary, this exploratory study found that Bai speakers distinguish between tense/lax tones on the basis of three acoustic cues: F0, F1, and spectral tilt. Speakers show the greatest homogeneity in their use of F0 as a cue to individual tones, and to tense/lax pairs of tones: across all seven speakers, tense tones have higher F0 than their lax counterparts. Speakers also systematically employ F1 to distinguish between tense and lax tones. In general, across speakers, tense tones have a higher F1 than lax tones. While there are individual differences, the most constricted tones (31+, 21) usually have the highest F1, and the least constricted (31) usually have the lowest. Similarly, for all spectral tilt measures employed (H1-H2, H1-A1*, H1-A2*, and H1-A3*), tense tones tend to have lower spectral tilt values than lax tones, and the most constricted tones (31+, 21) usually have the lowest values, and the least constricted (31) the highest. However, while all speakers use these three cues, individuals differ in the way they distribute these cues among the tones, and in the degree and range of laryngeal constriction they use in producing the tones. These findings suggest that while the phonological system strongly influences the use of acoustic cues, individual portrayals of that phonology differ.

It is possible that given the wide range of articulatory choices at a speaker's disposal in producing laryngeal constriction (i.e. the differing combinations of ventricular fold incursion, larynx raising, tongue and epiglottis retraction, aryepiglottic fold compression, and pharyngeal narrowing, and the differing qualities produced by such combinations, as described in [5]), there is greater scope for individual variation in the production of contrasts based on this feature compared to contrasts that are based exclusively on pitch or vowel formants. A Bai speaker could opt, for example, to produce tense tones with differing degrees of larynx raising, a choice that would tend to result in higher pitch for tense tones relative to their lax counterparts, as well as a more open jaw setting, which would raise F1. This strategy might be sufficient to produce the tense/lax contrast, while seldom resulting in audible harshness or, as a result, low spectral tilt measurements. Another speaker could adopt a more constricted setting across all tones, but employ a greater degree of laryngeal constriction on tense tones, resulting in audible harsh voice and/or aryepiglottic trilling on some tones. Such differences in articulatory strategies may account for the variability in individuals' spectral tilt measurements, while also accounting for the tendency for the F0 and F1 results to pattern similarly across speakers.

The primary intention of this study is to generate baseline adult data for comparison with infant utterances in a study of Bai infants' acquisition of the register tone system in the second and third years of life. This study has several implications for infants' acquisition of the relevant acoustic cues (F0, F1 and spectral tilt). Current research on first language acquisition suggests that in the first year of life (especially the first six months), infants can distinguish between all speech sounds that are employed in languages of the world. Towards the end of the first year, infants lose this sensitivity in favour of developing the ability to distinguish between the sounds used in their ambient language, a result that has been found for non-tonal [18] [19] and tonal [20] [21] [22] contrasts. Thus, like infants learning other tone languages, Bai-learning infants are likely to remain sensitive to pitch differences that are used to distinguish tones. However, in learning Bai, a register tone language, infants also need to remain sensitive to the relationship of these pitch differences to the inter-related cues of vowel quality and spectral tilt, which are used in varied ways by adult speakers of the language. Currently, no research exists on the acquisition of register tone languages, so how Bai infants achieve this perceptual organization remains an open question. Our auditory and acoustic studies of Bai infants' utterances in the first year of life [10] [11] [12] show that Bai-learning infants' babbling includes a higher incidence of, and greater variability in, laryngeal constriction than the babbling of English-learning infants. It remains to be seen, however, how this feature develops in the second and third years of life, and how it interacts with the infants' use of pitch and vowel quality.

Research on infants' and young children's tone production demonstrates that tone acquisition is a protracted developmental process. While infants begin to produce tones in the second and third years of life [23][24][25][26][27][28], their production is not fully adult-like [29][30][31][32], and shows little improvement in accuracy up to the age of five [33]. The latter studies are all based on infants' production of pitch. Our own study will focus on the previously unexplored relationships between pitch, vowel quality, and spectral tilt in the acquisition of register tones. It is likely that Bai infants, like learners of other tone languages, will not fully acquire the Bai tone system until at least the age of five. However, it will be of considerable interest to see which of the eight contrasting tones Bai infants begin to produce in the second and third years of life, and which of the relevant acoustic cues (F0, F1, and spectral tilt) they begin to correlate in their acquisition of the tense/lax contrast.

5. Acknowledgements

The authors would like to thank the Social Sciences and Humanities Research Council of Canada for their support of this research. We would also like to thank Dr. Jerold Edmondson for the use of field recordings of Bai made in Yunnan, China in 2008.

6. References

[1] Esling, J. H. and J. A. Edmondson, "The laryngeal sphincter as an articulator: Tenseness, tongue root and phonation in Yi and Bai", in A. Braun and H. Masthoff [Eds.], Phonetics and its Applications, 38-51, Franz Steiner Verlag, 2002.

[2] Xu, L. and Zhao, Y., "Baiyu gaikuang" (in Chinese), in Zhongguo Yuwen, 321-325, 1964.

[3] Xu, L. and Zhao, Y., "Baiyu jianzhi" (in Chinese), Beijing, 1984.

[4] Edmondson, J. and Li, S., “Voice quality and voice quality change in the Bai language of Yunnan Province”, in Q. Dai [Ed.], Linguistics of the Tibeto-Burman Area, 17(2):49-68, 1994.
[5] Edmondson, J. A., and Esling, J. H., “The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies”, Phonology 23:157–191, 2006.
[6] Kuang, J., “Production and perception of the phonation contrast in Yi”, Unpublished manuscript, University of California, Los Angeles, 2011.
[7] Brunelle, M., “Dialect experience and perceptual integrality in phonological registers: Fundamental frequency, voice quality and the first formant in Cham,” J. Acoust. Soc. Am. 131:3088–3102, 2012.
[8] Esposito, C., “An acoustic and electroglottographic study of White Hmong tone and phonation”, J. Phonetics, 40(3):466-476, 2012.
[9] Garellek, M., Keating, P., Esposito, C., and J. Kreiman, “Voice quality and tone identification in White Hmong, J. Acoust. Soc. of Am., 133(2):1078-1089, 2013.
[10] Benner, A., Grenon, I., & Esling, J.H., “Infants’ phonetic acquisition of voice quality parameters in the first year of life”, Proc. of the 16th Intl. C. of the Phonetic Sciences, Saarbrücken, Germany, www.icphs2007.de, 2007, accessed on 28 Dec. 2013.
[11] Benner, A., “Production and perception of laryngeal constriction in the early vocalizations of Bai and English infants”, Proc. Cdn. Acoust. Assn. Annual Conf., 2010.
[12] Benner, A., & Grenon, I., “The relationship between laryngeal constriction and vowel quality in infants learning English and Bai,” Proc. of the 17th Intl. C. Phonetic Sciences, Hong Kong, http://www.icphs2011.hk, 2011, accessed 29 Dec. 2013.
[13] Esling, J.H. & Benner, A., “Laryngeal-pharyngeal ontogeny: Speech in the first several months”, Paper presented at the Intl. Child Phonology Conf., Minneapolis, MI, 2012.
[14] Esling, J.H., “University of Victoria phonetic database (version 4.0)”, Victoria, BC & Lincoln Park, NJ, 1999.
[15] Boersma, P. & Weenink, D., “Praat: Doing phonetics by computer [computer program], (version 5.3.51)”, 2013, http://www.praat.org/, retrieved 2 June 2013.
[16] Esling, J.H., “There are no back vowels: The laryngeal articulator model”, Cdn. J. of Ling., 50:13-44, 2005.
[17] Moisik, S.R. & Esling, J.H., “Evaluating the vowel space effects of larynx height using laryngeal ultrasound”, Cdn. Acoust., 39:180-181, 2011.
[18] Werker, J.F. & Curtin, S. (2005). “PRIMIR: A developmental framework of infant speech processing”, Lang. Learning & Dev., 1(2):197-234, 2005.
[19] Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P., “Infants show a facilitation effect for native language phonetic perception between 6 and 12 months”, Developmental Science, 9(2):F13-F21, 2006.
[20] Mattock, K. & Burnham, D., “Chinese and English infants’ tone perception: Evidence for perceptual reorganization”, Infancy, 10(3):241-265.
[21] Mattock, K., Molnar, M., Polka, L., & Burnham, D., “The developmental course of lexical tone perception in the first year of life”, Cognition, 106(3):1367-1381, 2008.
[22] Yeung, H., Kenny, H., & Werker, J.F., “When does native language input affect phonetic perception? The precocious case of lexical tone”, J. of Memory and Lang., 68(2):123-139, 2013.
[23] Clumeck, H., “Studies in the acquisition of Mandarin phonology”, Unpublished doctoral diss., U. of California, Berkeley, 1977.
[24] Clumeck, H., “The acquisition of tone”, in G.H. Yeni-Komshian, J.F. Kavanaugh, & C.A. Ferguson [Eds.], Child phonology: Vol. 1, Production, 257-275, NY, Academic Press, 1980.
[25] Li, C.N. & Thompson, S.A., “The acquisition of tone in Mandarin-speaking children”, J. of Child Lang., 4:185-199, 1977.
[26] Zhu, H. & Dodd, B, “The phonological acquisition of Putonghua (Modern Standard Chinese)”, J. Child Lang., 27:3-42, 2000.
[27] Ota, M., “The development of lexical pitch accent systems: An autosegmental analysis”, Cdn. J. Ling., 48(3/4):357-383, 2003.
[28] To, C.K.S., Cheung, P.S.P., & McLeod, S., “A population study of children’s acquisition of Hong Kong Cantonese consonants, vowels, and tones”, J. Speech, Lang, & Hearing Research, 56(1):103-122, 2013.
[29] Wong, P., Schwartz, R., & Jenkins, J., “Perception and production of lexical tones by 3-year-old Mandarin-speaking children”, J. Speech, Lang., & Hearing Research, 48(5):1065-1079, 2005.
[30] Yang, J., & Lee, H.-T., “Lexical variation and rime-tone correlation in early tonal acquisition: A longitudinal study of Mandarin Chinese”, Int. Symp. Tonal Aspects Lang., 126-131, 2006.
[31] Wong, P., “Acoustic characteristics of three-year-olds’ correct and incorrect monosyllabic Mandarin lexical tone productions”, J. Phonetics, 40(1):141-151, 2012a.
[32] Wong, P., “Monosyllabic Mandarin tone productions by 3-year-olds growing up in Taiwan and the United States: Interjudge reliability and perceptual results”, J. Speech, Lang. & Hearing Research, 55(5):1423-1437, 2012b.
[33] Wong, P., “Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan”, J. Acoust. Soc. Am., 133(1):434-443, 2013.