Behavioral and subcortical signatures of musical expertise in Mandarin Chinese speakers

Caitlin Dawson[1]*, Mari Tervaniemi[2], Daniel Aalto[3,4]

1 Cognitive Brain Research Unit, Faculty of Medicine, University of Helsinki, Helsinki, Finland, **2** CICERO Learning Network, Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland, **3** Communication Sciences and Disorders, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Canada, **4** Institute for Reconstructive Sciences in Medicine, Misericordia Community Hospital, Edmonton, Canada

* cadawson4@gmail.com

Abstract

Both musical training and native language have been shown to have experience-based plastic effects on auditory processing. However, the combined effects within individuals are unclear. Recent research suggests that musical training and tone language speaking are not clearly additive in their effects on processing of auditory features and that there may be a disconnect between perceptual and neural signatures of auditory feature processing. The literature has only recently begun to investigate the effects of musical expertise on basic auditory processing for different linguistic groups. This work provides a profile of primary auditory feature discrimination for Mandarin speaking musicians and nonmusicians. The musicians showed enhanced perceptual discrimination for both frequency and duration as well as enhanced duration discrimination in a multifeature discrimination task, compared to nonmusicians. However, there were no differences between the groups in duration processing of nonspeech sounds at a subcortical level or in subcortical frequency representation of a nonnative tone contour, for f_o or for the first or second formant region. The results indicate that musical expertise provides a cognitive, but not subcortical, advantage in a population of Mandarin speakers.

Introduction

The plastic effects of musical training on the brain have gained great interest in the research community [1]. Musical training has been shown to be associated with perceptual benefits in lower frequency discrimination thresholds for pure tones [2,3] and faster and more accurate detection of small pitch changes [4] not only in nonspeech sounds but also in a foreign language [5], compared to nonmusicians. Musicians have shown enhanced mismatch negativity (MMN) to slightly detuned chords, indicating more precise detection of frequency deviations [6]. On a subcortical level, musicians show enhanced phase locking and pitch representation in the frequency following response (FFR) in both musical and speech sounds [7,8], enhanced representation of spectral content which contains vocal emotion [9,10], and enhanced

differentiation of speech sounds by encoding of the second formant [11]. The magnitude of brainstem responses to tuned and detuned chords was also related to perceptual differences in pitch discrimination between musicians and nonmusicians, indicating a link between behavioral performance and subcortical plasticity [12].

Auditory plasticity has been shown over short periods of time in schoolchildren participating in both formal and informal musical activities, indicating that experience-based effects of music are not limited to adult professional musicians but that musical experience promotes maturation of the auditory system [13,14,15] and auditory plasticity over the lifespan is sensitive to behavioral needs. The current view is that musical training promotes efficiency through corticofugal tuning which emphasizes features that are trained and/or are useful for the current task demands [16].

Native speakers of tone languages, which encode lexical pitch contrasts, show perceptual benefits for frequency and interval change detection as well as discrimination of nonnative linguistic tone contrasts compared to English speakers, even after training [17,18]. Mandarin speakers have shown stronger pitch representation and smoother pitch tracking to Mandarin tones as well as stronger representation of the second harmonic [19]. Pitch tracking of Mandarin Chinese and Thai speakers to linguistic tone contours was more accurate than that of English speakers, indicating a transfer effect between tone languages [20]. Moreover, pitch representation is enhanced to musical and nonmusical sounds, speech stimuli, and iterated ripple noise, which suggests an effect that is not specific to the speech context [21,22,23].

Effects of musical training and tone language are very similar, and many studies have equivocated them. However, recent attempts to disentangle the effects have shown a much more complex picture. Cooper and Wang [24] separated tone and non-tone speakers and musicians and nonmusicians in both linguistic groups (English and Thai) and taught them a new tone language (Cantonese). They found no clear advantage for tone language learning from the native tone language speaking musicians; rather, English-speaking musicians had the most advantage in learning Cantonese. The Thai speakers experienced tone confusion which impeded their learning of the new Cantonese tone contours, while the musicians in both linguistic groups performed better than the nonmusicians.

Language effects have been shown not only with tone languages but also quantity languages, like Finnish, which encode lexical duration. Previous studies have shown that native speakers exhibit enhanced perceptual, cortical, and duration processing at subcortical level [25,26, 27,28]. The interaction of these effects with musical training, however, is more complex, and the effects of musical expertise within linguistic groups are unclear. Enhanced MMNs and perceptual detection for duration deviants was found for Finnish speaking nonmusicians and French speaking musicians, but enhanced MMNs were found for frequency deviants only in French speaking musicians [29]. Likewise, Finnish speakers with greater musical sophistication have shown enhanced perceptual frequency discrimination, but not duration discrimination, and no enhanced subcortical duration discrimination, compared to those with less musical sophistication [30]. These studies indicate a specific effect of native language phonological patterns in the effects of musical expertise within the linguistic group.

Other research has shown an interesting disconnect between perceptual and neural effects when music and language are investigated in combination. Bidelman, Gandour, and Krishnan [31] found enhanced subcortical representation of pitch sequences in both musicians and Chinese speakers but only corresponding perceptual pitch discrimination advantages for the musicians, indicating that cognitive benefits of auditory training may arise only for behaviorally relevant tasks.

On the other hand, Hutka et al. [32] found enhanced perceptual pitch discrimination for both musicians and Cantonese speakers, compared to nonmusicians, but only enhanced

MMNs for pitch and timbre deviants in musicians. The authors interpret this as musical training having broader benefits to auditory processing than language, which is more specific. The divergence of results between several studies suggest that music and language may have different mechanisms or effects on plasticity; i.e. they do not appear to be clearly additive.

Moreover, there is a lack of linguistic group control in the language and music literature and little knowledge about the effects of musical expertise within different linguistic groups, particularly tone language speakers. If musical training and native language possibly have different mechanisms or interacting effects, then they must be adequately controlled in future research. This study attempts to contribute to the illumination of the separate effects of musical expertise and native language by investigating the effects of musical expertise on native speakers of a tone language (Mandarin Chinese). It uses both perceptual auditory feature discrimination tasks and brainstem recording designed to spotlight onset and sustained responses for subcortical duration and frequency signatures in order to form a thorough profile of the effects of musical expertise in Mandarin speakers.

Methods

Participants

57 native Mandarin Chinese speaking adults aged 18–35 participated in behavioral data collection (21 males, 28 nonmusicians, 29 musicians; Table 1).

55 of them also participated in the auditory brainstem response (ABR) data collection (20 males, 26 nonmusicians, 29 musicians). No participants had any experience with Finnish and spoke primarily Mandarin Chinese at home for the first 15 years of life. Some studies have shown connections between auditory discrimination and intelligence [33,34,35,36], but for practical reasons, it was not possible to conduct large-scale intelligence testing.

Musicians were defined as having more than 6 years of formal musical training and weekly musical practice, and nonmusicians were defined as having fewer than 2 years of musical training and no regular musical hobbies.

Participants were recruited by student telephone phone and email lists within Beijing Normal University and were compensated for their time. They gave written consent according to

Table 1. Description of participants: group designation, age, gender, Gold-MSI scores, and primary instruments of musicians.

	Total	Mean age	Gender		Mean Gold-MSI
			M	F	
Nonmusicians	**29**	**21.9**	**9**	**20**	**54.6**
Musicians	**28**	**20.5**	**12**	**16**	**97.6**
Piano	9	*Western Instruments*			
Violin	1				
Electric keyboard	1				
Guitar	5				
Electric guitar	2				
Drums	2				
Saxophone	1				
Voice	2				
Bamboo flute	1	*Traditional Instruments*			
Yangqin	1				
Erhu	1				
Guzheng/koto	2				

the Declaration of Helsinki and the ethical review committees of both the University of Helsinki and Beijing Normal University.

Procedure

The full experiment took 2 hours and all participants completed the brainstem recording first. The recording consisted of two blocks of a passive listening task, counterbalanced between participants in order to avoid any attentional issues that may affect data quality (boredom, movements, etc.). The first block contained two synthesized short sounds (see section: Stimuli) presented at 55 dB sound pressure level (SPL). The second block contained one natural consonant-vowel (CV) speech contour, /*puu*/, extracted from a longer Finnish word /*puuro*/ which means "porridge," presented at 65 dB SPL. There were a total of 6000 sweeps for each short stimulus (3000 per polarity) and 4000 sweeps for the speech stimulus.

For brainstem recording, a one-channel setup was used with one active channel at Cz online referenced to linked mastoids with a forehead ground at the hairline and four vertical and horizontal electrooculography (EOG) electrodes. A ±30 μV thresholding process was applied for artifact rejection. Data was collected in a shielded room using a Neuroscan SynAmps[2] Scan 4.5 system with a sampling rate of 20 kHz in AC mode/Gain 2010 and online open filter 10–3000 Hz with 6 dB roll-off. Sound stimuli were presented binaurally with shielded circumaural Sennheiser HD 419 headphones.

The behavioral experiment consisted of four listening tests modified from Kaernbach [37]. Participants listened to sounds with headphones presented from a laptop with sound calibrated to 65dB SPL. There were three adaptive single-feature tasks in which one sound feature was adjusted at a time (intensity, frequency, or duration) in order to find the 75% accuracy threshold for each feature. During each trial, two sounds were played in sequence and the participant was asked to press a key on the laptop to choose which sound was louder, higher, or longer (Intensity Test, Frequency Test, Duration Test, respectively). Correct answers increased the task difficulty by one step and incorrect answers reduced task difficulty by three steps (one-up three-down procedure), to find an accuracy rate of 75%. These tasks took about 10 minutes each. Then, a multifeature task asked again which sound was longer (duration), but all three features were varied randomly. This task took 20 minutes and terminated after 300 trials.

Stimuli

The first block of the ABR section consisted of two synthesized narrowband gamma-filtered stimuli, one at 162 Hz and on at 216 Hz, both presented at 55dB (SPL). A sawtooth wave of each pitch was narrow band filtered using a fourth order polynomial gammatone filter with centre frequency 3141.56 Hz; then, average intensities were normalized. Each stimulus is about 25ms in length with a 25ms silent buffer before and after the sound for an interstimulus interval (ISI) of about 50ms (the lengths are not actually absolute since the duration of the stimuli depend somewhat on the periodicity of the frequencies). The short stimuli were presented in alternating polarities and randomized.

The second block of the ABR section consisted of one CV syllable, /*puu*/, which means "tree," recorded from an adult female native Finnish speaker and cut from the longer word /*puuro*/, which means "porridge." The tone contour ranges in fundamental frequency (f_o) from 169 to 233 Hz and lasts 340ms long with a 20ms silence before and 30ms silence after, presented in a single polarity at 65 dB with a total of 4000 sweeps. Finnish does not have a system of lexical tones as Mandarin does, but instead a lexical duration contrast, in which vowels and consonants have a long and short version, e.g. *tuli*, "fire," *tuuli*, "wind," and *tulli*,

"customs." The long vowels are co-signaled by a tone contour which has a slight initial rise followed by a long fall which aids in recognition of duration contrasts. Mandarin has four lexical tones: high level, high rising, low falling-rising, and high falling. Thus, the tone contour used here came from a natural spoken language but represented a totally unfamiliar contour to Mandarin speakers.

The behavioral stimuli were synthesized in the same way as the short nonspeech sounds used for brainstem recording but were longer since they were used for perceptual judgments. The standard sounds were 150ms long, 65 dB, and 162 Hz. The behavioral tasks were created within custom Matlab functions to be within the range of human speech syllables in intensity, frequency, and duration. The three features were either held constant or varied adaptively or randomly, depending on the task. The adaptive tasks automatically terminated after 51 reversals; the multifeature task had 300 trials.

Analysis

Psychoacoustic tasks

The behavioral analysis used estimates from a logistic regression model that were fitted to the binary response data to calculate Weber fractions that represent discrimination thresholds for each auditory feature, using the equation $ln(3)/k$ where k is the GLM estimate. For the duration modulation test, *generalized Weber fractions* use the same calculation and represent the extent to which duration is judged longer, given an increase in each specific feature (intensity, frequency, or duration). Additional effects were calculated: the intensity ratio, which is the (absolute value of the) ratio of generalized Weber fractions for the intensity dimension over the duration dimension and represents the extent to which participants were influenced by variation in intensity when making the duration judgment (a larger ratio corresponds to more influence). The frequency ratio is the same calculation for the influence of frequency on duration judgment, and the duration ratio is the ratio of Weber fractions of duration discrimination from the simple task to the complex task, which represents the difference in performance between the simple and complex tasks (a smaller ratio corresponds to decrement in performance from simple to complex task). It is expected that all participants decrease in performance between the simple and complex task since ignoring distracting features is a more difficult task.

Subcortical responses

For analysis of the short ABR stimuli, data was preprocessed with band-pass filters at 80Hz and 4000Hz and an artifact rejection threshold of 30 μV and epochs of 15ms prestimulus and 30ms poststimulus. Due to a technical error, it was not possible to separate responses to the two different stimuli; therefore, the results show group grand averages. Wave V peak amplitudes and latencies were extracted with a custom Matlab thresholding algorithm designed to detect peaks within a designated time window as a percentage of total peak size, which is a conservative measure to take higher-amplitude noise into consideration. Wave V is thought to be generated by the inferior colliculus, which is a waystation for corticofugal connections and is an important integration point for incoming afferent and efferent information. The amplitude of wave V indicates precision in the temporal tuning of a population of neurons responding to sound [38]. It has been shown to reflect subcortical experience-based plasticity from auditory training and is affected by learning and language disorders [39,40,41]. It has previously been shown that wave V amplitude reflects enhanced duration processing at a subcortical level associated with quantity language experience [28], so the current study was interested in possible duration processing enhancement at the subcortical level due to musical expertise.

Responses to the speech stimulus were bandpass filtered from 80–1000 Hz. The analysis was mainly concerned with the sustained portion of the response (post-20ms). Waveforms for each subject were averaged before further analysis.

FFR analysis was conducted by means of a sliding window short-term autocorrelation function which allocated 40 ms time bins shifted by 1ms, creating 283 overlapping bins. For the pitch tracking analysis, each bin was autocorrelated (cross-correlated with itself) and the peak autocorrelation value (expressed as a number between 0 and 1, excluding the first lag which is 1) was identified for each bin, representing the periodicity strength of each time bin. Then, these peak values were averaged for each participant to determine the participant's pitch strength over the entire course of the response.

A short-term spectral analysis was also conducted using the same sliding window function. A Fast Fourier Transform (FFT) was applied to the windowed bins (Hanning window, bins zero-padded to 1 second to increase spectral resolution). From this, it was possible to extract the f_o contour from the spectrogram by identifying the frequency which shows the peak magnitude for each time bin. Thus, this is the measure of pitch tracking in terms of frequency. These peak magnitude frequencies per subject were then cross correlated with the stimulus itself (which has undergone the same short-term FFT process) to obtain the FFT pitch tracking measure (expressed as a cross correlation coefficient between 0 and 1) per participant.

Musical expertise

For measures of musical expertise used in correlations, the current study uses the generalized score of the self-report questionnaire from the Goldsmiths Musical Sophistication Index (Gold-MSI) [42]. As a full evaluation it consists of the self-report questionnaire and a battery of listening tests including melodic memory, beat perception, and sound similarity. The self-report questionnaire alone has been validated using objective listening tests and is an effective measure of musical ability [43]. The self-report inventory scores participants along five factors of musical engagement: active engagement, perceptual abilities, musical training, singing abilities, and emotional engagement. These factors are weighted together to create the generalized musical sophistication score. The Gold-MSI is equally useful for evaluating the musical sophistication of people who are highly formally trained, untrained, or have casual musical experience.

Statistical analysis

Since the distributions were not normal, nonparametric methods were used. A series of Mann-Whitney-Wilcoxon tests were run to compare the results of each test between music groups and Bonferroni corrected for multiple comparisons within effect type.

An additional comparison was done of pitch tracking in responses to the speech stimulus with a restricted frequency window of 100Hz around the first and second formants. A further analysis correlated Gold-MSI general sophistication scores with all of the previous effects: behavioral single-feature frequency, intensity, and duration discrimination, multifeature duration discrimination, frequency ratio, duration ratio, intensity ratio, wave V amplitude and latency, autocorrelation pitch tracking, FFT pitch tracking for f_o, F_1, and F_2.

Results

Perceptual effects

Musicians showed enhanced single-feature discrimination for both frequency and duration and for duration in the complex task compared to the non-musicians (for descriptives, see

Table 2. Mean and standard deviation of Weber fractions for musicians and nonmusicians for each perceptual variable of interest.

Weber fractions	Mean		Std dev	
	Musicians	Nonmusicians	Musicians	Nonmusicians
Frequency	0.5	7.42	0.76	18.4
Duration	0.025	0.066	0.016	0.13
Intensity	1.14	1.59	0.67	1.06
Complex duration	0.031	0.057	0.026	0.056
Frequency ratio	156.28	-1259.4	631.45	6596.13
Intensity ratio	1474.97	233.69	6280.8	965.24
Duration ratio	0.86	1.14	0.33	1.17

Table 2). They also showed a trending difference in single-feature intensity discrimination and frequency ratio, which did not reach significance at the corrected level (Table 3).

Subcortical effects

There were no differences between musicians and nonmusicians for either peak amplitude or latency of wave V in onset responses to the short nonspeech stimuli (Fig 1), nor for either auto-correlation pitch tracking or FFT pitch tracking of fundamental frequency in responses to the speech stimulus (Fig 2, Table 4).

Further analysis

Formant pitch tracking. The FFT pitch tracking sliding window algorithm was run again on 100 Hz windows around the average first (397–497) and second (700–800) formant frequencies as identified by Praat. The sliding window was also run on the original speech stimulus with the same restrictions and the results were cross correlated. Pitch tracking of the formants was also not significantly different between musicians and nonmusicians (Table 5).

Correlations with musical sophistication. A further analysis determined whether the results would be different with another measure of musical expertise, namely, the Gold-MSI general musical sophistication score. This score takes into consideration formal musical training but also factors which are unrelated to training and which may be due to aptitude or social or environmental conditions. All previously used perceptual and neural measures were correlated with the general musical sophistication index score, and the results mirror those of the music groups although some of the trends do not reach the corrected significance level (Tables 6 and 7). Moreover, the group music score means were significantly different, indicating that the participants were accurately assigned to musician and nonmusician groups (W = 812,

Table 3. Mann-Whitney-Wilcoxon test for perceptual effects between musicians and nonmusicians.

Effect	W value	P value
Simple frequency	145	0.000087*
Simple duration	199	0.0022*
Simple intensity	269	0.067
Complex duration	216	0.0035*
Frequency ratio	526	0.0279
Intensity ratio	479	0.16
Duration ratio	338	0.51

Bonferroni-corrected alpha level at 0.007.

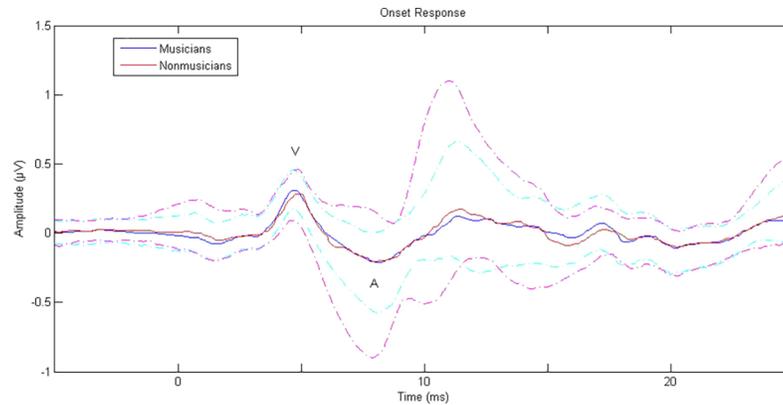* indicates significance at the Bonferroni-corrected level.

Fig 1. Onset response of musicians (blue, dark) and nonmusicians (red, light) to synthesized nonspeech sounds showing V-A complex. Dashed lines represent one standard deviation.

$p = 9.52 \times 10^{-11}$). The difference is likely due to the fact that the two measures of musical expertise emphasize slightly different factors.

Discussion

This work investigated the basic perceptual and subcortical auditory profiles of Mandarin speaking musicians and nonmusicians. Mandarin speaking musicians showed more accurate single-feature discrimination for both frequency and duration and a stronger influence of frequency on duration discrimination in a complex auditory environment. No subcortical effects were found.

Perceptual effects

There was no effect of duration ratio, which means that there was no group-based difference in the relationship between the single-feature duration task and the multifeature duration task. In general, participants decline in accuracy between the simple and complex tasks due to the increase in processing load from the addition of distracting features. It might be expected that
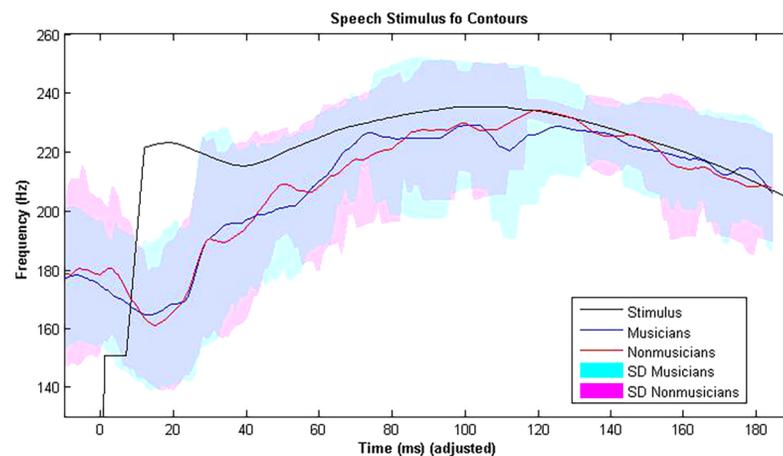

Fig 2. f_o contours of musicians (blue, dark) and nonmusicians (red, light) to a natural speech sound /puu/. The grey line shows the contour of the original stimulus, and dashed lines represent one standard deviation.

Table 4. *Mann-Whitney-Wilcoxon test for subcortical effects between musicians and* nonmusicians.

Effect	W value	P value
Autocorrelation pitch tracking	377	0.99
FFT pitch tracking f_o	402.5	0.69
Wave V peak amplitude	334	0.68
Wave V peak latency	312	1

Bonferroni-corrected alpha level at 0.0125.

* indicates significance at the Bonferroni-corrected level.

musicians would perform better in the complex task (showing less decrement in performance) than nonmusicians due to their superior processing skills. However, it may also be argued that enhancement in processing of low level single features could lead to an overall increase in system efficiency, which promotes integration of low level features. This appears to be what happened in a population of musically diverse Finnish speakers, whose linguistically driven enhancement for duration processing was more degraded in the complex task for those with higher levels of musical sophistication [30]. Here, there was no difference in degradation of duration discrimination for Mandarin speakers with the addition of distracting features between musicians and nonmusicians. In fact, the musicians showed significantly more accurate duration discrimination in the complex task compared to the nonmusicians. In other words, both the Mandarin speaking musicians and nonmusicians showed a similar extent of degradation between the simple and complex task, but the musicians had an overall more accurate duration discrimination within both tasks.

There was a nonsignificant trend (at corrected level) of frequency ratio. Previous studies have found that Mandarin speakers are less affected by frequency when making duration judgments than quantity language speakers (Finnish and Estonian), with both the most accurate duration discrimination and the most influence of frequency on duration discrimination occurring for Finns [44]. The positive correlation indicates that the more musically sophisticated participants were *more* affected by frequency in their duration judgments than the less musically sophisticated. Although counterintuitive, this indicates an enhanced efficiency in the auditory system since psychoacoustically, frequency contributes to perceived duration [45,46]. By integrating features which are perceptually bound, musicians process sound more efficiently in real-world acoustic environments like music performance.

Subcortical effects

Both groups showed high variability in amplitude of the onset response. Both groups accurately followed the speech stimulus tone contour, however, FFT pitch tracking for both musicians and nonmusicians, while giving generally high cross correlation values, was similarly highly variable and contained octave jumps. It is likely that since the participants were all

Table 5. Mann-Whitney-Wilcoxon test for first and second formant pitch tracking between musicians and nonmusicians.

Effect	W value	P value
FFT pitch tracking F_1	346.5	0.77
FFT pitch tracking F_2	362.5	0.99

Bonferroni-corrected alpha level at 0.025.

* indicates significance at the Bonferroni-corrected level.

Table 6. Correlations between Gold-MSI generalized musical sophistication score and perceptual effects.

Correlations	S	rho	P value
Music x frequency	39652	-0.51	.000077*
Music x duration	37090	-0.34	0.012
Music x intensity	31945	-0.15	0.27
Music x complex duration	37825	-0.29	0.029
Music x frequency ratio	21020	0.28	0.036
Music x intensity ratio	26078	0.11	0.43
Music x duration ratio	30894	-0.11	0.41

Bonferroni-corrected alpha level at 0.007.

* indicates significance at the Bonferroni-corrected level.

https://doi.org/10.1371/journal.pone.0190793.t006

healthy adult native Mandarin speakers, there was a ceiling effect for subcortical frequency processing due to linguistic expertise.

The speech stimulus was chosen to represent a nonnative tone contour from a natural language. It is possible that the Mandarin speakers did not process the stimulus as linguistic, and/ or that the musicians processed it as musical, which would activate perceptual benefits from cognitively identifying the task demands in a musical context. Previous research has shown top-down effects of language or music on categorization (and further pitch processing) of sounds which are similar to natural language tone contours or musical notes [47,48]. It may be necessary to direct participants' "listening mode" with stimuli that could be ambiguously interpreted to be linguistic or musical. Additionally, further investigations could use a wider range of similar natural speech, musical, and speech like stimuli, such as instrumental and vocal contours, synthesized contours without phonemes, and iterated ripple noise in order to determine the effect of top-down organization of auditory domains.

Musical expertise

The correlational analysis with Gold-MSI scores showed the same pattern of effects as the cross-sectional analysis, which was expected since the group means were significantly different. However, the distribution of scores was not bimodal, as would be expected from groups which did not overlap in level of musical training (fewer than 2 years/6 or more years). The Gold-MSI is likely capturing additional features that are not directly associated with formal musical training and which may have a weak effect on the results.

Table 7. Correlations between Gold-MSI generalized musical sophistication score and subcortical effect.

Correlations	S	rho	P value
Music x Autocorrelation pitch tracking	27548	0.0062	0.96
Music x FFT pitch tracking f_o	25003	0.098	0.48
Music x FFT pitch tracking F_1	23952	0.087	0.53
Music x FFT pitch tracking F_2	25773	0.018	0.90
Music x Wave V peak amplitude	19513	0.063	0.66
Music x Wave V peak latency	17492	0.16	0.27

Bonferroni-corrected alpha level at 0.008

* indicates significance at the Bonferroni-corrected level.

https://doi.org/10.1371/journal.pone.0190793.t007

Participants indicated their main instrument on the Gold-MSI (Table 1). Of the 28 musicians, 21 indicated Western instruments, 5 indicated traditional Chinese instruments, and 2 indicated voice. The traditional instruments included *guzheng* (Chinese zither), *koto* (Japanese instrument similar to the *guzheng*), *yangqin* (a hammered dulcimer), *erhu* (a two-stringed fiddle), and bamboo flute. Previous research has shown that there are differences in auditory feature processing between different kinds of instrumentalists and musical styles [49,50,51]. Here, it is possible that different styles or cultures of music training could emphasize different auditory features enough to influence the results. Unfortunately, the Western and traditional groups here were too different in number to compare in a statistically meaningful way. However, musical culture remains an interesting question for the future and could be investigated by focusing on style of musical expertise as a design factor.

Limitations

As mentioned above, it was not possible to statistically compare musicians trained in traditional or Western musical styles. It would be of particular interest to compare musicians trained in different tonal systems or on fixed- and movable pitch instruments or vocalists since regular practice of a tonal system with smaller or larger frequency differences between notes may influence discrimination patterns.

One of the main limitations of this work is the lack of a multifeature frequency discrimination task. In the future, some of these questions could be addressed by a more complete set of perceptual tasks, especially since the Mandarin speakers show music-based effects for both frequency and duration.

Some recent research has indicated genetic factors in auditory feature processing and musical aptitude heritability [52,53,54,55,56]. Future studies should consider the impact of genetic differences across major linguistic groups and the effect that difference may have in comparing auditory processing between the groups.

Conclusions

Knowledge about early auditory processing plasticity is becoming more granular and effects specific to certain sound environments are becoming clearer. Future investigations must take into consideration the differences between language environments and musical environments in their effects in tuning the auditory system. Additionally, in order to gain a more complete picture of the plasticity of the auditory system, musicality evaluations should be carefully considered as well as other factors like genetics/aptitude, socio-cultural differences in music attitudes, and behavioral task demands. Musical expertise appears to confer mainly perceptual advantages *within* linguistic groups. The transfer between language and music effects happen on an early level of processing, but responses are still modulated by behavioral goals which drive efferent connections as well as a holistic pressure to efficiency in the full system.

Supporting information

S1 Table. Data values. Data values for each participant used in statistical analyses and demographic information.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Caitlin Dawson, Mari Tervaniemi, Daniel Aalto.

Data curation: Caitlin Dawson.

Formal analysis: Caitlin Dawson.

Funding acquisition: Caitlin Dawson, Mari Tervaniemi.

Investigation: Caitlin Dawson.

Methodology: Caitlin Dawson, Daniel Aalto.

Project administration: Mari Tervaniemi.

Supervision: Mari Tervaniemi, Daniel Aalto.

Writing – original draft: Caitlin Dawson.

Writing – review & editing: Caitlin Dawson, Mari Tervaniemi, Daniel Aalto.

References

1. Kraus N, Chandrasekaran B. Music training for the development of auditory skills. Nature reviews. Neurosci. 2010 Aug 1; 11(8):599.

2. Kishon-Rabin L, Amir O, Vexler Y, Zaltz Y. Pitch discrimination: Are professional musicians better than non-musicians?. J Basic Clin Physiol Pharmacol. 2001; 12(2):125–44.

3. Moreno S, Bidelman GM. Examining neural plasticity and cognitive benefit through the unique lens of musical training. Hear Res. 2014 Feb 28; 308:84–97. https://doi.org/10.1016/j.heares.2013.09.012 PMID: 24079993

4. Tervaniemi M, Just V, Koelsch S, Widmann A, Schröger E. Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. Exp Brain Res. 2005 Feb 1; 161 (1):1–0. https://doi.org/10.1007/s00221-004-2044-5 PMID: 15551089

5. Marques C, Moreno S, Castro SL, Besson M. Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. J Cogn Neurosci. 2007 Sep; 19 (9):1453–63. https://doi.org/10.1162/jocn.2007.19.9.1453 PMID: 17714007

6. Koelsch S, Schröger E, Tervaniemi M. Superior pre-attentive auditory processing in musicians. Neuroreport. 1999 Apr 26; 10(6):1309–13. PMID: 10363945

7. Wong PC, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci. 2007 Apr; 10(4):420. https://doi.org/10.1038/nn1872 PMID: 17351633

8. Lee KM, Skoe E, Kraus N, Ashley R. Selective subcortical enhancement of musical intervals in musicians. J Neurosci. 2009 May 6; 29(18):5832–40. https://doi.org/10.1523/JNEUROSCI.6133-08.2009 PMID: 19420250

9. Musacchia G, Sams M, Skoe E, Kraus N. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc Natl Acad Sci. 2007 Oct 2; 104(40):15894–8. https://doi.org/10.1073/pnas.0701498104 PMID: 17898180

10. Strait DL, Kraus N, Skoe E, Ashley R. Musical experience and neural efficiency–effects of training on subcortical processing of vocal expressions of emotion. Eur J Neurosci. 2009 Feb 1; 29(3):661–8. https://doi.org/10.1111/j.1460-9568.2009.06617.x PMID: 19222564

11. Parbery-Clark A, Tierney A, Strait DL, Kraus N. Musicians have fine-tuned neural distinction of speech syllables. Neuroscience. 2012 Sep 6; 219:111–9. https://doi.org/10.1016/j.neuroscience.2012.05.042 PMID: 22634507

12. Bidelman GM, Krishnan A, Gandour JT. Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. Eur J Neurosci. 2011 Feb 1; 33(3):530–8. https://doi.org/10.1111/j.1460-9568.2010.07527.x PMID: 21198980

13. Shahin A, Roberts LE, Trainor LJ. Enhancement of auditory cortical development by musical experience in children. Neuroreport. 2004 Aug 26; 15(12):1917–21. PMID: 15305137

14. Putkinen V, Tervaniemi M, Huotilainen M. Informal musical activities are linked to auditory discrimination and attention in 2–3-year-old children: an event-related potential study. Eur J Neurosci. 2013 Feb 1; 37(4):654–61. https://doi.org/10.1111/ejn.12049 PMID: 23167769

15. Putkinen V, Tervaniemi M, Saarikivi K, Ojala P, Huotilainen M. Enhanced development of auditory change detection in musically trained school-aged children: a longitudinal event-related potential study. Dev Sci. 2014 Mar 1; 17(2):282–97. https://doi.org/10.1111/desc.12109 PMID: 24283257

16. Strait DL, Kraus N, Parbery-Clark A, Ashley R. Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. Hear Res. 2010 Mar 31; 261(1):22–9.

17. Giuliano RJ, Pfordresher PQ, Stanley EM, Narayana S, Wicha NY. Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. Front Psychol. 2011; 2.

18. Wayland RP, Guion SG. Training English and Chinese listeners to perceive Thai tones: A preliminary report. Lang Learn. 2004 Dec 1; 54(4):681–712.

19. Krishnan A, Xu Y, Gandour J, Cariani P. Encoding of pitch in the human brainstem is sensitive to language experience. Cogn Brain Res. 2005 Sep 30; 25(1):161–8.

20. Krishnan A, Gandour JT, Bidelman GM. The effects of tone language experience on pitch processing in the brainstem. J Neurolinguistics. 2010 Jan 31; 23(1):81–95. https://doi.org/10.1016/j.jneuroling.2009.09.001 PMID: 20161561

21. Krishnan A, Xu Y, Gandour JT, Cariani PA. Human frequency-following response: representation of pitch contours in Chinese tones. Hear Res. 2004 Mar 31; 189(1):1–2.

22. Swaminathan J, Krishnan A, Gandour JT. Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. Neuroreport. 2008 Jul 16; 19(11):1163. https://doi.org/10.1097/WNR.0b013e3283088d31 PMID: 18596621

23. Krishnan A, Swaminathan J, Gandour JT. Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. J Cogn Neurosci. 2009 Jun; 21(6):1092–105. https://doi.org/10.1162/jocn.2009.21077 PMID: 18702588

24. Cooper A, Wang Y. The influence of linguistic and musical experience on Cantonese word learning. J Acoust Soc Am. 2012 Jun; 131(6):4756–69. https://doi.org/10.1121/1.4714355 PMID: 22712948

25. Nenonen S, Shestakova A, Huotilainen M, Näätänen R. Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. Cogn Brain Res. 2003 May 31; 16(3):492–5.

26. Tervaniemi M, Jacobsen T, Röttger S, Kujala T, Widmann A, Vainio M, Näätänen R, Schröger E. Selective tuning of cortical sound-feature processing by language experience. Eur J Neurosci. 2006 May 1; 23(9):2538–41. https://doi.org/10.1111/j.1460-9568.2006.04752.x PMID: 16706861

27. Kirmse U, Ylinen S, Tervaniemi M, Vainio M, Schröger E, Jacobsen T. Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. Int J Psychophysiol. 2008 Feb 29; 67(2):131–43. https://doi.org/10.1016/j.ijpsycho.2007.10.012 PMID: 18160160

28. Dawson C, Aalto D, Šimko J, Putkinen V, Tervaniemi M, Vainio M. Quantity language speakers show enhanced subcortical processing. Biol Psychol. 2016 Jul 31; 118:169–75. https://doi.org/10.1016/j.biopsycho.2016.06.003 PMID: 27297179

29. Marie C, Kujala T, Besson M. Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. Cortex. 2012 Apr 30; 48(4):447–57. https://doi.org/10.1016/j.cortex.2010.11.006 PMID: 21189226

30. Dawson C, Aalto D, Šimko J, Vainio M, Tervaniemi M. Musical sophistication and the effect of complexity on auditory discrimination in Finnish speakers. Front Neurosci. 2017; 11.

31. Bidelman GM, Gandour JT, Krishnan A. Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. Brain Cogn. 2011 Oct 31; 77(1):1–0. https://doi.org/10.1016/j.bandc.2011.07.006 PMID: 21835531

32. Hutka S, Bidelman GM, Moreno S. Pitch expertise is not created equal: cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music. Neuropsychologia. 2015 May 31; 71:52–63. https://doi.org/10.1016/j.neuropsychologia.2015.03.019 PMID: 25797590

33. Lynn R, Wilson RG, Gault A. Simple musical tests as measures of Spearman's g. Personal Individ Differ. 1989 Jan 1; 10(1):25–8.

34. Sergeant D, Vhatcher G. Intelligence, social status and musical abilities. Psychol Music. 1974 Oct; 2(2):32–57.

35. Mosing MA, Pedersen NL, Madison G, Ullén F. Genetic pleiotropy explains associations between musical auditory discrimination and intelligence. PLOS ONE. 2014 Nov 24; 9(11):e113874. https://doi.org/10.1371/journal.pone.0113874 PMID: 25419664

36. Deary IJ. Intelligence and auditory discrimination: Separating processing speed and fidelity of stimulus representation. Intelligence. 1994 Apr 30; 18(2):189–213.

37. Kaernbach C. Simple adaptive testing with the weighted up-down method. Atten Percept Psychophys. 1991 May 1; 49(3):227–9.

38. Rønne FM, Dau T, Harte J, Elberling C. Modeling auditory evoked brainstem responses to transient stimuli. J Acoust Soc Am. 2012 May; 131(5):3903–13. https://doi.org/10.1121/1.3699171 PMID: 22559366

39. Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N. Auditory training improves neural timing in the human brainstem. Behav Brain Res. 2005 Jan 6; 156(1):95–103. https://doi.org/10.1016/j.bbr.2004.05.012 PMID: 15474654

40. Russo NM, Skoe E, Trommer B, Nicol T, Zecker S, Bradlow A, Kraus N. Deficient brainstem encoding of pitch in children with autism spectrum disorders. Clin Neurophysiol. 2008 Aug 31; 119(8):1720–31. https://doi.org/10.1016/j.clinph.2008.01.108 PMID: 18558508

41. Hornickel J, Kraus N. Unstable representation of sound: a biological marker of dyslexia. J Neurosci. 2013 Feb 20; 33(8):3500–4. https://doi.org/10.1523/JNEUROSCI.4205-12.2013 PMID: 23426677

42. Müllensiefen D, Gingras B, Stewart L, Musil J. The Goldsmiths Musical Sophistication Index (Gold-MSI): Technical Report and Documentation v0. 9. London: Goldsmiths, University of London. 2011 Dec 1.

43. Müllensiefen D, Gingras B, Musil J, Stewart L. The musicality of non-musicians: an index for assessing musical sophistication in the general population. PLOS ONE. 2014 Feb 26; 9(2):e89642. https://doi.org/10.1371/journal.pone.0089642 PMID: 24586929

44. Simko J, Aalto J, Lippus P, Włodarczak M, Vainio M. Pitch, perceived duration and auditory biases: Comparison among languages. In18th International Congress of Phonetic Sciences, Glasgow, Scotland, UK, August 10–14, 2015 2015. University of Glasgow.

45. Henry FM. Discrimination of the duration of a sound. J Exp Psychol. 1948 Dec; 38(6):734. PMID: 18893188

46. Pisoni DB. Fundamental frequency and perceived vowel duration. J Acoust Soc Am. 1976 Apr; 59(S1):S39–.

47. Bidelman GM, Gandour JT, Krishnan A. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. J Cogn Neurosci. 2011 Feb; 23(2):425–34. https://doi.org/10.1162/jocn.2009.21362 PMID: 19925180

48. Weidema JL, Roncaglia-Denissen MP, Honing H. Top–down modulation on the perception and categorization of identical pitch contours in speech and music. Front Psychol. 2016; 7.

49. Slater J, Azem A, Nicol T, Swedenborg B, Kraus N. Variations on the theme of musical expertise: cognitive and sensory processing in percussionists, vocalists and non-musicians. Eur J Neurosci. 2017 Apr 1; 45(7):952–63. https://doi.org/10.1111/ejn.13535 PMID: 28177157

50. Tervaniemi M, Janhunen L, Kruck S, Putkinen V, Huotilainen M. Auditory profiles of classical, jazz, and rock musicians: genre-specific sensitivity to musical sound features. Front Psychol. 2015; 6.

51. Vuust P, Brattico E, Seppänen M, Näätänen R, Tervaniemi M. The sound of music: differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. Neuropsychologia. 2012 Jun 30; 50(7):1432–43. https://doi.org/10.1016/j.neuropsychologia.2012.02.028 PMID: 22414595

52. Oikkonen J, Huang Y, Onkamo P, Ukkola-Vuoti L, Raijas P, Karma K, Vieland VJ, Järvelä I. A genome-wide linkage and association study of musical aptitude identifies loci containing genes related to inner ear development and neurocognitive functions. Mol Psychiatry. 2015 Feb; 20(2):275. https://doi.org/10.1038/mp.2014.8 PMID: 24614497

53. Selinger L, Zarnowiec K, Via M, Clemente IC, Escera C. Involvement of the serotonin transporter gene in accurate subcortical speech encoding. J Neurosci. 2016 Oct 19; 36(42):10782–90. https://doi.org/10.1523/JNEUROSCI.1595-16.2016 PMID: 27798133

54. Kanduri C, Kuusi T, Ahvenainen M, Philips AK, Lähdesmäki H, Järvelä I. The effect of music performance on the transcriptome of professional musicians. Sci Rep. 2015; 5.

55. Pulli K, Karma K, Norio R, Sistonen P, Göring HH, Järvelä I. Genome-wide linkage scan for loci of musical aptitude in Finnish families: evidence for a major locus at 4q22. J Med Genet. 2008 Jul 1; 45(7):451–6. https://doi.org/10.1136/jmg.2007.056366 PMID: 18424507

56. Ukkola LT, Onkamo P, Raijas P, Karma K, Järvelä I. Musical aptitude is associated with AVPR1A-haplotypes. PLOS ONE. 2009 May 20; 4(5):e5534. https://doi.org/10.1371/journal.pone.0005534 PMID: 19461995