

Distinction between Mandarin Tones 2 and 3 for L1 and L2 Listeners

Yung-hsiang Shawn Chang
University of Illinois at Urbana-Champaign

Among the four Mandarin lexical tones, Tone 2 and Tone 3 have been found to be the most confusable perceptually for both native speakers and non-native speakers. This study investigates whether the overall tone duration serves as a perceptual cue and whether L1 and L2 listeners weigh the cue in the same manner in distinguishing the two tones. In doing so, two perceptual experiments using duration-normalized stimuli and non-normalized stimuli respectively were conducted on L1 and L2 Mandarin listeners. The results show that native listeners do not heavily rely on duration for the Tone 2- Tone 3 distinction, whereas L2 listeners give more perceptual weight to duration. The overall tone duration is concluded to be an important parameter for L2 tone training and acquisition.

1. Introduction

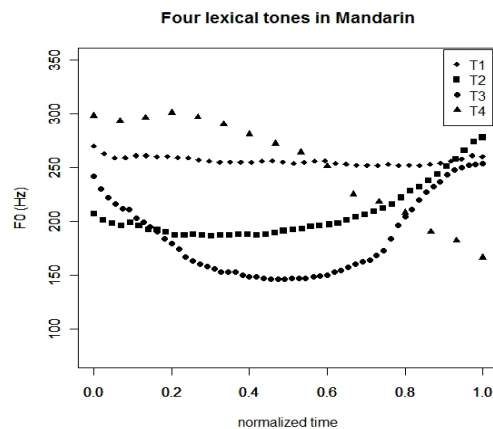
Mandarin is a tonal language and uses pitch to distinguish lexical meanings. Perception studies on Mandarin tones have found the distinction between Tone 2 (T2) and Tone 3 (T3) to be the most difficult for native speakers (Chuang et al. 1972; Shen & Lin 1991; Zue 1976) and non-native speakers (Kiriloff 1969; Wang et al. 1999; Shih et al. 2010) alike. While the F0 turning point and the F0 difference between the tonal onset and the turning point are often manipulated to study the perception of T2 and T3 (e.g., Blicher et al. 1990; Moore & Jongman 1997; Shen 1993), one temporal cue—the overall tone duration—has not received much attention in tone perception studies. On the one hand, T3 is found to be consistently longer than T2 such that duration is suggested to be a perceptually relevant acoustic cue for tone distinction (Blicher et al. 1990). On the other hand, Shih (2007) argued that in conversational speech, T3 is most likely to become the shortest tone because its rising tail is not realized. The conflicting views of overall duration serving as a cue for perception motivates the current study to investigate whether duration indeed plays a role in T2-T3 distinction and whether the durational cue is weighed in the same manner for native (L1) and second language (L2) speakers. To this end, three research questions are asked: 1. Does duration normalization, which keeps all other spectral cues intact, affect the accuracy of T2-T3 distinction? 2. If the performance difference of distinction of duration-normalized vs. non-normalized T2-T3 pairs is reflected in accuracy, is it also reflected in reaction time measures? 3. Do native

speakers and non-native speakers exhibit the same perceptual patterns in T2-T3 distinction?

1.1. Acoustic characteristics of Mandarin T2 and T3

Mandarin T2 and T3 have been traditionally characterized as having a high-rising pitch contour and a low falling-rising pitch contour respectively (Chao 1948). When produced in citation form, T2 and T3 appear distinguishable acoustically, as seen in Figure 1, where all tones are time normalized. In the literature, high confusability for the T2-T3 pair is suggested to be due to similar F0 contours between the two tones (Kirilloff 1969; Chuang *et al.* 1972; Gandour 1978; Li and Thompson 1978) or a similar rise in F0 during the final portion of the two tones (Blicher *et al.* 1990).

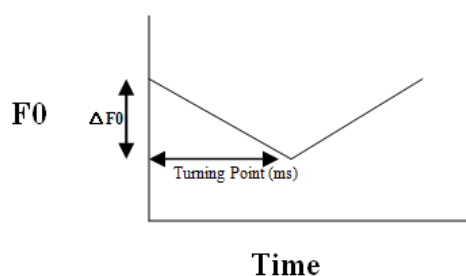
Figure 1. Four lexical tones in Mandarin



To further characterize the two tones, besides both having a concave shape, T2 and T3 was argued by Shen (1990) to differ to a great extent in the timing of the turning point in the pitch contour (schematized in Figure 2). The turning point is claimed to occur close to the onset in Tone 2 and toward the middle of Tone 3. Shen & Lin (1991) tested this claim by having subjects listen to stimuli that varied in the timing of the turning point. They concluded timing dose constitute a perceptual cue for differentiating T2 and T3. Shen & Lin also found that the degree of the initial fall (i.e., F0 difference between the tonal onset and the turning point, shorthand as $\Delta F0$) to be correlated with the turning point, as they reported perception of stimuli with mismatched turning point and $\Delta F0$ to incur more errors. Moore & Jongman (1997) further investigated this potential covariance of the two parameters by using synthetic stimuli, whose turning point and $\Delta F0$ were systematically manipulated, in a perception experiment. They found that both acoustic

dimensions trigger categorical identification and “operate in tandem as perceptual cues to tones” (p. 1871).

Figure 2. Schematized turning point and the degree of the initial fall ($\Delta F0$) for a contour tone (figure reproduced from Moore & Jongman 1997)



While most production and perception studies suggest that turning point and $\Delta F0$ are the main acoustic cues for T2-T3 distinction, Blicher et al. (1990) found syllable duration to also have an effect on perception. In fact, an intrinsic durational difference among four Mandarin tones has been noted as early as in Lin (1965), with T3 being the longest, and T4 being the shortest. An acoustic study by Shen in 1990 confirmed T3 to be consistently longer than T2 when produced in isolation. Assuming that listeners may use this durational difference to enhance the perceptual contrast between T2 and T3, Blicher et al. (1990) varied T2-T3 along a duration continuum and studied whether lengthening would shift the perceptual boundary. Their findings show that listeners tend to produce T3 responses when a stimulus is longer in duration. That is, lengthening may constitute a cue in favor of T3 perception. However, Shih (2007) challenged the view that duration can be a reliable cue for tone identification by pointing out the divergent reports in the literature on Mandarin tone duration. She argued that T3 is longer than T2 only when produced in isolation; in conversational speech, T3 most likely becomes the shortest among all four tones because its rising tail is not realized.

A difference in intrinsic amplitude was found for the four Mandarin tones in Chuang et al. (1997), with Tone 3 having the lowest amplitude, and Tone 4 the highest. In this regard, the contribution of the amplitude contour to Mandarin tone identification has also been studied in the literature. The findings diverge with respect to the effect of amplitude contours on tone perception. Lin (1988) manipulated the amplitude contour but found no effect in the presence of F0. In the absence of F0, however, Whalen & Xu (1992) found amplitude contours to be a useful cue for their subjects in identifying Mandarin tones, although their results do not particularly generalize the effect of amplitude to T2-T3 distinction. Fu et al. (2003) also studied how the amplitude cue contributes to tone perception. They concluded that the amplitude cue only contributes to T3 and T4

discrimination and that use of amplitude in tone recognition has a large variability across listeners.

Besides F0 and temporal envelop cues, Yang (1989) reported that vowels have an effect on T2-T3 distinction, although such an effect was not observed in discrimination of other tone pairs. Yuan (2003) followed up on this study using selective adaption tests to see if tone identification is influenced by repeated presentation of a stimulus that shares similar acoustic features. He found that only the *da2* and *da3* adaptors, but not *bi2* and *bi3*, had a selective adaptation effect on the identification test on the *da2-da3* continuum. Based on this discrepancy, Yuan suggested that tone perception may not be independent from segmental perception.

One acoustic attribute oftentimes associated with T3 that is considered extralinguistic in Mandarin tones is creakiness (or laryngealization). Yu (2010) reported in her corpus study that laryngealization occurred most frequently for T3 and 68% of the time across speakers. While Davison (1991) and Yu suggest laryngealization as a feature to aid tonal perception, Yang (2011)'s study provides evidence for categorical perception in discriminating all tonal pairs over continua, varying voice qualities but holding F0 constant. That is, voice quality can contain functions beyond auditory enhancement in Mandarin tone perception.

1.2. Native vs. non-native speakers' perception of lexical tones

Many studies found that only native listeners exhibit categorical perception of lexical tones, whereas non-native listeners do not (e.g., Hallé *et al.* 2004; Xu *et al.* 2006). Therefore, for tones to be used in speech communication, they must be perceived by the hearer as linguistic objects instead of musical pitches. Before learners acquire phonemic boundaries between tones, any within category phonetic variation may be misperceived as linguistically relevant. In addition to a psychoacoustic approach of studying lexical tone perception, a flux of recent research pursue the same line of research using neurophysiological methods and has reported neural response differences between native and non-native speakers of tone languages. For example, in Zatorre & Gandour's (2008) fMRI study, they found that English speakers who have no prior experience with a tone language exhibit a lack of left-hemisphere dominance in the processing of lexical tones. In contrast, there is strong left-hemisphere involvement in native speakers' processing of lexical tones. However, the processing difference between native speakers and non-native speakers is not strictly dichotomous but has scalar dimensions. Many studies (e.g., Bent 2005; Huang 2004) that investigate native vs. non-native speakers' perception of lexical tones found that L1 and L2 listeners weigh various acoustic cues differently, which results in different perceptual patterns. In distinguishing Mandarin tones, Gandour (1978) pointed out that tone height and direction are the most important dimensions. He found that his native speaker subjects attended more to the direction dimension whereas the non-native counterparts directed more attention to the pitch height of the stimuli. However, the non-native like perceptual system is not unmodifiable. Many auditory training studies show that given intensive laboratory training, L2 learners improved in

perception and production of non-native sounds as their cue weighting approximates that of native speakers', both at segmental and suprasegmental levels. Especially with regards to Mandarin tone perception, Wang et al. (1999) reported significant improvement for their American English subjects over a short training period. And the suprasegmental perceptual modification was even retained six months after training.

1.3. Operationalization of the study

The current study investigated whether duration is used as a perceptual cue for Mandarin T2-T3 distinction in addition to F0 information. Although Blicher et al. (1990) has reported that syllable lengthening would enhance the percept of T3 for native listeners, we are not certain as to whether when the durational difference between T2 and T3 is removed, T2-T3 discriminability would be similarly affected for native and non-native listeners. To answer this question, two auditory discriminations tasks were conducted on native listeners and L2 learners of Mandarin. Our hypothesis was that listeners would have a lower accuracy score or a longer reaction time in distinguishing the stimuli with the duration cue removed. To avoid the practice effect being a confounding factor, we gave the task with unaltered stimuli (where T3 was naturally longer than T2) first. If the practice effect is strong, we might see both groups perform better in the second task or we might see no statistically significant performance difference between the two tasks. If we observe significantly lower accuracy score and/or longer reaction time in performing the second task, we could conclude that the effect of duration as an auditory enhancement is robust enough to override any practice effect.

2. Methodology

2.1. Auditory stimuli

Six native speakers of Mandarin (3 male and 3 female) were recruited to record the stimuli for this study. They were graduate students at the University of Illinois at Urbana-Champaign, aged between 25 and 30. All recordings were conducted in a sound-treated booth in the Phonetics Lab at UIUC. An AKG C520 head-worn condenser microphone was used to record acoustic signals onto a Marantz PMD570 recorder. The stimuli were two syllables *ma* and *da* carrying four Mandarin tones. Each syllable-tone combination was repeated five times in a randomized order. The stimuli were visually presented to the speakers in E-prime.

All recordings were later screened and only one female speaker's data was used because 1. she had no creaky voicing in her T3 production, which could be a potential cue for T2-T3 distinction, and 2. her T3 production was fully realized. We did not use more than one speaker's data as perception of tone has been suggested to be a talker-contingent process (Moore & Jongman 1997). At the end, only two sweeps of each syllable-tone combination was chosen out of five repetitions to limit the number of stimuli pairing.

Duration normalization was performed using the PSOLA function in Praat. This method preserves the spectral structure of the syllables regardless of modified duration values. The duration changes were made consistently in that the longer syllable was shortened to match the duration of the shorter syllable within a tone minimal pair. More specifically, *ma2* and *ma3* were all duration-normalized to 470 ms while *da2* and *da3* to 370 ms. The stimuli were normalized for RMS amplitude at 60 dB. At the end of post-processing, the stimuli were screened by two native speakers for accuracy and all of them were correctly identified.

2.2. Subjects for the discrimination tasks

Eight Mandarin native speakers and 8 non-native speakers participated in the discrimination tasks. The native speakers included 3 males and 5 females, aged between 24 and 30. The non-native speakers were intermediate-level Mandarin learners. They included 2 males and 6 females, aged between 20 and 24.

2.3. Experiment procedure

The experiment consisted of 2 forced-choice discrimination tasks. Task 1 used the stimuli where T3 was naturally longer than T2; task 2 used duration-normalized stimuli. In each task, 30 items (16 T2-T3 pairs and 14 non-T2-T3 pairs as distractors) were randomized and presented in E-Prime. All our stimuli can be represented by various lexical items. Since we were not particularly looking into the lexical frequency effect on the T2-T3 distinction, all stimuli were presented auditorily and no orthographic information was provided. The inter-stimulus interval was set at 1.3 seconds, in which there was a 1 second beep. Both accuracy and reaction time were logged automatically in E-Prime. Subjects were asked to press the button Y when the tones in the word pairs were the same and N for different tones. The subjects were encouraged to respond as quickly and accurately as possible.

3. Results

For the purpose of this paper, only the T2-T3 discrimination data were analyzed. Table 1 shows the percentage accuracy scores in the two tasks and for both native and non-native listener groups. We see that the native group's performance almost reached the ceiling. Non-native group had a rather high accuracy as well, although with a greater variance. The result is generally in line with Shih & Lu's (2010) tone training study in that the only few tone discrimination mistakes native listeners made were all on T2-T3 pairs and that T2-T3 distinction yielded most of the errors for L2 learners of Mandarin. In task 2, both native and non-native groups scored lower than in Task 1, where the durational cue was present. For statistical analysis, a mixed-design ANOVA was run with group as a between-subject variable, and task type as a within-subject variable. The results revealed significant main effects of group ($F(1,14)= 8.34, P=.012$), task ($F(1, 14)=33.51, p <0.001$), and task*group interaction ($F(1, 14)=12.064, P<.01$). Pairwise comparisons with a Bonferroni correction exploring the task*group interaction effect

showed a significant difference between two groups in Task 2 ($p < .05$), but not in Task 1. For the native listener group, there was no significant difference between the accuracy scores in Task 1 and Task 2 ($p=0.0796$), while for the non-native group, a significant difference was observed ($p < .001$). The results suggest that our intermediate-level learners performed no differently from their native counterparts in Task 1, where the T3 tokens were all naturally longer than T2. In contrast, in Task 2, where duration normalization was performed on the stimuli, L2 learners did significantly worse than L1 speakers.

Table 1. Accuracy score (%) for native and non-native listeners in 2 tasks

	Native	Non-native
Task 1	0.99 (SD=0.02)	0.95 (SD=0.06)
Task 2	0.97 (SD=0.047)	0.86 (SD=0.07)

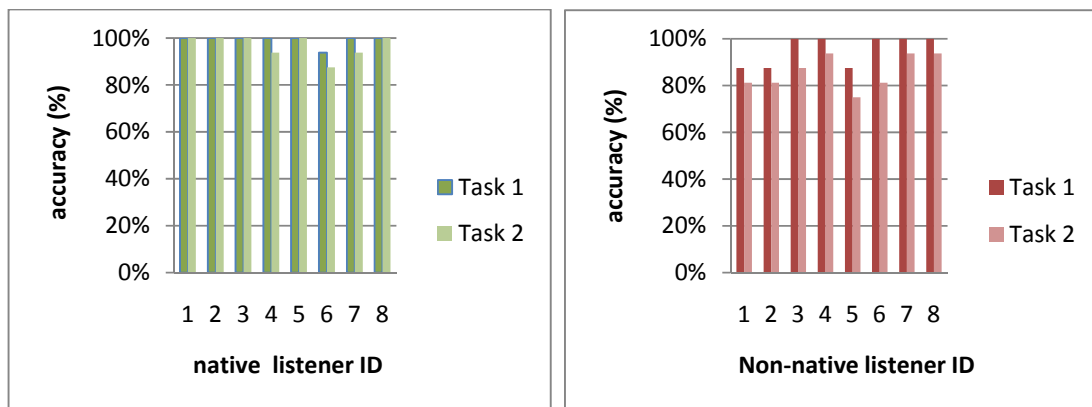
In terms of reaction time, the descriptive statistics are presented in Table 2. The reaction time data exhibits the same pattern as the accuracy data in that the native listener group has a lower mean reaction time than the non-native group in both tasks. And more reaction time was needed for both native and non-native listeners when the durational cue for T2-T3 distinction was absent. Note that the standard deviation is fairly large for both groups in both tasks. We reason that some listeners had decided on the answer as soon as the second word finished playing, while other listeners could have taken their time to retrieve the first word they heard after hearing the second word. Once again, a mixed-design ANOVA was conducted on the reaction data. The results showed a significant main effect of group ($F(1,14)= 4.695$, $P < .05$) and task ($F(1, 14)= 12.099$, $p < .01$), but no significant task*group interaction. That is, native listeners had a shorter reaction time than their non-native counterparts in both tasks. Absence of the durational cue in distinguishing T2 and T3 would induce longer reaction time for both groups.

Table 2. Reaction time (ms) for native and non-native listeners in 2 tasks

	Native	Non-native
Task 1	514.6 (SD=110.69)	720.23 (SD=217.12)
Task 2	573.17 (SD=128.46)	754.85 (SD=234.05)

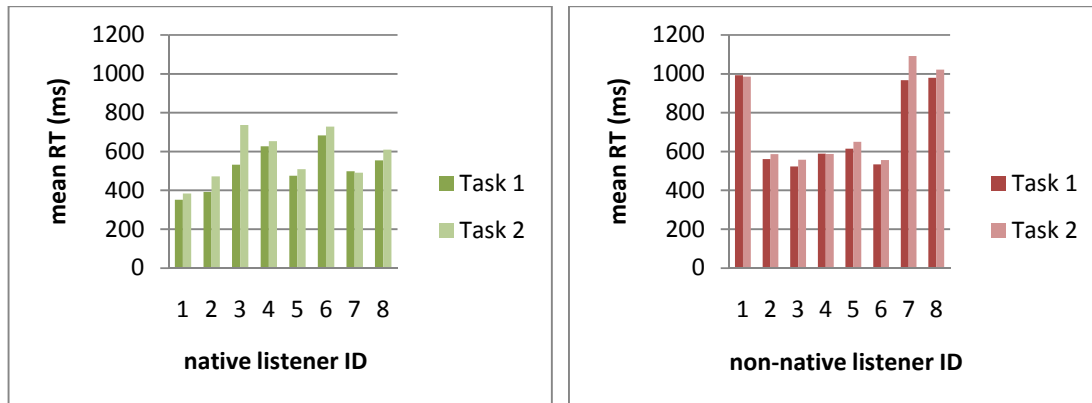
Besides the group mean performance, we were also interested in within-group variability to see if there is any perceptual pattern within or across individuals in the two tasks. In figure 3, all listeners' scores in Task1/Task2 were plotted back-to-back. We see that almost all of our native listeners reached the ceiling in the first task. In the second task where the stimuli were duration normalized, three listeners scored lower. As for our non-native listeners, five out of eight scored a hundred percent in the first task. No one got all the items correct in the second task; in fact, all non-native listeners' scores dropped in Task 2.

Figure 3. Individual accuracy scores (%) in both tasks



The individual reaction time data was plotted in Figure 4. For both native and non-native listeners, their reaction time mostly ranged between 400-600 ms, except for non-native listeners 1, 7, 8. Note that non-native listeners 7 and 8 had the highest accuracy scores compared to other non-native listeners. Therefore, there could be some accuracy-speed trade-offs for these two listeners. From Figures 3 and 4, we generally observe a longer reaction time across subjects in the second task. Only native listener 7 and non-native listeners 1 and 4 show the opposite pattern. However, the reaction time difference between the 2 tasks for these speakers was noticeably smaller than for all other speakers such that this divergence may be negligible.

Figure 4. Individual reaction time (ms) in both tasks



Although orthographic information was purposely not provided for the auditory stimuli in the experiment, we seem to observe a lexical frequency effect on the T2-T3 distinction. In analyzing the T2-T3 discrimination data, we pooled the data of *ma* and *da*. Upon a closer look at the *ma* and *da* data respectively, we see more errors on *da2-da3* pairs for both native listeners and non-native listeners. We have two speculations regarding this pattern. First, the lexical frequency difference between the *ma* and *da* syllables may play a role. Also, although the stimuli were correctly identified by two native speakers, whether all tone productions were the equally same/different has not been acoustically quantified. That is, it may well be the case that the pitch contours of *da2-da3* were more similar, and hence more confusable, than *ma2-ma3*.

4. Discussion and Conclusion

Previous literature on Mandarin tone perception generally agrees upon F0 being the most important cue in the T2-T3 distinction. Given that T3 is consistently longer than tone 2 in citation form, Blicher et al. (1990) tested whether duration difference can be used to enhance the perceptual contrast. While Blicher et al. did find syllable lengthening to auditorily enhance the percept of T3, the current study took a step further to investigate the role of duration in T2-T3 distinction. Our results show that native speakers' accuracy did not suffer in discriminating two duration-normalized tones. However, the reaction data indicates that native listeners would take significantly more time to respond when the durational cue was not available. In contrast, non-native listeners' performance exhibited significantly lower accuracy and significantly longer reaction time where the durational cue was absent. Our findings suggest that native and non-native listeners weigh the durational cue differently. For native listeners, syllable duration may be an enhancing cue that facilitates T2-T3 distinction. For non-native listeners, however, duration serves more as a primary cue in addition to F0. This psychoacoustic account for the performance difference between native and non-native listeners as well as natural stimuli and duration-normalized stimuli is compatible with previous L2 acquisition

literature in that native speakers and non-native speakers give different weight to various acoustic properties in perception (e.g., Bohn 1995; Flege et al. 1997).

In terms of L2 acquisition of lexical tones, some pedagogical implication can be drawn from the different perceptual patterns observed in our L1 and L2 listeners. It is not uncommon that in teaching Mandarin T3, language instructors generally exaggerate the syllable duration in order to fully realize its F0 contour. As a result, length differences can be taken as one of the primary cues for distinction between T2 and T3. Indeed, our study found that duration only appeared to be an enhancing cue for our native listeners, whereas L2 listeners' discrimination accuracy dropped significantly in the absence of the durational cue. Therefore, giving duration-normalized T2-T3 pairs may allow L2 learners to improve in perception as their cue weighting is shifted from duration to the F0 dimensions.

At last, we acknowledge that some native speakers do consistently produce T3 shorter than T2 even in isolation, as their T3 is not fully realized. It is unclear to us whether they also perceive T3 to be shorter than T2. In future studies, it will be of our interest to include another experiment where T3 tokens are shorter than T2 and see whether this durational warping will hinder or facilitate the discrimination of T2 and T3.

REFERENCES

- Alexander, J., Wong, P. C. M., & Bradlow, A. (2005). Lexical tone perception by musicians and non-musicians. *Proceedings of Interspeech 2005*.
- Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, 18(1), 37-49.
- Bent, T. (2005). *Perception and Production of Non-Native Prosodic Categories*. Unpublished Ph.D. dissertation, Northwestern University, Chicago.
- Bohn, O.-S. (1995). Cross-language speech perception in adults: first language transfer doesn't tell it all. In Strange, W. (ed.). *Speech perception and linguistic experience: theoretical and methodological issues*. Timonium, MD: York Press.
- Chao, Y. R. (1948). *Mandarin Primer*. Cambridge: Harvard University Press.
- Chuang, C.-K., & Hiki, S. (1972). Acoustical Features and Perceptual Cues of the Four Tones of Standard Colloquial Chinese. *The Journal of the Acoustical Society of America*, 52 (1A), 146.

- Davison, D.S. (1991). An acoustic study of so-called creaky voice in Tianjin Mandarin. *Working Papers in Phonetics, UCLA*, 78. 50–77.
- Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In *Speech perception and linguistic experience: issues in cross-language research* (W. Strange, ed.), pp. 233–277. Baltimore: York Press.
- Flege, J. E., Bohn, O-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470.
- Fu, Q.J., Zeng, F.G. 2000. Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech Language and Hearing* 5, 45-58.
- Gandour, J. (1978). The Perception of Tone. In V. A. Fromkin (Ed.), *Tone: A Linguistics Survey* (pp. 41-76). New York: Academic Press.
- Gandour, J. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics*, 12(2).
- Hallé, P. A., Chang, Y., & Best, C. T. 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32 (3). pp. 395-421.
- Howie, J. M (1976). *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge University Press, Cambridge.
- Huang, T. (2004). Language-specificity in auditory perception of Chinese tones. Unpublished Doctoral Dissertation, Ohio State University, Columbus.
- Kiriloff, C. (1969). On the Auditory Perception of Tones in Mandarin. *Phonetica*, 20(2-4), 63-67.
- Li, C. N., & Thompson, S. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2), 185-199.
- Lin, M.-C. (1965). Yingao xianshiqi yu Putonghu shengdiao yingao texing [The pitch indicator and the pitch characteristics of tones in Standard Chinese]. *Shengxue Xuebao* [Chinese Journal of Acoustics] 2 (1), 8–15.
- Lin. M.-C. (1988). Pulonghua silcngdiao de shengxue texing he zhijue zhengzhao [The acoustic characteristics perceptual cues of tones in Standard Chinese]. *Zhongguo*

- Yuwen [Chinese Language], 204. 182- 193.
- Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of Acoustical Society of America*, 102, 1864-1877.
- Shen, X-N. S. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics*, 18, 281-295.
- Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech*, 34, 145-156.
- Shen, X. S., Lin, M., & Yan, J. (1993). F0 turning point as an F0 cue to tonal contrast: A case study of Mandarin tones 2 and 3. *Journal of the Acoustical Society of America*, 93, 2241.
- Shih, C. (2007). *Prosody learning and generation*. Berlin: Springer.
- Shih, C., Lu, H.-Y., Sun, L., Huang, J.-T., & Packard, J. (2010). An Adaptive Training Program for Tone Acquisition. In *Proceedings of Speech Prosody 2010*.
- Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, 106, 3649-3658.
- Wang, Y., Sereno, J. A., Jongman, A. & Hirsch, J. (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *J. Cogn. Neurosci.* 15, 1019–1027.
- Whalen, D.H., Xu, Y. 1992. Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica* 49, 25-47
- Xu, Y., Gandour, J., & Francis, A. (2006). Effects of language experience and stimulus complexity on categorical perception of pitch direction. *Journal of the Acoustical Society of America*, 120 (2). pp. 1063-1074.
- Yang, R. (2011). The Phonation Factor in the Categorical Perception of Mandarin Tones. In *Proceedings of ICPHS XVII*.
- Yang, Y. (1989). Yuanyin he Hanyu Shengdiao de Ganzhi Wenti [Vowels and the Perception of Mandarin Tones]. *XinLi XueBao [Acta Psychologica Sinica]*, 34

(1).

- Yu, K. M. (2010). Laryngealization and features for Chinese tonal recognition. In *Proceedings of Interspeech 2010*.
- Yuan, J. (2003). Selective Adaptation Tests on Mandarin Tone2 and Tone3. *Proceedings of ICPhS XV*, 1719-1722.
- Zatorre, R. J., & Gandour, J. T. (2008) Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society Biological Sciences*, 363 (1493). 1087-1104.
- Zue, V. W. (1976). Some perceptual experiments on the Mandarin tones. *Journal of Acoustical Society of America*, 60 (1), 45.