Speech Prosody 2016
31 May - 3 June 2016, Boston, USA

Cantonese Tone Discrimination using Amplitude Envelope: Implications for Cochlear Implants

Yining V. Zhou, Brett A. Martin

Speech-Language-Hearing Sciences, Graduate Center, City University of New York, USA

yzhou@gradcenter.cuny.edu, bmartin@gc.cuny.edu

Abstract

The purpose of the current study is to investigate whether amplitude envelope alone can cue the auditory discrimination of any two lexical tones in Cantonese, and to tease apart the relative contributions of three acoustic factors to lexical tone perception using amplitude envelope. Signal-correlated noise stimuli were re-synthesized based on six Cantonese lexical tones produced naturally. Thirty native listeners of Cantonese and thirty native listeners of American English were presented pairs of the stimuli, and were instructed to report whether each pair consisted of identical or different Cantonese lexical tones. The results indicated that native listeners of Cantonese discriminated each of the thirty-six lexical tone pairs in Cantonese significantly above chance and with greater accuracy and shorter reaction time than the native listeners of English. The relative contributions of three acoustic factors and one linguistic factor were delineated. These findings could potentially help improve the encoding of lexical tone contrasts for lexical tone perception in cochlear implant users, who typically have difficulties with lexical tone perception due to the limited capacity of cochlear implants in conveying F0.

Index terms: lexical tones, tone perception, phonology of tones, amplitude envelope, Cantonese, cochlear implant

1. Introduction

Amplitude envelope (AE) of a sound is defined as the amplitude fluctuation in the waveform of the sound between 2 and 50 Hz during the production of the sound [1]. Similarly, AE of a lexical tone (abbreviated as ***tone*** hereafter) is defined as the amplitude fluctuation in the waveform of the tone between 2 and 50 Hz during the production of tone [2][3][4]. It reflects the overall rising, falling or steady trend of amplitude fluctuation throughout the production of a tone [2][3]. In Cantonese, for example, a falling tone has a falling AE, whereas a rising tone has a rising AE [2].

The role of AE in tone perception has been investigated in Mandarin [3][4] and Cantonese [2]. Native listeners of Mandarin or Cantonese are able to use the AE cue alone to identify tones in their native language with above-chance accuracy [2][3][4]. These findings have far-reaching clinical implications. Cochlear implant users who speak a tone language have difficulties perceiving tones in a tone language due to the limitations of currently available commercial speech processing strategies in encoding F0 [5][6][7][8]. Experimental speech processing strategies have demonstrated promising clinical results in improving tone perception in cochlear implant users by enhancing the encoding of the AE cue [2].

The findings from research on Cantonese tone perception using AE were of special interest. A tone language uses either tone contour, tone height or both for phonemic contrasts [9]. Cantonese uses both tone contour and tone height for phonemic contrasts. That is, there are three tone contours that are phonemically contrastive: falling, rising, and level. Relative tone height is also used for phonemic contrasts, yielding six phonemically contrastive tones in Cantonese: Tone 1 (High Falling), Tone 2 (High Rising), Tone 3 (Upper Middle Level), Tone 4 (Low Falling), Tone 5 (Low Rising), and Tone 6 (Lower Middle Level) [2][9][10][11]. Thus, if AE alone can cue tone perception in Cantonese, it should be able to cue tone perception in any language. However, this needs to be tested empirically across various families of tone languages.

Research on tone perception using the AE cue is still in its infancy, with several unanswered questions. First, F0 was found to co-vary with AE in Mandarin and Cantonese tones [2][3][4], and this covariability was hypothesized to be the reason why native speakers could identify lexical tones in their native language using the AE cue alone [3]. However, in these previously published studies, a tone with a higher degree of F0-AE covariability was not necessarily identified with greater accuracy than a tone with a lower level of F0-AE covariability [2][3][4]. This suggested that the F0-AE covariability might not be the only variable affecting tone perception using the AE cue alone [2]. In fact, other factors have been reported in previous studies on tone perception using the AE cue alone [2][3]. For example, some tone types were identified with significantly higher accuracy than others [2][3][4], and some carrier syllables yielded better tone identification accuracy [3][4]. Yet, it was unclear whether these factors were acoustic or linguistic in nature.

Another unanswered question in previous studies on tone perception using AE was that the most often mis-identified tone was the rising tone in one study [3], but dipping tone in another [4], even though both studies involved only the four Mandarin tones and were comparable in all aspects of the two studies. These inconsistent findings indicated that the research on tone perception using AE leaves much to be desired.

Another issue in previous studies on tone perception using AE was the use of a word identification paradigm with signal-correlated noises (SCNs) as stimuli. These SCNs contained nothing but the AE of a Mandarin or Cantonese word. Yet, upon hearing each SCN, participants were to identify, from a list of written Chinese characters differing only in tones, the character that had the same tone as the SCN. This was a very difficult task. In a post-experiment

 10.21437/SpeechProsody.2016-247

survey [2], some participants complained that the task was almost impossible.

One way to alleviate this problem would be to use a pair-wise tone discrimination paradigm. This paradigm would only require listeners to report whether two auditorily presented SCNs contain the same tone or not. Another advantage of this paradigm is the possibility of using the results of tone discrimination to determine whether certain pairs of tone contrasts are acoustically more confusable than others. The advantage of this paradigm is increased when nonnative speakers with no exposure to tone languages are also involved because nonnative speakers make their decisions on acoustic basis only, whereas native speakers' decisions are both acoustically and phonemically based [12]. Thus, a comparison of the native and nonnative speakers' tone discrimination using the AE cue alone can help determine whether tone contrasts in a tone language are acoustically and phonemically distinguishable [12].

Previous studies on tone discrimination also suggested that the interstimulus interval (ISI) could affect phonemic or acoustic processing of tones [13]. A pair of tones presented with an ISI of 500 ms or less between the tones tended to processed or discriminated on an acoustic basis, whereas an ISI of 1500 ms or more would favor phonemic processing [13]. Thus, the relative contribution of acoustic versus linguistic information to tone discrimination can also be teased apart by manipulating the ISI.

The purpose of the current study is to use a Cantonese tone discrimination experiment to investigate the above-mentioned unanswered questions from previous studies on tone perception using AE, and to tease apart the effects of various variables on tone perception using AE. Four specific questions will be answered: (1) Can every possible pair of tone contrasts in Cantonese be differentiated auditorily using the AE cue alone? (2) Are certain tone pairs acoustically more confusable than others? (3) To what extent do acoustic variables affect tone discrimination using AE? (4) Are there any non-acoustic factors in tone discrimination using AE?

2. Methods

2.1 Participants

69 adults between 25 and 40 years of age participated in the current study. Two native speakers of Cantonese (one male and one female) produced Cantonese tones for the current study. Another seven native speakers of Cantonese served as naïve judges to rate the perceptibility of these Cantonese tones. Thirty additional native speakers of Cantonese participated in the tone discrimination study. Thirty English-speaking adults with no formal or informal learning experience of a tone language also participated in the tone discrimination study. All participants had normal hearing with pure-tone air conduction thresholds of 25 dBHL or better bilaterally at octave frequencies between 250 Hz and 8000 Hz [14]. They also had Type A tympanograms and present ipsilateral acoustic reflexes at 90dB HL at 1000 Hz.

2.2 Stimuli

The stimuli used in the current study were Cantonese tones produced in isolation with the carrier syllable /ji/ or /wai/. The use of tones produced in isolation as stimuli was to control for several variables affecting Cantonese tone perception [15][16]. Using the carrier syllables /ji/ and /wai/ would facilitate comparison with previous studies which also used these syllables [7][8]. Second, these two carrier syllables could be combined with the six Cantonese tones to constitute twelve common words in Cantonese [7][8]. Lastly, a significant effect of carrier syllable was found in previous studies using these carrier syllables [7][8], which suggested that the tone perception using the AE cue alone could be affected by carrier syllables. Thus, it might be desirable to use multiple carrier syllables to investigate the effects of carrier syllables on tone discrimination using AE.

Stimulus recording was done in a sound-attenuated booth. Each of the twelve written Chinese characters representing the twelve target syllables was presented on a computer screen to each talker who was instructed to read each character aloud, as if he/she were reading the letters on an eye exam chart during a vision examination. Each character was presented ten times, in a random order with the constraint that no character be presented twice in a row. Thus, the two talkers each produced ten tokens of each target syllable, and, the two tokens with the least deviation from the ten-token average were selected as target syllable prototypes. With two tokens of twelve target syllables produced by two talkers in the current experiment, a total of forty-eight target syllable prototypes were selected.

These syllable prototypes were then equalized in duration to remove the duration cue using a two-step procedure. The grand average duration of all of the tone tokens produced by both talkers were computed, and the duration of each of the forty-eight target syllables was equalized to the grand average using a linear interpolation method [3]. The impact of this duration equalization on the perceptibility of the target syllables was assessed using a 6AFC identification task, and each of the seven native Cantonese judges performed with 90% or higher accuracy (chance level = 16.67%). Thus, the forty-eight target prototypes were accepted and were later transformed into forty-eight SCNs using the procedure in [2][3]. Each SCN contained only the AE of the original tone prototype. For ease of discussion, these SCNs were labeled ***processed tones***.

With six tone categories in Cantonese, thirty-six different tone pairs were possible, including thirty ***Different*** pairs (e.g., Tone 1 vs. Tone 2), and six ***Same*** pairs (e.g., Tone 1 vs. Tone 1). For each ***Different*** pair, only the token with the least deviation from the ten-token average was used. For each ***Same*** pair, two different tokens of the same tone type were used. The AEs of any two tokens in a ***Same*** pair were similar but not identical as shown in Figure 1, which would favor phonemic processing instead of acoustic processing in the Cantonese participants in the current tone discrimination experiment [17].

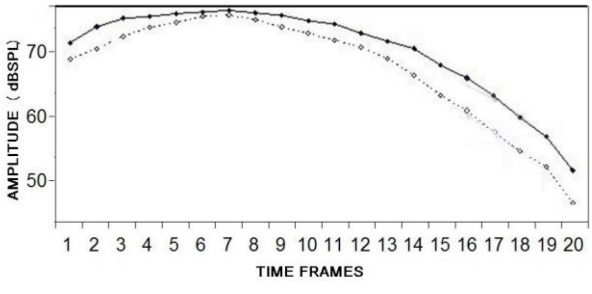

Figure 1: AEs of two tokens of Tone 1 (High Falling) by male talker with carrier syllable /ji/

With each of the two carrier syllable yielding thirty-six tone pairs, there were a total of seventy-two tone pairs in each talker block. Within each talker block, each of the seventy-two tone pairs occurred ten times randomly. Thus, there were 720 pairs of stimuli in each talker block. There was a 1500-msec silence between the two tones in each pair, and a 4-second interval between the onsets of two adjacent pairs.

There were two experimental conditions in the current experiment. Thus, there were a total of four blocks (2 talkers x 2 conditions) of stimuli. In Condition One, the original amplitude (or intensity level) of each of the tone pair was preserved. In Condition Two, the rms of each tone pair was equalized to the grand mean rms of all the tone pairs in Condition One in order to eliminate the relative AE height differences among the six Cantonese tones. Thus, a comparison of the results in both conditions could reveal the independent contribution of the relative AE height to Cantonese tone perception.

2.3 Experimental procedure

The four blocks of stimuli were presented in a random order to each participant who was instructed to press either the left or the right button of the computer mouse upon hearing each pair of tone contrasts. For half of the listeners in each language group, the left button on the mouse was labeled "Same Tone", whereas the right button was labeled "Different Tones." For the remaining listeners, the labels on the buttons were reversed. All participants were right-handed, and decided on their own to use their right hands to press the mouse buttons. The whole experiment lasted approximately three hours per listener, including a brief break between blocks.

2.4 Data analysis

Sixty participants' percent correct measures in the current study were analyzed using a 2 (groups) x 2 (conditions) x 2 (carriers) x 2 (talkers) x 36 (tone pairs) repeated measures ANOVA. To quantify the effects of different acoustic variables on tone discrimination using AE, three acoustic measures were used: F0-AE Covariability Index, AE Contour Dissimilarity Index, and AE Height Difference Index. The F0-AE Covariability Index was included because it was hypothesized to be the reason why AE could cue tone perception [3]. It was estimated using Pearson's linear correlation between the F0 contour and the AE contour of a tone [2][3]. The AE Height Difference Index measured the mean difference between two tones in a pair in terms of AE height. The AE Contour Dissimilarity Index measured the dissimilarity between the AE contours of any two tones in a pair, and was estimated using Pearson's linear correlation and computed as 1 minus Pearson's r (1 – Pearson's r).

3. Results and discussion

3.1 Tone discrimination using AE

Cantonese participants' average percent correct measures for each tone pair in Condition One (where stimuli's original intensity level was preserved) are shown in Figure 2. Performance ranged from 0.85 to 0.92 (mean = 0.88; SD = 0.05) and was significantly above the chance level of 0.5 ($p < 0.01$). This indicated that AE was effective in cueing Cantonese tone discrimination in native speakers. There were no significant differences among the thirty-six pairs of tone contrasts ($p < 0.01$).

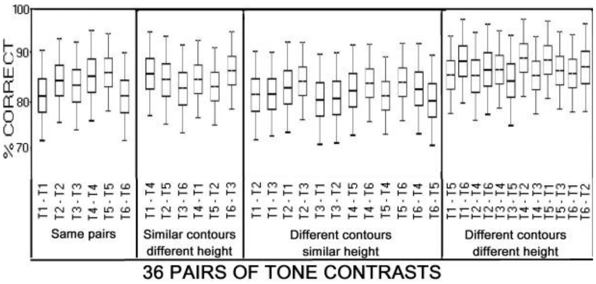

Figure 2: Cantonese tone discrimination in Native speakers using AE (boxes: 95% confidence intervals, whiskers: ±2SD)

In Condition Two, the rms of all stimuli was equalized to the grand mean rms to eliminate the relative AE height differences among the six Cantonese tones. Cantonese participants' performance deteriorated significantly only for the tone pairs containing tones with similar contours but different height before the rms equalization (mean = 0.04, SD = 0.01, $p < 0.01$). This was expected because the pairs containing tones with similar contours but different height (e.g., ***High*** Falling Tone vs. ***Low*** Falling Tone) became more like ***Same*** pairs after the rms equalization, as illustrated in Figure 3. Indeed, native speakers perceived these pairs as ***Same*** pairs 85% of the time (SD = 0.04, $p < 0.01$). This suggested that the rms equalization had indeed removed the relative AE height differences among the six tones, and consequently eliminated the AE height cue in Condition Two. A comparison of the native speakers' performance in both conditions suggested that the AE height cue was indispensable for differentiating tones with similar contours but different height in Cantonese. The relative contributions of AE height and AE contour are quantified in the next section.

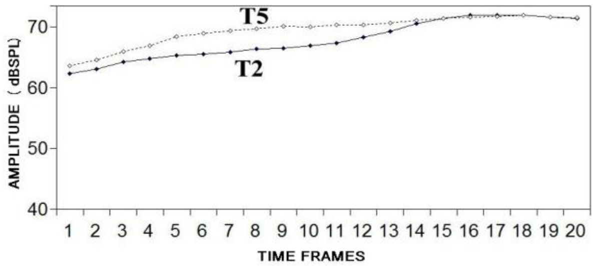

Figure 3: AEs of ***High*** Falling Tone vs. ***Low*** Falling Tone after rms equalization

3.2 Acoustic factors in tone discrimination using AE

The relative contributions of these three acoustic indices to the Cantonese participants' percent correct performance in the tone discrimination experiment were measured using a multiple regression analysis. The AE Contour Dissimilarity Index accounted for 32% of the variance, and it was statistically significant ($t_{(8639)} = 63$; $p < 0.01$). A higher AE Contour Dissimilarity Index for a pair of tones was associated with a higher accuracy rate with which the tone pair was differentiated using AE in the current study. The AE Height Difference Index accounted for 8% of the variance, and it was statistically significant ($t_{(8639)} = 7$; $p < 0.01$). Specifically, a higher AE Height Difference Index for a tone pair was associated with a higher accuracy rate with which the tone pair was differentiated using AE in the current study. The F0-AE Covariability Index had the least, but nonetheless statistically significant contribution ($t_{(1, 8639)} = 27$; $p < 0.01$), accounting for 0.4% of the variance.

Together, the three factors accounted for approximately 37.5% of the variance. This suggested that the AE contour dissimilarity between any two tones played a greater role in tone discrimination using AE than the AE height difference and the F0-AE covariability. That is, the AE contour dissimilarity between two tones was the most salient acoustic cue in Cantonese tone discrimination, and two tones with greater AE contour dissimilarity were discriminated with greater accuracy. However, since all three acoustic variables accounted for only 37.5% of the variance in Cantonese tone discrimination using AE, there could be other factors.

3.3 Other factors in tone discrimination using AE

The Cantonese participants discriminated all tone pairs in Condition One with significantly greater accuracy ($F_{(1, 29)} = 2432$; $p < 0.01$) and shorter reaction time ($F_{(1, 29)} = 1743$; $p < 0.01$) than the native speakers of English, as illustrated in Figure 4. This might be attributable to the Cantonese participants' linguistic experience with tones in their native language. As reported in previous studies of tone discrimination involving native and non-native speakers, native speakers used both linguistic and acoustic processing whereas the nonnative speakers relied solely on acoustic processing [12].

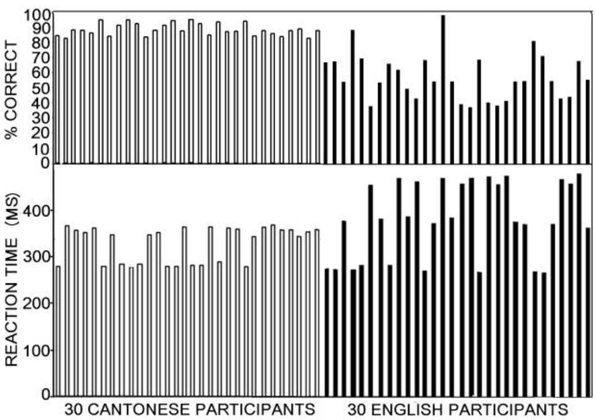

Figure 4: Cantonese tone discrimination using AE

In addition, the tone discrimination results suggested that categorical perception of tones might have been involved. The evidence came from the discrimination of the ***Same*** pairs. As illustrated in Figure 1, each ***Same*** pair in the current study consisted of two different tokens of a Cantonese tone produced with the carrier syllable by the same Cantonese talker. The acoustics of the two tokens were very similar, but not identical. Twenty of the thirty English participants mistakenly considered these two tokens as representing two different Cantonese tones at significantly above chance level ($t_{(19)} = 2.08$; $p = 0.028$), whereas all of the Cantonese participants correctly perceived these two tokens as representing the same Cantonese tone with significantly above chance level (mean = 0.83, SD = 0.045, $p < 0.01$). This suggests that the Cantonese participants perceived these two tokens phonemically, instead of using pure acoustic criteria as in the case of the native speakers of English. The Cantonese participants' phonemic perception of Cantonese tones may have somewhat constrained the acoustical factors discussed in the previous section. These findings were consistent with results from previous studies in favor of categorical perception of naturally produced tones [18].

4. Summary and Conclusion

This study shed light on all four of the research questions of the current study. First, the AE cue alone could cue tone discrimination. This finding was a unique contribution of the current study to the research on tone perception using AE because this was the first tone discrimination study using AE.

Second, AE could cue auditory discrimination of any possible pair of tone contrasts in Cantonese with equal effectiveness. This was another unique contribution of the current study to the research on tone perception using AE. As discussed above, tone languages in the world use only tone height, tone contour or a combination of both for phonemic tonal contrasts [9]. Cantonese uses both tone height and tone contour for phonemic contrasts [9][10][11]. Therefore, it is theoretically possible for AE to cue tone perception in any tone language.

Third, the current study teased apart the relative contribution of F0-AE covariability, AE contour and AE height to Cantonese tone perception using AE. This was the third unique contribution of the current study to the research of tone perception using AE. Specifically, AE contour seemed to be the most important acoustic cue for Cantonese tone perception using AE, whereas the relative AE height played a less important but still crucial role in Cantonese tone perception using AE. This finding is consistent with results from previous studies on perception of naturally produced tones [11]. F0-AE covariability played a weak but statistically significant role in Cantonese tone discrimination using amplitude envelope alone. Together, these three acoustical factors accounted for 37.5% of the variances in the current Cantonese tone discrimination.

Lastly, linguistic factors such as experience with tone language and phonemic processing of tones seemed to play an important role in native speakers of Cantonese in the current experiment, and somewhat constrained the operation of the acoustic factors. This was the fourth unique contribution of the current study to the research on tone perception using AE because previously published studies did not investigate the contribution of these factors to tone discrimination using AE.

The current study has at least one clinical implication. As discussed above, cochlear implant users who speak a tone language have difficulties perceiving tones and therefore overall speech in their native language due to the limitations of currently available commercial speech processing strategies in encoding F0. Experimental speech processing strategies have demonstrated promising results in improving tone perception in cochlear implant users by enhancing the encoding of the AE cue [5]. Since the current study suggests that it is theoretically possible for AE to cue tone perception in any tone language, these experimental speech processing strategies could possibly be adapted to enhance the encoding of the AE cue in any tone language, and thus improve tone perception in cochlear implantees who speak any tone language.

5. Acknowledgements

The authors thank Glenis Long, Winifred Strange, Douglas Whalen, Puisan Wong and two anonymous reviewers for their helpful suggestions and comments.

6. References

[1] Rosen, S., "Temporal information in speech: acoustic, auditory and linguistic aspects", Philosophical Transactions: Biological Sciences, 336, 367-373, 1992.

[2] Zhou, Y. V., and Martin, B.A., "The role of amplitude envelope in Cantonese lexical tone perception: Implications for cochlear implants," Proceedings of the 6th International Speech Prosody International Conference, 629, 2012.

[3] Fu, Q. J., and Zeng, F. G., "Identification of temporal envelope cues in Chinese tone recognition", Asia Pacific J. Speech, Language and Hearing, 5, 45-47, 2000.

[4] Kuo, Y., Rosen, S., and Faulkner, A., "Acoustic cues to tonal contrasts in Mandarin: Implications for cochlear implants", Journal of the Acoustical Society of America, 123(5), 2815-2824, 2008.

[5] Luo, X., and Fu, Q. J., "Enhancing Chinese tone recognition by manipulating amplitude envelope: implications for cochlear implants", Journal of Acoustical Society of America, 116(6), 3659-3667, 2004.

[6] Fu, Q. J., Zeng, F. G., Shannon, R. V., and Soli, S. D., "Importance of tone envelope cues in Chinese speech recognition", Journal of Acoustical Society of America, 104(1), 505-510, 1998.

[7] Yuen, K. C., Tong, M. C. F., van Hasselt, C., Yuan, M., Lee, T, and Soli, S. D., "Cantonese lexical tone recognition from frequency-specific temporal envelope and periodicity components in the same versus different noise band carriers", Cochlear Implants International, 10(S1), 148-158, 2009.

[8] Yuan, M., Lee, T., Yuen, K.C.P., Soli, S.D., Tong, M.C.F. & van Hasselt, C.A., "Band-specific temporal periodicity enhancement for Cantonese tone perception with noise-excited vocoder," Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 694-697, 2007.

[9] Yip, M., Tone. Cambridge: Cambridge University Press, 2002.

[10] Fok, Y.Y., "A Perceptual Study of Tones in Cantonese", Hong Kong: University of Hong Kong Press, 1974.

[11] Gandour, J. T., "Perceptual dimensions of tone: evidence from Cantonese", Journal of Chinese Linguistics, 9 20-36, 1981.

[12] Lee, Y.S., Vakoch, D.A., & Wurm, L.H., "Tone perception in Cantonese & Mandarin: A cross-language comparison," Journal of Psycholinguistic Research, 25, 527-542, 1996.

[13] Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P., "Perception of lexical tones across languages: Evidence for a linguistic mode of processing," International Conference on Spoken Language Processing, 1996, 2514-2517, 1996.

[14] American Speech-Language-Hearing Association, "Guidelines for Audiologic Screening," Rockville, MD, 1997.

[15] Wong, P. and Diehl, R., "Perceptual normalization for inter- and intra-talker variation in Cantonese level tones", Journal of Speech, Language, and Hearing Research, 46, 413-421, 2003.

[16] Khouw, E., and Ciocca, V., "Perceptual correlates of Cantonese tones", Journal of Phonetics, 35(1), 104-117, 2007.

[17] Strange, W., Speech Perception & Linguistic Experience: Issues in Cross-Language. Timonium, MD: York Press, 1995.

[18] Zheng, H. Y., Minett, J.W., & Wang, W. S. Y., "The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study", Language and Cognitive Processes, 1, 1-31, 2010.