Article

Musicians have larger memory spans for Mandarin tones but not segments

Lijun Yu[1], Xiaonuo Li[2], Hua Yu[1], Zhuoya Cui[1], Wenchen Liao[1], Sha Li[2], Yu Peng[1] and Zhaoxin Wang[1]

Abstract

It is unclear how music training leads to superior memory of language. In the present study, we investigated whether musically trained adults (musicians) have superior segmental and tonal loops using non-musical Mandarin stimuli. Forty-three musicians and thirty-nine demographically matched non-musically trained adults (non-musicians) participated in this study, all native Chinese. Memory spans, typical indicators of capacity of the phonological loop, were measured in both visual and auditory modalities in the following three conditions: (1) a segmental condition, defined as different syllables with the same high level tone; (2) a tonal condition (suprasegmental condition), defined as the same syllable with different tones; and (3) a mixed condition, defined as different syllables with different tones. The results revealed a main effect of condition. Memory spans of the tonal conditions were significantly smaller than those of segmental and mixed conditions, regardless of group and modality. Moreover, a significant condition by group interaction was found. Musicians outperformed non-musicians in the tonal conditions, but not in the segmental or mixed conditions in both modalities. These findings suggest that there are tonal and segmental loops for Mandarin, and musicians, compared to controls, have larger memory spans for Mandarin tones but not segments.

Keywords

Mandarin, memory, musician, phonological loop, segment, tone

Music and spoken language are two main ways that humans communicate vocally and express themselves (Chen-Hafteck, 1997). On one hand, music and language have much in common. Both domains rely primarily on the auditory modality, and involve the perception and production of sound (Rebuschat, 2011). Further, both can utilize manipulations in tone/pitch to convey meaning (Baddeley, 1986). They require memory capacity for storing representation (chords,

[1]East China Normal University, China
[2]Shanghai Conservatory of Music, China

Corresponding author:
Zhaoxin Wang, Key Laboratory of Brain Functional Genomics (MOE and STCSM), Institute of Cognitive Neuroscience, School of Psychology and Cognitive Science, East China Normal University, 3663, North Zhongshan Rd, Shanghai, P.R. 200026, China.
Email: zxwang@nbic.ecnu.edu.cn

words, etc.) and the ability to combine these representations by means of a system of rules or structural schemata (Jackendoff, 2009). On the other hand, music and language are processed independently and maintain dissociable mental processing systems (Schulze & Koelsch, 2012; Stewart, Walsh, Frith, & Rothwell, 2001). A widespread belief is that linguistic processing takes place within the left hemisphere of the brain in most people, while music processing occurs in the right hemisphere (Nicholson et al., 2003; Schulze & Koelsch, 2012). Besides, most of the neuropsychological findings have also emphasized separate processing domains for music and language. The most well-known example is the dissociation of performance between amusia (a music perception deficit in the absence of difficulties with language) and aphasia (language processing difficulty in the absence of amusia) (Albouy, Schulze, Caclin, & Tillmann, 2013). The interplay of common and distinct mechanisms involved in music and language processing has been investigated in a growing number of research studies in this field.

There is intriguing evidence for a better verbal memory in musicians compared to controls, as observed in a number of tasks, such as digit span (Fujioka, Ross, Kakigi, Pantev, & Trainor, 2006), reproduction of previously memorized nouns or two-digit numbers (Brandler & Rammsayer, 2003), recognition (Cohen, Evans, Horowitz, & Wolfe, 2011), and immediate and delayed recall of word lists (Jakobson, Lewycky, Kilgour, & Stoesz, 2008). As the advantage for musicians' verbal memory disappeared when articulatory suppression was introduced, Franklin et al. suggested that musicians may have a better phonological loop (Franklin et al., 2008).

However, both segmental units (such as vowels or consonants) and suprasegmental units (such as tones) are used in language, specifically in tonal languages (Lee & Hung, 2008). Previous studies have shown that there might be a dissociation between segment and tone of language. For example, accumulating pieces of evidence from behavioral (Ferrand & Segui, 1998; Meijer, 1996), lesion (Cappa, Nespor, Ielasi, & Miozzo, 1997; Laganaro, Vacheresse, & Frauenfelder, 2002), and imaging studies (Liu et al., 2006; Luo et al., 2006) have demonstrated separate storage and processing systems for segments and tones. The right frontal cortex is more involved in the processing of the tone of Chinese (Liu et al., 2006; Luo et al., 2006). Moreover, tone is also a musically relevant parameter as it allows definition of the melodic aspects of a musical sequence. One well-documented finding is that music training not only leads to better perception of musical pitches (Magne, Schön, & Besson, 2006; Posedel, Emery, Souza, & Fountain, 2012), but also improves processing of language tones. For example, music training leads to better perception of Chinese tones (Bidelman, Hutka, & Moreno, 2013; Gottfried, 2007), and musicians detect tone violations in both music and language better than non-musicians (Magne et al., 2006). Results from electrophysiological studies have also revealed that the latency of the N2/N3 component to tone variations (Fujioka et al., 2006; Moreno et al., 2009) was shorter in musicians than in non-musicians. Furthermore, effects of tonal language on music were also reported. For example, tonal language background (Cantonese) is associated with higher auditory perceptual performance for music listening (Bidelman et al., 2013), and speakers of tonal languages, like Mandarin Chinese, are more likely to have absolute pitch than speakers of non-tonal languages (Deutsch, Henthorn, Marvin, & Xu, 2004, 2006). These results suggested that there are bidirectional influences between musical pitch and language tone (but not segment). It has been suggested that there are both tonal (for music) and phonological (for language) loops (Schulze & Koelsch, 2012), and previous studies reported that musicians had a better memory for musical pitch (Cohen et al., 2011; Williamson, Baddeley, & Hitch, 2010). Thus, it is possible that music training may have different effects on segmental and tonal loops of language.

However, most previous studies used western languages such as English (Cohen et al., 2011), French (Baddeley, 1983; Marques et al., 2007), Portuguese (Besson, Schon, & Moreno, 2007),

and German (Roden, Grube, Bongard, & Kreutz, 2013). Typically, a word in these languages comprises several syllables. In linguistics, each syllable is deemed to include at least two phonological units: segmental units such as a vowel and a consonant and suprasegmental (or prosodic frame) units such as tone or stress (Liu et al., 2006). Mandarin Chinese is a tonal language. One advantage of Chinese is that most morphemes are monosyllabic (Koelsch et al., 2009), thus it is easier to control the number of syllables. The second advantage is that word meaning is partially determined by tones. Therefore the tonal information can be manipulated naturally, without confounding factors, such as emotion associated with tone, being introduced. Thus, tonal languages provide a particularly valuable window to understand the interplay between music and language processes (Zatorre & Gandour, 2008). Tonal languages also allow investigation of the hypothesized phonological loop advantages in musicians compared to controls, and possible contributions of segmental as well as tonal loops. To date, only one group has used Mandarin stimuli (Chan, Ho, & Cheng, 1998; Ho, Cheung, & Chan, 2003), and they reported that musicians outperformed non-musicians in memorizing 16 two-character Chinese word lists. As a two-character word consists of at least two syllables and tones, whether musical training has different effects on segmental and tonal loops remains unanswered.

To address this question, the current study investigates memory spans of segment and tone in musically trained adults (musicians) and non-musically trained adults (non-musicians) using Mandarin stimuli. The memory task is similar to the forward digit span task of the Wechsler Memory Scale, a typical task used as a measure of the phonological loop (Alexander, 2005), with one difference: only four stimuli were applied to create stimulus sequences in each condition, as there were four tones typically used in Mandarin Chinese. Different segmental units with the high-level tone were used in the segmental condition, while the same segmental units with different tones were used in the tonal condition. We hypothesized different memory spans for tones and segments and a better performance in the tonal condition, but not in the segmental condition for musicians compared to controls. A mixed condition in which each stimulus has a different segmental unit as well as a different tonal unit was also included to further investigate the contribution of segments and tones in memory. As both segment and tone information were included, we hypothesized that memory span of the mixed condition should be most pronounced. A better tone capacity was expected to lead to a better performance in the mixed condition.

Method

Participants

Forty-three musicians (13 males, aged from 18 to 24, with a mean age of 20.0, SD = 1.5) with a minimum of 5 years of music training (range = 5–19 years, mean = 13.8 years) from the Shanghai Conservatory of Music participated in this experiment. All musicians were pianists, some of them also studied vocal arts (17 participants), pipa (Chinese lute, 1 participant), or erhu (Chinese violin, 1 participant). Thirty-nine demographically matched students without a self-reported history of music training (14 males, aged from 17 to 24, with a mean of 20.7, SD = 2.0) from East China Normal University also participated in the study as non-musicians. There were no significant differences in age ($t = -1.9$, $p = 0.07$) or educational level. All participants were native speakers of Chinese and right-handed with normal hearing and normal or corrected-to-normal vision. They were paid to compensate for their time. Written informed consent was obtained from all participants and the present study was approved by local ethics committee of East China Normal University.

Procedure

A 2 × 2 × 3 design was applied. Memory spans were measured in musicians and nonmusicians, in both the auditory and visual modalities, and in the following three conditions: a segmental, a tonal (suprasegmental), and a mixed condition. In the segmental condition, digits from monosyllabic 0–9 with the high level tone, i.e., 1 (yī), 3 (sān), 7(qī), and 8 (bā), were selected. The high level tone was used because it is the most stable one among the four tones of Mandarin Chinese. In the tonal condition, the segment of "yi" with different tones was used. "Yi" was selected because it is widely used in Mandarin in all four tones. For example, yī, the high level tone, could mean one (一), clothes (衣), obey (依), or medicine (医); yí, the low rising tone, could mean aunt (姨), doubt (疑), move (移), or eh (咦); yǐ, high rising falling tone, could mean almost (已), ant (蚁), chair (椅), or second (乙); and yì, high falling tone, could mean recall (忆), means (意), easy (易), or wings (翼), etc. Thus, all these stimuli were very familiar to the participants. In the mixed condition, 3 (sān), 0 (líng), 9 (jiǔ), and 4 (sì) were used; each of them has a different segment and a different tone.

For the auditory modality, 32-bit auditory female sounds, one for each stimulus, were synthesized individually by Neospeech Lily TTS (text-to-speech) engine with standard Mandarin. The F0 and formant structure of these stimuli (created in Praat, V5.3.86) are displayed in Figure 1. For the visual modality, numbers or Chinese phonetic alphabet format (for non-number stimuli) were presented. Each stimulus was presented in black-and-white, with a visual angle of about 3.0°×2.0°.

Stimuli were presented via E-PRIME. There was a 3-trial practice session in each condition. Each trial began with a 1-s cue, the to-be-remembered stimuli of a sequence were then played/presented sequentially at a speed of 1 second per item (stimulus onset asynchrony = 1s). In each sequence, stimuli were pseudo-randomly assigned; the same stimuli did not appear consecutively. No obvious meaningful expression could be derived from these sequences. Participants were instructed to report their answers orally in the original order. Each condition began with a sequence of four items. If the recall was right for one trial, the number of items in a sequence was increased by one. A failed trial led to another trial with the same number of items. The inter-trial interval of approximately 5 seconds was chosen by one experimenter. If participants failed to recall two trials, the number of correctly recalled items in the last sequence was taken as memory span (Wechsler, 1997). The two modalities and the three conditions were blocked in counterbalanced order across participants.

Data analysis

Data from two participants (1 musician, 1 non-musician) were excluded because their performances exceeded ±3 standard deviations (SD) from the mean in at least one condition. A 2 (group: musician, non-musician) × 2 (modality: visual, auditory) × 3 (condition: segmental, tonal, mixed condition) repeated measures ANOVA was applied using SPSS (Version 16). *T*-tests were used for post hoc analysis. The significance threshold was set at $p = 0.05$.

Results

Mean memory spans and standard deviations are shown in Figure 2. In the repeated measures ANOVA, a main effect of condition was found, $F(2,156) = 240$, $p < 0.001$, $\eta^2 = 0.754$. Post hoc paired *t*-tests revealed that there were significant differences between the segmental, tonal, and
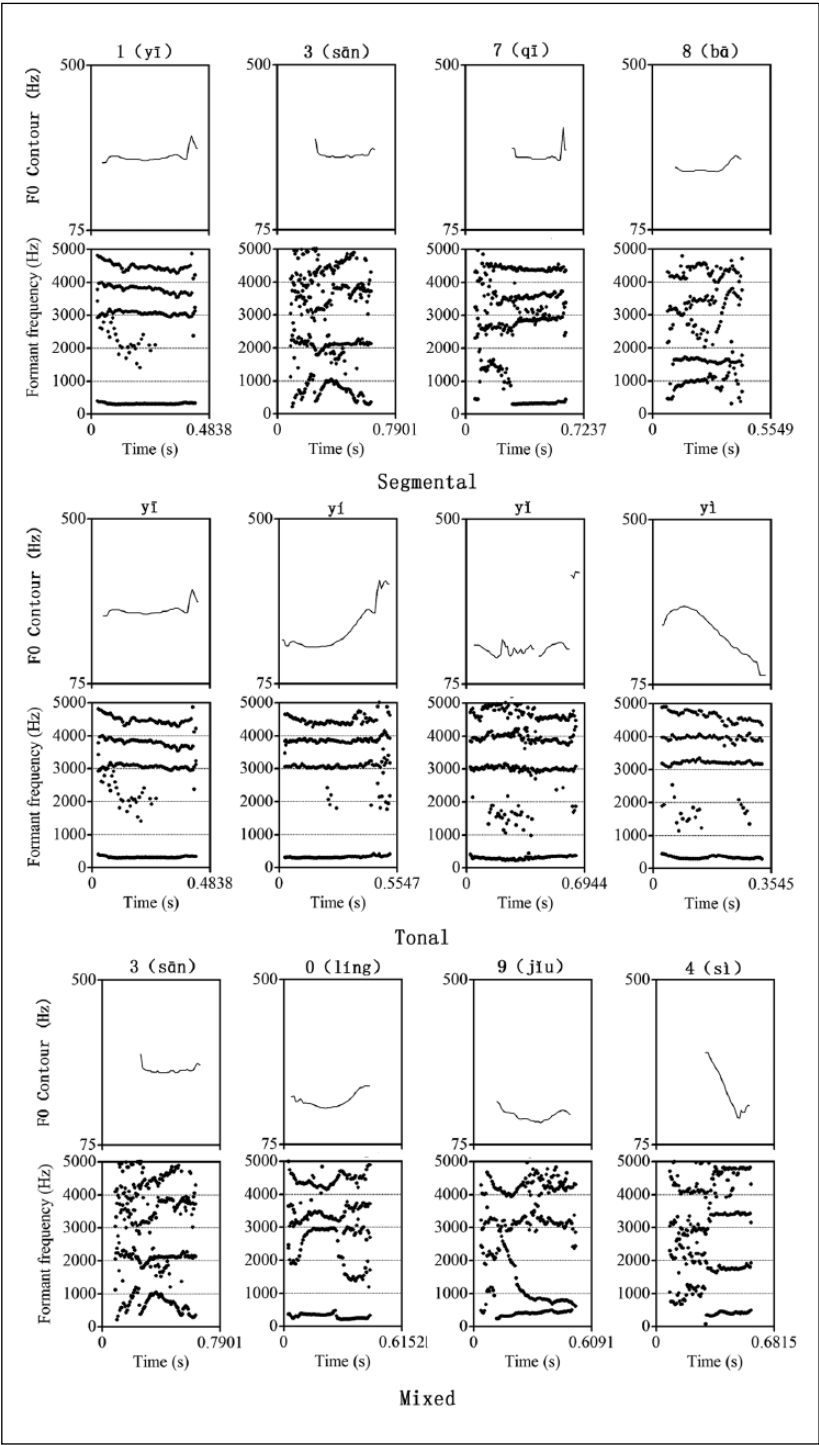
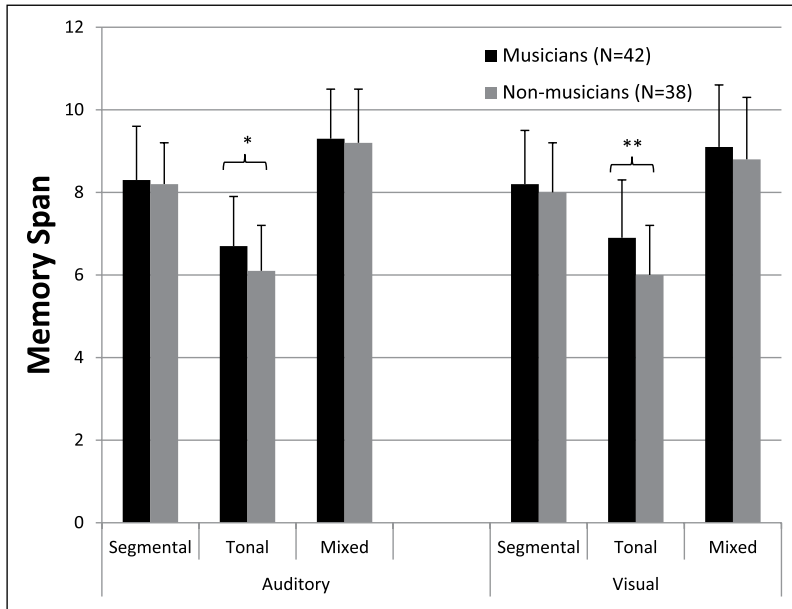**Figure 1.** F0 contour and formant structure of each stimulus.

Figure 2. Memory spans of segmental, tonal, and mixed conditions.
*p = 0.03, **p = 0.005, two-tailed. Error bar represents SD.

mixed conditions in both modalities of both groups ($ts > 6.3$, $ps < 0.001$, two-tailed). Moreover, a significant group by condition interaction was found, $F(2,156) = 3.08$, $p = 0.049$, $\eta^2 = 0.038$. Post hoc independent *t*-tests revealed that musicians outperformed non-musicians in the tonal condition in both modalities (auditory modality: $t(78) = 2.2$, $p = 0.03$, visual modality: $t(78) = 2.9$, $p = 0.005$, two-tailed), but not in other conditions ($ts < 0.7$, $ps > 0.4$). Otherwise, the ANOVA did not reveal a significant difference between the musicians and non-musicians, $F(1,78) = 3.24$, $p = 0.084$, $\eta^2 = 0.038$, or modalities, $F(1,78) = 3.0$, $p = 0.087$, $\eta^2 = 0.037$, although a trend seemed to emerge. No other significant main effects or interactions were detected.

Discussion

The most intriguing finding of the present study is that musicians outperformed non-musicians in the tonal condition, but not in the segmental condition in both modalities. This is the first report of a superior tonal span for non-musical stimuli in musicians. These results are in line with Williamson et al.'s finding that musicians compared to non-musicians performed better in musical pitch recall but not letter recall (Williamson et al., 2010). However, the better performance of musicians in Williamson's studies might be attributed to group differences in the perception of musical pitches. First, musicians are more familiar with musical pitches, and familiarity can enhance the memory span (Hulme, Roodenrys, Brown, & Mercer, 1995). Second, musicians performed better in musical pitch perception (Zatorre & Gandour, 2008). Thus, less effort may be needed during the encoding processes, whereas more resources may be used during memory processes. In this study, differences in familiarity as well as perceptual characteristics were minimized (although they may not vanish due to interplay between music and language) by using non-musical language materials. Note that no acoustical signals were

given during the encoding process in the visual modality. These results also suggest that music training facilitates not only perception, but also the internal representation of Mandarin tones.

The superior memory span for tonal information in musicians is most likely due to the fact that musicians have long-term training in memorizing long and complex music pieces, resulting in superior memory for musical pitches (Williamson et al., 2010). As musical and linguistic tone perception share some common neural correlates (Bidelman et al., 2013; Liu et al., 2006; Luo et al., 2006), memory enhancement occurs only in the tonal loop, while memory for segment is almost unaffected. Findings in the present study are in line with the idea that language tones and music share some common neural correlates. Indeed, Peretz et al. reported a case of acquired amusia in the absence of aphasia, i.e., participant G.L., who displayed an inability to discriminate between pairs of sentences in which the prosodic focus was shifted (Peretz, 1993).

We did not find any significant between-group difference in the mixed condition. These results seem not to support our hypothesis that better memory for tone may lead to better memory of the mixed condition. One possibility is that while numbers were used in the segment and mixed condition, stimuli in the tonal condition were not exact numbers (despite having the same segmental sequence (i.e. yi) as the number 1). Thus, the memory spans of the mixed conditions were more related with the other digit condition, but less affected by the tonal condition. Second, considering our finding of enhanced capacity of segmental compared to tonal memory in both groups, and left versus right hemispheric specialization for linguistic (Zhang & Zhu, 2011) versus tonal processing (Liu et al., 2006; Luo et al., 2006) respectively, it is possible that the mixed condition and the segmental condition, but not the tonal condition, recruited similar neural correlates. For example, using the same Chinese materials, Yu reported that the left frontal gyrus is more engaged memorizing four-item segmental sequences than four-digit tonal sequences (Yu, 2015). The third possibility is that tonal information may play a more important role with multiple syllabic materials, presumably by creations of melody-like prosodies (Nicholson et al., 2003). While the current study used monosyllabic stimuli, previous studies employed materials with multiple segmental/tonal information (Besson et al., 2007; Fujioka et al., 2006; George & Coch, 2011).

We also noted that the tonal condition was the most difficult one among the three conditions according to memory span. Previous studies have shown that musical expertise influenced the neural processing of both tonal and segmental stimuli. Electrophysiologically, larger amplitude P300 to tones (Cappa et al., 1997; George & Coch, 2011; Marie, Delogu, Lampis, Belardinelli, & Besson, 2011) as well as segmental variations (Marie et al., 2011) were observed in musicians. The P300 component is thought to be related to task difficulty (Polich, 2007), with larger P300 amplitudes in "easier" tasks (Cappa et al., 1997; Nittono, Nageishi, Nakajima, & Ullsperger, 1999) as well as higher musicality scores (Cappa et al., 1997). Further, decreased P300s have been observed in response to increasing memory load (Hyde & Peretz, 2004). Thus, the level of difficulty of the memory task might be modulated by musical expertise. It is possible that musicians might be better on the most difficult tonal condition, while a ceiling effect may occur in the segmental and mixed conditions. These possibilities should be addressed in further studies, for example, imaging studies.

Limitations

There were several limitations of the present study. First, all stimuli were chosen from 0–9 (or the same syllable in the tonal condition) to closely mimic the digit span task. This resulted in only four digits with the same high level tone, including yī and qī that with the same vowels. A phonological similarity effect (Baddeley, 1983) may have occurred and resulted in the smaller

capacity of the syllable loop. Yet, the differences between groups are probably not affected by this, as this phonological similarity applies to both groups. Second, findings of the present study may not be generalized to children (Besson et al., 2007; Ho et al., 2003). For example, Roden et al. reported that children receiving music training specifically benefit in the phonological loop with a one-syllable word span test in German (i.e. limited segmental/tonal information) (Roden et al., 2013), although the one-syllable German words with diphthong used in Roden et al.'s study (personal communication) were comprised of multiple tones. Third, the interplay between language and music may be stronger in tonal languages (Bidelman et al., 2013). Whether this effect can be found in western languages deserves further investigation. Finally, with the current study we cannot answer whether better tone-processing aptitude in musicians is caused by the training per se or by other factors such as inherent ability (Alexander, 2005). Future longitudinal study may help to address this issue.

Conclusion

Beyond Schulze's suggestion that there is both a tonal and a phonological loop for music and language respectively (Schulze & Koelsch, 2012), our results support the suggestion that there are different loops for Mandarin segments and tones. Our findings suggest that musicians have a superior phonological loop for Mandarin tones but not segments.

Acknowledgements

Funding

References

Albouy, P., Schulze, K., Caclin, A., & Tillmann, B. (2013). Does tonality boost short-term memory in congenital amusia? *Brain research*, *1537*, 224–232. doi:10.1016/j.brainres.2013.09.003

Alexander, J. A. (2005). *Lexical tone perception in musicians and nonmusicians*. Paper presented at the Interspeech 2005—Eurospeech—Ninth European Conference on Speech Communication and Technology, Lisbon, Portugal.

Baddeley, A. (1983). Working memory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *302*(1110), 311–324.

Baddeley, A. (1986). *Working memory*. Oxford: Oxford University Press.

Besson, M., Schon, D., & Moreno, S. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience*, *25*, 399–410.

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLoS One*, *8*(4), e60676.

Brandler, S., & Rammsayer, T. H. (2003). Differences in mental abilities between musicians and non-musicians. *Psychology of Music*, *31*(2), 123–138. doi:10.1177/0305735603031002290

Cappa, S. F., Nespor, M., Ielasi, W., & Miozzo, A. (1997). The representation of stress: Evidence from an aphasic patient. *Cognition*, *65*(1), 1–13. doi:10.1016/S0010–0277(97)00024–3

Chan, A. S., Ho, Y.-C., & Cheng, M.-C. (1998). Music training improves verbal memory. *Nature*, *396*, 128.

Chen-Hafteck, L. (1997). Music and language development in early childhood: Integrating past research in the two domains. *Early child development and care*, *130*(1), 85–97.

Cohen, M. A., Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). Auditory and visual memory in musicians and nonmusicians. *Psychonomic bulletin & review*, *18*(3), 586–591.

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2004). Tone language and absolute pitch: Prevalence among American and Chinese conservatory students. *Journal of the Acoustical Society of America*, *116*(4), 2580–2580. doi:10.1121/1.1744751

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: prevalence differences, and evidence for a speech-related critical period. *Journal of the Acoustical Society of America*, *119*(2), 719–722. doi:10.1121/1.2151799

Ferrand, L., & Segui, J. (1998). The syllable's role in speech production: Are syllables chunks, schemas, or both? *Psychonomic bulletin & review*, *5*(2), 253–258. doi:10.3758/Bf03212947

Franklin, M. S., Sledge Moore, K., Yip, C. Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, *36*(3), 353–365. doi:10.1177/0305735607086044

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, *129*(10), 2593–2608. doi:10.1093/brain/awl247

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, *49*(5), 1083–1094. doi:10.1016/j.neuropsychologia.2011.02.001

Gottfried, T. L. (2007). Music and language learning: Effect of musical training on learning L2 speech contrasts. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 221–237). Amsterdam and Philadelphia: John Benjamins.

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

Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology*, *86*(4), 527–536. doi:10.1111/j.2044–8295.1995.tb02570.x

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, *15*(5), 356–360. doi:10.1111/j.0956–7976.2004.00683.x

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception*, *26*(3), 195–204. doi:citeulike-article-id:4013427

Jakobson, L. S., Lewycky, S. T., Kilgour, A. R., & Stoesz, B. M. (2008). Memory for verbal and visual material in highly trained musicians. *Music Perception*, *26*(1), 41–55.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Muller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: An FMRI study. *Human Brain Mapping*, *30*(3), 859–873. doi:10.1002/hbm.20550

Laganaro, M., Vacheresse, F., & Frauenfelder, U. H. (2002). Selective impairment of lexical stress assignment in an Italian-speaking aphasic patient. *Brain and language*, *81*(1–3), 601–609. doi:10.1006/brln.2001.2550

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

Liu, L., Peng, D., Ding, G., Jin, Z., Zhang, L., Li, K., & Chen, C. (2006). Dissociation in the neural basis underlying Chinese tone and vowel production. *Neuroimage*, *29*(2), 515–523. doi:10.1016/j.neuroimage.2005.07.046

Luo, H., Ni, J.-T., Li, Z.-H., Li, X.-O., Zhang, D.-R., Zeng, F.-G., & Chen, L. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences*, *103*(51), 19558–19563.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, *18*(2), 199–211.

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience*, *23*(10), 2701–2715.

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than non-musicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453–1463.

Meijer, P. J. (1996). Suprasegmental structures in phonological encoding: The CV structure. *Journal of Memory and Language*, *35*(6), 840–853.

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.

Nicholson, K. G., Baum, S., Kilgour, A., Koh, C. K., Munhall, K. G., & Cuddy, L. L. (2003). Impaired processing of prosodic and musical patterns after right hemisphere damage. *Brain Cognition*, *52*(3), 382–389. doi:S0278262603001829 [pii]

Nittono, H., Nageishi, Y., Nakajima, Y., & Ullsperger, P. (1999). Event-related potential correlates of individual differences in working memory capacity. *Psychophysiology*, *36*(6), 745–754.

Peretz, I. (1993). Auditory atonalia for melodies. *Cognitive Neuropsychology*, *10*(1), 21–56. doi:10.1080/02643299308253455

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. doi:10.1016/j.clinph.2007.04.019

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

Rebuschat, P. (2011). *Language and music as cognitive systems*: Oxford University Press.

Roden, I., Grube, D., Bongard, S., & Kreutz, G. (2013). Does music training enhance working memory performance? Findings from a quasi-experimental longitudinal study. *Psychology of Music*, *42*(2), 284–298. doi:10.1177/0305735612471239

Schulze, K., & Koelsch, S. (2012). Working memory for speech and music. *Annals of the New York Academy of Sciences*, *1252*, 229–236. doi:10.1111/j.1749–6632.2012.06447.x

Stewart, L., Walsh, V., Frith, U., & Rothwell, J. (2001). Transcranial magnetic stimulation produces speech arrest but not song arrest. *Annals of the New York Academy of Sciences*, *930*(1), 433–435.

Wechsler, D. (1997). *Wechsler adult intelligence scale* (3rd ed.). San Antonio, TX: Psychological Corporation.

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Memory and Cognition*, *38*(2), 163–175. doi:10.3758/MC.38.2.163

Yu, L. (2015). *Dissociation and the mechanism for the short-term memory of segment and tone in Mandarin* (unpublished Master's dissertation). East China Normal University, Shanghai, China.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1087–1104.

Zhang, Q., & Zhu, X. (2011). The temporal and spatial features of segmental and suprasegmental encoding during implicit picture naming: An event-related potential study. *Neuropsychologia*, *49*(14), 3813–3825. doi:10.1016/j.neuropsychologia.2011.09.040