THE EFFECT OF MUSICALITY ON CUE SELECTION IN PITCH PERCEPTION BY ENGLISH AND MANDARIN SPEAKERS

Aletheia Cui, Jianjing Kuang

Department of Linguistics, University of Pennsylvania
acui@sas.upenn.edu, kuangj@ling.upenn.edu

ABSTRACT

Musicality, the aptitude for pitch and rhythm processing, is associated with language processing and acquisition. Previous studies have found that listeners integrate spectral and f0 cues in pitch perception. In this study, we investigate whether musicality and language background affect the cue integration process in pitch perception. Speakers from both a tone language (Mandarin) and a non-tone language (English) were recruited. The subjects first participated in a pitch perception experiment to test the effects of spectral slope and f0 on their pitch judgments, and they then took the Montreal Battery of Evaluation of Musical Abilities. The results show that subjects with higher musicality scores were more likely to rely on f0 in relative pitch judgment, while subjects with lower musicality scores were more affected by differences in spectral slope. There were no differences between the two language groups.

Keywords: Musicality, pitch perception, tone language

1. INTRODUCTION

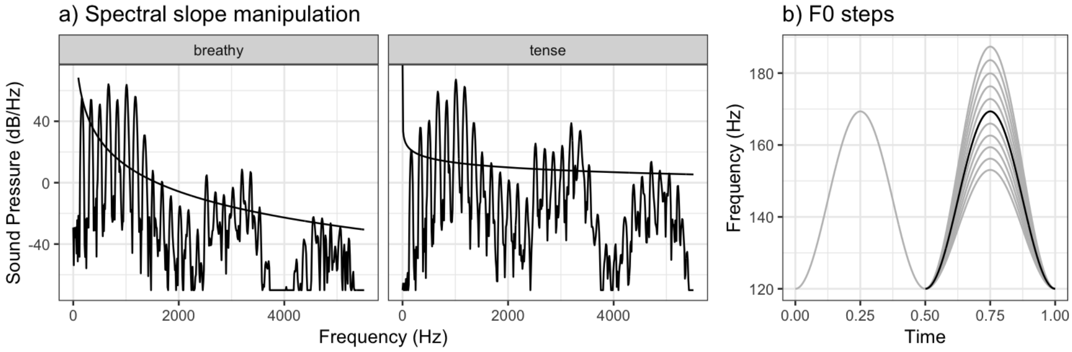
Music and speech share similar acoustic cues. While studies in hemispheric lateralization show that music processing and speech processing occur in the right and left hemispheres respectively [3, 26], music and speech may share similar processing mechanisms [21, 28]. Indeed, musicality, or the aptitude for processing rhythm and pitch, has been found to facilitate language acquisition and processing. For example, musicians generally are faster and better at learning second languages than non-musicians [16, 18, 23]. Musical training is also helpful for first language acquisition, and it has been shown that music training can improve linguistic processes such as pitch processing [19], segmentation [5], reading ability [19, 25], and verbal memory [8].

Since musicality is beneficial for language learning, people with poor musicality and especially people with congenital amusia (a neurogenetic disorder in pitch perception) might face major disadvantages in language processing, especially for tone and intonation. However, it remains controversial how much language processing can be affected by amusia. For example, although people with amusia experience impaired pitch processing in music [11, 1], most amusic Mandarin speakers have little trouble perceiving tones [27, 10], and only a subset of them show impairment in tone perception. However, even this subset of speakers had normal tone production [20, 17]. An interesting question thus arises: How are these speakers able to acquire normal tone production if they struggle to perceive pitch cues?

Nan et al. [20] proposed two possible explanations: 1) production and perception of pitch have different neural pathways, and 2) amusic individuals use non-pitch based cues to guide their production of pitch. The first hypothesis is unlikely, since auditory feedback is crucial for language acquisition. If speakers cannot hear pitch, it would be not possible for them to develop reliable production. The second hypothesis is not necessary, because speech cues are often highly redundant, and pitch perception and production can be mapped onto multiple acoustic cues. Therefore, in this study, we propose an alternative hypothesis: People who struggle with f0 discrimination can use acoustic cues that co-vary with f0 to acquire pitch distinctions.

It has been well-established that voice quality co-varies with pitch, and higher pitch is naturally associated with tenser voice [9, 24]. Acoustically, tenser voice has a flatter spectrum [6]. In pitch perception, previous studies have shown that spectral slope manipulations can cause significant shift in pitch perception, with tenser voice judged as being higher in pitch [14, 15]. Moreover, Kuang [14] found that non-musicians were more affected by spectral slope manipulations than musicians. In this study, instead of using self-reported musician labels, we examined whether continuous musicality scores could be a predictor of pitch perception strategies. To test our hypothesis, we ran a pitch perception experiment with varying spectral slope and f0 cues as well as a musicality test. We predict that subjects with higher musicality scores would be more likely to rely on f0 in pitch perception, whereas less musical subjects

Figure 1: a) The spectral slope is flattened to created tenser-sounding phonation. b) The f0 contours for the stimuli. The second peak is a continuum of 11 f0 steps.



would be more influenced by spectral cues. We recruited both tone and non-tone language speakers, as exposure to contrastive pitch differences in tone languages might lead to better pitch processing.

2. METHODS

2.1. Participants

Seventy-one native English speakers (age range 18-25; mean 19.82) and 44 native Mandarin speakers (age range 18-50; mean 24.95) were recruited to participate in a pitch perception experiment and then a musicality experiment.

2.2. Experiment 1: Pitch perception

The pitch perception experiment tested each subject's reliance on f0 and spectral slope cues in a relative pitch judgment task. This experiment was adapted from the procedures in [13]. The same stimuli and task were used but the number of stimulus repetition was reduced.

2.2.1. Stimuli

The stimuli were resynthesized from the natural production of a "ma-MA-ma" sequence of a male speaker. The original phonation of the speaker constituted the "breathy" voice quality in this experiment. A "tense" version of the "ma-MA-ma" sequence was created using TANDEM-STRAIGHT [12] so that the spectrum was 6 dB/octave greater than the breathier version. Figure 1a illustrates the spectral slope manipulation. Each stimulus consisted of two continuous "ma-MA-ma" utterances in each of the four possible spectral slope combinations (Table 1). For example, for the BT condition, the listener would hear a breathy "ma-MA-ma" followed immediately by a tense "ma-MA-ma".

In addition to spectral slope manipulations, the f0 contour was also modified (Figure 1b). The lowest f0 was the same for both "ma-MA-ma" peaks (120 Hz). While the maximum value of the first peak was kept constant at 169.34 Hz, the second peak was an 11-step continuum varying from 153.06 Hz to 187.36 Hz (0.35 semitone/step). At step 6, the f0 of the second peak was equal to the f0 on the first peak. In total, there were 44 distinct stimuli after spectral slope and f0 manipulations (4 spectral slope conditions $\times$ 11 f0 steps).

2.2.2. Procedure

The listeners participated in these experiments in a sound proof booth. The stimuli were played through Sennheiser HD 280 Pro headphones. The subjects were instructed to think of each "ma-MA-ma" as a word, and upon hearing a stimulus with two "ma-MA-ma" words, they were asked to do a forced-choice classification to indicate which of the two "words" sounded higher in pitch. The presentation of the stimuli was randomized, and each stimulus was presented 5 times.

2.3. Experiment 2: Musicality

The Montreal Battery of Evaluation of Musical Abilities (MBEMA) was used to evaluate each subject's musicality [22].

Table 1: Summary of the four spectral slope conditions.

Peak 1	Peak 2	Intended spectral slopes
Tilted	Tilted	Breathy + Breathy (BB)
Flat	Flat	Tense + Tense (TT)
Tilted	Flat	Breathy + Tense (BT)
Flat	Tilted	Tense + Breathy (TB)

Figure 2: Second peak higher response rates by English and Mandarin speakers. X-axis = f0 steps, y-axis = proportion of responses where subject chosen peak 2 to be higher.

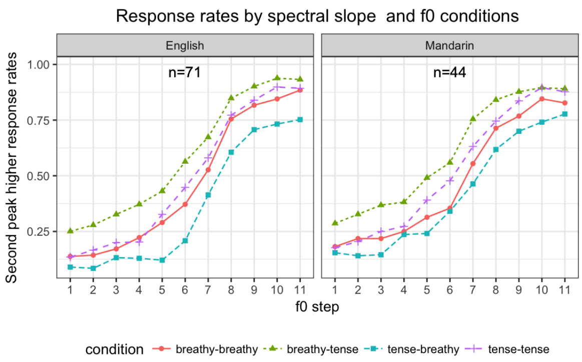


Table 2: Summary statistics of musicality scores for each of the three tests. Speakers from both language groups had similar scores.

		English	Mandarin
melody	mean	0.88	0.87
	sd	0.11	0.09
rhythm	mean	0.89	0.88
	sd	0.09	0.09
memory	mean	0.88	0.90
	sd	0.09	0.08
total	mean	2.65	2.64
	sd	0.21	0.21

2.3.1. Stimuli

The MBEMA stimuli were short musical phrases played with different instruments. The phrases were either played by itself or paired with another musical phrase depending on the task.

2.3.2. Procedure

There were three tasks, evaluating the subject's ability to identify differences in melody and rhythm and their memory of musical phrases. Each task consisted of 20 stimuli. The melody test played two consecutive melodies and asked the subject to identify whether the melodies were the same or different. The rhythm test also played two consecutive musical phrases, but the rhythm might be different between the two. Lastly, the memory test played only one melody, and and it asked the listener to indicate whether they have heard this melody in the previous tasks. For each subject, the score for each test was calculated as the percentage of correct answers for that test. An overall musicality score was calculated by summing up the scores for all three tests.

3. RESULTS

3.1. Pitch perception

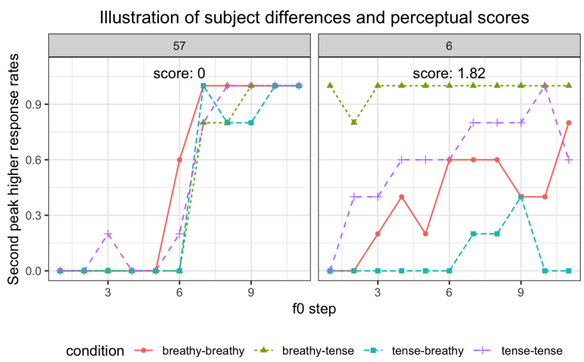
Figure 2 plots the rates at which subjects chose the second peak to be higher, and it shows the overall group results for English and Mandarin speakers. Both groups exhibit similar shifts for "breathy-tense" and "tense-breathy" spectral slope conditions. When the first peak was breathy and the second tense, listeners were more likely to select the second peak as higher. However, when the first peak was tense and the second breathy, listeners chose the first peak as higher in pitch at higher rates. The main effects of spectral slope conditions were evaluated using an MCMC generalized linear mixed-effects model in R [7]. Separate models were built for English and Mandarin speakers. Spectral slope conditions and f0 steps were the fixed factors, and subjects were the random intercepts. The main effects of the spectral slope conditions are summarized in Table 3.

3.2. Musicality

Table 2 summarizes the musicality scores for each test for the English and the Mandarin speakers. Both groups have similar scores across all three tests. A t-test confirms that group means between the two language groups do not differ (t = -0.275, df = 93.156, p = 0.784).

3.3. Relationship between musicality and pitch perception

Figure 3: The perceptual shift score quantifies the subject's pitch perception strategy. On the left, subject 57 has a low score and relies heavily on f0, while subject 6, with a much higher score, shows significant influence of spectral slope.

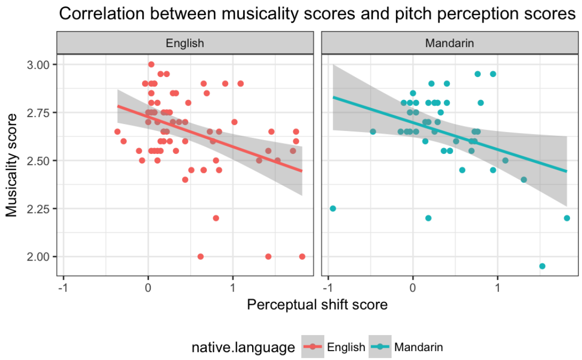


To investigate whether there is correlation between musicality and individual pitch perception

Table 3: Main effects of spectral slope for every pair of conditions for English and Mandarin speakers. Means of regression coefficients are followed by 95% highest posterior density intervals in square brackets and p-values.

		BB	**TT**	**BT**
English	**TT**	0.11 [0.03, 0.18], p = 0.006		
	BT	0.53 [0.44, 0.62], p < 0.001	0.69 [0.46, 0.89], p < 0.001	
	TB	-0.47 [-0.56, -0.38], p < 0.001	-0.97 [-1.20, -0.64], p < 0.001	-1.12 [-1.28, -0.97], p < 0.001
		BB	**TT**	**BT**
Mandarin	**TT**	0.22 [0.11, 0.34], p < 0.001		
	BT	0.62 [0.49, 0.74], p < 0.001	0.54 [0.37, 0.74], p < 0.001	
	TB	-0.29 [-0.41, -0.18], p < 0.001	-0.71 [-0.94, -0.52], p < 0.001	-1.69 [-2.24, -1.15], p < 0.001

Figure 4: Correlation of musicality and perception scores by English and Mandarin speakers.



strategies, we calculated a perceptual shift score for each subject. This score was intended to quantify how much each subject was affected by spectral slope in their judgment of relative pitch, and it was calculated as the sum of the mean differences of BT and TB from BB and TT: Shift $= (\overline{BT} - \overline{BB}) + (\overline{BT} - \overline{TT}) + (\overline{BB} - \overline{TB}) + (\overline{TT} - \overline{TB})$.

Figure 3 illustrates how this score quantifies perceptual shift. Subject 57 has a very low score of 0, and their pitch judgment shows no influence of spectral slope condition. For all spectral slope conditions, they judged the first peak to be higher for f0 steps 1-5, and their responses changed categorically at step 6. On the other hand, subject 6 has a much higher perceptual shift score, and as can be observed from the plot, they almost always judged the second peak to be higher for BT and second peak to be lower for TB. For these perceptual shift scores, there is no significant difference (t = -0.828, df = 95.752, p = 0.41) between English (mean = 0.466, sd = 0.535) and Mandarin speakers (mean = 0.384, sd = 0.503).

The perceptual shift scores from the pitch perception experiment are plotted against each subject's musicality score in Figure 4. The results from English and Mandarin speakers both show significant negative correlations between the musicality scores and the perceptual shift scores. For English speakers, the correlation is -0.392 (t = -3.51, df = 68, p = 0.00079), and for Mandarin speakers, the correlation is -0.340 (t = -2.34, df = 42, p = 0.024).

4. DISCUSSION

In this study, we investigate the relationship between musicality and pitch perception strategies in speech. This study replicated previous findings that listeners use both spectral slope and f0 cues in pitch processing [13, 15]. Overall, listeners judged pitch to be higher when the higher frequencies of the spectrum were boosted. On the individual level, there are differences in pitch processing strategies since some listeners were more affected by spectral slope cues than others. We hypothesized that subjects with higher musical aptitude were more likely to rely on f0, and indeed, the subjects who scored higher on the musicality test showed less perceptual shift as the result of differences in spectral slope cues. These findings have implications for language acquisition. Although some speakers may be worse at f0 discrimination, we demonstrated that other cues also matter in pitch perception, and speakers who experience difficulty in discriminating f0 can rely on additional spectral cues in acquiring linguistic distinctions. Therefore, subjects with amusia or tone agnosia may acquire normal tone production by relying on cues that co-vary with f0.

Moreover, we tested whether language experience had effects on cue preferences in pitch processing. We found that English and Mandarin speakers had similar musicality and perceptual shift scores, indicating that being a native speaker of a tone language does not influence cue integration in pitch processing. This is not surprising, as it has been found that tone language speakers have similar rates of amusia as non-tone language speakers [20]. Moreover, tone language speakers do not perform better on pitch discrimination tasks than non-tone language speakers [2, 4]. Overall, our results support the idea that musicality influences linguistic processing, but the opposite relationship was not found.

5. REFERENCES

[1] Ayotte, J., Peretz, I., Hyde, K. 2002. Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain* 125(2), 238–251.

[2] Bent, T., Bradlow, A. R., Wright, B. A. 2006. The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance* 32(1), 97–103.

[3] Bever, T. G. 1975. Cerebral asymmetries in humans are due to the differentiation of two incompatible processes: Holistic and analytic. *Annals of the New York Academy of Sciences* 263(1), 251–262.

[4] DiCanio, C. T. 2012. Cross-linguistic perception of itunyoso trique tone. *Journal of Phonetics* 40(5), 672–688.

[5] François, C., Chobert, J., Besson, M., Schön, D. 2012. Music training for the development of speech segmentation. *Cerebral Cortex* 23(9), 2038–2043.

[6] Gobl, C., Chasaide, A. N. 2013. Voice source variation and its communicative functions. *The Handbook of Phonetic Sciences* 378–423.

[7] Hadfield, J. D., others, 2010. Mcmc methods for multi-response generalized linear mixed models: the mcmcglmm r package. *Journal of Statistical Software* 33(2), 1–22.

[8] Ho, Y.-C., Cheung, M.-C., Chan, A. S. 2003. Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology* 17(3), 439–450.

[9] Hollien, H. 1974. On vocal registers. *Journal of Phonetics* 2, 125–143.

[10] Huang, W.-T., Liu, C., Dong, Q., Nan, Y. 2015. Categorical perception of lexical tones in mandarin-speaking congenital amusics. *Frontiers in psychology* 6, 829–838.

[11] Kalmus, H., Fry, D. 1980. On tune deafness (dysmelodia): frequency, development, genetics and musical background. *Annals of human genetics* 43(4), 369–382.

[12] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., Banno, H. 2008. Tandem-straight: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation. *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE 3933–3936.

[13] Kuang, J., Guo, Y., Liberman, M. 2016. Voice quality as a pitch-range indicator. *Proceeding of Speech Prosody*.

[14] Kuang, J., Liberman, M. 2015. The effect of spectral slope on pitch perception. *Sixteenth Annual Conference of the International Speech Communication Association*.

[15] Kuang, J., Liberman, M. 2016. The effect of vocal fry on pitch perception. *Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 5260–5264.

[16] Lee, C.-Y., Hung, T.-H. 2008. Identification of mandarin tones by english-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America* 124(5), 3235–3248.

[17] Liu, F., Chan, A. H., Ciocca, V., Roquet, C., Peretz, I., Wong, P. C. 2016. Pitch perception and production in congenital amusia: Evidence from cantonese speakers. *The Journal of the Acoustical Society of America* 140(1), 563–575.

[18] Milovanov, R., Pietilä, P., Tervaniemi, M., Esquef, P. A. 2010. Foreign language pronunciation skills and musical aptitude: A study of finnish adults with higher education. *Learning and Individual Differences* 20(1), 56–60.

[19] Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., Besson, M. 2008. Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cerebral Cortex* 19(3), 712–723.

[20] Nan, Y., Sun, Y., Peretz, I. 2010. Congenital amusia in speakers of a tone language: association with lexical tone agnosia. *Brain* 133(9), 2635–2642.

[21] Patel, A. D. 2003. Language, music, syntax and the brain. *Nature neuroscience* 6(7), 674–681.

[22] Peretz, I., Gosselin, N., Nan, Y., Caron-Caplette, E., Trehub, S. E., Béland, R. 2013. A novel tool for evaluating children's musical abilities across age and culture. *Frontiers in systems neuroscience* 7, 30–40.

[23] Posedel, J., Emery, L., Souza, B., Fountain, C. 2012. Pitch perception, working memory, and second-language phonological production. *Psychology of Music* 40(4), 508–517.

[24] Roubeau, B., Henrich, N., Castellengo, M. 2009. Laryngeal vibratory mechanisms: the notion of vocal register revisited. *Journal of voice* 23(4), 425–438.

[25] Tierney, A., Kraus, N. 2013. Music training for the development of reading skills. In: *Progress in Brain Research* volume 207. Elsevier 209–241.

[26] Witelson, S. F., Pallie, W. 1973. Left hemisphere specialization for language in the newborn: Neuroanatomical evidence of asymmetry. *Brain* 96(3), 641–646.

[27] Yang, W.-x., Feng, J., Huang, W.-t., Zhang, C.-x., Nan, Y. 2014. Perceptual pitch deficits coexist with pitch production difficulties in music but not mandarin speech. *Frontiers in Psychology* 4, 1024–1034.

[28] Yu, M., Xu, M., Li, X., Chen, Z., Song, Y., Liu, J. 2017. The shared neural basis of music and language. *Neuroscience* 357, 208–219.