The Role of Auditory Feedback in Mandarin Tone Production in Native and Non-native Mandarin Speakers

Honor Thesis

Yiran Chen

Vanderbilt University

Abstract

The current study examines the role of auditory feedback in Mandarin tone production among native and non-native speakers of Chinese through two production tasks where participants are asked to read and pronounce pseudo-words that share characteristics with Mandarin words. Every participant was tested in both tasks, with and without auditory feedback. In the Tonal Pseudo-word Task, participants are asked to read nonsensical Mandarin words with consonants and vowels shared in English. In the Lexical Pseudo-word Task, participants read nonsensical pseudo-mandarin words without tonal markings and were asked not to worry about the tones. The tones produced were masked by a low-band filter as using Praat and judged independently by native Chinese speakers for accuracy. The quality of the consonants and vowels that are unique to Mandarin were evaluated by experts' judgment. The design of the study was based in part on earlier studies of the role of auditory feedback in singing tunes from memory (Erdemir and Rieser, 2016; Beck, Rieser, and Erdemir, 2017), which showed that expert musicians depended less on accurate tone production than did non-musicians. The goal of the present study is to find out if native Mandarin speakers rely less on auditory feedback than college students who are studying Mandarin.

Results showed that non-native speakers are significantly more affected than native speakers and that non-native speakers find the tonal task more challenging than the unfamiliar phoneme task. The original hypothesis that native speaker of a tonal language, comparable to expert singer, has better motor representation of tone production so that they are resilient to the deprivation of auditory feedback.

The Role of Auditory Feedback in Mandarin Tone Production in Native and Non-native Mandarin Speakers

It fits experiences from everyday life to think that we rely on auditory feedback when we speak. For example, people tend to speak more loudly when wearing headphones and when struggling to hear over the telephone, people often raise their voice, anticipating reciprocal adjustments. This study is focused on the effect of disrupted auditory feedback on the pitch when native and non-native speakers of Mandarin Chinese are asked to speak pseudo-words that resemble Mandarin, either with or without the tonal component.

Previous research indicates that changes in auditory feedback induce changes in pitch during speech production. For example, demonstrations of the Lombard effect suggest that people raise intensity and pitch when auditory feedback of the speech is disrupted (Patel, 2008). People with acquired hearing losses show deteriorations in their speech in terms of pitch, including an overall higher frequency and abrupt and uncontrolled intonation in speech (Cowie, 1992). These results indicate that pitch while speaking can be altered through changes in auditory feedback.

In addition to the effects of auditory feedback on pitch while speaking, additional research shows that auditory feedback plays a role in pitch production while people sing. One of these studies (Erdemir and Rieser, 2015) served as a basis for the present study. The participants in this study were asked to sing a “Happy Birthday to You” in one condition with normal auditory feedback and in another where their ability to hear themselves was blocked by a masking stimulus. Half of the participants were musically trained singers and the other half were not musically trained. The results showed that the pitch production of participants in both groups benefited from the presence of auditory feedback, but that there was a much greater improvement in the singing by the musically untrained participants.

Results from these two and other previous studies of pitch when speaking and when singing suggest that auditory feedback plays an important role in the production of pitch variation. In linguistic context, besides prosodic features like intonation, there is a specific phenomenon that consists of pitch variation. Tonal language is a specific type of language, in which different tones of the same word lead to different meanings. For example, Mandarin is a tonal language. In Mandarin, *yi* in Tone 1 means “clothes”, means “aunt” in Tone 2, means “chair” in Tone 3 and means “difference” in Tone 4. Lexical tones are defined both by registers (frequency of pitch) and pitch contour (Deutsch, 2012). The four lexical tones in Mandarin are demonstrated clearly by the Figure 1 cited below (Jiang et al, 2012). The study of tonal language and auditory feedback would yield more in-depth insight on pitch perception and production in linguistic context: Do speakers rely on auditory feedback in the same way to produce tonal language they rely on their auditory feedback in producing prosody or singing?

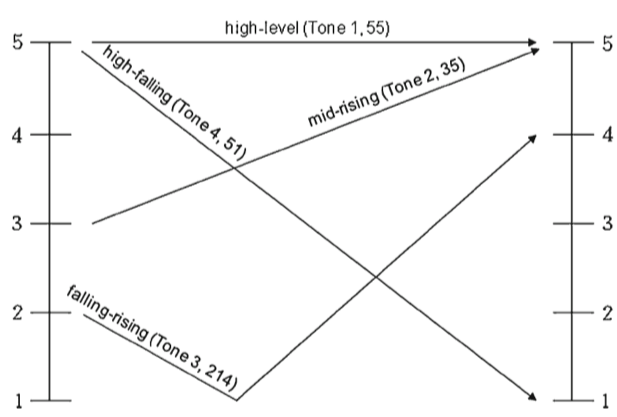

Figure 1 Pitch Contour in four Mandarin Tones

Auditory feedback has been shown to play an important role in Mandarin tone production. For example, Jones (2002) showed that native Chinese speakers compensated for sudden auditory perturbation by lowering their pitches when producing the first tone (the high flat one). It is reasonable to hypothesize that absence of auditory feedback would contribute to variation of the frequency and contour of the pitches even in native speakers.

However, native speakers and non-native speakers may approach speaking Mandarin differently, regardless of the proficiency of the learners. The exposure to the language in the first few months after birth already makes a difference in identification and discrimination of crucial phonetic information of that language. For example, Mattock and Burnham compared the speech tone discriminations of Chinese infants and English infants over the course of one year (2009). The results showed that English infants' discrimination of lexical tones declined between 6 and 9 months of age due to lack of exposure. Callan (2004) used neuro-imaging to compare the brain activation of native English speakers and Japanese English-learners when they discriminated "r-l" contrasts. His results suggested that different neurological pathways were being used by native speakers and learners when processing phonetic information. Thus, a question arises: Do native speakers and non-native speakers use their auditory feedback differently when producing tonal languages?

If lexical tone processing shares common mechanisms with music processing, we could be informed about this question by the research done on singing without auditory feedback. Several studies showed that musical pitch perception and lexical pitch perception employ common mechanisms. Jiang and his colleagues showed that Chinese amusic people were impaired in a tone distinguishing task compared to control (2012). As we can see, deficit in musical pitch perception would affect the ability in perceiving lexical tones, which support some shared mechanism between the two. Another study showed that there is higher percentage of absolute pitch, the ability to identify or re-create a given note without any reference, among Chinese conservatory students than their American counterparts (Deutsch, 2006). The theory suggests that the early intensive exercise on the fine-grained pitch distinction, required for native speakers of Chinese, make them more skilled at pitch detection and labeling.

After reviewing evidence on the commonality between lexical and musical pitch perception, it is a natural next step to examine whether the similar expert resilience would be embodied in lexical pitch production. Revisiting the musical pitch production by Erdemir and Rieser, we can see that found that, when asked to sing with and without auditory feedback, professional singers performed significantly better than instrumentalists and non-musicians (2016). According to the results, less reliance on auditory feedback might be explained by their superior ability to utilize kinesthetic feedback. If expertise in a specific pitch-variation production task contributes to higher ability to utilize kinesthetic feedback, which will be orosensory feedback in this case, then it would be reasonable to hypothesize that non-native speakers would be more heavily affected by the deprivation of auditory feedback on their tone production.

This current study was designed to investigate how the absence of auditory feedback affects the speech, especially the production of Mandarin tones, of native speakers of Mandarin and students learning Mandarin in an American university with the following hypotheses. First of all, I hypothesized that non-native speakers’ performance would be more severely affected by the deprivation of auditory feedback than native speakers. As the Erdemir and Reiser’s study suggested, native speakers, having higher expertise in Mandarin production, could rely on their motor representation more to succeed at the production.

My second hypothesis, informed by anecdotal experiences as well as research on second language phonological acquisition (Callan, 2004; Mattock, 2009), is that the tonal task is will be more challenging for non-native speakers because the unfamiliarity with treating pitch variation as phonemic contrasts.

Therefore, inferred from the potential difficulty of the tonal task, my third hypothesis is that on different production task, the importance of auditory feedback would differ. It is less directly informed through the current literature. However, I am curious whether non-native speakers rely on auditory feedback even more when asked to produce the tonal feature in a word unit. Granted that there are prosodic pitch variation on phrasal level, it is still less common and salient to semantic comprehension. Therefore, I hypothesized that for non-native speakers, the absence of auditory feedback would lead to worse performance in a tonal task than a non-tonal task, such as unfamiliar phonemes.

Furthermore, I am curious about the nature of the expertise of native speakers: Is it the fact that they have been speaking Mandarin for a longer period of time that matters? Or is it the fact that they are exposed to the language at a very early period of their life? Although this study is not designed to examine acquisition of tone production, I still want to address whether the critical period has a determining role on the way we use our auditory feedback and motor feedback during speech production. If it is years of experience that rendered native speakers more resilient to the deprivation of auditory feedback, non-native speakers who have longer experience with Chinese would be more resilient than their inexperienced peers because their ability to utilize their motor feedback improved over years. However, it could also be the case that there is only a critical period in which we could learn to better synthesize auditory and motor feedback required to produce a tonal language. In that case, after a period of familiarizing themselves with the pronunciation rules in Chinese, non-native speakers wouldn't gain any more chance to learn how to utilize their motor representation even when they learn Chinese for longer. My hypothesis is that experienced native speakers would perform better at the production tasks and will be more resilient to the disruption of auditory feedback.

My next hypothesis regards the commonality between music and lexical training. Just as Deutsch's study with conservatory students in two countries suggest that Chinese students' exercise on lexical tones is transferrable onto musical pitch skill, I hypothesized that non-native speakers who had more years of music training would perform better at the tonal tasks.

Method

Design Of the Study

The goal of the study was to investigate the role of auditory feedback in speaking Mandarin, for native speakers and for students learning Mandarin. To assess this, adults were audio-recorded while attempting to read Mandarin-like pseudo words out loud. The accuracy of their spoken productions was scored independently by two native Mandarin speakers.

There were three independent variables. One was Mandarin Expertise. About half of the participants were native speakers and the others were undergraduate students who had taken two or more courses in Mandarin. A second was the presence or absence of Auditory Feedback. Half of the subjects participated first in the Auditory Present condition and the others participated first in the Auditory Absent Condition. And a third was the task. In one task participants were asked to read pseudo-words that followed the rules of Mandarin and involved tonal inflections (this is the Tonal Inflection Task). In the other task participants were asked to read pseudo-words that followed the rules of Mandarin but involved little or no tonal inflection. Half of the subjects participated first in the Tonal Inflection Task and the others participated first in the Non-Tonal Task.

Thus, the design of the study was a 2 (Mandarin Expertise) by 2 (Auditory Feedback Present or Absent) by 2 (Tonally Inflected Pseudo-words or Not-inflected Pseudo-words) design, with repeated measures on the last two factors.

Participants

A total of 39 Vanderbilt undergraduate students participated. Twenty were native Mandarin speakers and 19 were students who had taken more than two semesters of Chinese in courses offered at Vanderbilt. Only students who have taken Chinese more than 2 semesters will be considered because we want to ensure that they have enough instruction and practice to know the rules of producing Mandarin tones.

Non-native speaker participants were recruited through the researchers' own connection with two student organizations on campus. Some participants are residents in the McTyeire International House, which is a language interest house where residents dedicate one hour everyday to practice a language. Ten of the nineteen non-native speakers are residents in the Chinese language Hall. None of the participants speak another tonal language. The researcher is also a volunteer in the Language Corner Program under Vanderbilt Undergraduate Chinese Association, which is dedicated to help Chinese learners with Chinese and conversational skills. The other nine non-native speakers were recruited through the program. The native speakers (20) are all International Chinese students at Vanderbilt. They were recruited based on the connection of the researcher.

Materials

Babble mask

Following Erdemir and Rieser (2016), in one condition people were unable to hear themselves because their hearing was blocked with an auditory masking stimulus called a Babble mask. It is a .wav sound file that is synthesized from unintelligible sound of 20 adults talking simultaneously. The babble-mask has been shown to be an effective auditory feedback blocker in Erdemir and Rieser's study about singing without auditory feedback (2015).

The loudness of the Babble Mask was set individually. Before the testing, each participant wore the headphones with the babble mask set at 40dB loudness. While they spoke they were asked to increase the loudness until they could not longer hear themselves. As in earlier studies, the self-determined loudness settings averaged 85 dB.

Pseudo-word Stimuli in Tonal Task

All the auditory stimuli in this task are pseudo words in Chinese, which means that they are made up from phonemes in Mandarin and English combined following phonetic rules in Mandarin. The International Phonetic Alphabet Mandarin and General American was used to make sure that the phonemes are present in both Mandarin and in English. This is intended to control the subjects' familiarity with the presented phonemes to make sure that if the non-native speakers performed worse on the tonal task, their worse performance would not be explained by their difficulties with the phonemes themselves. All the pseudo words were bi-syllabic words in order to inhibit the participants to associate the pronunciation with single syllable words that they are familiar with. The reason behind is that it is highly possible that a single syllable sound would make sense, more so to native speakers than non-native speakers because native speakers have a bigger vocabulary. However, the arbitrary combination of two syllables is less likely to have a meaning, therefore prohibiting participants to have a semantic association. When designing the list of stimuli, it is ensured that every possible combination of two tones is represented, except for third-third combination because we don't want the established tone sandhi to complicate the rating process. The stimuli list is attached in Appendix A.

Pseudo-word Stimuli in the Unfamiliar Phoneme Task

All the auditory stimuli in this task were also pseudo words in Mandarin Chinese. However, in the non-tonal task, all the pseudo words were made of phonemes that are unique to Mandarin. Only the existing combination of phonemes are employed. This is meant to be a control task to examine how the non-native speakers would do when the tonal feature of Mandarin is not emphasized. The stimuli list is included in Appendix B.

Procedures

The tests lasted 10-15 minutes and were conducted in a variety of study rooms in residence halls, classrooms and conference room that were convenient for the participants. The test rooms were generally quiet and free of distractions. The researcher was equipped with a laptop to record the spoken words, a pair of noise canceling headphones, and a portable microphone, together with lists of pseudo-words for the participants to read.

When the participants completed the consent process, the researcher collected background information in the format of structured interview which included questions about their language experience and musical background. Then, they were given instructions of the first condition of the task. The orders in which the subjects will go through the tonal task or the non-tonal task, and respectively with blocked auditory feedback or natural auditory feedback were varied, so that approximately half participated in the natural auditory feedback first for both the Tonal Task and the Non-tonal Task. The Babble Mask was played on a multitrack media player, namely Audacity, and the recording was set up on another track. In each condition, participants are instructed to read at a rate that was comfortable for them and to try to read consistently at this rate.

During the natural auditory feedback trials, participants just read the stimuli off the printed paper in a natural speed. In blocked auditory feedback condition they read while hearing the babble mask that had been individually adjusted in volume, as described above. After the entire experiment, participants were debriefed.

After the experiment, the researcher first typed the participants' answers to the biographical questions about their languge experience and musical training. The researcher then rmeasured the duration of each trial and the individually noted the intensity of the recordings. After this, the volume of the recordings was adjusted to the same level through Praat. The identity of the speakers was masked through the "change gender" function in Praat. After the condition information was recorded, each recording was numbered and randomized so that the researcher and the other rater (both native speakers of Mandarin Chinese) were unaware of the auditory feedback condition. The pseudo-word spoken for each trial was coded by the scorer as either showing a disfluency (these included prolonged hesitations, repetitions, and corrections) and the frequencies of the disfluencies were tallied.

In order for that the non-native accents would not affect the ratings, a band filter was put on so that the raters focused only on the pitch contour of the utterances. The raters are asked to mark down, with tone symbols in Mandarin, the tones that they thought they heard in the utterances. The raters' answers were then compared to the original stimuli list. For the Unfamiliar Phoneme Task, two raters listen and transcribe the spoken words they heard in recordings in pinyin. Again, their answers were compared to the original stimuli sheet and incongruencies were counted as errors. Since the words in the non-tonal task are not written with tone markers, all the pinyin that have the same form with a word in English were eliminated from analysis (e.g. *can, song*). After the errors were scored, the percentages of correct pronounciations were calculated.

Results

As expected the speakers spoke more rapidly and more loudly (this is known as the Lombard effect; see Pick, Siegel and Fox, 1989) in the Auditory Feedback Absent condition. In addition, the presence or absence of auditory feedback did not significantly influence the instances scored as disfluencies. Since the overall speed of speaking, loudness of speaking and speech disfluencies are not a focus of the experiment, they are not discussed further in this paper.

Each participant underwent 20 repeated trials in each of four conditions when auditory feedback was present or absent and the pseudo-word stimuli were either tonal or not-tonal. The spoken words were scored independently as correct or incorrect by two independent judges, both of whom were native speakers of Mandarin. The judgments were easy and the inter-rater agreement was 99.8%.

The dependent variable consisted of the proportions of spoken words that were scored as correct. The percentage correct scores were submitted to a Mandarin Expertise (native speaker of Mandarin versus non-native) by Auditory Feedback (present or absent) by Pseudo-word Type (Tonal versus Non-tonal) analysis of variance with repeated measures on the second two factors.

The proportion correct scores were submitted to a three way analysis of variance (Mandarin Expertise, Auditory Feedback, and Pseudo-word Type) with repeated measures on the last two factors. The results of the anova showed statistically significant main effects for Mandarin Expertise, Auditory Feedback and Pseudo-word Type. And in addition, the results showed a statistically significant Mandarin Expertise x Auditory Feedback interaction and a statistically significant Mandarin Expertise x Pseudo-word Type interaction. Each of these is discussed below. The test of the possible three-way interaction did not approach statistical significance.

The results of the analysis of variance showed that there is a main effect of expertise, meaning that on average, native speakers do better under all conditions than non-native speakers ($F(1, 37)= 17.859$, $p=.000$). In addition, there was a statistically significant main effect of auditory feedback, showing subjects spoke with fewer errors with auditory feedback than without it ($F(1,37)= 16.779$, $p=.000$). Finally, the main effect of Pseudo-word Type was statistically significant, ($F(1,37)=6.539$, $p=.015$). From the means, one can see that people generally did better at the non-tonal task (the mean for the Tonal Pseudo-words was 89.66, and the mean for the Unfamiliar Phoneme Pseudo-words was 94.62).

Two of the three possible two-way interactions were statistically significant. First of all, the effect of auditory feedback on production accuracy significantly differed for the native speakers and non-native speakers ($F(1, 37)=6.37$, $p=.016$). In addition there was a statistically significant interaction between Pseudo-word Type and Mandarin Expertise ($F(1, 37) =8.678$, $p=.006$). As shown in Figure 3, when auditory feedback is collapsed, non-native speakers did better at the Unfamiliar Phoneme task. However, there is no significant interaction between task and auditory feedback, which means that on the same group of people, the auditory feedback doesn't help one task more than the other ($F(1,37)=1.193$, $p=.282$). The three-way interaction is not significant either, which means the degree of either interaction between two variables doesn't differ on the two different levels of third variable.

I designed experimental tests of the possible effects of auditory feedback and pseudo-word type on the pronounciations by native speakers and non-native speakers. But in addition, I wished to explore the possible relationships of musical training (especially to tone production) and of the number of years of studying Mandarin by the non-native speakers. Correlation have been performed respectively to examine years of musical training and performance on the tonal task, but no significant correlation was found ($r_{aud}= .156$, $p_{aud}= .523$; $r_{no\ aud}= .238$ $p_{no\ aud}=. 327$). For non-native speakers, we didn't find a significant correlation between number of years of Chinese course taken and performance on the tonal task ($r_{aud}= .036$, $p_{aud}= .884$; $r_{no\ aud} =.029$, $p_{no\ aud} = .905$).

Discussion

It is evident that native speakers would perform better than non-native speakers in general and that people would perform better with auditory feedback than without. Mean comparison shows that the deprivation of auditory feedback didn't significantly impair native speakers' production, however did impair non-native speakers'. Two other important hypotheses regarding the interaction of variables were supported. Statistic results showed that deprivation of auditory would affect both group to different extent and that when auditory is deprived, non-native speakers perform better in the non-tonal task than in the tonal task. However, null result is found for the hypothesis that in non-native speakers, the role auditory feedback would be different in both tasks.

On one hand, consistent with the singing without hearing experiment (Erdemir & Rieser, 2016), in pitch production task, novices rely on auditory feedback more than the experts. In our study, it is supported that non-native speakers benefited more from being able to hear themselves while speaking Mandarin. It is also true that non-native speakers find the tonal task more challenging than the non-tonal task. Therefore, it is also possible that the pitch component differentiate lexical tones from other non-native phonological contrasts. In other words, the nature of phonological contrasts makes production of some novel sounds harder to learn than others even when they are all non-native to the second language learners.

However, on the other hand, statistical analysis didn't support the third hypothesis that auditory feedback would affect the participants differently in the two tasks. It is possible that I failed to find the results because of ceiling effect- both native and non-native speakers performed well on the tasks. However, looking at the means, I cannot detect noticeable in that direction: the accuracy rate dropped for 3% in both conditions for non-native speakers. In other words, although tonals are harder to produce for non-native speakers, auditory feedback is not any more important for them during the production to improve them.

The clear difference between native and non-native speakers makes me wonder what factors can account for the dramatic expert resilience. It is undisputed that native speakers have more exercise and experience with Mandarin production. However, if the difference is solely attributed to the experience, there should be a correlation between years of experience with Chinese and nonnative speakers' performance in the tasks. Admittedly, null results can't lead to any conclusion, but the current results in this study on the correlation between years of Chinese and performance is inconsistent with what the theory would predict. In fact, there is not any trend of improvement on the tasks as experience with Chinese increases in the current non-native sample. The results are consistent with the early exposure theory that states that the intensive experience with the fine-grained tonal feature during critical period is the only way to acquire the quality of tone production and the way to synthesize auditory feedback and kinesthetic feedback. In future, a study with a bigger sample, of a bigger range with non-native speakers' experience would provide a more definite answer to this question.

There are several limitations to this study. First of all, there is overall a ceiling effect in the data: both native and non-native speakers did very well on both tasks. In order to address the ceiling effect and potentially answer the question on whether auditory feedback's importance differs by task, I will need to increase variability in both groups. It can either be achieved through improving sensitivity of the ratings. Instead of making the "right/wrong judgment", I can ask raters to provide an evaluation of accuracy on a scale from 1 to 5. I can also intentionally recruit non-native speakers from different level of Chinese courses or report to have different amount of experience with Chinese.

In addition, I didn't address certain confounds. Both groups of our participants can speak both English and Mandarin Chinese, respectively as their L1 and L2. It is possible that the L2 proficiency of native Mandarin speakers are higher than L2 proficiency of native English speakers because native Mandarin speakers' English is good enough for them to receive higher education in English. Therefore, there is concern that that the L2 proficiency could be a potential confound because it may represent more deliberate training on pronunciation and experience with learning and comparing phonetics. In future studies, it is better to incorporate proficiency tests in both language and match the L2 level of both groups with some mathematical tools.

References

Beck, S., Rieser, J., and Erdemir, A. (2017). Singing without hearing: A comparative study of children and adults singing a familiar tune. **Psychomusicology: Music, Mind and Brain, 27**, 122-131.

Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native-and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory–auditory/orosensory internal models. NeuroImage, 22(3), 1182-1194.

Cowie, R., & Douglas-Cowie, E. (1992). Postlingually acquired deafness: speech deterioration and the wider consequences (Vol. 62). Walter de Gruyter.

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical perioda). The Journal of the Acoustical Society of America, 119(2), 719-722.

Erdemir, A., & Rieser, J. J. (2016). Singing Without Hearing. Music Perception: An Interdisciplinary Journal, 33(5), 546-560.

Jiang, C., Hamm, J., Lim, V., Kirk, I., & Yang, Y. (2012). Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. Memory & Cognition, 40(7), 1109-1121.

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: studies of Mandarin tone production. Journal of Phonetics, 30(3), 303-320.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. Infancy, 10(3), 241-265.

Patel, R., & Schell, K. W. (2008). The influence of linguistic content on the Lombard effect. Journal of Speech, Language, and Hearing Research, 51(1), 209-20.

Pick, H. L., Siegel, G., and Fox, P. (1989). Inhibiting the Lombard effect. Journal of the Acoustical Society of America, **85**, 894.

Xu, Y., Larson, C. R., Bauer, J. J., & Hain, T. C. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *The Journal of the Acoustical Society of America*, *116*(2), 1168-1178.

Longer Bibliography

Beck, S. L. (2014). *DIFFERENCES IN USE OF AUDITORY FEEDBACK DO NOT ACCOUNT FOR AGE-RELATED CHANGES IN CHILDREN'S SINGING* (Doctoral dissertation, Vanderbilt University).

Burnett, T. A., Senner, J. E., & Larson, C. R. (1997). Voice F 0 responses to pitch-shifted auditory feedback: a preliminary study. *Journal of Voice*, *11*(2), 202-211.

Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native-and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory–auditory/orosensory internal models. NeuroImage, 22(3), 1182-1194.

Cowie, R., & Douglas-Cowie, E. (1992). Postlingually acquired deafness: speech deterioration and the wider consequences (Vol. 62). Walter de Gruyter.

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical perioda). The Journal of the Acoustical Society of America, 119(2), 719-722.

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395-421.

Jiang, C., Hamm, J., Lim, V., Kirk, I., & Yang, Y. (2012). Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. Memory & Cognition, 40(7), 1109-1121.

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: studies of Mandarin tone production. Journal of Phonetics, 30(3), 303-320.

Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). *Perception and production of Mandarin Chinese tones*. na.

Larson, C. R. (1998) Cross-modality influences in speech motor control: the use of pitch shifting for the study of F0 control, Journal of Communication Disorders, 31, 498–503.

Larson, C. R., Burnett, T. A., Kiran, S. & Hain, T. C. (2000) Effects of pitch-shift velocity on voice F0 responses, Journal of the Acoustical Society of America, 107, 559–564.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. Infancy, 10(3), 241-265.

Patel, R., & Schell, K. W. (2008). The influence of linguistic content on the Lombard effect. Journal of Speech, Language, and Hearing Research, 51(1), 209-20.

Saffran, J. R. (2003). Absolute pitch in infancy and adulthood: The role of tonal structure. Developmental Science, 6(1), 35-43.

Wu, J. L., & Yang, H. M. (2003). Speech perception of Mandarin Chinese speaking young children after cochlear implant use: effect of age at implantation. *International journal of pediatric otorhinolaryngology*, *67*(3), 247-253.

Xu, Y., Larson, C. R., Bauer, J. J., & Hain, T. C. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *The Journal of the Acoustical Society of America*, *116*(2), 1168-1178.

Yip, M. (1995) Tone in east Asian languages. In The handbook of phonological theory (J. A. Goldsmith, editor). Cambridge: Blackwell

Figures

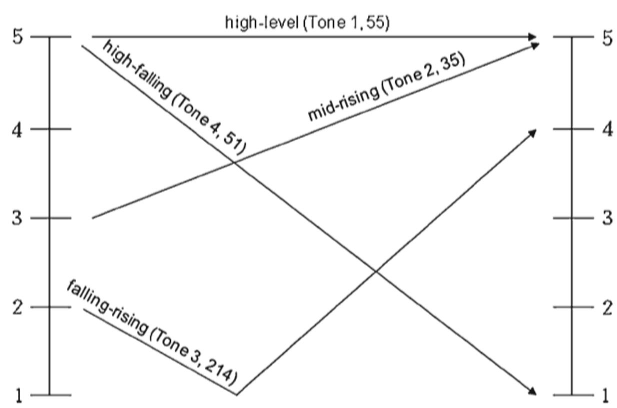

Figure 1 Pitch Contour in Mandarin Tones

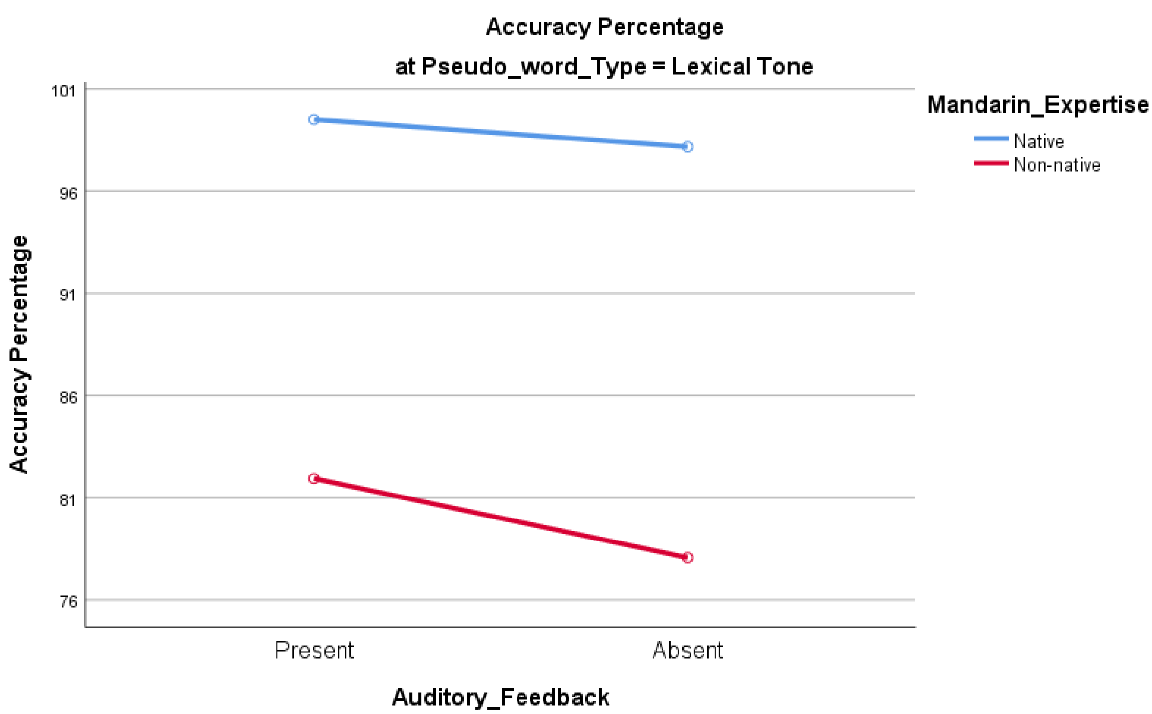

Figure 2. Accuracy of performance of native and non-native speakers with and without auditory feedback in the tonal task

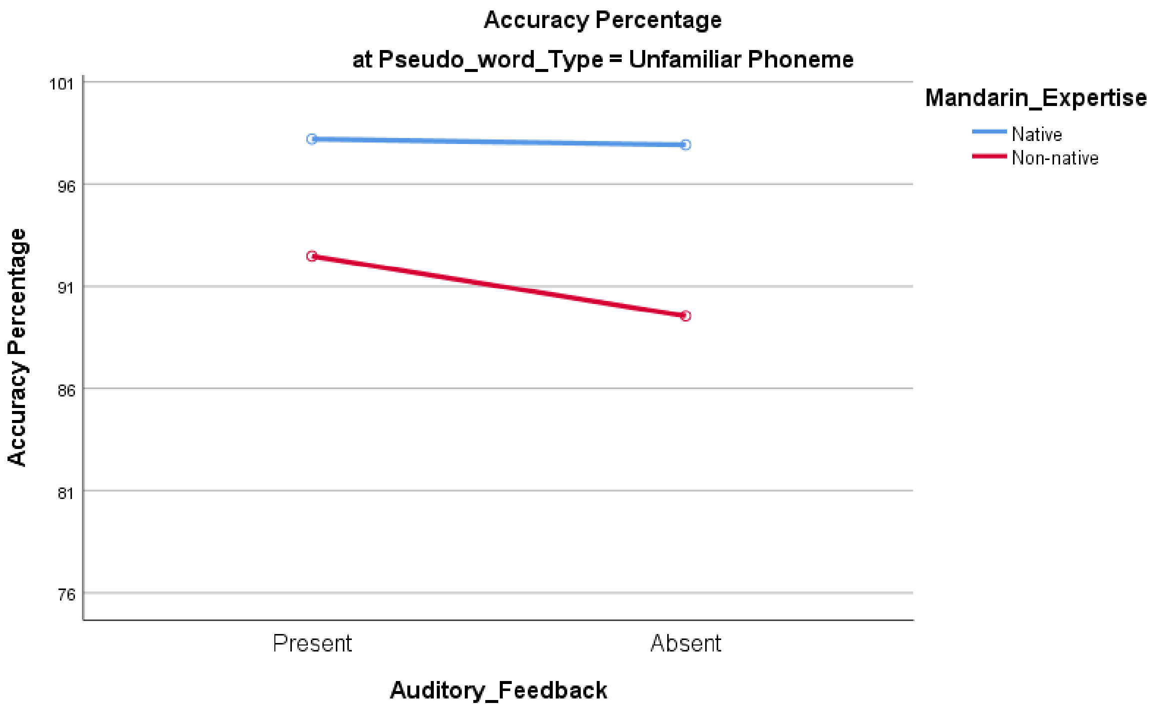

Figure 3. Accuracy of performance of native and non-native speakers with and without auditory feedback in the unfamiliar phoneme task

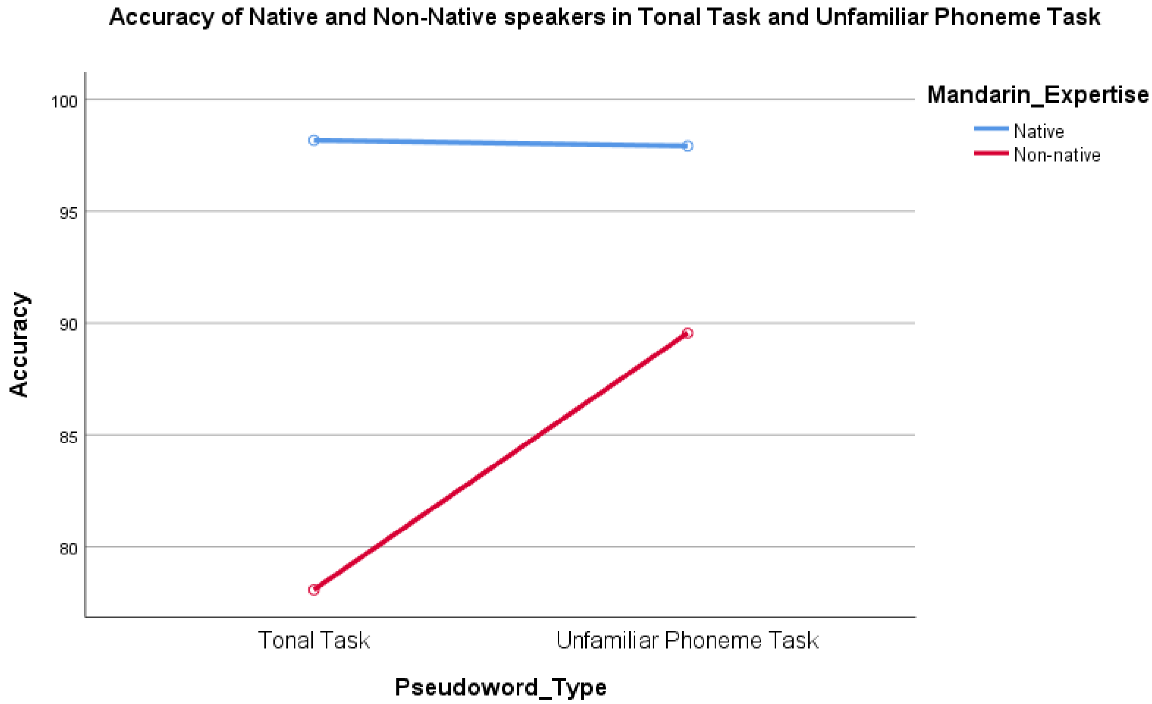

Figure 4. Accuracy of performance of native and non-native speakers in the tonal task and the unfamiliar phoneme task

Appendix A: Stimuli for Tonal Task

pāi dī

féi tāo

nǐ fā

gài mī

bēi mái

náo yé

pǐ bái

taì méi

dāo bǐ

péi dǎi

wài bǎo

māo nài

kǎ bài

dí kào

kài fà

Appendix B: Stimuli for Non-Tonal Task

1. zhuo,
2. zheng,
3. san,
4. cun,
5. shan,
6. jiong,
7. qian,
8. yong,
9. quan
10. zhua,
11. yuan,
12. shen,
13. chuang
14. xiang,
15. zun,
16. jian,
17. suo,
18. jun,
19. qia,
20. zhuan,
21. zhui,
22. chan,
23. xiong,
24. chuo,
25. xin,
26. qin,
27. chui,
28. cong,
29. yun,
30. juan
31. zi,
32. zuan
33. qi,
34. xi,
35. qing,
36. suan,
37. zang,
38. zhan,
39. jiang,
40. jing,
41. cuo,
42. sen,
43. xing
44. zhen,
45. seng,
46. chun,
47. zan,
48. qiang,
49. shuo,
50. jia,
51. shi,
52. cheng,
53. cang,
54. yue,
55. ji,
56. ci,
57. zeng,
58. si,
59. xia,
60. song,
61. sun,
62. chen,
63. can,
64. zhi,
65. zuo,
66. chi,
67. jin,
68. zhuang
69. xian,
70. cuan
71. sheng,
72. chuan,