10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Musical Memory and Pitch Discrimination Abilities as Correlates of Vocal Pitch Control for Speakers with Different Tone and Musical Experiences

Li-Hsin Ning

National Taiwan Normal University, Taiwan
lihsin@ntnu.edu.tw

Abstract

This paper investigates whether there is a correlation between pitch-related abilities (musical memory and pitch discrimination) and vocal responses to pitch perturbation in the following five groups: L2 learners of Mandarin (beginner or advanced), native Mandarin speakers without formal musical training, native Mandarin speakers specializing in instrument playing and native Mandarin speakers specializing in vocal performance. In the musical memory task, the participants had to judge whether two musical phrases were the same. In the adaptive pitch discrimination task, they had to listen to a series of two tones and judge whether the second tone was higher or lower than the first one. In the production task, they had to vocalize /a/ at a steady pitch under perturbed auditory feedback.

Musical memory and pitch discrimination abilities had to do with tone and musical experiences. Mandarin-speaking musicians and vocalists performed the best, followed by native Mandarin speakers without formal musical training. The performance of the L2 groups was the worst; however, L2 advanced learners were still better than L2 beginners. It turns out that the ability to control vocal pitch in face of pitch perturbation was correlated with musical memory ability and pitch discrimination ability.

Index Terms: music memory ability, pitch discrimination ability, pitch-shift responses

1. Introduction

Music and speech share some acoustic features, such as fundamental frequency (F0), spectral characteristics, duration, and intensity. Research on the link between music and speech has shown that there is an overlapping network of music and language in our brain [1]. The transfer of the two domains (music and language) can be bi-directional. The use of tonal contours in language enhances the acuity of pitch perception [2]. On the other hand, musical training can facilitate identification of emotion in prosody [3].

The motor system also plays an important role in speech perception. For example, speech perception of the tritone paradox has a close connection to listener's vocal range [4]. It has been shown that stimulating the larynx pre-motor area with TMS (transcranial Magnetic Stimulation) lead to a faster reaction time for discriminating a pitch shift of small interval ($\pm$ 3%) [5]. To examine the correlation between pitch perception and production at the behavioral level, the present study employed the pitch-shift paradigm. In this paradigm, a short and artificial change in pitch is fed back to speakers through the headphones while they are vocalizing vowels like /a/ at a steady pitch. Typically, speakers would compensate for pitch perturbation by changing the vocal pitch in an opposite direction to the pitch-shift stimulus [6]. The compensation, also called pitch-shift responses, has been regarded as a corrective mechanism that is executed in our motor system to reduce the perceived error in pitch.

The magnitudes of pitch-shift responses have to do with tone experiences or musical experiences. Research has shown that tone speakers or vocalists were able to suppress pitch-shift responses, suggesting that they were more stable in maintaining vocal pitch and less affected by the artificial perturbation [7-8]. In the present study, we would like to explore the correlation between this susceptibility to pitch perturbation and pitch perception abilities in speakers with different tone and musical experiences. It is expected that smaller pitch-shift responses would be related to more acute pitch perception.

2. Methods

2.1. Participants

Five groups of participants were recruited, including seven L2 beginners of Mandarin (L2 BEGINNER for short), seven L2 advanced learners of Mandarin (more than 5 years of learning; L2 ADVANCED for short), fifteen native Mandarin speakers without formal musical training (MANDARIN for short), fifteen native Mandarin speakers specializing in instrument playing (MUSICIAN for short), and fifteen native Mandarin speakers specializing in vocal performance (VOCALIST for short). All participants passed a hearing test at 20 dB hearing level bilaterally at 125, 250, 500, 750, 1000, 2000, 3000, and 4000 Hz. They were all university students (age: 20~35) and were paid for their participation.

2.2. Materials and procedures

Two perception tasks (the musical memory task and the adaptive pitch discrimination task) and one production task (the pitch-shift task) were administered to all participants. In the musical memory task (http://jakemandell.com/tonedeaf/), there were 36 trials. In each trial, two musical phrases, each with a duration of 2-8 seconds, were played with a 2-second interstimulus interval. Within each trial, the two monophonic musical phrases had the same timbre (the same musical instrument) but may differ in one or some of the musical notes. Overall, the 36 trials contained melodies of keyboard instruments, percussion, stringed instruments, wind instruments, and electronic musical instruments. After listening to the two musical phrases, the participants made a judgment by clicking on the buttons on the webpage (as shown in Figure 6). No accuracy feedback was provided during the test. An overall accuracy was calculated and presented on the computer screen at the end of the task.

 10.21437/SpeechProsody.2020-125

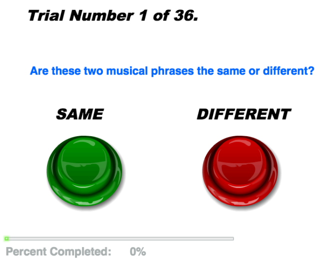

Figure 1: *Illustration of the musical memory task.*

In the adaptive pitch discrimination task (http://jakemandell.com/adaptivepitch/), the participants listened to a series of two short pure tones (300 ms each) and had to judge if the second tone was higher or lower than the first one (as shown in Figure 2). The first trial started from 96 Hz pitch difference. When participants made three consecutive correct responses, the pitch difference was reduced to half of the pitch difference in the previous trial (i.e., 48 Hz, advancing to a more difficult level). If they made one erroneous response, the pitch difference would double (i.e., regressing to the easier level). In this case, the procedure was adaptive. Unlike the musical memory task, the adaptive pitch discrimination task provided accuracy feedback on the screen during the test. In order to compare the results with the musical memory task, we covered the computer screen with a cardboard. Participants responded by pressing the up and down arrow keys, and were allowed to replay the sound by pressing the spacebar.

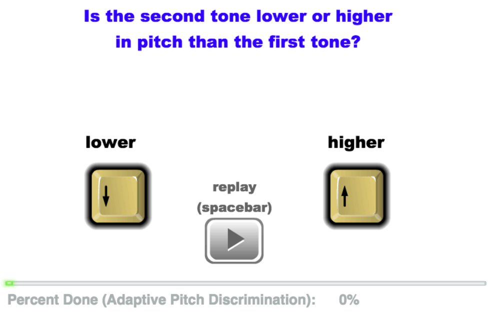

Figure 2: *Illustration of the adaptive pitch discrimination task (Current pitch difference: 96 Hz).*

Finally, in the pitch-shift task, the participants were asked to vocalize /a/ with a steady and comfortable pitch for 3-5 seconds. There were 60 vocalizations in total. During each vocalization, they would hear a short (200 ms) artificial change in pitch that randomly appeared at 1000-1500 ms after vocal onset. The pitch shift signal was generated by Eventide H7600, with a magnitude of either 50 cents or 100 cents, and fed back to the headphone simultaneously during the vocalization. Each participant's own voice signal was amplified with a McLELLAND MAR-16P headphone amplifier and played back to him or her with approximately 10 dB SPL louder than his/her own voice to reduce the possible influence of bone conduction. The direction of the pitch shift was either up (higher than the participant's pitch) or down (lower than the participant's pitch). This gave rise to four experimental conditions: 50 up, 50 down, 100 up, and 100 down. The participants were instructed to ignore the artificial change and to keep their pitch steady.

2.3. Data analysis

Two of the L2 beginners failed to complete the pitch discrimination task. Thus, their data were all eliminated from the data analysis.

Mandell [9] provided the interpretations of the perception scores at the end of the tasks. The interpretation of the performance on musical memory (Table 1) was based on the data from 61,036 subjects (MEAN = 73.9 %, SD = 9.99, see Figure 3). The interpretation of the performance on adaptive pitch discrimination (Table 2) was based on the data from 11,761 subjects (MEAN = 3.98 Hz, see Figure 4). Our participants' scores were recorded for data analysis.

Table 1: Interpretation of the musical memory score

Score	Performance
Above 90%	Exceptional performance
Above 80%	Very good performance
Above 70%	Normal performance
Above 60%	Low-normal performance
Below 55%	Possible pitch perception or memory deficit

Table 2: Interpretation of the adaptive pitch discrimination score

Score	Performance
Less than 0.75 Hz	Exceptional ear
Less than 1.5 Hz	Very good
Less than 6 Hz	Normal
Less than 12 Hz	Low-normal
Greater than 16 Hz	Possible pitch perception deficit

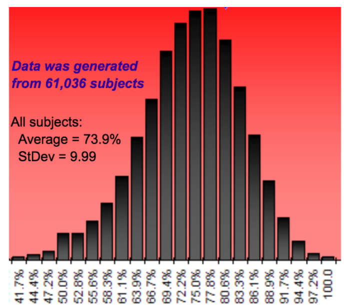

Figure 3: *Histogram of the musical memory scores from 61,036 subjects (Mandell [9]).*

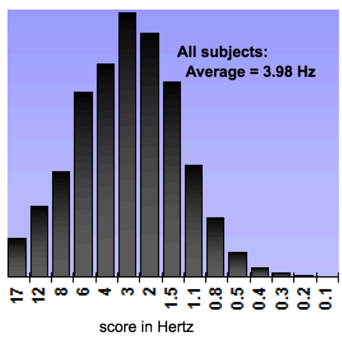

Figure 4: *Histogram of the adaptive pitch discrimination scores from 11,761 subjects (Mandell [9]).*

As for the pitch-shift task, the signals were time-locked to the onset of the pitch-shift stimulus and aggregated by the direction (up/down) and magnitude (50/100 cents) of the pitch-shift stimulus. The analysis window included a 200 ms pre-pulse period, a 200 ms pulse period, and a 1000 ms post-pulse period (see Figure 5). Only compensatory responses (where the direction of the vocal response was opposite to the direction of the pitch-shift stimulus) were included for two reasons: (i) compensatory responses (77%) were more frequent than following responses (23%); and (ii) compensatory responses suggest the correction mechanism is employed when speakers monitor their vocal motor commands. Then, the peak amplitude of the averaged response contour (i.e., the absolute maximum) in each condition (50 up, 50 down, 100 up, and 100 down) was retrieved for further statistical analysis.

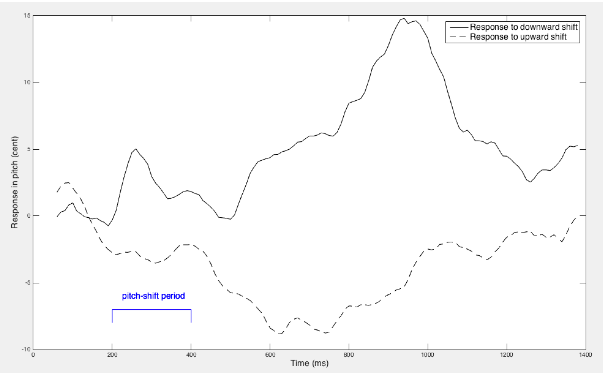

Figure 5: *Illustration of the pitch-shift responses.*

Repeated measures ANOVA was used to examine whether there was a significant group effect in the perception and production measures. Pearson correlations were conducted to investigate if perception abilities were correlated to the abilities to maintain a steady pitch under pitch perturbation. Finally, linear discriminant analysis was employed to explore if the perception and production measures can successfully classify or predict the participants' group.

3. Results

The averaged perception scores and the peak amplitudes of the pitch-shift responses are displayed in Table 3. The numbers in parentheses are the standard errors of the mean.

There was a significant group effect in the musical memory task ($F(4,52) = 11.560$, $p < .001$), the adaptive pitch discrimination task ($F(4,52) = 7.416$, $p < .001$), the 50 up pitch-shift condition ($F(4,52) = 2.970$, $p < .05$), the 50 down pitch-shift condition ($F(4,52) = 2.672$, $p < .05$), and the 100 down pitch-shift condition ($F(4,52) = 2.482$, $p < .05$). However, no significant group effect was found in the 100 up pitch-shift condition ($F(4,52) = 1.336$, $p = .269$). Post-hoc comparisons with Tukey's Honest Significant Difference (Tukey's HSD) method show that in general Mandarin speakers with musical training (either instrumental or vocal) performed better on the pitch perception and vocal pitch control than Mandarin speakers without musical training, followed by L2 learners. L2 ADVANCED learners performed slightly better than L2 BEGINNER in most of the tasks.

The results of Pearson correlations were depicted in Figure 6, with the starts indicating the level of significance. As you can see from the figure (particularly, the two blue rectangles), the perception abilities were significantly correlated with the abilities to control vocal pitch under small pitch perturbation (50 up and 50 down conditions, $p < .05$ at least). The better the perception abilities, the smaller the corrective pitch-shift responses.

Finally, the linear discriminant analysis included the perception and production measures as the predictor variables and the group as the outcome variable. The results were depicted in Figure 7. The circles were Euclidean distances from the center with the radius of 2. The cluster of MUSICIAN overlapped with the cluster of VOCALIST. Both had some distance away from the cluster of MANDARIN, suggesting that musical training mattered. The cluster of L2 BEGINNER was most distant from that of the speakers with musical training. The cluster of L2 ADVANCED learners somewhat overlapped with the cluster of MANDARIN, suggesting that they might be approaching the native-like level on pitch perception and vocal pitch control.

Table 3: Summary of the perception and production measures

Group / Task	L2 BEGINNER	L2 ADVANCED	MANDARIN	MUSICIAN	VOCALIST
Perception measures					
Musical memory (100 %)	63.88 (5.56)	74.21 (5.01)	75.75 (10.11)	84.40 (6.95)	86.55 (5.87)
Adaptive pitch discrimination (Hz)	24.48 (12.42)	18.34 (12.93)	12.68 (13.20)	4.01 (2.76)	5.94 (3.12)
Production measures: absolute peak amplitude					
50 up (cents)	42.08 (7.23)	51.73 (5.37)	31.12 (5.78)	14.99 (1.72)	10.54 (2.02)
50 down (cents)	54.86 (8.32)	32.78 (5.88)	27.04 (6.88)	17.80 (2.21)	11.58 (1.86)
100 up (cents)	56.44 (13.00)	42.60 (3.01)	44.76 (7.06)	28.78 (3.36)	20.92 (3.68)
100 down (cents)	61.33 (12.06)	38.39 (5.16)	32.87 (6.35)	22.47 (2.67)	27.98 (2.99)

Note. Numbers in parentheses represent standard errors.

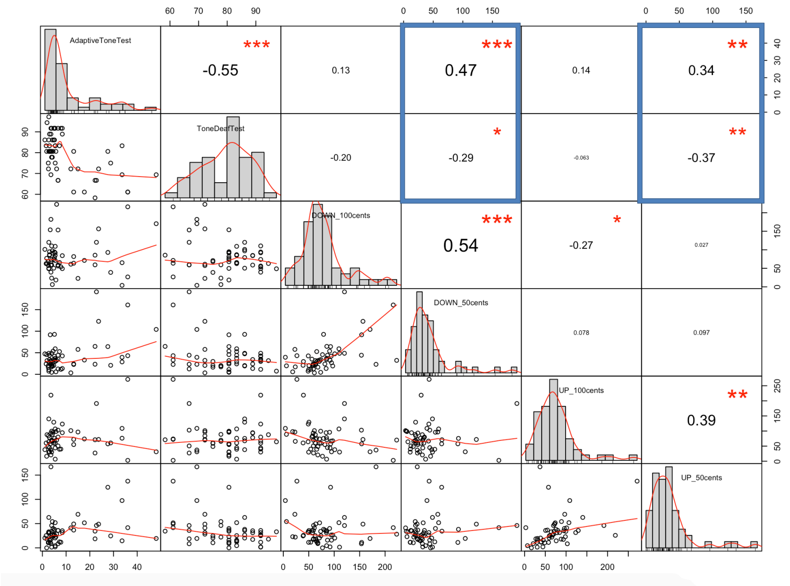

Figure 6: *Visualization of the correlation matrix.*

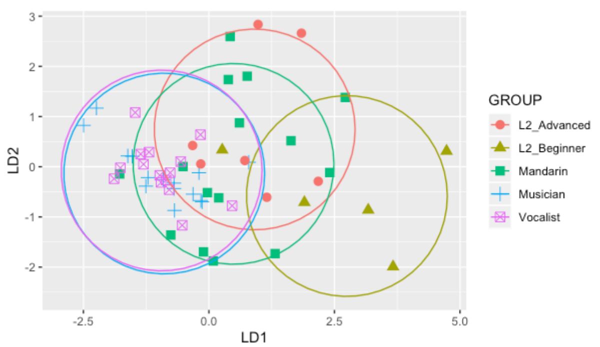

Figure 7: *Plot of linear discriminant scores from the first two discriminant functions.*

4. Discussion and conclusion

Our results indicate that tone and musical experiences mattered in both pitch perception and vocal pitch control. In general, speakers with tone or musical experiences had better pitch perception results and were less susceptible to pitch perturbation (i.e., smaller pitch-shift responses). The less susceptibility to pitch perturbation suggests that they were able to ignore the artificial change in pitch while maintaining their voice at an intended pitch level. Notice that overall L2 ADVANCED learners performed better than L2 BEGINNER. It is unclear whether our L2 ADVANCED learners had good pitch perception abilities prior to tone learning or whether their pitch perception abilities were improved because of tone learning. However, based on the growing evidence that speech and music are strongly correlated, the latter one is highly likely.

Pitch perception abilities (including musical memory and adaptive pitch discrimination) were significantly correlated to the abilities to control vocal pitch (particularly for small perturbation). This suggests that whether you can suppress pitch-shift responses (i.e., a more steady pitch) particularly for a small perturbation (i.e., 50 cents) has to do with whether you would be able to detect and memorize pitch differences. In other words, our perception and production for pitch are interdependent.

The above-mentioned correlation appeared in the small pitch perturbation (50 cents), but not in the large pitch perturbation (100 cents). So far it is unclear why this is the case. It is speculated that pitch-shift responses are reflex-like correction tuned for small deviation [10]. Therefore, in face of large perturbation, the response strategies may be different and less reflex-like.

We used the perception measures and production measures to classify the participants. The classification results suggest that there was a gradient difference among the groups in terms of their perception and production. While there was an overlap between MUSICIAN and VOCALIST, L2 BEGINNER was most dissimilar to the speakers with musical training. MANDARIN and L2 ADVANCED resided in the middle, with the former closer to MUSICIAN and VOCALIST. The overlap between MANDARIN and L2 ADVANCED suggests that the L2 ADVANCED learners had performed like a tone speaker in terms of pitch perception and vocal pitch control. Since the L2 ADVANCED learners all had more than 5 years of learning Mandarin, this implies that achieving a native-like pitch behavior may require more than 5 years of training.

5. Acknowledgements

The author gratefully acknowledges the assistance of Chien-Min Kuo in the data collection.

6. References

[1] M. Bangert, T. Peschel, G. Schlaug, M. Rotte, D. Drescher, H. Hinrichs, H. Heinze, E. Altenmuller, “Shared networks for

auditory and motor processing in professional pianists: Evidence from fMRI conjunction," *Neuroimage*, vol. 3, p.p. 917–926, 2006.

[2] I. Peretz, S. Nguyen, S. Cummings, "Tone language fluency impairs pitch discrimination," *Front. Psychol,* vol. 2, No. 145. 2011.

[3] W. Thompson., E. Schellenberg, G. Husain, "Decoding speech prosody: Do music lessons help?" *Emotion*, vol. 4, p.p. 46–64, 2004.

[4] D. Deutsch, T. North, L. Ray, "The tritone paradox: Correlate with the listener's vocal range for speech," *Music Perception*, vol. 7, p.p. 371-384, 1990.

[5] A. D'Ausilio, I. Bufalari, P. Salma, P. Busan, L. Fadiga, "Vocal pitch discrimination in the motor system," *Brain and Language*, vol. 118, p.p. 9-14, 2011.

[6] T. A. Burnett, M. B. Freeland, C. R. Larson, "Voice F0 responses to manipulations in pitch feedback," *J. Acoust. Soc. Am*, vol. 103, No. 6, p.p. 3153– 3161, 1998.

[7] L.-H. Ning, T. M. Loucks, C. Shih, "The effects of language learning and vocal training on sensorimotor control of lexical tone," *Journal of Phonetics*, vol. 51, p.p. 50-69, 2015.

[8] J. M. Zarate, R. J. Zatorre, "Experience-dependent neural substrates involved in vocal pitch regulation during singing," *NeuroImage*, vol. 40, p.p. 1871–1887, 2008.

[9] J. Mandell, "Adaptive pitch test: Accurately measure your pitch perception abilities," Retrieved 03.12.19, from 〈http://tonometric.com/adaptivepitch/〉.

[10] H. Liu, C. R. Larson, "Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex," *Journal of the Acoustical Society of America*, vol. 122, No. 6, p.p. 3671–3677, 2007.