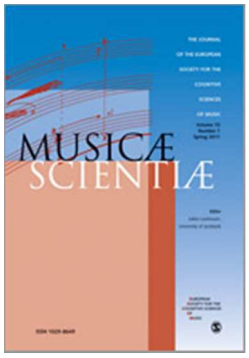


For Review Only

The benefit of musical and linguistic expertise on language acquisition in sung material

Journal:	*Musicae Scientiae*
Manuscript ID:	MS-12-0005.R2
Manuscript Type:	Original Manuscript
Keywords:	Musical expertise, transfer effects, artificial language learning, songs, executive function
Abstract:	Music learning and training appear to have large cross-domain transfer effects: they are beneficial in various cognitive domains including language. The present study aimed at examining the role of music expertise on how musical and linguistic information contained in songs is used. However, as the superiority of musicians could be attributed to improvements in executive functions (e.g., Bialystok, & DePape, 2009), we tried to isolate the role of music training by comparing music experts to both non-experts and language experts. To this aim we used the tasks proposed by Schön, Boyer, Moreno, Besson, Peretz, and Kolinsky (2008), who compared artificial language learning (ALL) based on spoken sequences to ALL based on sung sequences. These authors concluded that songs, more than speech, allow fast and strong learning. In contrast to Schön et al. (2008), the benefit of songs on ALL was not found among non-experts in our study. However, there was a highly significant interaction between type of expertise and materials. The music experts were the only group to benefit from the matching of linguistic and music information, hence showing a different profile than language experts. The present data thus confirms the specificity of transfer effects linked to music expertise.

SCHOLARONE™
Manuscripts

The benefit of musical and linguistic expertise on language acquisition in sung material

Pauline Larrouy-Maestri[1], Jacqueline Leybaert[2], & Régine Kolinsky[3,4]

1. Logopédie de la Voix, Département de Psychologie, Université de Liège, Belgium

2. Laboratoire Cognition, Langage et Développement, Centre de Recherche Cognition & Neurosciences, Université Libre de Bruxelles (U.L.B.), Belgium

3. Fonds de la Recherche Scientifique-FNRS

4. Unité de Recherches en Neurosciences Cognitives, Centre de Recherche Cognition & Neurosciences, Université Libre de Bruxelles (U.L.B.), Belgium

Correspondence concerning this article should be addressed to Pauline Larrouy-Maestri, Logopédie de la voix, Département de Psychologie: cognition et comportement, Université de Liège, B-38, Rue de l'Aunaie, 30, 4000 Liège, Belgium. Electronic mail may be sent to pauline.larrouy@ulg.ac.be.

For Peer Review

Abstract

Music learning and training appear to have large cross-domain transfer effects: they are beneficial in various cognitive domains including language. The present study aimed at examining the role of music expertise on how musical and linguistic information contained in songs is used. However, as the superiority of musicians could be attributed to improvements in executive functions (e.g., Bialystok, & DePape, 2009), we tried to isolate the role of music training by comparing music experts to both non-experts and language experts. To this aim we used the tasks proposed by Schön, Boyer, Moreno, Besson, Peretz, and Kolinsky (2008), who compared artificial language learning (*ALL*) based on spoken sequences to ALL based on sung sequences. These authors concluded that songs, more than speech, allow fast and strong learning.

In contrast to Schön et al. (2008), the benefit of songs on ALL was not found among non-experts in our study. However, there was a highly significant interaction between type of expertise and materials. The music experts were the only group to benefit from the matching of linguistic and music information, hence showing a different profile than language experts. The present data thus confirms the specificity of transfer effects linked to music expertise.

Keywords: Musical expertise, transfer effects, artificial language learning, songs

For Peer Review

Introduction

Songs combine melody and lyrics and thus constitute an ideal tool to observe the relationship between the language and music domains (Schön, Gordon, & Besson, 2005). The interactions between these two dimensions of songs have been debated in both perception and memory. Although the data support partial autonomy of the processing of the linguistic and musical dimensions of songs (e.g., Besson, Faïta, Peretz, Bonnel, & Requin, 1998; Hébert, Racette, Gagnon, & Peretz, 2003; Kolinsky, Lidji, Peretz, Besson, & Morais, 2009; Peretz, Kolinsky, Tramo, Labrecque, Hublet, Demeurisse, & Belleville, 1994; Racette & Peretz, 2007; Samson & Zatorre, 1991), most researchers acknowledge the intervention of some general, domain-independent, mechanisms. Among these are the learning mechanisms relying on the extraction of statistical information.

In an artificial language learning (*ALL*) situation, Saffran and colleagues have shown that adults (Saffran, Aslin, & Newport, 1996a) and infants (Saffran, Aslin, & Newport, 1996b) can use the statistical properties of syllable sequences to extract words from continuous speech, in particular the transitional probabilities between syllables (*TPs*). TPs are defined as the probability with which one syllable (x) predicts the following one (y): TP (y|x) = frequency (xy) / frequency (x). Within any language, TP from one unit of sound (e.g., a syllable) to the next is generally highest when the two units follow one

For Peer Review

another within a word, whereas the TPs that span a word boundary are relatively low. TPs are thus exploited for the discovery of significant units, e.g., words, in a continuous input stream. With adults, this has been shown in ALL situations in which listeners were exposed to a continuous synthesized speech stream generated from six trisyllabic nonsense "words" (such as "bupada" and "dutaba") that were concatenated together (Saffran et al., 1996a). Their task was to discover the "words" of that artificial language. The speech stream contained no cues to the new "word" boundaries, except for the statistical properties, which distinguished what we henceforth will call *TP-words* from the sequences spanning TP-word boundaries: TPs within the TP-words were higher than TPs between the TP-words. Following 21 minutes of exposure to this speech stream, adults demonstrated on a forced-choice test that they could distinguish sequences of sounds that were "words" of the artificial language (e.g., "bupada") from sequences of sounds made up of the same syllables but not forming TP-words (e.g., "daduta").

This mechanism of TP extraction seems particularly valuable in the earliest stages of language acquisition, in the absence of top-down knowledge to help the perceiver to segment the stream, and hence constitutes a "first window" (Thiessen & Saffran, 2003) that allows the discovery of significant units in a continuous input stream. In addition, using the same statistical structure as in Saffran et al. (1996b) with a stream in which each syllable

For Peer Review

had been "translated" into a note, Saffran, Johnson, Aslin, and Newport (1999) showed that a similar learning mechanism operates with musical stimuli.

In a previous study combining the linguistic and musical dimensions, Schön, Boyer, Moreno, Besson, Peretz, and Kolinsky (2008) compared ALL based on spoken sequences to ALL based on sung sequences. They observed that, compared to spoken sequences, a consistent mapping of linguistic and musical information enhanced learning with a sung sequence; in other words, the extraction of linguistic TPs was better when linguistic and melodic TP dips (i.e., boundaries) coincided. Most importantly, this effect could not be accounted for only by the enhancement of motivation related to music in songs, since it was larger when linguistic and musical TPs were consistent than when there were inconsistent, i.e., when their boundaries did not coincide. The authors concluded that learning a new language, especially in the first phase during which it is necessary to segment new words, benefits from the structural redundancies between the melody and the lyrics in songs.

In the present study, we examined whether music expertise may enhance such a benefit. Indeed, music training seems to induce cross-domain transfer effects. Music training has positive effects on some aspects of foreign language processing, by increasing sensitivity to non-native lexical stress (Kolinsky, Cuvelier, Goetry, Peretz, & Morais, 2009), by influencing the processing of prosody (e.g., Magne, Schön, & Besson, 2006;

For Peer Review

Moreno, Marques, Santos, Santos, Castro, & Besson, 2008; **Marques, Moreno, Castro, & Besson, 2007;** Schön, Magne, & Besson, 2004) and linguistic tone patterns (Delogu, Lampis, & Belardinelli, 2010; **Marie, Delogu, Lampis, Belardinelli, & Besson, 2011**), and by facilitating perception and production of the phonology of a second language (Milovanov, Huotilainen, Välimäki, Esquef, & Tervaniemi, 2008; Milovanov, Pietilä, Tervaniemi, & Esquef, 2010; Slevc & Miyake, 2006).

However, the mechanisms by which musical experience affects cognitive processing are unclear (for a review, see e.g., Peretz, 2009, François, Tillmann, & Schön, 2012). **Transfer effects from music to language can result both from common processing of acoustic features shared by music and speech (such as pitch and duration) and from more general mechanisms (see discussions in e.g. Besson, Chobert, & Marie, 2011, and Kraus & Chandrachekaran, 2010).**

In particular, as music performance demands high levels of control through the need for selective attention and inhibition, switching, updating and monitoring, the superiority of musicians could be attributed to improvements in executive functions. A recent training study on preschool children (Moreno, Bialystok, Barac, Schellenberg, Cepeda, & Chau, 2011) showed a causal beneficial effect of music training on executive functions. Yet, at least in adults, Bialystok and De Pape (2009) showed that some of these

For Peer Review

effects are not specific to music expertise. Comparing musicians, bilinguals and non-experts, they showed that both musicians and bilinguals were superior to monolinguals with no musical experience on a spatial conflict task. In addition, in a task pertaining to the auditory domain, the musicians outperformed the others, including the bilinguals: their response times were faster and their ability to resolve conflict was better when the task was based on pitch. Hence, extended musical experience enhances executive control and more specialized auditory control mechanisms.

It therefore seems necessary to isolate the specific role of music training from more general effects of expertise by comparing music experts to both non-experts and people presenting another kind of expertise. In the present study, we compared the use of statistical information contained in songs between people with musical expertise (henceforth, *music experts*) to *non-experts* as well as to people with formal expertise in language, in this case, speech therapists (henceforth, *language experts*). Speech therapists develop high sensitivity to linguistic information. Indeed, their training and daily practice include fine phonemic discrimination and the ability to retrieve words from patients' stream speech, sometimes with very low intelligibility.

We used the three materials designed by Schön et al. (2008): spoken sequences (without any melodic variation), sung consistent sequences (in which linguistic and musical

For Peer Review

TPs are redundant) and sung inconsistent sequences (in which linguistic and musical TPs have different, contradictory boundaries). These three materials were presented to independent groups of participants, in order to avoid cross-materials transfer of learning. All groups where then tested on the same spoken items. We predicted that all groups would perform above chance with the second material, indicating that the consistent mapping of linguistic and musical information enhances learning an artificial language based on pseudowords. In addition, we expected that music and language experts would obtain better performance than the non-experts. Furthermore, a different profile between music and language experts across the three experiments would allow differentiating the role of music expertise from more general training effect. Indeed, if all the experts show the same profile across conditions, we could conclude to a general training effect. By contrast, if music and language experts show different profiles **(i.e. superiority of the music experts for the sung consistent sequences and superiority of the language experts for the spoken sequences),** the effect of music or of language expertise cannot be reduced to a general effect of expertise.

For Peer Review

Experiment 1: Spoken sequences

Method

Participants. A total of 48 native French speakers participated in the study. They were assigned to one of three groups of expertise, each including 16 participants. In order to avoid a double expertise, speech therapists who were musicians or multilingual musicians were not included in any group.

As in Kolinsky et al. (2009), non-experts (aged 29 to 60 yrs, average: 44.75 yrs; 9 men and 7 women) had never followed solfeggio lessons, never learned to read and write music, and were totally unable to do so. Non-experts had less than three years of music practice, and those who had some practice had stopped it for at least six years.

Music experts (aged 19 to 69 yrs, average: 28.06 yrs; 5 men and 11 women), all educated in classical music, were recruited at the Conservatoire Royal of Mons (Belgium) and among professionals. They were either instrumentalists or singers, had begun music lessons between 5 and 12 years of age (on average, at 8.36 yrs) and had between 8 and 52 years of music training (on average, 16.5 yrs).

Language experts (aged 24 to 51 yrs, average: 28.5 yrs; 2 men and 14 women) were graduate or undergraduate speech therapists. Whereas every speaker can be considered as a language expert, we decided to choose speech therapists regarding to their formation

For Peer Review

focuses on linguistic information and the particular development of phonemic discrimination and word segmentation. Note that this profession is particularly feminine, explaining the high proportion of women in the group.

None of the participants had any hearing or memory problems. They all obtained at least 95% correct responses in a test of phonological memory requiring the repetition of pseudo-words of one to five syllables (Mousty, Leybaert, Alegria, Content, & Morais, 1994).

Material and procedure. The AL was the same monotone continuous stream of concatenated synthesized syllables as the one used by Schön et al. (2008) in their first experiment, with no acoustic cues indicating TP-words boundaries. These trisyllabic pseudowords (i.e., *TP-words*) were signalled only by dips in TPs, which ranged from 0.31 to 1.0 within TP-words, and from 0.1 to 0.2 between TP-words. There were six different TP-words: /gimysy/, /mimosi/, /pogysi/, /pymiso/, /sipygy/, and /sysipi/, synthesized using the Mbrola speech synthesizer (http://tcts.fpmc.ac.be/synthesis/mbrola.html).

As in Schön et al. (2008), participants were first exposed to a seven-minute continuous stream, being asked only to listen carefully without trying to analyze it. Then, during the test phase, participants were invited to perform a two-alternative forced-choice task that included 36 trials. Each test trial consisted of two trisyllabic spoken items: one

For Peer Review

was a TP-word, the other a *part-word*, consisting of either the last syllable of a TP-word plus the first two syllables of the next TP-word, or of the last two syllables of a TP-word plus the first syllable of the next word. There were six different part-words: /sypogy/, /sogimy/, /pygymi/, /pisipy/, /mosigi/, and /gysimi/. On each trial, participant had to indicate which of the two items was most likely to be a word from the AL. All stimuli in both phases were presented through headphones (Panasonic RP – HTX7). The testing session was individual and required approximately forty minutes for each participant.

Results

The distribution of percentages of TP-word choices is presented in Figure 2. Although the analysis of variance (*ANOVA*) run on the raw number of TP-word choices did not show a significant effect of expertise, F (2, 45) = 1.6, $p > .10$, only language experts presented a better than chance performance, with 57.12% TP-word choices, on the average, $t(15) = 2.21$, $p < .025$. Non-experts and music experts only reached 50.17 and 51.22% average TP-word choices, respectively. Their performance did not differ from chance level, both $ts < 1$.

Experiment 2: Sung consistent sequences

The second experiment was identical to the first one, except that the syllables of the continuous stream were sung by the synthesizer rather than spoken. As in the second

For Peer Review

experiment of Schön et al. (2008), there was a consistent mapping of linguistic and musical statistical information (see Figure 1a) leading to a superposition of linguistic and melodic TPs (see Figure 1b). Since we wanted to test language learning, the testing phase was identical to Experiment 1, using spoken and not sung items.

Method

Participants. A total of 56 native French-speaking volunteers, different from the first experiment, took part in this experiment. Participants were assigned to one of the three groups of expertise according to the same criteria as in Experiment 1. There were 20 non-experts (aged 20 to 59 yrs, average: 35.55 yrs; 5 men and 15 women),16 language experts (aged 23 to 34 yrs, average: 25.87 yrs; 2 men and 14 women) and 20 music experts (aged 19 to 68 yrs, average: 32.65 yrs; 7 men and 13 women). The music experts had begun music lessons between 3 and 12 years of age (on average, at 7.7 yrs), with 9 to 47 years of music training on an instrument (on average, 18.7 yrs). As in Experiment 1, all participants obtained at least 95% of correct response in the repetition of pseudo-words test (Mousty et al., 1994).

Material and procedure. The material was the same as the one used by Schön et al. (2008) in their second experiment. Each of the 11 syllables was associated with a specific pitch (see Figure 1a). Therefore each word was always sung on the same melodic

For Peer Review

contour (see Figure 1b). The syllable-pitch association was stable and hence there was no contour change within each word. For instance, the TP-word /sysipi/ rising F5G5A5, could be followed by a higher pitch, e.g., /mimosi/ E6Db6G5 (thus still with a rising contour) or by a lower pitch, e.g., /sipygy/ G5B5C6 (thus with a change in contour). In addition, the mean pitch interval within TP-words was not significantly different from the mean interval between TP- words.

Procedure was the same as in Experiment 1. The testing phase was identical to the previous experiment, using spoken and not sung items.

Results

The distribution of percentages of TP-word choices is presented in Figure 3. Only music experts performed above chance, with 61.25 % TP-word choices, on the average, $t(19) = 3.92$, $p = .0005$. Non-experts and language experts only reached an average level of TP-word choice of 51.81 and 47.57 %, respectively. Their performance did not differ from chance level, both ts ≤ 1. The ANOVA run on the raw number of TP-word choices showed a significant effect of expertise, $F(2, 53) = 9.87$, $p < .0005$, with music experts performing significantly better than both non-experts ($p = .01$), and language experts ($p < .0005$), according to Scheffé's tests.

For Peer Review

Experiment 3: Sung inconsistent sequences

In the present experiment, while each syllable was still sung, there was variable syllable-pitch mapping. Hence, as in the third experiment of Schön et al. (2008), linguistic and musical boundaries, both of them signaled by TP dips, no longer matched (see Figure 1c).

Method

Participants. A total of 48 native French-speaking new volunteers took part in this experiment. Participants were assigned to one of the three groups of expertise, each including 16 participants, according to the same criterions as in the former experiments. The non-experts were aged 30 to 66 years (average: 45.13 yrs; 8 men and 8 women) and the language experts were aged 22 to 51 years (average: 28.94 yrs; 1 man and 15 women). The music experts, aged 19 to 69 years (average: 26.88 yrs; 8 men and 8 women) had begun music lessons between 4 and 12 years of age (on average, at 7.81 yrs) and had between 10 and 35 years of music training (on average, 16.56 yrs). As in the former experiments, all participants obtained at least 95% correct responses in the repetition of pseudo-words test (Mousty et al., 1994).

Material and procedure. The material was the same as the one used by Schön et al. (2008) in their third experiment. The musical and linguistic TPs were preserved but they

For Peer Review

were not matching any more (see Figure 1c) and participants could not segment the stream of speech on the basis of the musical feature. The testing phase was identical to the previous experiments, including only spoken items.

Results

The distribution of percentages of TP-word choices is presented in Figure 4. Only language experts performed above chance, with 60.42% TP-word choices, on the average, $t(15) = 3.61$, $p < .005$. Non-experts and music experts only reached an average level of TP-word choice of 51.91 and 53.13%, respectively. Their performance did not differ from chance level, both $ts < 1$. The effect of expertise was significant in the ANOVA run on the raw number of TP-word choices, $F(2, 45) = 3.14$, $p = .05$. However, Scheffé's tests only revealed a trend for language experts to perform better than non-experts, $p = .08$.

Cross-experiments analysis

We performed an ANOVA with two between-subjects factors, expertise and material (i.e., experiment: spoken, sung consistent, or sung inconsistent sequence). Neither the main effect of expertise, $F(2,143) = 2.21$, $p > .10$, nor the main effect of the experiment, $F < 1$, were significant but, as expected from the different response patterns reported above for each experiment, there was a highly significant interaction between these factors, $F(4, 143) = 5.29$, $p = .001$, which is illustrated in Figure 5**. Note that for the age and the**

For Peer Review

gender of the participants, we controlled a posteriori that these variables did not affect the interaction between expertise and material.

In fact, material did not affect non-experts' performance, $F < 1$, whereas it did affect the performance of both music experts, $F(2, 49) = 3.22$, $p < .05$ and language experts, $F(2, 45) = 6.23$, $p < .005$, but in different ways. Separate ANOVAs confirmed that, for music experts, the sung consistent sequences led to better performance than the spoken sequences, $F(1, 49) = 5.53$, $p < .025$, and (although to a lesser extent) the sung inconsistent sequences, $F(1, 49) = 3.62$, $p = .06$. On the contrary, for language experts, better performance was observed for both the spoken, $F(1, 45) = 6.38$, $p < .025$, and sung inconsistent sequences, $F(1, 45) = 11.55$, $p < .005$, in comparison to the sung consistent sequences.

Discussion

In the present study, we presented either spoken or sung sequences to different groups of adults: non-experts, music experts and language experts. Our results show a highly significant interaction between experiment (i.e. material) and expertise. **Only music experts showed the pattern of results obtained by Schön et al. (2008),** with higher scores for sung consistent sequences, in which music and linguistic TPs coincide, compared to other sequences. Neither non-experts nor language experts, **who did not have any music**

For Peer Review

expertise, benefited significantly from the congruence of musical and linguistic statistical information. In addition, for consistent sequences, music experts outperformed the other participants. **This results are thus in line with the hypothesis of the music experts superiority, but only when musical and linguistic information are consistent.**

The absence of any effect of the redundancy of musical and linguistic TPs in non-experts may seem surprising in comparison to the results reported by Schön et al. (2008). One should however note that in Schön et al., participants' performance for the sung consistent sequences were significantly different from chance level but not extremely high (64% of correct responses). Possible differences in age, gender, **musical background** and level of motivation between the participants examined by Schön et al. and the ones examined in the present study may perhaps explain this difference. In addition, several studies on non-experts have reported interactions between the linguistic and melodic dimensions of songs, at both the behavioral and electrophysiological levels (e.g., Gordon, Schön, Magne, Astésano, & Besson, 2010; Kolinsky et al., 2009; Lidji, Jolicoeur, Kolinsky, Moreau, Connolly, & Peretz, 2010). Thus, our results do not imply that the non-experts of the present study were totally insensitive to the musical information contained in the sung sequences. Instead, it might be the case that the learning measure we used was not sensitive enough. Indeed, François and Schön (2010) examined musically untrained participants

For Peer Review

exposed to 15 min sung consistent sequences. They observed that although their performance in a musical learning test was not significantly different from chance, electrophysiological measures (event-related potentials, *ERPs*) revealed a significant familiarity effect, showing that in fact participants learned the musical structure better than their **overt** behavior seemed to suggest. Future studies would thus benefit from combining the ALL situation with these kinds of measures, or with on-line behavioral measures of word learning like the click detection task, as this measure (slower detection of clicks located within TP-words than between them) is sensitive to ALL learning after just two minutes familiarization (Gómez, Bion, & Mehler, 2011).

Furthermore, in future ALL studies on sung material, it would be interesting to present music and language experts with both a linguistic and a musical learning test. Recently, François and Schön (2011) observed an effect of music expertise on ERPs not only in a linguistic test similar as the one used here, but also in a music test. Examining also language experts in both tasks would clarify the causes underlying the higher sensitivity of music experts to the superposition of musical and linguistic TPs. As a matter of fact, music experts may have been more sensitive to the musical structure than the other participants, which would explain their higher sensitivity to the superposition of musical and linguistic TPs.

For Peer Review

In any case, our results show that the effect of music expertise cannot be reduced to a general effect of expertise on executive control. Had it been the case, we would have observed better performance for musicians and language experts compared to other participants whatever the material, which was clearly not the case. Much on the contrary, music experts seem to have been disturbed when the musical TPs did not fit with the linguistic ones, as was the case in the inconsistent sequences used in Experiment 3. This supports the notion that they were more sensitive or paid more attention to the musical structure than the other participants (musical TPs, or musical contours, or both), leading to greater interference in case of mismatch.

Unexpectedly, language experts were better with both the sung inconsistent sequences and the spoken sequences than with the sung consistent sequences. Their expertise as speech therapists includes fine phonemic discrimination and the ability to retrieve words from patients' corpus, often with very low intelligibility and/or deviant forms of language. **This may lead to predict that they may be better than non-experts with spoken sequences. Yet, a superiority of language experts compared to non-experts was only observed (as a trend) with the sung inconsistent sequences. The exact reason (i.e., why the redundancy of the sung consistent sequences deteriorated rather than benefited their performance) should be matter of future research.** In future work,

For Peer Review

it would also be interesting to check whether such a response pattern would differ as a function of specific language expertise, contrasting for example speech therapists to translators, linguists, phoneticians, or to 'informal" language experts like bilinguals.

In any case, the results observed with the language experts argue for the specificity of music expertise, which aid learning an artificial language with sung material.

For Peer Review

Acknowledgements

Pauline Larrouy-Maestri is now PhD student under the supervision of Dominique Morsomme, who kindly supported the writing of this paper. The authors thank Sean Hutchins for his insightful comments on earlier draft of the paper.

For Peer Review

Funding

This research was supported by a FRFC grant 2.4633.66.F ("Mental representations of music and language in singing and nature of their interactions") as well as by a FER grant from Université Libre de Bruxelles ("MIEL: La Musique dans ses Interactions avec les Emotions et le Langage"). The last author is Research Director of the *Fonds de la Recherche Scientifique – FNRS*.

For Peer Review

References

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of Training between Music and Speech: Common Processing, Attention, and Memory. *Frontiers in Psychology, 2*, 94.

Besson, M., Faïta, F., Peretz, I., Bonnel, A.-M., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science, 9*(6), 494-498.

Bialystok, E., & DePape, A.-M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 565-574.

Delogu, F., Lampis, G., & Belardinelli, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology, 22*(1), 46-61.

François, C., & Schön, D. (2010). Learning of musical and linguistic structures: Comparing event-related potentials and behavior. *NeuroReport,* 21(14), 928-932.

François, C., & Schön, D. (2011). Musical expertise boost implicit of both musical and linguistic structures. *Cerebral Cortex, 21*(10), 2357-2365.

For Peer Review

Francois, C., Tillmann, B., & Schon, D. (2012). Cognitive and methodological considerations on the effects of musical expertise on speech segmentation. *Annals of the New York Academy of Sciences, 1252*, 108-115.

Gómez, D., Bion, R. A. H., & Mehler, J. (2011). The word segmentation process as revealed by click detection. *Language and Cognitive Processes*, *26*(2), 212-223.

Gordon, R. L., Schön, D., Magne, C., Astésano, C., & Besson, M. (2010). Words and melody are intertwined in perception of sung words: Eeg and behavioral evidence. *PLoS ONE, 5(3).*

Hébert, S., Racette, A., Gagnon, L., & Peretz, I. (2003). Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain, 126*(8), 1838-1850.

Kolinsky, R., Cuvelier, H., Goetry, V., Peretz, I., & Morais, J. (2009). Music training facilitates lexical stress processing. *Music Perception, 26*(3), 235-246.

Kolinsky, R., Lidji, P., Peretz, I., Besson, M., & Morais, J. (2009). Processing interactions between phonology and melody: Vowels sing but consonants speak. *Cognition, 112*(1), 1-20.

Kraus N, Chandrasekaran B. (2010) Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11*, 599-605.

For Peer Review

Lidji, P., Jolicoeur, P., Kolinsky, R., Moreau, P., Connolly, J. F., & Peretz, I. (2010). Early integration of vowel and pitch processing: A mismatch negativity study. *Clinical Neurophysiology, 121*, 533-541.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological appoaches. *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 Segmentation and Tonal Processing in Mandarin Chinese. *Journal of Cognitive Neuroscience, 23*(10), 2401-2415.

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

Milovanov, R., Huotilainen, M., Valimaki, V., Esquef, P., & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. *Brain Research, 1194*, 81-89.

For Peer Review

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

Mousty, P., Leybaert, J., Alegria, J., Content, A., & Morais, J. (1994). BELEC : Une batterie d'évaluation du langage écrit et de ses troubles. In J. Grégoire & B. Piérart (Eds.), *Evaluer les troubles de la lecture : Les nouveaux modèles théoriques et leurs implications diagnostiques* (pp.127-145). Bruxelles : De Boeck.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science,* 22(11), 1425-1433.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2008). Musical Training Influences Linguistic Abilities in 8-Year-Old Children: More Evidence for Brain Plasticity. *Cerebral Cortex*, *19*(3), 712-723.

Peretz, I. (2009) Music, Language and Modularity Framed in Action. *Psychologica Belgica*, *49*, 157-175.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain, 117*, 1283-1301.

For Peer Review

Racette, A., & Peretz, I. (2007). Learning lyrics: To sing or not to sing? *Memory & Cognition, 35*(2), 242-253.

Saffran, J., Aslin, R. N., & Newport, E. L. (1996a). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*(4), 606-621.

Saffran, J., Aslin, R. N., & Newport, E. L. (1996b). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Saffran, J., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27-52.

Samson, S., & Zatorre, R. J. (1991). Recognition memory for text and melody of songs after unilateral temporal lobe lesion: Evidence for dual encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(4), 793-804.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition, 106*(2), 975-983.

Schön, D., Gordon, R., & Besson, M. (2005). Musical and linguistic processing in song perception. *Annals of the New York Academy of Sciences, 1060*, 1-11.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41*(3), 341-349.

For Peer Review

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science, 17*(8), 675-681.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39*(4), 706-716.

For Peer Review

Figure Captions

Figure 1. Illustration of the sung sequences used in Experiments 2 and 3. (a): correspondence syllable / sound used for Experiment 2; (b): consistent sequences used in Experiment 2, in which the pitches used for a given word are always the same; (c): inconsistent sequences used in Experiment 3, in which pitches were variable for a given word.

Figure 2. Summary of the distribution of TP-word choices (in percentages) observed in Experiment 1, separately for each group of expertise. The bottom and top of the boxes represents the 25th and 75th percentile (lower and upper quartiles), respectively, with a line at the median. Errors bars represent the lowest and highest scores within 1.5 interquartile range (IQR).

Figure 3. Summary of the distribution of TP-word choices (in percentages) observed in Experiment 2, separately for each group of expertise. The bottom and top of the boxes represents the 25th and 75th percentile (lower and upper quartiles), respectively, with a line at the median. Errors bars represent the lowest and highest scores within 1.5 interquartile range (IQR).

Figure 4. Summary of the distribution of TP-word choices (in percentages) observed in Experiment 3, separately for each group of expertise. The bottom and top of the

For Peer Review

boxes represents the 25th and 75th percentile (lower and upper quartiles), respectively, with a line at the median. Errors bars represent the lowest and highest scores within 1.5 interquartile range (IQR).

Figure 5. Average percentages of TP-word choices observed in Experiments 1 to 3, separately for each material (i.e., experiment) and group. Errors bars represent standard deviations. 50% represents the chance level.

For Peer Review

Figure 1a

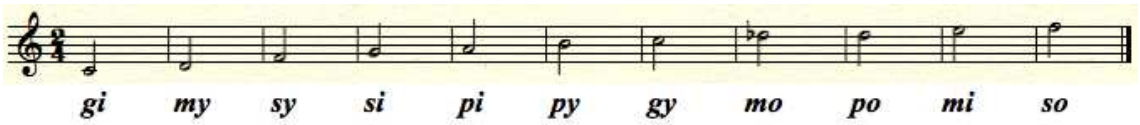


Figure 1b

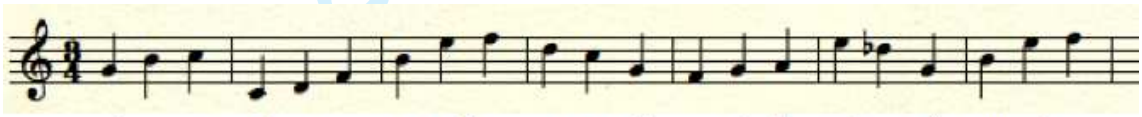
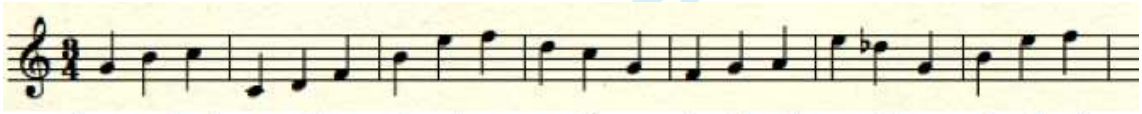
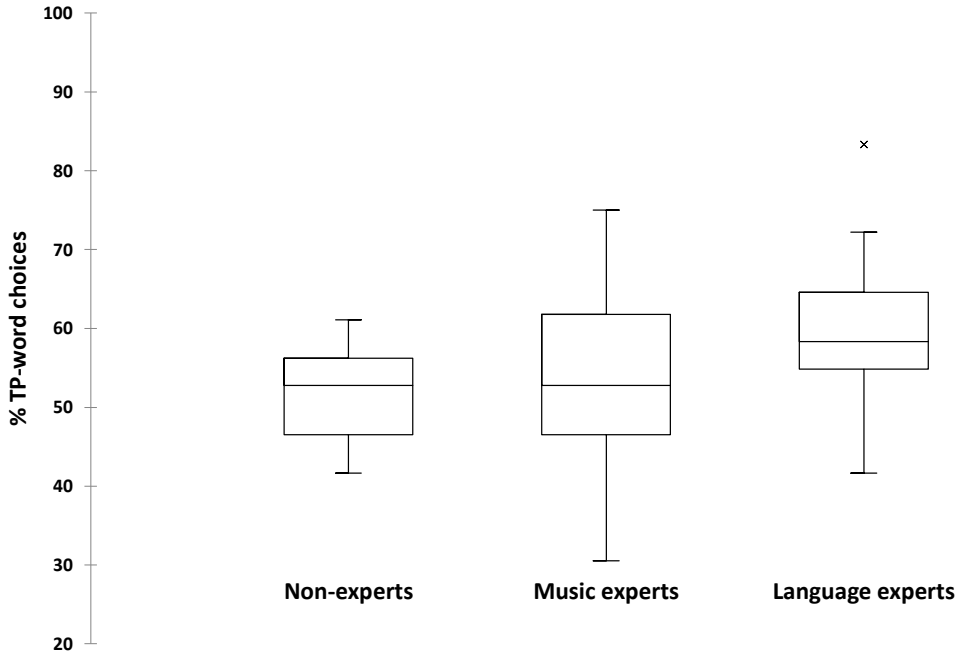


Figure 1c



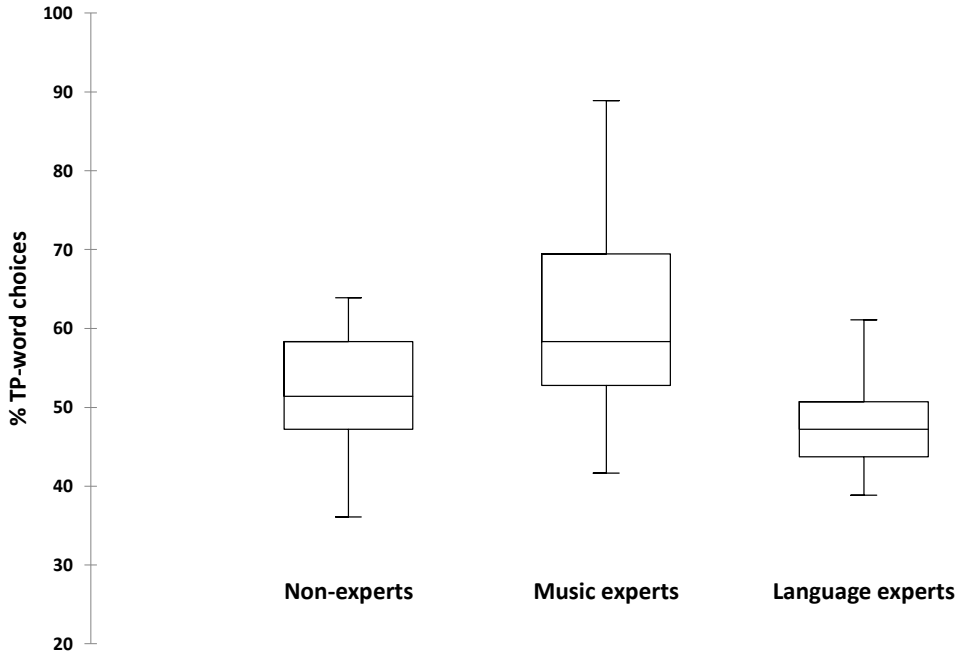
For Peer Review

Figure 2



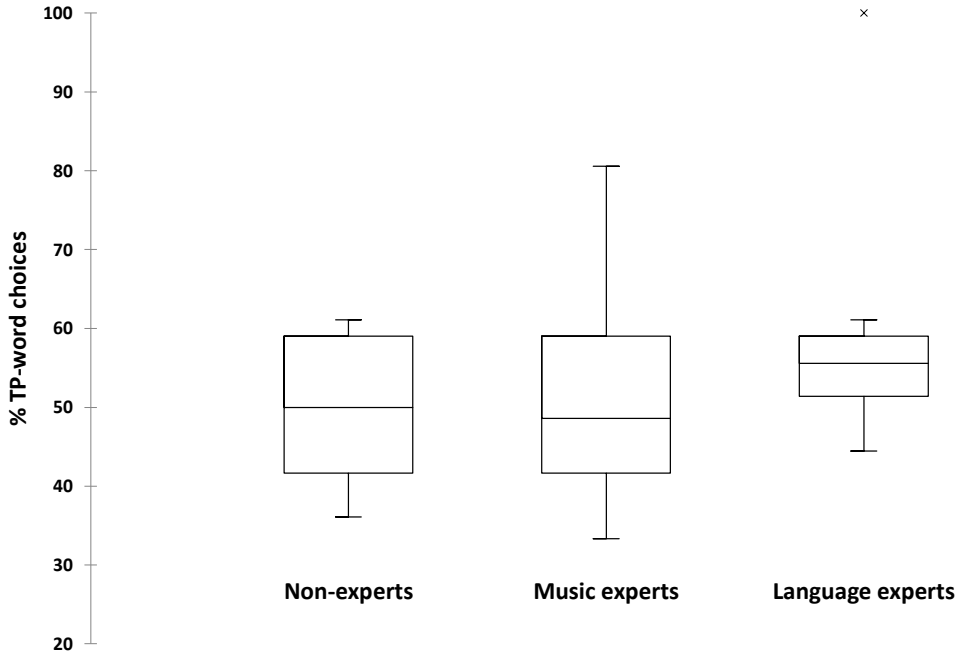
Review

Figure 3



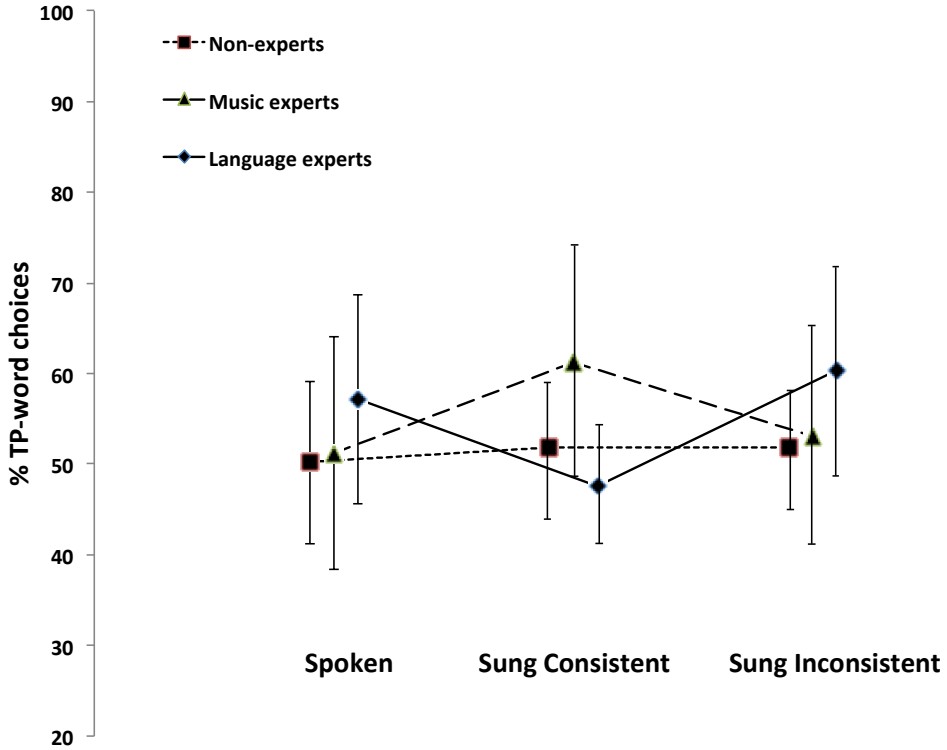
Review

Figure 4



Review

Figure 5


100
90
80
70
60
50
40
30
20
% TP-word choices
Non-experts
Music experts
Language experts
Spoken
Sung Consistent
Sung Inconsistent