Tartu Ülikool

Sotsiaalteaduste valdkond

Psühholoogia instituut

Emily Ellervee

Musicality and perception of Estonian as a foreign language by Chinese speakers

Uurimistöö

Juhendaja: Siqi Lyu, PhD

Jooksev pealkiri: Musicality and perception of language

Tartu 2024

Musicality and perception of Estonian as a foreign language by Chinese speakers

Abstract

This study compared Chinese musicians' and non-musicians' performance on a same-different discrimination task to examine whether and how musical experience influenced tonal language speakers' perception of the Estonian quantity. Musicians outperformed non-musicians in the duration condition regardless of the stimulus type. This indicates that musicality benefits perception of cues not present in native language. In the conditions where pitch cue was present, musicians and non-musicians performed similarly. This could be because language experience shapes the brain to be more sensitive to the perception of the linguistically relevant cues, which in the case of Chinese speakers is pitch. Both groups were better at perceiving word stimuli than pure tone stimuli. Taken together, results suggest that native language experience overrides musicality, at least at the attentive level.

Keywords: non-native language perception, musicality, Chinese speakers, Estonian quantity

Musikaalsus ja eesti keele kui võõrkeele tajumine hiina keele kõnelejate poolt

Lühikokkuvõte

Selles uuringus võrreldi Hiina muusikute ja mittemuusikute sooritust sama-erinev diskrimineerimisülesandes, et uurida, kas ja kuidas muusikakogemus mõjutas tonaalse keele kõnelejate taju eesti väldetest. Muusikud edestasid kestuse tingimuses mittemuusikuid olenemata stiimuli tüübist. See tulemus viitab sellele, et musikaalsus aitab tajuda vihjeid, mida emakeeles ei esine. Tingimustes, kus esines helikõrgus, oli muusikute ja mittemuusikute sooritus sarnane. Selline tulemus võib tuleneda sellest, et keelekogemus muudab aju tundlikumaks keeleliselt oluliste vihjete tajumise suhtes, mis hiina keele kõnelejate puhul on helikõrgus. Mõlemad rühmad tajusid paremini sõnalisi stiimuleid kui lihthelilisi stiimuleid. Kokkuvõttes näitavad tulemused, et emakeelekogemus alistab musikaalsuse, vähemalt tähelepanelikul tasandil.

Märksõnad: võõrkeeletaju, musikaalsus, hiina keele kõnelejad, eesti keele välted

Introduction

Music and language are cognitive systems that form “hierarchically structured sequences according to syntactic principles” (Patel, 2003), and are produced using the same organs (i.e., the larynx). As there are rules around the way letters or characters form words which then form sentences, there are similar rules for music, where singular notes create chords, which in turn create chord progressions (Patel, 2003).

The Overlap between Music and Language

The degree of similarities between the two systems has been widely researched from the neuroanatomical perspective. Maess et al. (2001) and Schön et al. (2010) both found overlap in music and language processing in certain parts of the temporal and frontal brain regions, including the Broca’s area and its right-hemisphere homologue. In both studies, although the left hemisphere was found to be more involved with language processing and the right hemisphere more involved with musical processing, both hemispheres were found to be considerably active in both domains. Following this finding, Maess et al. (2001) and Schön et al. (2010) both theorized about a possible network for processing complex auditory input that is not music or language specific.

The overlap between the two cognitive systems has given way to theories of facilitatory cross-domain transfer effects. For example, the Overlap, Precision, Emotion, Repetition, Attention (OPERA) hypothesis proposes that, through musical training, neural plasticity drives the shared networks to function with higher precision than is needed for speech, which in turn benefits the language perception (Patel, 2011). For that reason, musical education might be helpful in the perception of languages with different prosodic systems, accustoming the person to the note frequency and time relations in music which they can then use to perceive the pitch and temporal cues, respectively.

Effect of Musicality on the Perception of Pitch

The role of musicality has been studied most widely in the perception of lexical pitch or tone, predominantly among *non-native* speakers’ perception of tonal languages (e.g., Alexander et al., 2005; Lee & Hung, 2008; Zhao & Kuhl, 2015). In tonal languages like Mandarin Chinese (Mandarin or Chinese hereafter), pitch is used lexically to discriminate word meanings (Lin, 2007). As shown by Alexander et al. (2005), when asked to discriminate between the four Mandarin lexical tones from the twenty monosyllabic Mandarin words, English-speaking musicians outperformed non-musicians in both identification and discrimination accuracy, and their discrimination accuracy was comparable to that of native Mandarin speakers. In another study, Zhao and Kuhl (2015) found English-speaking musicians to score higher than English and even Mandarin non-musicians in pitch discrimination task, although their lexical tone continuum perception manner was closer to that of English non-musicians than native Mandarin speakers.

Musicality has been found to be helpful even in *native* tonal language speakers' perception of pitch, although the results are inconclusive. To test if these effects were only specific to musical experience, Nan et al. (2018) compared three groups (piano training, reading training control, no-contact control) of Mandarin-speaking children between the ages of 4 and 5 after training for 6 and 12 months in respective assignments. The results showed that the piano training group had the most enhancements in neural processing of both the musical pitch and the lexical tone, as well as behavioral sensitivity to auditory word discrimination. Similar results have been found mostly in studies focusing on children (e.g., Besson et al., 2007; Chobert et al., 2014; Moreno et al., 2009), who have a possible advantage over adults in terms of neural plasticity. Studies conducted on adult musician and non-musician participants, however, showed no significant behavioral enhancements in lexical tone processing in native speakers of tonal languages (Mok & Zuo, 2012; Tang et al., 2016). Cooper and Wang's (2012) and Maggu et al.'s (2018) studies both confirmed the notion that, when it comes to lexical pitch processing in adults, the effects of tone language and musical backgrounds are not additive.

Effect of Musicality on the Perception of Duration

The research on how musical experience affects the perception of temporal cues like duration in language is rather sparse. According to the Precise Auditory Timing Hypothesis (PATH), musical training contributes to a more precise perception of timing and rhythm in acoustic events, which in turn leads to enhanced sensitivity in the perception of speech timing (Tierney & Kraus, 2014). Cooper et al. (2017) found that English musicians outperformed non-musicians in discriminating between-category and within-category vowel lengths across speaking rates, performing on par with native Thai listeners. Similarly, in a study by Sadakata and Sekiyama (2011), Dutch musicians outperformed non-musicians in the discrimination of Japanese temporal cues, such as the duration of stop consonants. Musicality was found to give an advantage in pre-attentive processing of duration in Finnish (Milovanov et al., 2009) and French (Chobert et al., 2014) children as well.

The majority of the reviewed literature on the perception of temporal cues only examines languages such as Finnish and Japanese or uses monosyllables as speech stimuli. Therefore, the conclusions that can be drawn from the literature are limited. A study that looks at the temporal cue in another quantity language, e.g., Estonian, is needed to fill the gaps in the existing literature and reduce the inconsistencies.

Estonian is a quantity language where both duration and pitch cues exist in its phonetic system (Lippus et al., 2009). Estonian uses a three-way contrastive quantity system in disyllabic words – short (Q1), long (Q2), and overlong (Q3). As identified by Lehiste (1997), the duration ratios are 2:3 for Q1, 3:2 for Q2, and 2:1 for Q3, with the pitch peak in Q1 and Q2 being in the second half of the first syllable and in the beginning of the first syllable in Q3 (Lehiste, 1997, as cited in Lippus & Pajusalu, 2009). Estonian speakers use both duration and pitch cues for the quantity perception. Lippus et al. (2009) have found that Estonian speakers use the context of vowel duration ratios and pitch movement to co-signal quantity contrast, with pitch as a secondary cue being crucial specifically for Q2/Q3 differentiation (Lippus et al., 2009; Lippus et al., 2011; Lippus et al., 2013).

The nature of Estonian quantities has repeatedly been found as hard to grasp for non-native speakers. Leppik et al. (2019) found that Spanish L1 speakers had trouble distinguishing the long (Q2) and overlong (Q3) quantities. Similar results have been found in L1 Russian, Latvian, and Finnish speakers (Meister & Meister, 2012; Meister & Meister, 2013; Meister & Meister, 2014). Whether and how the musical background among non-native Estonian speakers affects the perception of the Estonian quantity remains to be explored.

Effect of Musicality on the Perception of Non-linguistic Stimuli

When it comes to the musicality effect on the perception of linguistic and non-linguistic stimuli, the results are inconclusive. Research comparing musicians' and non-musicians' perception of non-linguistic stimuli focuses in large part on the pre-attentive processing. Using Mandarin monosyllables and harmonic sounds as stimuli, Martínez-Montes et al. (2013) found that the musician group showed stronger neural responses than non-musicians to non-linguistic stimulus deviants as compared to linguistic stimulus deviants; however, no significant group differences were found for duration deviants in linguistic stimuli. In a mismatch negativity (MMN) study by Milovanov et al. (2009), enhanced processing of duration in both linguistic (monaural Finnish vowel /ö/) and non-linguistic stimuli was found in Finnish subjects with greater musical aptitude as compared to the less-advanced subject group. Frey et al. (2022) used a training group that received psychoacoustic musical training on pitch and a control group that received psychoacoustic training on intensity to test the music-to-language transfer effects at the pre-attentive processing level. They found that pitch-based musical training had no music-to-language transfer effects on the pre-attentive processing of pitch in linguistic stimuli. However, musical training did have a positive influence on attentive perception of pitch in linguistic stimuli and pre-attentive perception of non-linguistic pitch.

The Present Study

The current study aims to examine if musical experience among tonal language speakers has a positive effect on perceiving the Estonian quantity. To this end, I will include two groups of participants, i.e., Chinese musicians and Chinese non-musicians, and adopt a same-different discrimination task (McGuire, 2010). Given that the Estonian quantity is a complex combination of both duration and pitch cues, three conditions will be examined, including the duration only, the pitch only, and the pitch plus duration conditions. Non-linguistic stimuli (pure tones) will also be included in the experiment. Specifically, I have the following hypotheses.

Firstly, musicality gives the person an advantage in the perception of note frequency and time relations in non-linguistic stimuli (Frey et al., 2022; Martínez-Montes et al., 2013; Milovanov et al., 2009). Thus, I postulate that musicians will outperform non-musicians in the perception of all three conditions (pitch, duration, pitch+duration) in pure tone stimuli (H1).

Secondly, it has been proposed that musical experience interacts with native language phonology in a way that only enhances the processing of acoustic events lacking in one's native language (Dawson et al., 2017; see also Mok & Zuo, 2012 and Tang et al., 2016). Based on that, I predict musical experience to have a positive effect only on the perception of linguistic duration (H2) but not linguistic pitch (H3) among Chinese speakers, as both groups of Chinese speakers should be well-attuned to perceive the pitch information due to their native language experience.

Lastly, regarding the pitch+duration condition of the present study, it is unclear which cue will override the other. If pitch overrides duration, then I will expect the linguistic pitch+duration condition to have the same results as the linguistic pitch condition, i.e., musicians will perform similarly to non-musicians. If duration overrides pitch, then I will expect the pitch+duration condition to have the same results as the linguistic duration condition, i.e., musicians will outperform non-musicians (H4).

Method

Participants

The study included 22 musician and 22 non-musician native Chinese speakers. Table 1 shows the sex and age of the participants. Non-musicians were selected from a larger pool of non-musician participants that were included in a study by Lyu et al. (2023) to match the sex and age of the musicians. The musicians had received more than seven years of formal training in any musical instrument or vocal outside of school activities and were still having at least weekly practice at the time of the experiment. The non-musicians had no self-reported prior musical education.

Table 1

The sex and age of participants

Musicality	Female	Male	Min. Age	Max. Age	Mean Age (*SD)*
Musicians	13	9	18	28	20.50 (2.22)
Non-musicians	13	9	18	25	20.86 (1.52)
Total	26	18	18	28	20.68 (1.89)

Note. Female = number of female participants; Male = number of male participants; Min. Age = minimum age of participants in years; Max. Age = maximum age of participants in years; *SD* = standard deviation.

Participants were recruited through flyers that were distributed on social media and around the campus of Zhejiang University of Technology in China (ZJUT). The study was approved by the Institutional Review Board, College of Foreign Languages, ZJUT. Participants gave their written consent before the experiment and had the freedom to withdraw from the experiment at any time. The experiment was non-invasive and did not cause harm to the participants. The personal data collected from the participants were pseudonymized to ensure confidentiality. After the experiment, participants received a gift card as a reward.

Materials

Four word stimuli were created for the experiment in Praat software (Boersma, 2001). The stimuli were created from two base Estonian words: *saada* “to send” (Q2) and *saada* “to get” (Q3). The base words were recorded by a female native Estonian speaker in a quiet room with a sampling frequency of 44.1k Hz using Praat. The first vowel of the two base words was manipulated to be either short (170 ms) or long (290 ms). Other consonant and vowel lengths were kept the same within each condition (Table 2).

Table 2

Physical features of the Estonian stimuli (ms)

Stimulus	Consonant 1	Vowel 1	Consonant 2	Vowel 2
	S	A	D	A
Q2_170	90	170	74	115
Q2_290	90	290	74	115
Q3_170	100	170	90	68
Q3_290	100	290	90	68

Note. Q2_170: Estonian long quantity (Q2) with a short first-vowel manipulation (170 ms),

Q2_290: Estonian long quantity (Q2) with a long first-vowel manipulation (290 ms),

Q3_170: Estonian overlong quantity (Q3) with a short first-vowel manipulation (170 ms),

Q3_290: Estonian overlong quantity (Q3) with a long first-vowel manipulation (290 ms).

The pure tone version of the four stimuli was additionally created by extracting the pitch contour from each of the word stimuli and then creating a sine wave from the extracted pitch contour in Praat. The pure tone stimuli assembled the physical features of the corresponding word stimuli except that they did not carry any consonants or vowels. There were four pure tone stimuli in total.

Procedure

The AX experiment was conducted in the Cross-linguistic Studies and Brain Science Lab at ZJUT and was part of a series of experiments (see Lyu et al., 2023, for details). Upon arriving at the lab, participants completed a background questionnaire on paper, followed by an EEG experiment and a behavioral AX experiment, which altogether lasted around 2.5 hours. The experiments were conducted in a quiet, dimmed, and electrically shielded room. Only the results of the AX experiment were analysed in this paper.

The AX experiment was conducted using the E-Prime 3.0 software (Psychology Software Tools, Pittsburgh, PA). Participants were presented through earphones with a pair of audio stimuli with an interstimulus interval of 300ms, after which they had 1000ms to answer if the pairs presented were the same or different using designated keys on a keyboard – “f” for same and “j” for different. The instructions were given to the participants in their native language.

For the experiment, word pairs were created using all possible combinations among the four stimuli. Out of the 12 “different” pairs, 4 pairs each accounted for the differences in duration, pitch, and pitch+duration. 4 pairs made up the “same” pair condition. To balance the number of “different” and “same” pairs presented, the “same” pairs and “different” pairs were repeated different times, adding up to 120 pairs of word stimuli presented. Table 3 details the combinations of pairs achieved and the condition each combination accounts for, as well as the number of repetitions used on each pair.

Table 3

Word pairs produced from the four types of stimuli

A	X	Repetitions	Pair type
Q2_170	Q2_290	5	Different (Duration)
Q2_170	Q3_170	5	Different (Pitch)
Q2_170	Q3_290	5	Different (Pitch+Duration)
Q2_290	Q2_170	5	Different (Duration)
Q2_290	Q3_170	5	Different (Pitch+Duration)
Q2_290	Q3_290	5	Different (Pitch)
Q3_170	Q2_170	5	Different (Pitch)

A	X	Repetitions	Pair type
Q3_170	Q2_290	5	Different (Pitch+Duration)
Q3_170	Q3_290	5	Different (Duration)
Q3_290	Q2_170	5	Different (Pitch+Duration)
Q3_290	Q2_290	5	Different (Pitch)
Q3_290	Q3_170	5	Different (Duration)
Q2_170	Q2_170	15	Same
Q2_290	Q2_290	15	Same
Q3_170	Q3_170	15	Same
Q3_290	Q3_290	15	Same

Note. Q2_170: Estonian long quantity (Q2) with a short first-vowel manipulation (170 ms), Q2_290: Estonian long quantity (Q2) with a long first-vowel manipulation (290 ms), Q3_170: Estonian overlong quantity (Q3) with a short first-vowel manipulation (170 ms), Q3_290: Estonian overlong quantity (Q3) with a long first-vowel manipulation (290 ms).

The same manipulation was done to the pure tone stimuli, adding up to 240 pairs presented to each participant. Each type of stimuli was divided equally into two blocks of 60 pairs, with the pure tone stimuli being presented in blocks 1 and 3 and word stimuli in blocks 2 and 4. Within each block, the pairs were presented in a randomized order. Between each block, the participants were allowed to have a short break. The AX experiment lasted about 15 minutes in total.

Results

The raw data was converted into d-primes using the hit rate – subject responded correctly “j” for different to a different pair – and false alarm rate – subject responded incorrectly “j” for different to a same pair – of each subject. In hit and false alarm rate calculations, only the trials in which the subject had answered within the 1000ms time frame were included. From the musician group, a total of 86 out of 5280 trials were excluded (1.63%), 66 out of 5280 trials (1.25%) were excluded from the non-musician group.

D-primes were calculated using Excel formula for d-primes described in the UCLA Phonetics Lab document for d-prime analysis (Keating, 2005) for each subject in all three conditions (pitch, duration, pitch+duration) and in both stimulus types (word, pure tone). Due to Excel's limitations, the values of the specified mean were set to 0.01 and the specified standard deviation were set to 0.99, which set the highest d-prime value to 4.65.

To compare the differences in the within-subject variable (stimulus type) between the two groups, d-prime scores were then analyzed using three repeated measures analyses of variance (or RMANOVAs), one for each condition. In each RMANOVA, the two levels of stimulus types (word vs pure tone) were included as a within-subject factor, and musicality was set as the between-subject factor. Because all RMANOVAs passed the assumption checks, parametric tests were used. For effect size, $\eta_p{}^2$ statistic was used.

Table 4 summarizes musicians' and non-musicians' d-prime score means and standard deviations for both stimulus types in all three conditions.

Table 4

D-prime mean scores and standard deviation

Condition	Stimulus type	Musicality	*N*	*M*	*SD*
Pitch	Word	Musician	22	4.17	0.51
		Non-musician	22	3.97	0.54
	Pure tone	Musician	22	3.88	0.74
		Non-musician	22	3.63	0.81
Duration	Word	Musician	22	2.91	0.96
		Non-musician	22	2.08	0.97
	Pure tone	Musician	22	2.15	0.87
		Non-musician	22	1.68	0.81
Pitch+duration	Word	Musician	22	4.16	0.47
		Non-musician	22	4.02	0.54
	Pure tone	Musician	22	3.79	0.74
		Non-musician	22	3.77	0.72

Note. *N* = number of subjects; *M* = d-prime mean score; *SD* = standard deviation.

There was a significant main effect of stimulus type in all three conditions: F (1, 42) = 9.53, $p = .004$, $\eta_p{}^2 = .19$ for pitch; F (1, 42) = 27.22, $p < .001$, $\eta_p{}^2 = .39$ for duration; F (1, 42) = 11.66, $p = .001$, $\eta_p{}^2 = .22$ for pitch+duration. In all conditions, participants performed better in perceiving word stimuli than pure tone stimuli, regardless of the musical background. There was a main effect of musicality in the duration condition, where musicians performed significantly better than non-musicians, *F* (1, 42) = 6.92, *p* = .012, $\eta_p{}^2$ = .14, but there was no effect of musicality in either pitch (*p* > .1) or pitch+duration conditions (*p* > .6). No significant interaction effects between musicality and stimulus type were found in any of the three conditions, *p* > .7, for pitch; *p* > .1 for duration; *p* > .5 for pitch+duration. The results are illustrated in Figure 1.

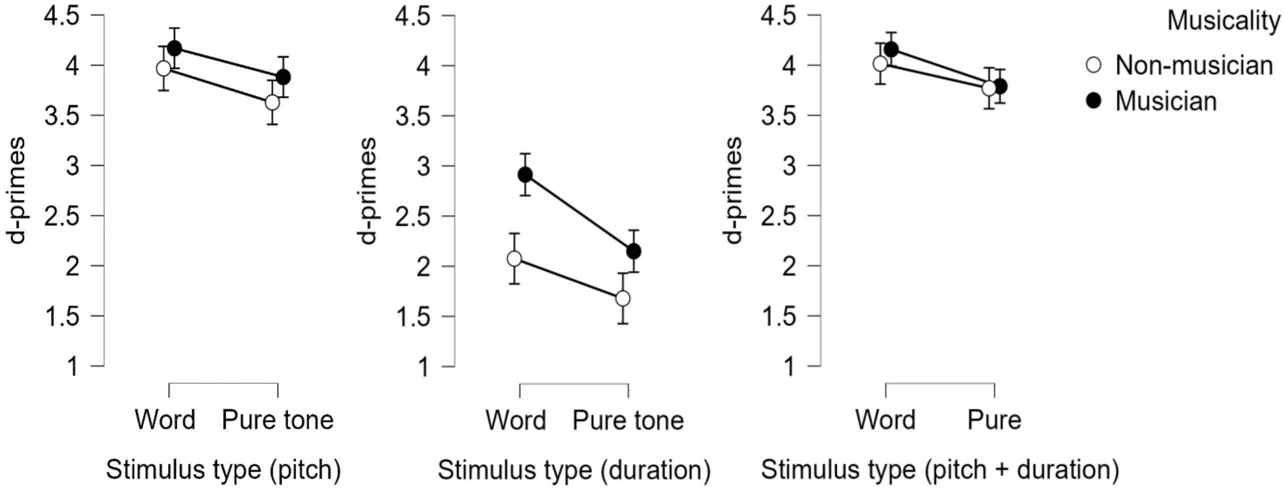

Figure 1. D-prime mean scores of musicians and non-musicians in all three conditions (pitch, duration, pitch+duration) in both stimulus types (word, pure tone).

Discussion

The current study set out to examine whether and how musical experience among tonal language speakers had a positive effect on perceiving the Estonian quantity. This was done using two groups of participants, Chinese musicians and Chinese non-musicians, and a same-different discrimination task (McGuire, 2010). The experiment used three conditions (pitch, duration, pitch+duration) and both word and pure tone stimuli were included. I hypothesized that musicians would outperform non-musicians in perception of non-linguistic pure tone stimuli in all three conditions (pitch, duration, pitch+duration) (H1). I expected musical experience to have a positive effect only on the perception of linguistic duration (H2) but not linguistic pitch (H3). Regarding the linguistic pitch+duration condition, I decided to leave it open because it is unknown which cue overrides the other (H4).

Duration (H1 and H2)

In the duration condition, musicians were significantly better in perceiving pure tone stimuli than non-musicians. However, in the pitch and pitch+duration conditions, there was no significant difference in the performances of the two groups in perceiving the pure tone stimuli. This means my H1 was only partially confirmed. In the duration condition, musicians outperformed non-musicians in perceiving word stimuli as well. With that, H2 is confirmed. Together, the results suggest that Chinese musicians are better than non-musicians in the perception of duration in both linguistic and non-linguistic contexts.

Previous studies examining the effect musicality has on perception of duration focus on languages such as Finnish, Thai, and Japanese. Sadakata and Sekiyama (2011) found that Dutch musicians discriminated Japanese temporal cues, such as the duration of stop consonants, better than non-musicians. English musicians were found to perform better than non-musicians and on par with native Thai speakers in discriminating between-category and within-category vowel lengths across speaking rates in Cooper et al. (2017). Enhanced neural processing of both linguistic and non-linguistic duration was found even in native quantity-language (Finnish) speaking children with more musical aptitude (Milovanov et al., 2009). The current study sought to find if these results could be replicated using a quantity language such as Estonian. The results are consistent with the previous literature. Musicians outperformed non-musicians in the duration condition regardless of the stimulus type.

More importantly, this result supports Dawson et al.'s (2017) proposition that musical experience enhances only the processing of acoustic events lacking in one's native language. Mandarin is a tonal language; pitch is used lexically to discriminate word meanings (Lin, 2007). Whereas both Chinese musicians and non-musicians are sensitive to differences in pitch, musicians have the advantage of a more precise perception of temporal information in sounds, which is less important in the Chinese language system. As shown in the current study, Chinese musicians are better at perceiving duration in both Estonian words and their non-linguistic equivalents.

Pitch (H3 and H4)

Musicians performed on par with non-musicians in the pitch condition of both word and pure tone. This confirms my H3: musicality does not have significant positive effect on Chinese speakers' perception of linguistic pitch. Regarding my H4, I found no significant difference between the performance of musicians and non-musicians. This indicates that for Mandarin speakers, if presented with the pitch and the duration cues at the same time, the duration cue is overridden by the pitch cue.

My results showed that Chinese musicians did not necessarily outperform non-musicians in perceiving pitch-related information in a foreign language. The results match those of Mok and Zuo (2012) and Tang et al. (2016). In adult native tonal language speakers, musicality does not affect lexical tone processing at the behavioral/attentive level. A possible reason for that is because the effects of language experience and musical background are not additive (Maggu et al., 2018). Language processing is automatic on a neural level. It is not far-fetched to speculate that one's native language also shapes their brain to enhance the processing of linguistically necessary cues, like pitch in case of Mandarin speakers. Krishnan and Gandour's literature review (2010) supports this claim; native tone language speakers have been found to have higher sensitivity to specific pitch contours relevant to their language in the brainstem.

Noteworthy, in their study on Chinese musicians' and non-musicians' perceptions of Chinese tone, although Tang et al. (2016) did not find behavioral evidence (i.e., accuracy) for musicality enhancing perception of pitch in native Mandarin speakers, there was a significant difference in the response time and electrical responses between the two groups. Tang et al. (2016) concluded from these findings that the discrimination task lacked sensitivity to detect individual differences in lexical tone processing for musicians. In the present study, the enhancements might not be noticeable in the measurement we examined, i.e., the discrimination sensitivity. However, musicality might have had a positive effect on a different behavioral measure, like response time, and/or the pre-attentive level.

The musicians performed similarly to non-musicians in the linguistic pitch+duration condition, suggesting that the duration cue was overridden by the pitch cue. Chinese non-musicians can make use of their native language knowledge about pitch to perceive the Estonian pitch+duration stimuli to the same extent as musicians. Language experience affects the brain on a structural level, sensitizing the speakers to acoustic events relevant to their language (Chen et al., 2018; Krishnan & Gandour, 2010; Liu et al., 2023; Peng et al., 2013). For Mandarin speakers, the ability to perceive pitch is ingrained in their neural networks, making it automatic. In the context of the current study, this could mean that the language experience is more influential in language processing, or even general acoustic perception, than musicality.

In conclusion, for Mandarin speakers, musicality has an advantage when it comes to the perception of duration, both linguistic and non-linguistic. However, musicality does not significantly impact the perception of either stimulus type if the duration cue is combined with the pitch cue. Language experience shapes the brain to be more sensitive to the perception of the linguistically relevant cues, which in the case of Mandarin speakers is pitch. This indicates that, among Chinese speakers, native language background overrides the musical background when perceiving pitch in a foreign language, at least at the behavioral/attentive level.

Additional findings

Literature focusing on the behavioral effects musicality might have on the perception of non-linguistic stimuli is sparse. So is literature comparing the effects of native language experience and musicality on auditory perception. The results of my study help to fill the gap. In this study, I found that both musicians and non-musicians were better at perceiving word stimuli than pure tone stimuli in all three conditions. This could stem from humans' inherent sensitivity to speech as opposed to non-speech sounds. It has been found that newborns prefer human speech over non-speech, even when controlled for acoustic complexity (Dehaene-Lambertz et al., 2010; Vouloumanos & Werker, 2007). Dehaene-Lambertz et al. (2010) found this sensitivity to be present on a subcortical/neural level as well, with speech having higher left-hemisphere temporal activation compared to music. Similarly, Parviainen et al. (2004) found N100m responses for linguistic sounds to be left-lateralized, whereas for non-speech sounds, the responses were more equally distributed across the hemispheres. Binder et al. (2000) found superior temporal sulcus to have stronger responses to speech (words, pseudowords, and reversed speech) than to non-speech (white noise and pure sine wave tones). Speech perception being an automized process was also confirmed by Johnson and Ralston (1994): linguistic stimuli were easier to categorize and they were categorized faster than non-linguistic stimuli. This heightened sensitivity to linguistic stimuli offers an explanation to my finding. Humans are wired to be good at linguistic perception, so the linguistic stimuli get processed faster and more efficiently than non-linguistic stimuli. Brain's equivalent response to linguistic stimuli, be it actual words, pseudowords, or reversed speech (Binder et al., 2000) suggests this sensitivity to linguistic stimuli could possibly extend to non-native language stimuli as well.

Limitations and future research

Despite the significant musicality main effect in the duration condition, visual inspection of Figure 1 shows another difference among conditions. Both subject groups did the worst in the duration condition compared to the other two conditions (i.e., pitch and pitch+duration). Whereas the scores for the pitch and pitch+duration conditions range between 3.63 and 4.17 for both groups in both stimulus types, the mean score range for the duration condition is 1.68 to 2.91. The current study does not allow for further significant effects to be seen. This study focused on musicality, the difference among conditions was not relevant and thus the conditions were not considered as a variable in the analysis. By conducting a more complex statistical analysis that includes the condition variable, more can be inferred about the way musicality interacts with different auditory cues.

Current findings are best accompanied with electrophysical evidence to see if the current AX-discrimination task is sensitive enough to measure effects of musicality on perception of language. Frey et al. (2022) found musical training to be beneficial on attentive perception of pitch in linguistic stimuli and pre-attentive perception of non-linguistic pitch, but not on pre-attentive processing of pitch in linguistic stimuli. Musicians' behavioral discrimination of the non-linguistic pitch changes in the study by Tervaniemi et al. (2005) was faster and more accurate compared to non-musicians. However, the two groups' neural responses did not differ except for N2b being larger for musicians than non-musicians.

There is a possibility that, similarly to the idea proposed by Tang et al. (2016), musicality does influence the perception of pitch and/or non-linguistic sounds, but with the current design, the effects would be noticeable only on a subcortical level. The behavioral task might not have been sensitive enough to catch that. The current design was able to catch a significant difference in the condition both groups had differences in prior experience with (duration), so we know it is sensitive enough for testing when it comes to perception of languages with different prosodies. Both Tang et al. (2016) and Tervaniemi et al. (2005) found one behavioral difference between the two groups to be the response time, which was not included in this study. Taking response time into account in the future will add depth to our understanding of how musicality affects auditory perception on a behavioral level.

My first hypothesis was only partially confirmed, musicians did not outperform non-musicians in the tasks using pure tone stimuli in two of the three conditions. There is a certain possibility that it was due to the non-linguistic stimuli (pure tones) that was used. Nikjeh et al. (2009) found for example that physiological detection of pure tones in P1 was slower for musicians than non-musicians. For harmonic tone complexes, both groups had similar P1 latencies and musicians had shorter MMN and P3 latencies, indicating higher automaticity and attentional efficiency in processing. Musicians perceiving pure tones as irrelevant sensory stimuli and harmonic sounds as familiar and relevant stimuli was suggested as a possible explanation for the finding. Testing, whether changing the non-linguistic stimuli, for example to harmonic sound complexes, makes a difference in musicians' performance, might shed additional light on my results in the context of H1.

Specificity of musical experience is something to expand on in further research as well. Slater et al. (2017) found that for percussionists, enhancements in perception happened more in the acoustic features associated with duration, whereas vocalists had advantages in pitch perception. Current study did not differentiate between the instruments of the musicians, which could have influenced the results. Controlling for musicians' instruments or focusing on the differentiation of the instruments as a variable will deepen our understanding of effects musical background has on linguistic perception.

Choosing Estonian language is advantageous because of its prosody. Having both the pitch and the duration cue lends itself to more in-depth analysis of how musicality can influence the perception of different cues. However, to test my proposition that musicality is most helpful in the perception of languages with opposing prosody, further research into Mandarin speakers' perception of other languages with similar and differing prosodies is needed.

References

Alexander, J. K., Wong, P. C. M., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. *Proc. Interspeech 2005*, 397–400. https://doi.org/10.21437/interspeech.2005-271

Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical training on pitch processing in music and language. *Restorative neurology and neuroscience. 25*, 399-410.

Binder, J. R. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, *10*(5), 512–528. https://doi.org/10.1093/cercor/10.5.512

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

Chen, A., Peter, V., Wijnen, F., Schnack, H. G., & Burnham, D. K. (2018). Are lexical tones musical? Native language's influence on neural response to pitch in different domains. *Brain and Language, 180–182*, 31–41. https://doi.org/10.1016/j.bandl.2018.04.006

Chobert, J., François, C., Velay, J., & Besson, M. (2014). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex*, *24*(4), 956–967. https://doi.org/10.1093/cercor/bhs377

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *Journal of the Acoustical Society of America*, *131*(6), 4756–4769. https://doi.org/10.1121/1.4714355

Cooper, A., Wang, Y., & Ashley, R. (2017). Thai rate-varied vowel length perception and the impact of musical experience. *Language and Speech*, *60*(1), 65–84.

Dawson, C., Aalto, D., Šimko, J., Vainio, M., & Tervaniemi, M. (2017). Musical sophistication and the effect of complexity on auditory discrimination in Finnish speakers. *Frontiers in Neuroscience*, *11*. https://doi.org/10.3389/fnins.2017.00213

Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allirol, L., Dubois, J., Hertz-Pannier, L., & Dehaene, S. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and Language*, *114*(2), 53–65. https://doi.org/10.1016/j.bandl.2009.09.003

Frey, A., Barbaroux, M., Dittinger, E., & Besson, M. (2022). Effects of psychoacoustic training on the pre-attentive processing of harmonic sounds and syllables. *Journal of Speech Language and Hearing Research*, 1–13. https://doi.org/10.1044/2022_jslhr-21-00441

Johnson, K., & Ralston, J. (1994). Automaticity in speech perception: Some speech/nonspeech comparisons. *Phonetica*, *51*(4), 195–209. https://doi.org/10.1159/000261975

Keating, P. (2005). *D-prime (signal detection) analysis*. The UCLA Phonetics Lab. http://phonetics.linguistics.ucla.edu/facilities/statistics/dprime.htm

Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language*, *110*(3), 135–148. https://doi.org/10.1016/j.bandl.2009.03.005

Lee, C., & Hung, T. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *Journal of the Acoustical Society of America*, *124*(5), 3235–3248. https://doi.org/10.1121/1.2990713

Leppik, K., Lippus, P., & Asu, E. L. (2019). *The production of Estonian vowels in three quantity degrees by Spanish L1 speakers* [Conference paper]. International Congress of Phonetic Sciences 2019, Melbourne, Australia.

Lin, Y. (2007). *The sounds of Chinese*. Cambridge University Press. http://ci.nii.ac.jp/ncid/BA84071515

Lippus, P., & Pajusalu K. (2009). Regional variation in the perception of Estonian quantity. In Martti Vainio, Reijo Aulanko & Olli Aaltonen (Eds.), *Nordic Prosody. Proceedings of the Xth Conference, Helsinki 2008* (pp. 151–157). Frankfurt: Peter Lang.

Lippus, P., Asu, E. L., Teras, P., & Tuisk, T. (2013). Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian. *Journal of Phonetics*, *41*(1), 17–28. https://doi.org/10.1016/j.wocn.2012.09.005

Lippus, P., Pajusalu, K., & Allik, J. (2009). The tonal component of Estonian quantity in native and non-native perception. *Journal of Phonetics*, *37*(4), 388–396. https://doi.org/10.1016/j.wocn.2009.07.002

Lippus, P., Pajusalu, K., & Allïk, J. (2010). The role of pitch cue in the perception of the Estonian long quantity. In S. Frota, G. Elordieta, & P. Prieto (Eds.), *Prosodic Categories: Production, Perception and Comprehension* (1st ed., pp. 231–242). Springer Dordrecht. https://doi.org/10.1007/978-94-007-0137-3_10

Liu, J., Hilton, C. B., Bergelson, E., & Mehr, S. A. (2023). Language experience predicts music processing in a half-million speakers of fifty-four languages. *Current Biology*, *33*(10), 1916-1925.e4. https://doi.org/10.1016/j.cub.2023.03.067

Lyu, S., Põldver, N., Kask, L., Wang, L., & Kreegipuu, K. (2023). *Native language background affects the perception of duration and pitch* [Manuscript under review]

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience*, *4*(5), 540–545. https://doi.org/10.1038/87502

Maggu, A.R., Wong, P.C.M., Liu, H., Wong, F.C.K. (2018). Experience-dependent influence of music and language on lexical pitch learning is not additive. In: *Interspeech, 2018,* 3791-3794. doi: 10.21437/Interspeech.2018-2104

Martínez-Montes, E., Hernández-Pérez, H., Chobert, J., Morgado-Rodríguez, L., Suárez-Murias, C., Valdés-Sosa, P., & Besson, M. (2013). Musical expertise and foreign speech perception. *Frontiers in Systems Neuroscience*, 7, 84.

McGuire, G. L. (2010). Perception a brief primer on experimental designs for speech perception research. *UCSC Linguistics Research Center*, 77–92. https://www.semanticscholar.org/paper/Perception-A-Brief-Primer-on-Experimental-Designs-McGuire/3e07b5a3c0feef01a11c37822d8b8197af8fd059

Meister, E., & Meister, L. (2013). Native and non-native production of Estonian quantity degrees: Comparison of Estonian, Finnish and Russian subjects". In E. L. Asu and P. Lippus, eds. *Nordic prosody: proceedings of the Xith conference,* 235–243. Frankfurt am Main: Peter Lang Verlag.

Meister, E., & Meister, L. (2014). Estonian quantity degrees produced by Latvian subjects. *Linguistica Lettica. 22*, 85–106.

Meister, L., & Meister, E. (2012). The production and perception of Estonian quantity degrees by native and non-native speakers. In: *Interspeech 2012: 13th Annual Conference of the International Speech Communication Association,* September 9–13, 2012. Portland, Oregon: International Speech Communication Association, 2012, 886–889.

Milovanov, R., Huotilainen, M., Esquef, P. A. A., Alku, P., Välimäki, V., & Tervaniemi, M. (2009). The role of musical aptitude and language skills in preattentive duration processing in school-aged children. *Neuroscience Letters*, *460*(2), 161–165.

Mok, P. K. P., & Zuo, D. (2012). The separation between music and speech: Evidence from the perception of Cantonese tones. *The Journal of the Acoustical Society of America*, *132*(4), 2711–2720.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex, 19(3)*, 712–723.

Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C., Dong, Q., Gabrieli, J. D. E., & Desimone, R. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(28). https://doi.org/10.1073/pnas.1808412115

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). Preattentive cortical-evoked responses to pure tones, harmonic tones, and speech: influence of music training. *Ear And Hearing*, *30*(4), 432–446. https://doi.org/10.1097/aud.0b013e3181a61bf2

Parviainen, T., Helenius, P., & Salmelin, R. (2004). Cortical differentiation of speech and nonspeech sounds at 100 ms: implications for dyslexia. *Cerebral Cortex*, *15*(7), 1054–1063. https://doi.org/10.1093/cercor/bhh206

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681. https://doi.org/10.1038/nn1082

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00142

Peng, G., Deutsch, D., Henthorn, T., Su, D., & Wang, W. S. (2013). Language experience influences nonlinguistic pitch perception. *Journal of Chinese Linguistics*, *41*(2), 447–467.

Psychology Software Tools. (2021). E-Prime (Version 3.0) [Computer software]. Psychology Software Tools. https://pstnet.com/products/e-prime/

Sadakata, M., & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: A cross-linguistic study. *Acta Psychologica*, *138*(1), 1–10. https://doi.org/10.1016/j.actpsy.2011.03.007

Schön, D., Gordon, R. L., Campagne, A., Magne, C., Astésano, C., Anton, J., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage*, *51*(1), 450–461. https://doi.org/10.1016/j.neuroimage.2010.02.023

Slater, J., Azem, A., Nicol, T., Swedenborg, B., & Kraus, N. (2017). Variations on the theme of musical expertise: cognitive and sensory processing in percussionists, vocalists and non-musicians. *European Journal of Neuroscience*, *45*(7), 952–963. https://doi.org/10.1111/ejn.13535

Tang, W., Xiong, W., Zhang, Y., Dong, Q., & Nan, Y. (2016). Musical experience facilitates lexical tone processing among Mandarin speakers: Behavioral and neural evidence. *Neuropsychologia*, *91*, 247–253.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, *161*(1), 1–10.

Tierney, A., & Kraus, N. (2014). Auditory-motor entrainment and phonological skills: precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00949

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164. https://doi.org/10.1111/j.1467-7687.2007.00549.x

Zhao, T., & Kuhl, P. K. (2015). Effect of musical experience on learning lexical tone categories. *Journal of the Acoustical Society of America*, *137*(3), 1452–1463. https://doi.org/10.1121/1.4913457

Käesolevaga kinnitan, et olen korrektselt viidanud kõigile oma töös kasutatud teiste autorite poolt loodud kirjalikele töödele, lausetele, mõtetele, ideedele või andmetele.

Olen nõus oma töö avaldamisega Tartu Ülikooli digitaalarhiivis DSpace.

Emily Ellervee