Exp Brain Res (2017) 235:3619–3630
DOI 10.1007/s00221-017-5080-7

RESEARCH ARTICLE

Music and speech distractors disrupt sensorimotor synchronization: effects of musical training

Anita Białuńska[1,2] · **Simone Dalla Bella**[1,3,4,5]

Received: 15 November 2016 / Accepted: 2 September 2017 / Published online: 9 September 2017

Abstract Humans display a natural tendency to move to the beat of music, more than to the rhythm of any other auditory stimulus. We typically move with music, but rarely with speech. This proclivity is apparent early during development and can be further developed over the years via joint dancing, singing, or instrument playing. Synchronization of movement to the beat can thus improve with age, but also with musical experience. In a previous study, we found that music perturbed synchronization with a metronome more than speech fragments; music superiority disappeared when distractors shared isochrony and the same meter (Dalla Bella et al., PLoS One 8(8):e71945, 2013). Here, we examined if the interfering effect of music and speech distractors in a synchronization task is influenced by musical training. Musicians and non-musicians synchronized by producing finger force pulses to the sounds of a metronome while music and speech distractors were presented at one of various phase relationships with respect to the target. Distractors were familiar musical excerpts and fragments of children poetry comparable in terms of beat/stress isochrony. Music perturbed synchronization with the metronome more than speech did in both groups. However, the difference in synchronization error between music and speech distractors was smaller for musicians than for non-musicians, especially when the peak force of movement is reached. These findings point to a link between musical training and timing of sensorimotor synchronization when reacting to music and speech distractors.

Keywords Sensorimotor synchronization · Timing · Musical training · Music interference · Speech interference

Introduction

Humans tend to move naturally and spontaneously to the beat of music. In contrast, moving to speech stresses is rare. This discrepancy is likely to result from the regular, rhythmical structure of music, as opposed to temporally irregular speech stresses (Dalla Bella et al. 2013). The rhythmical properties of speech are rich like in music (Liberman and Prince 1977; Patel and Daniele 2003; Patel 2008). However, while music is universally characterized by regular isochronous beats (Drake and Bertrand 2001; Stevens and Byron 2009) affording a synchronized motor response (Large and Jones 1999; Large and Palmer 2002; London 2012) this is not true for speech. Beat perception (i.e., the perception of predictable sequences of strong and weak beats in music) is supported and reinforced by the properties of the musical structure consisting of temporal patterns with multiple embedded periodicities (i.e., meter; Large and Jones 1999; Large and Palmer 2002; London 2012). Stresses in speech similarly evoke a subjective impression of isochrony (Lehiste 1977), but the claim that stressed syllables underlie

---

✉ Anita Białuńska
anita.bialunska@gmail.com

✉ Simone Dalla Bella
simone.dalla-bella@umontpellier.fr

[1] Department of Cognitive Psychology, WSFiZ, Warsaw, Poland

[2] Department of Rehabilitation Sciences, College of Health Sciences, University of Texas at El Paso, 1101 Campbell Street, El Paso, TX 79902, USA

[3] EuroMov, University of Montpellier, 700 Avenue du Pic St Loup, 34090 Montpellier, France

[4] Institut Universitaire de France, Paris, France

[5] International Laboratory for Brain, Music and Sound Research (BRAMS), Montreal, Canada

an isochronous pulse (Abercrombie 1967; Pike 1945) is not empirically supported (Arvaniti 2009; Dauer 1983; Patel 2008). First, inter-stress intervals in spoken utterances show high variability, much higher than inter-beat intervals in performed expressive music (Dauer 1983; Repp 1998). Moreover, the metrical structure of conversational speech (based on rhythmic prominence of syllables, words, and phrases), albeit showing a hierarchical structure (Liberman and Prince 1977), is less strict and regular than musical meter (Selkirk 1984). Note, however, that greater metrical regularity is typical of other registers generally referred to as "metrical speech", such as poetry declamation, or in a group context, choral speaking and prayers (Cummins 2009).

In a previous study (Dalla Bella et al. 2013), we tested the tendency of music and metrical speech (i.e., spoken utterances with a regular stress pattern and metrical structure) to attract movement in a finger tapping task (sensorimotor synchronization, SMS; e.g., Drake et al. 2000; Jones and Pfordresher 1997; Lidji et al. 2011; Villing et al. 2011; for reviews, Repp 2005; Repp and Su 2013). An interference paradigm was used in which participants tap their finger to the sounds of a metronome (targets) while simultaneously periodic distractors are presented, but at one of various temporal separations from the targets (Dalla Bella et al. 2013; Hove et al. 2013; Repp 2003; Repp and Penel 2004). It was found that both music and speech distractors disturb synchronization to the beat. However, this effect is more visible for music than for metrical speech (i.e., music attracts movement more than speech). The strength of this effect depends on the temporal regularity of musical beats as opposed to irregular speech stresses and on pitch-related factors. Music's perturbing effect is reduced when beat/stress isochrony and average pitch are controlled. The critical factor accounting for the difference between music and speech is likely to be music's metrical structure. Indeed, when the speech distractor is manipulated so as to have an isochronous beat and a music-like metrical structure speech attracts movement as much as music does. In sum, auditory sequences with a regular beat structure (i.e., a periodic beat and a regular meter) attract movement irrespective of the stimulus domain (music or speech).

The dominance of music over speech in attracting movement was observed in participants who did not have any formal musical training or particular musical experience. The main goal of the present study is to assess whether the distractor effect previously observed is dependent on musical training. In general, musicians possess extensive and explicit knowledge of musical pitch and rhythmic structure due to long-lasting formal musical training and active musical listening (Bigand 2003; Kraus and Chandrasekaran 2010; Hallam et al. 2014). This knowledge allows them, for example, to extract the tonality of a musical excerpt, recognize a given chord or pitch interval, or name a piece's composer. This explicit musical knowledge is typically not mastered by non-musicians. The difference between musicians and non-musicians, however, is significantly reduced when considering implicit knowledge of music. There is a consistent body of evidence showing that musicians and non-musicians alike acquire significant implicit musical knowledge (e.g., tonal knowledge) by mere music listening (Bigand 1997; Bigand et al. 1999; Tillmann et al. 2001). Non-musicians use the same principles as musicians during music perception for example by internalizing complex statistical regularities through passive exposure to musical stimuli (Bigand 2003). This implicit knowledge allow non-musicians to perform complex analyses of musical structure (e.g., to develop a hierarchical system of tonal expectancies), thus displaying listening skills which are characteristic of music experts.

Similarities and differences linked to musical training were observed in a few studies on SMS. When asked to tap to the beat of rhythmic sequences, people with formal musical training are typically more accurate (i.e., they tap closer in time to the stimulus) and more precise (i.e., they show lower variability of the inter-tap interval both in synchronization and continuation tapping) than non-musicians (Repp and Doggett 2007). Moreover, musicians display the greatest sensitivity to changes in the rate of the stimuli to which they synchronize. Musicians detect smaller changes of the inter-stimulus interval and quickly adjust their taps to the tempo changes of the stimuli than non-musicians, thus showing lower tapping variability after the presentation of the last stimulus (Repp 2010). Two online correction mechanisms are involved in SMS, which afford sustained and long-lasting synchronization to a beat. The first mechanism—phase correction—is responsible for adjusting the onsets of each successive tap; it does not affect the tapping rate. In contrast, the second mechanism—period correction—adjusts the interval of the internal timekeeper resulting in a modification of the tapping rate (Repp 2006). Phase correction is thought as being a relatively automatic process while period correction is largely consciously controlled. Phase correction is tested with tasks in which participants synchronize with an isochronous sequence of sounds embedding a perturbation (i.e., an interval lasting longer or shorter than the regular inter-stimulus interval; see Repp 2000, 2001). Notably, motor reaction to perturbation (i.e., variation in stimulus-tap asynchrony following the perturbation) occurs even when the perturbation is not detectable (i.e., subliminal). This response is thought to reflect automatic phase correction processes (Repp 2000, 2001; Thaut et al. 1998). Accordingly, phase correction is most effective in reaction to small changes and is independent of attention modulation (Repp and Keller 2004). In addition, when the subjects are instructed to not react to a detectable event onset shift in the sequence of the pacing stimuli, they could reduce the effect of phase correction, but not suppress it completely (Repp

2002; Repp and Keller 2004). In contrast, period correction in reaction to tempo changes in a sequence is more effective when the change of tempo is noticeable than when it is not perceptible (Repp 2001). The action of period correction can be observed in particular in a synchronization-continuation tapping task with tempo perturbation introduced immediately before the continuation phase. In this phase, the inter-tap intervals start to be adjusted to the new inter-stimulus intervals while variability of the stimulus-tap asynchronies decreases over time. Period correction is affected by attention and it is possible to stop this process voluntarily (Repp and Keller 2004).

Musicians show faster phase correction following a tempo change in the pacing sequence than non-musicians. However, phase correction following a simple phase shift (i.e., without a tempo change) does not seem to depend on musical training (Repp 2010). Note that permanent phase shift responses of participants who had even occasional musical training are faster than those of participants with no instrument training, provided that the perturbation is detected. These findings suggest that phase and period correction are simultaneously engaged following the detection of a phase shift or the possibility of cognitive control during phase correction in case of participants with musical experience. An interesting finding is the effect of musical training in synchronization/perturbation detection tasks on the phase correction response described by Repp (2010). Greater task experience slows phase correction in participants with long-lasting musical training and/or practice. On the other hand, this effect was not tested in non-musicians because all of them were novice in SMS task. It is worth noting that fast phase correction is not always the most efficient response strategy, especially when the phase shift in the stimulus sequence is local and does not entail a persistent change in stimulus rate. Non-reacting to such perturbation requires strategic control of movement timing during SMS, which can be improved with practice on SMS. Overall, these findings suggest that formal musical training may affect the performance of paced tapping tasks by fostering intentional and conscious control of error correction mechanisms. This is critical in SMS for achieving high precision of synchronization and perceptual sensitivity to timing changes.

The aim of the present study is to examine the effect of musical training on SMS in the presence of music and speech distractors. As done in previous studies (Dalla Bella et al. 2013) we adopt an interference paradigm using finger tapping in which music and speech are used as distractors. Participants are asked to synchronize with an isochronous sequence (i.e., a metronome) while distractors showing a periodic beat/stress pattern are presented at one of various temporal phase offsets (Dalla Bella et al. 2013; Hove et al. 2013; Repp 2003). We expect music distractors to interfere with synchronization more than speech distractors, as observed in our previous study. We also anticipate the effect of distractors to be influenced by musical training. Maintaining synchronization to the metronome in the presence of a distractor is likely to put particular demands on implicit and explicit error correction mechanisms, as compared to a standard SMS task. It is expected that musicians, because of their fine-tuned correction mechanisms (in particular period correction), will have less difficulties than non-musicians in keep synchronization to the metronome in the presence of the distractors. This advantage is likely to be more visible with music distractors than with speech distractors, due to musicians' experience with playing in ensembles. This may facilitate focusing the attention on a single auditory stream while ignoring simultaneous musical performances.

Methods

Participants

Two groups of native Polish speakers took part in the study. The first group ('non-musicians') was formed by 26 students without formal musical training from the University of Finance and Management in Warsaw (22 females; mean age 21.7 years, range 19–42). The second group ('musicians') consisted of 14 active musicians recruited from the Warsaw community (11 females; mean age 27 years, range 23–47) with at least 12 years of formal musical training ($M = 12.9$ years, range 12–17 years). Nobody among musicians played percussions, contrabass or bass guitar. All participants were right handed, except two participants, one in each group. Participants did not report any history of hearing impairment. All participants gave informed consent according to the procedures approved by the Ethics Committee of the University of Finance and Management in Warsaw.

Stimulus materials

The target and distractor sequences were the same as those used in a previous study (Dalla Bella et al. 2013; Original condition). A target sequence included thirty-five 30-ms computer-generated tones with constant pitch (880 Hz sinusoids with a linear 17-ms down-ramp) and intensity. They were presented with an inter-onset interval (IOI) of 600 ms. The music distractors were three computer-generated well-formed musical fragments from familiar music written in binary meter (i.e., circus music, 'Sleigh ride', and Bee Gees' 'Stayin'Alive'). They included 29–33 musical beats, and were presented with an inter-beat interval of 600 ms. Speech distractors were well-formed fragments from three well-known excerpts of Polish children poetry 'Pstryk' and 'Lokomotywa' (Tuwim 1980); 'Na straganie (Brzechwa 1980). 'Pstryk' and 'Na straganie' were written in a binary meter (i.e., every second syllable was stressed), ‘Lokomotywa in a ternary meter (i.e., every third syllable was stressed).[1] Speech distractors included 28–32 stresses. Speech fragments were read by an actor who synchronized speech stresses to the sounds of a metronome (IOI = 600 ms) and recorded. The mean inter-stress interval of the recorded speech fragments was 598 ms (SD = 66 ms), showing that the actor was able to maintain the speech rate. Distractor stimuli were all normalized to the same maximum sound pressure level and the intensity was set at a comfortable level for each participant.

Following the first five sounds of the target sequence, the distractor was presented at a particular temporal separation from the sixth target sound. The first musical beat/speech stress of the distractor was obtained by the experimenter by visual inspection of the waveform and by listening to the stimuli. As in our previous study, 20 relative temporal separations between target and distractor were used (i.e., ‘relative phases’, called hereafter simply ‘phases’). At phase 0, the sixth sound of the target sequence and the first musical beat or speech stress of the distractor coincided (e.g., Dalla Bella et al. 2013; sound examples can be found at http://www.mpblab.vizja.pl/dallabella_et_al_plos1_stimuli.html for Experiment 1, Original condition). The sixth sound of the target sequence was aligned to the first musical beat or to the first speech stress. Note that target-distractor alignment at phase 0 usually led to comparable synchronization performance with musical and speech stimuli (Dalla Bella et al. 2013). The remaining 19 phases ranged from −50% of the IOIs (−300 ms) to +45% of the IOIs (+270 ms) with a step of 5% of the IOIs (30 ms). Negative and positive phases indicate that the musical beats and speech stresses occurred before and after target sounds, respectively.

Musical stimuli were generated with a Yamaha MidiRack synthesizer. Speech sequences were recorded with a Shure SM58 microphone onto a hard disk through Fostex D2424 LV 24 Track Digital Recorder (sampling rate = 44.1 kHz).

Procedure

Each participant was submitted to three conditions: a control condition (i.e., Target only condition), where only the target sequence was presented without a distractor, and two distractor conditions (Music condition and Speech condition). In the distractor conditions, distractor sequences (music or speech) were presented at a particular temporal separation between beats/speech stresses and metronome sounds. Participants sit in a quiet room in front of the computer monitor. Their motor response was measured by an isometric response, which is comparable to standard tapping but more sensitive to perception and action components of movement preparation (Białuńska et al. 2011; Jaśkowski and Włodarczyk 2006). For each stimulus sequence, participants put their index finger on the surface of a low-profile force transducer, and were asked to increase the finger’s pressure force in synchrony with the pacing stimulus (e.g., the target sequence alone in Target only condition). During all the sequences participants kept their finger in contact with the surface of the transducer.

The Target only condition was performed twice, before performing the two distractor conditions. In one block, the distractor + target stimuli sequences at all phases were presented in random order. The order of the distractors (i.e., music or speech) and the order of the blocks were counterbalanced across subjects. For each distractor type (i.e., music or speech) there were three blocks of trials, one for each of the three stimuli. Stimuli were presented binaurally over Sennheiser eH2270 headphones at the same comfortable intensity level. The experiment was run on Presentation software (Neurobehavioral Systems, Inc.) using a PC-compatible computer. Pressure force data was recorded by the force transducers, amplified (QuickAmps, BrainProducts Inc.), and stored onto the computer hard disk at a sampling rate of 250 Hz (BrainRecorder software, BrainProducts Inc.). The experiment lasted approximately 1 h and a half.

Analysis of force pressure data

Three measures were derived from the force pulses obtained in each sequence. Two of them, Time to threshold and Time to peak are measures of the asynchrony between the force pulses and the target stimuli (see Białuńska et al. 2011). Time to threshold (in ms) is the time interval between the stimulus onset and the moment at which force reaches a threshold value of 1.5 N.[2] Force pulses with an intensity

---

[1] Both the speech and music materials were highly controlled for familiarity. This was assessed for both the musical excerpts and speech fragments in a pilot experiment (see Dalla Bella et al. 2013). Thirty-two participants (13 musicians and 19 non-musicians; 28 females, mean age = 20.6 years, range 20–28 years) rated the distractors on a 10-point scale (1 = not familiar; 10 = very familiar). Note that the 13 musicians who participated in the pilot experiment also took part in the main experiment. Music and speech distractors did not differ in terms of familiarity (for music, mean rating = 6.7; for speech, mean rating = 7.0; $t < 1$). Moreover, ratings from musicians and non-musicians for music familiarity did not differ (for musicians, mean rating = 7.2; for non-musicians, mean rating = 6.4; $t < 1$).

[2] 1.5 Newton correspond to 300 μV, which is the force intensity typically needed to obtain a response using the computer keyboard or a response pad. One Newton is equivalent to the force needed to give a mass of 1 kg an acceleration of 1 m/s^2.

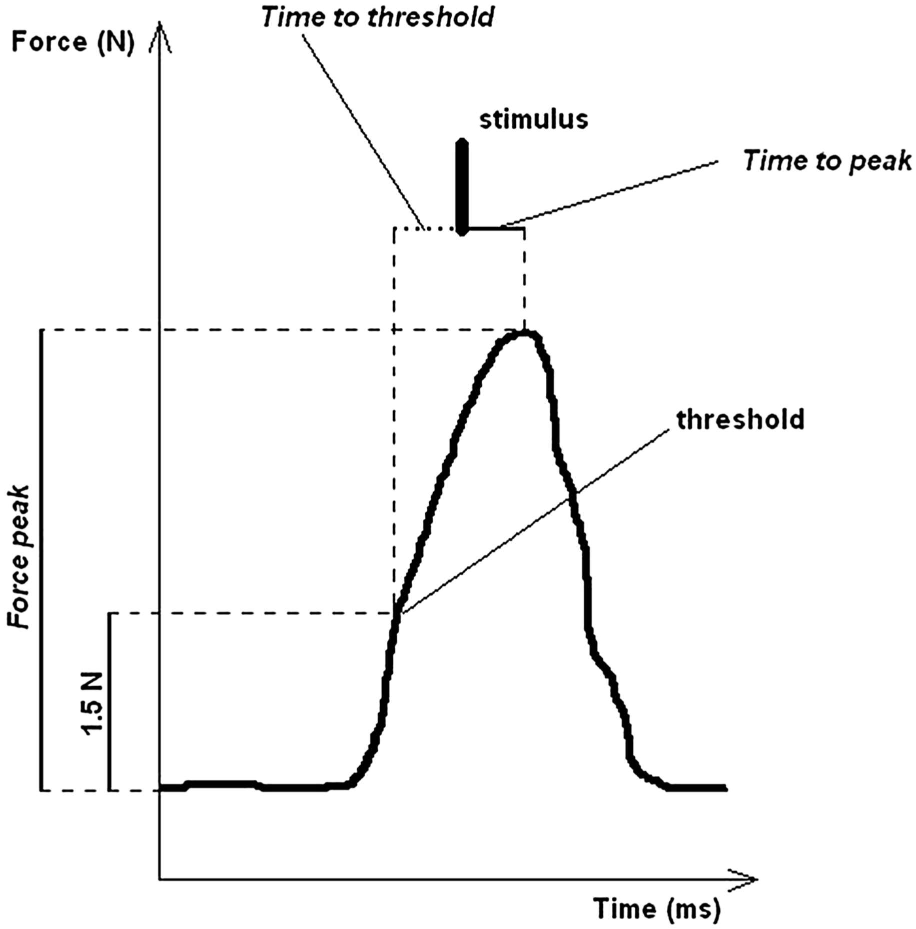

Fig. 1 Time to threshold, time to peak, and force peak for a typical force time course obtained in the SMS task (Białuńska et al. 2011)

below 1.5 N (i.e., weak intensity pressure) were treated as ‘missing taps’ and the corresponding force trajectories were discarded. Negative time to threshold indicates that the threshold is reached before the stimulus. Time to peak (in ms) is the time interval between the stimulus onset and the moment when force reaches its peak. Negative time to peak indicates that the force peak is reached prior to the stimulus. An additional measure, Force peak (in N) corresponds to the maximum force reached at the peak of the intensity trajectory.

The measures are illustrated in Fig. 1 for a typical force pulse obtained in the SMS task

Results

The force peaks produced by all musicians were all greater than 1.5 N. However, four out of 26 non-musicians were discarded because they produced less than 23 consecutive synchronized taps, i.e., the force peaks were smaller than 1.5 N (80% of the maximum number of taps) in the Target only condition. The remaining 22 non-musicians produced a very few force peaks (6% of all taps, on average) with intensity below 1.5 N. Taps corresponding to the first seven target sounds were not analyzed. After participants finished the tasks they often spontaneously reported that it was quite difficult to ignore the distractor sequences when they were not aligned with the target sequence. To assess whether the perturbation of tapping due to music and speech distractors followed a pattern across phases which is comparable to the one observed in previous studies (Repp 2003; Dalla Bella et al. 2013), asynchrony was computed as a measure of synchronization accuracy. For each tapping trial, the signed time differences between the target sounds and the taps separately for each measure of asynchrony (time to threshold and time to peak) were calculated (as in Dalla Bella et al. 2013). These differences indicate signed asynchronies, and for simplicity, are referred to as “asynchronies” throughout the paper. Mean asynchrony for each tapping trial was obtained for data in the Target only and in the distractor conditions separately. Mean asynchrony for time to threshold and time to peak with music and speech distractors as a function of distractor phase for musicians and non-musicians are illustrated in Figs. 2 and 3, respectively. Both music and speech distractors affected synchronization beyond normal tapping variability, as shown by several points falling beyond the normal

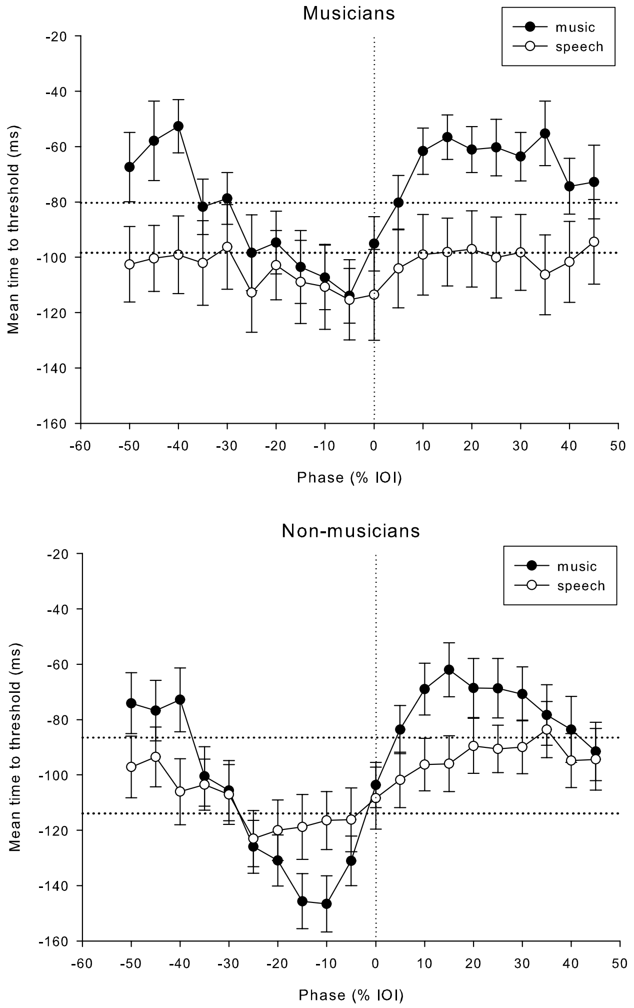

Fig. 2 Mean time to threshold with music and speech distractors as a function of the relative phase between distractors and target sounds. Error bars indicate SE of the mean. The horizontal dotted lines indicate ±SE of signed asynchrony obtained in the Target only condition

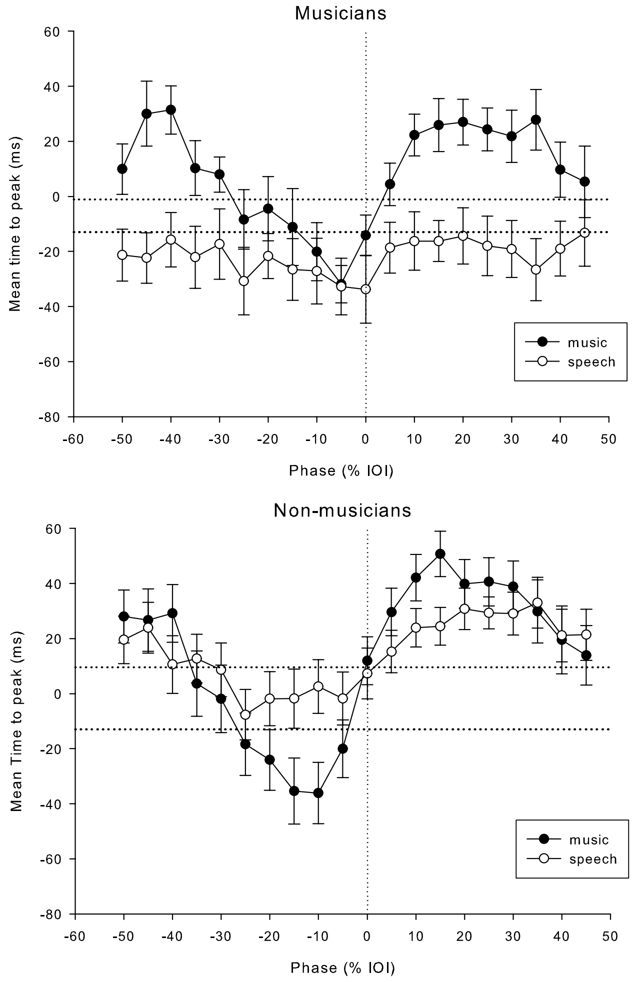

Fig. 3 Mean time to peak with music and speech distractors as a function of the relative phase between distractors and target sounds. Error bars indicate SE of the mean. The horizontal dotted lines indicate ±SE of signed asynchrony obtained in the Target only condition

variability around the mean (i.e., asynchrony obtained in the Target only condition ± standard error of asynchrony from this condition) represented in Figs. 2 and 3 by horizontal dotted lines. Asynchrony in the Target only condition calculated from a total of six trials did not significantly differ between the two groups, indicating comparable synchronization accuracy in terms of both time to threshold (non-musicians, asynchrony $= -100.2$ ms, SD $= 64.1$ ms; musicians, asynchrony $= -89.3$ ms, SD $= 33.9$; $t < 1$) and time to peak (non-musicians, asynchrony $= -1.7$ ms, SD $= 52.6$ ms; musicians, asynchrony $= -7.0$ ms, SD $= 33.9$; $t < 1$).

Results at phase 0 (baseline) where no interference was expected were analyzed first.

Mean asynchronies obtained at phase 0 in the Music and Speech conditions were entered in 2 (group) × 2 (distractor) mixed-design ANOVA, separately for time to threshold and time to peak, considering subjects as the random variable. Group (musicians vs. non-musicians) was the between-subjects factor and Distractor (music vs. speech) was the within-subjects factor. For all ANOVAs' results, the Greenhouse–Geisser correction for inhomogeneity of variance was applied whenever appropriate. Uncorrected degrees of freedom, epsilon value and probability level following correction, as well as power computed using G*Power 3.1.9.2 software (Faul et al. 2007) with $\alpha = .05$ are reported. For both groups, a tendency to synchronize more accurately in the Music condition than in the Speech condition was observed for time to threshold ($F(1.34) = 4.10$, $\eta_p^2 = .11$, $p = .05$, power $= .95$); however, for time to peak the difference did not reach significance. The Group × Condition interaction was not significant for both asynchrony measures. Next, we compared asynchrony in the Target only condition and in each distractor condition at phase 0, taken separately. Non-musicians showed no effect of interference when comparing distractor conditions to the Target only condition. In contrast, musicians showed an effect of interference at phase 0 only with speech distractors (time to threshold $t(13) = 2.33$, $d = .49$, $p < .05$, power = .40; time to peak $t(13) = 2.36$, $d = .74$, $p < .05$, power = .73). Thus, in musicians speech distractors presented at the same time as the target sounds led to greater asynchrony.

Mean asynchronies were entered in two 2 (group) × 2(distractor) × 20 (phase) mixed-design ANOVAs, separately for time to threshold and time to peak, considering subjects as the random variable. Group (musicians vs. non-musicians) was the between-subjects factor; Distractor (music vs. speech) and Phase (−50 to +45% of the IOI) were the within-subjects factors. Synchronization accuracy based on time to threshold differs depending on the distractor and phase as attested by a significant Distractor × Phase interaction ($F(19,646) = 10.75$, $\eta_p^2 = .24$, $\varepsilon = .33$, $p < .001$; power = .99), indicating that generally music was more distracting than speech at given phases. Differences between speech and music distractors at each phase were tested with Bonferroni-corrected t tests ($p < .05$). At phases −50, 45, −40, −10% and at +5, +10, +15, +20, +25, +30, +35 and +40% music distractors interfered with synchronization more than speech distractors did. Greater interference corresponded to a larger departure (larger or smaller asynchrony) from the performance obtained in the Target only condition. There was neither effect of Group nor other significant interactions. Similar effects were obtained when considering time to peak. Music interfered with synchronization more than speech, but this effect was not observed at all phases, as attested by a significant Distractor × Phase interaction [$F(19, 646) = 8.63$, $\eta_p^2 = .20$, $\varepsilon = .36$, $p < .001$, power = .99]. Unlike previous analyses, the Group × Phase × Distractor interaction reach significance [$F(19,646) = 1.62$, $\eta_p^2 = .05$, $p < .05$, power = .99]. The triple interaction was decomposed by running two 2 (distractor) × 20 (phase) ANOVAs, separately for each group. A significant Distractor × Phase interaction was found for both musicians [$F(19, 247) = 4.51$, $\eta_p^2 = .26$, $\varepsilon = .20$, $p < .01$, power = .99] and non-musicians [$F(19, 399) = 6.74$, $\eta_p^2 = .24$, $\varepsilon = .35$, $p < .001$; power = .99]. There were significant differences between music and speech for musicians at phases: −50, −45, −40, −35, −30, −25%, 0, +5, +10, +15, +20, +25, +30, +35 and +40%; these differences in the non-musicians group were significant only at four phases: −20, −15, −10 and +15%. In general, differences between music and speech are more often visible for musicians than for non-musicians.

To compare the overall perturbation of synchronization accuracy caused by the two distractors, we further computed the absolute value of the difference between asynchrony in the distractor conditions and the asynchrony in the Target only condition, averaged across phases. A similar measure was used in one of our previous studies (Dalla Bella et al. 2013). Note that in the past the performance obtained at phase 0 was taken as a reference to compute this difference. Here, however, because musicians showed an effect of interference at phase 0 with the speech distractors, it was more appropriate to take the performance in the Target only condition as a reference. These measures of absolute asynchrony, computed based on time to threshold and time to peak for music and speech distractors and for musicians and non-musicians are reported in Fig. 4. Greater absolute asynchrony indicates lower accuracy in synchronization to the metronome relative to the Target only condition. Absolute asynchronies were entered in separate 2 (group) × 2 (distractor) mixed-design ANOVAs, considering subjects as the random variable. Group (musicians vs. non-musicians) was the between-subjects factor, and Distractor (music vs. speech) was the within-subjects factor.

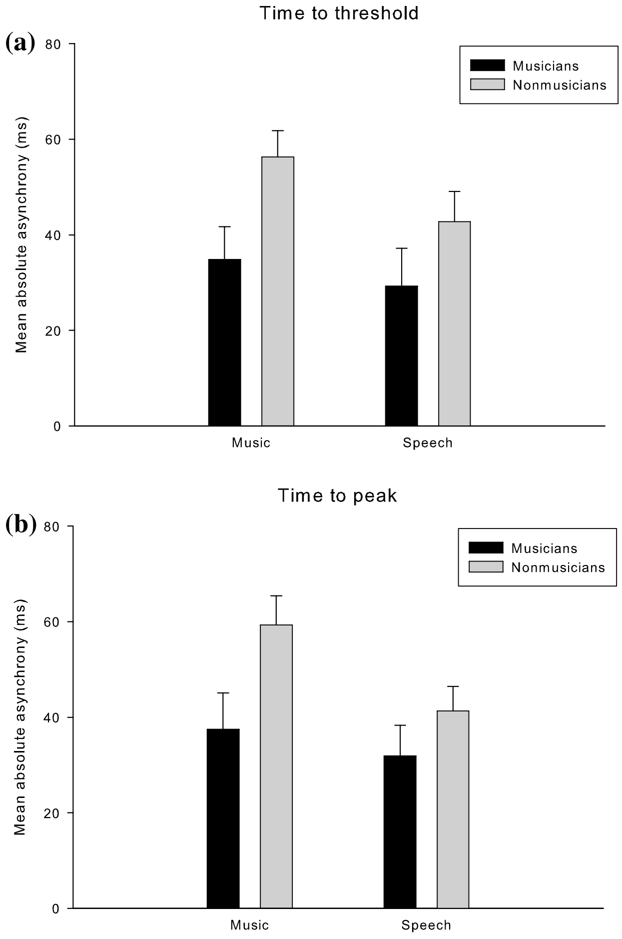

Fig. 4 Mean absolute asynchrony (i.e., synchronization accuracy) based on time to threshold (**a**) and time to peak (**b**) for music and speech distractors obtained for musicians and non-musicians. Error bars are SE of the mean

Absolute asynchrony computed based on time to threshold revealed that music distractors were more disrupting than speech distractors, as revealed by a main effect of Distractor [$F(1, 34) = 9.09$, $\eta_p^2 = .21$, $p < .01$, power = .83]. However, both the main effect of Group and the Group × Distractor interaction failed to reach significance. When absolute asynchrony was calculated based on time to peak, the interfering effect of the distractors was larger for non-musicians than for musicians, but this effect depended on distractor type, showed by a significant Group × Distractor interaction [$F(1, 34) = 7.19$, $\eta_p^2 = .18$, $p < .05$, power = .99]. Musicians were similarly affected by music and speech distractors ($t(13) = 1.00$, $d = .24$, $p = .33$), while non-musicians' synchronization was more disrupted by music than by speech distractors ($t(21) = 6.38$, $d = .59$, $p < .001$, power = .75).

To test whether music and speech distractors affected the force of pressure in the difference condition force peak was analyzed. No group difference was found in terms of force peak in the Target only condition ($M = 1038.1$ N, SD = 680.4 N for musicians and $M = 1096.7$ N, SD = 608.4 N for non-musicians; $t < 1$). In the distractor conditions, the force peak did not vary as a function of the Group or the type of Distractor [$F(1, 34) = .03$, $\eta_p^2 < .01$, $p = .86$ and $F(1.34) = .90$, $\eta_p^2 = .03$, $p = .35$, respectively]; moreover, there was no interaction between Distractor and Group [$F(1, 34) = .03$, $\eta_p^2 < .01$, $p = .87$]. In sum, for both musicians and non-musicians the force of pressure remained constant across all the conditions.

Next, we examined the differences in the time course of force pressure, namely the time interval between the moment when threshold is crossed and peak force is reached. With this analysis we could test whether synchronization accuracy at the moment when the finger pulses reach the maximum force in the presence of a distractor is compensated by a change in the force trajectory. Interestingly, this time interval generally was shorter for musicians than for non-musicians as revealed by a main effect of Group [$F(1, 34) = 8.76$, $\eta_p^2 = .21$, $p < .01$, power = .99]. Because the triple Group × Distractor × Phase interaction was significant [$F(19, 646) = 2.06$, $\eta_p^2 = .06$, $\varepsilon = .37$, $p < .05$, power = .99] two independent 2 (distractor) × 2 (phase) ANOVAs for each Group were performed. Musicians' needed the same amount of time to reach peak force from the moment when they crossed the threshold regardless of the distractor [$F(1, 13) = 1.07$, $\eta_p^2 = .08$, $p = .32$] and of the phase [$F(19, 247) = 1.31$, $\eta_p^2 = .09$, $p = .28$]. The Distractor × Phase did not reach significance, either [$F(19, 247) = 1.50$, $\eta_p^2 = .10$, $p = .23$]. In contrast, the time interval between the moment when a threshold is crossed and peak force is reached was longer for speech than for music distractors for non-musicians [$F(1, 21) = 5.14$, $\eta_p^2 = .20$, $p < .05$; power = .99]. Neither the main effect of Phase [$F(19, 399) = 2.25$, $\eta_p^2 = .10$, $\varepsilon = .22$, $p = .07$] nor the Distractor × Phase interaction [$F(19, 399) = 2.01$, $\eta_p^2 = .09$, $\varepsilon = .31$, $p = .07$] were significant for non-musicians.

We also checked whether non-musicians' performance in the Target only condition correlated with the average level of disruption due to the distractors (as reflected by absolute asynchrony). We found that time to threshold in the Target only condition was highly correlated with the same measure obtained in the presence of a distractor for most of the phases (with music distractors, rs > .41, ps < .05; with speech distractors, rs > .43, ps < .05). A similar relation was found for time to peak (with music distractors, rs > .43, ps < .05; with speech distractors, rs > .42, ps < .05).

Finally, to test the possibility that the familiarity of the musical and speech material affected participants' performance, we examined whether familiarity ratings were associated with the performance in the synchronization tasks for the subgroup of 13 musicians who performed both the pilot experiment and the main experiment. Correlations (Pearson, or Spearman when the distribution was not normal) were computed between musicians' familiarity ratings and average absolute asynchrony results calculated based both on time to threshold and time to maximum force pressure. None of these correlations reached significance (average correlation for music material = .18; for speech material = .12).

Discussion

The goal of the present study was to examine the role of musical expertise in SMS to a metronome in the presence of speech and music distractors. Both music and speech distractors perturbed SMS with a temporally predictable target sequence for both musicians and non-musicians. This effect of interference varied as a function of the relative phase between the target sounds and beat/stress onsets. In addition, music distractors were typically more disrupting than speech distractors. Interference was the greatest at relative phases around 100 ms, an effect particularly visible in non-musicians. In addition, interference was modulated by musical training and depended on the specific measure of the asynchrony between the movement and the target sounds (i.e., based on a force pressure threshold or on maximum force peak). In general, musicians are more resistant to interference than non-musicians, by showing greater synchronization accuracy (i.e., absolute tap-target sounds asynchrony) in the presence of both music and speech distractors across phases. This is observed with both measures of asynchrony (time to threshold and time to maximum force) based on the finger pressure response. However, a difference between music and speech in the two groups emerged when considering the time needed to reach peak force. Musicians ignored music distractors more effectively than non-musicians did; as a result, musicians synchronized the moment of force peak to the metronome sounds with the same level of accuracy for both distractors. In contrast, non-musicians' synchronization at the force peaks was more disrupted by music than by speech distractors. Finally, note that musicians and non-musicians did not differ in their motor response when they synchronized to the metronome in the absence of a distractor and in general in their response force. Thus, the aforementioned differences between the groups cannot be merely ascribed to general differences in synchronization skills, or response strategy.

Rhythmic movement is more strongly perturbed by music than by speech, confirming what found in a previous study (Dalla Bella et al. 2013; Exp. 1) with a standard tapping task. This effect was associated to differences between music and speech in terms of temporal regularity and pitch characteristics. Different hypotheses can account for the aforementioned effect of perturbation by a rhythmic distractor. The attraction of taps to periodic distractors can be mediated by correction mechanisms (e.g., phase correction), depending on the temporal integration of target and distractor sounds (Repp 2003, 2004). When the distractors and the target stimuli occur within a fixed temporal window of approximately 120 ms, they tend to be perceptually integrated, and thereby they can perturb error correction and disturb synchronization (Repp and Penel 2004). Even though this hypothesis is generally plausible, an additional assumption is needed to account for the observed differences between music and speech. This finding can be explained by assuming different temporal windows of integration depending on the stimulus domain, or on its features such as temporal regularity. It is well known that the temporal window of integration can vary as a function of the modality, sensory reliability, and the task (Colonius and Diederich 2010; Mégevand et al. 2013; Occelli et al. 2011; Roy et al. 2016). Yet, to our knowledge, to date there is no evidence in favor of an effect of the domain or temporal variability on the integration window. Thus, this hypothesis needs further investigation.

An alternative hypothesis is that tapping to a metronome requires attending to its periodic sounds; a process mediated by the entrainment of internal neurocognitive oscillations to the target stimuli (Dynamical Attending Theory, DAT; Fujioka et al. 2012; Jones 2009, 2010; Nozaradan et al. 2013). Distractor sequences compete with the target sequences in attracting listeners' attention; this effect depends on the stimulus temporal regularity (Large and Palmer 2002; Jones 2009). A stimulus with a complex and regular rhythmic pattern (e.g., musical meter) typically entrains internal oscillations more efficiently than a stimulus with lower temporal regularity (e.g., speech). Thus, music distractors, due to their regular temporal structure, may be particularly well-suited to entrain internal oscillators, and to attract tap movements.

A novel important finding of this study was to show that musical training modulates the interference of music and speech distractors on SMS. Why are musicians able to ignore music distractors more effectively than non-musicians? Moreover, why speech distractors lead to anticipating the beat regardless of the phase separation between the distractors and metronome sounds, and disturb tapping at phase 0 for musicians only? A possible explanation is that extensive musical training may make musicians more apt to ignore musical stimuli when presented in the context of a complex auditory scene. This skill might have ecological relevance for example in ensemble performance, when a musician has to avoid the interference of other instruments while still focusing on the conductor. In parallel, there is evidence that during ensemble performance musicians can selectively focus their attention on relevant information while relatively ignoring other information (Keller 2008; Keller et al. 2014). This form of divided attention, referred to as "prioritized integrative attending" assists individuals in integrating their own actions with others' actions, while maintaining autonomous control of their own movements. This possibility is compatible with the aforementioned DAT (Jones 2009), whereby the simultaneous segregation and integration of information from separate sources of musical rhythm required in ensemble performance is mediated by the dynamic allocation of attentional resources to multiple periodicities associated with metric structure (London 2012). Moreover, musicians are often trained to pay attention to a particular metrical level (e.g., corresponding to the quarter note) during ensemble performance. Thus, they might focus more attention than non-musicians to a metronome. This possibility is supported by neurophysiological evidence showing enhanced induced gamma band responses to different kind of tones (phase-locking and size) in musicians as compared to non-musicians; this brain response is associated with attention, expectation, memory and multisensory integration (Trainor et al. 2009). Thus, musical training develops top-down and attentional processes, and enhances coordination between body movements and auditory perception (Zatorre et al. 2007).

An additional outcome of musical training is that it increases sensitivity to properties of the auditory signal (e.g., rhythmical or/and melodic features) in a different domain, such as language (e.g., Besson et al. 2011; Bolger et al. 2014; Flaugnacco et al. 2015). These transfer effects of musical training to another domain are compatible with the OPERA hypothesis (Patel 2011). According to this hypothesis, musical training enhances the neural encoding of speech because of the partial sharing of auditory processing networks with music, when five conditions are met: overlap, precision, emotion, repetition, and attention. These conditions drive adaptive neural plasticity in auditory processing networks, leading to higher precision in processing speech signals.

It is known that musicians are better at processing rhythmic aspects of auditory stimuli (e.g., beat regularity) than

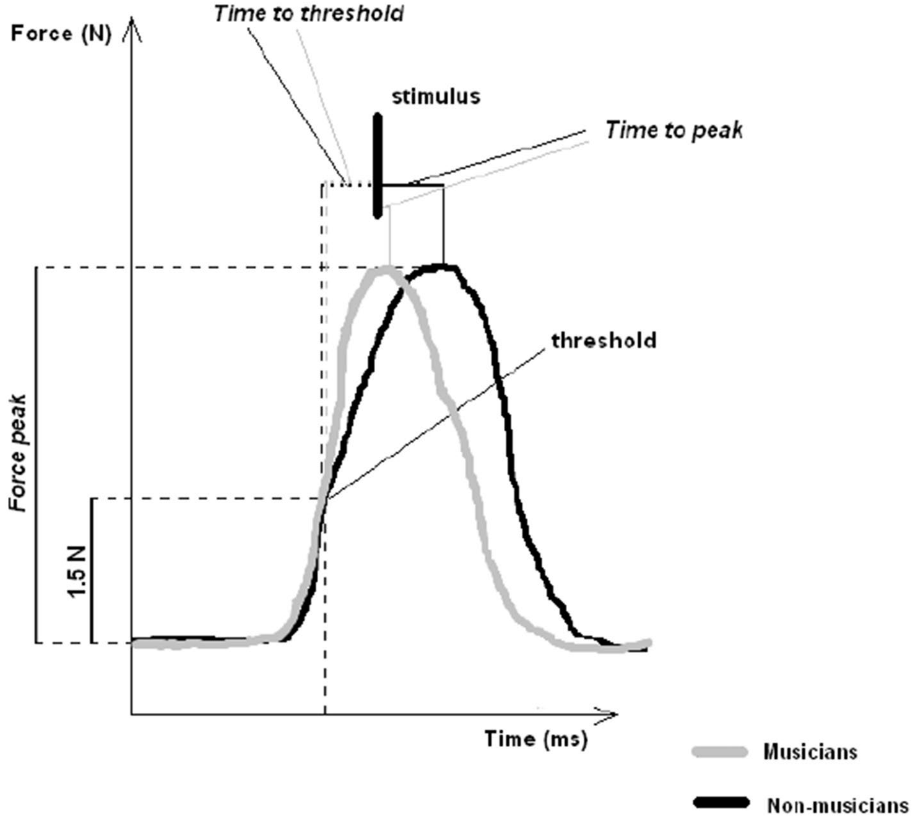

Fig. 5 An example of time courses of force pressure for musicians (gray line) and non-musicians (black line) obtained in one of the trial of speech condition

non-musicians (Kraus and Chandrasekaran 2010; Thompson et al. 2015), there is evidence that this advantage extends to speech processing (Slater and Kraus 2015). Note that music and speech distractors in our study differed in terms of temporal variability (i.e., only musical beats are perfectly isochronous). It is possible that musicians, due to fine-tuned rhythmic processing, were particularly sensitive to the slight deviations from isochrony present in the spoken sequences. Thus, extracting speech stresses may have required more attentional resources for musicians than for non-musicians because speech stimuli exhibited less regular but still rhythmical structure. This was visible in musicians' lower accuracy than non-musicians (i.e., they tapped earlier than the target sounds with speech distractors) even when targets and speech stresses were exactly aligned. Note that, in spite of this fact, musicians could still ignore speech distractors regardless of the phase separation.

Finally, an intriguing finding is that the interference caused by the music distractor varies depending on the measure of synchronization accuracy (based on time to threshold or time to peak force). The effect of musical training is visible when considering the time to peak force during SMS, but not in terms of final force pressure. This indicates that musicians and non-musicians used different force trajectories to reach the same pressure force at the time of synchronization, thus pointing to differences in movement dynamics between the two groups. In general, the moment of maximum force is reached earlier by musicians than by non-musicians as can be seen in Fig. 5. Thus, even though synchronization to the target sounds when crossing a threshold of force pressure is comparable in musicians and non-musicians, musicians are more rapid in achieving peak force. This may indicate more efficient movement preparation and execution in musicians irrespective of the distractor type. Yet, non-musicians need less time to reach force peak with music than with speech distractors. This finding suggests more efficient movement preparation and execution in the general population in the musical context. Hence, it is not excluded that the aforementioned differences linked to musical training might not just be the outcome of differences in beat perception (for a link between neural encoding of auditory rhythms and beat synchronization see Tierney and Kraus 2014; Woodruff Carr et al. 2014), but also in motor planning/sensorimotor integration. Note that different strategies in realizing a simple SMS task are not unusual, and can even be found in primates. Such differences in performing a synchronization-continuation task (vs. a phasic tapping task or a continuous cyclic movement) were observed in monkeys, given advantage of the former as a more effective, explicit timing behavior (Donnet et al. 2014). However, this issue is still open to future studies.

To sum up, musical training influences the ability to ignore a rhythmic distractor during a SMS task. This effect is dependent on the domain of the rhythmic stimulus. Even though music perturbs the synchronization with a metronome more than speech, this effect of interference is modulated by musical training. Musicians are generally more resistant to the deleterious effect of a music distractor than non-musicians. This effect may be due to differences between musicians and non-musicians in the allocation of attentional resources while tracking the beat, in the fine-tuning of beat perception processes, or in motor planning/preparation.

The study is not without limitations. For example, musicians are likely to be in general more exposed to music than non-musicians. Thus, it cannot be excluded that greater exposure to the stimulus material or to music of the same musical genre may explain the observed differences between musicians and non-musicians, rather than musical training. However, it is unlikely that musicians were more exposed to the specific musical stimuli used in this study, as shown by their comparable familiarity ratings, treated as an indirect measure of exposure. The main reason why we chose highly familiar stimuli which belong to a general repertoire, shared by both musicians and non-musicians, was to try to equalize as much as possible the familiarity of these stimuli across the two groups. The excerpts selected for the experiment are highly familiar, but not included in the standard repertoire (classical music), usually trained as part of formal musical education. Note also that differences in familiarity with the musical material were not associated with differences in the effect of the distractors, as tested in a subgroup of musicians. In sum, familiarity as a result of exposure to the specific musical material is unlikely to be responsible for the observed differences between musicians and non-musicians. Nevertheless, it is still possible that musicians were more exposed to the specific musical stimuli without leading to a change in familiarity ratings, or to music of the same musical genre. Hence, a further study should be conducted in the future, in which the effects of music and speech distractors are corroborated with unfamiliar material.

Acknowledgements We thank the Editor and two anonymous Reviewers for their helpful comments on the manuscript. The study was supported by a Junior grant from the Institute Universitaire de France to SDB, and a Marie Curie ITN grant from the European Union (7th Framework Programme) to AB.

References

Abercrombie D (1967) Elements of general phonetics. Aldine, Chicago

Arvaniti E (2009) Rhythm, timing and the timing of rhythm. Phonetica 66:46–63

Besson M, Chobert J, Marie C (2011) Transfer of training between music and speech: common processing, attention, and memory. Front Psychol 2:94

Białuńska A, Dalla Bella S, Jaśkowski P (2011) Increasing stimulus intensity does not affect sensorimotor synchronization. Psychol Res 75(1):43–53

Bigand E (1997) Perceiving musical stability: the effect of tonal structure, rhythm, and musical expertise. J Exp Psychol Hum Percept Perform 23:808–822

Bigand E (2003) More about the musical expertise of musically untrained listeners. Ann N Y Acad Sci 999:304–312

Bigand E, Parncutt R (1999) Perceiving musical tension in long chord sequences. Psychol Res 62:237–254

Bolger D, Coull JT, Schön D (2014) Metrical rhythm implicitly orients attention in time as indexed by improved target detection and left inferior parietal activation. J Cognitive Neurosci 26(3):593–605

Brzechwa J (1980) Brzechwa dzieciom. Warszawa, Nasza Księgarnia

Colonius H, Diederich A (2010) The optimal time window of visual-auditory integration: a reaction time analysis. Front Integr Neurosci 4:11–12

Cummins F (2009) Rhythm as an affordance for the entrainment of movement. Phonetica 66:15–28

Dalla Bella S, Białuńska A, Sowiński J (2013) Why movement is captured by music, but less by speech: role of temporal regularity. PLoS One 8(8):e71945

Dauer RM (1983) Stress-timing and syllable-timing reanalyzed. J Phon 11:51–62

Donnet S, Bartolo R, Fernandes JM, Cunha JP, Prado L, Merchant H (2014) Monkeys time their pauses of movement and not their movement-kinematics during a synchronization-continuation rhythmic task. J Neurophysiol 111:2138–2149

Drake C, Bertrand D (2001) The quest for universals in temporal processes in music. Ann N Y Acad Sci 930:17–27

Drake C, Jones MR, Baruch C (2000) The development of rhythmic attending in auditory sequences: attunement, reference period, focal attending. Cognition 77:251–288

Faul F, Erdfelder E, Lang A-G, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res 39:175–191

Flaugnacco E, Lopez L, Terribili C, Montico M, Zoia S, Schön D (2015) Music training increases phonological awareness and reading skills in developmental dyslexia: a randomized control trial. PLoS One 10(9):e0138715

Fujioka T, Trainor LJ, Large EW, Ross B (2012) Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. J Neurosci 32:1791–1802

Hallam S, Creech A, Varvarigou M, McQueen H, Gaunt H (2014) Does active engagement in community music promote enhanced quality of life in older people? Art Health 6(2):101–116

Hove MJ, Fairhurst MT, Kotz SA, Keller PE (2013) Synchronizing with auditory and visual rhythms: an fMRI assessment of modality differences and modality appropriateness. Neuroimage 67:313–321

Jaśkowski P, Włodarczyk D (2006) Task modulation of the effect of brightness on reaction time and response force. Int J Psychophysiol 61:98–112

Jones MR (2009) Musical time. In: Hallam S, Cross I, Thaut M (eds) The Oxford handbook of music psychology. Oxford University, Oxford, pp 81–92

Jones MR (2010) Attending to sound patterns and the role of entrainment. In: Nobre C, Coull JT (eds) Attention and time. Oxford University, New York, pp 317–330

Jones MR, Pfordresher PQ (1997) Tracking melodic events using joint accent structure. Can J Exp Psychol 51:271–291

Keller PE (2008) Joint action in music performance. In: Morganti F, Carassa A, Riva G (eds) Enacting intersubjectivity: a cognitive and social perspective to the study of interactions. IOS Press, Amsterdam, pp 205–221

Keller PE, Novembre G, Hove MJ (2014) Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. Philos T R Soc B 369:20130394

Kraus N, Chandrasekaran B (2010) Music training for the development of auditory skills. Nat Rev Neurosci 11:599–605

Large EW, Jones MR (1999) The dynamics of attending: how people track time-varying events. Psychol Rev 106(1):119–159

Large EW, Palmer C (2002) Perceiving temporal regularity in music. Cogn Sci 26:1–37

Lehiste I (1977) Isochrony reconsidered. J Phon 5:253–263

Liberman M, Prince A (1977) On stress and linguistic rhythm. Linguist Inq 8:249–336

Lidji P, Palmer C, Peretz I, Morningstar M (2011) Listeners feel the beat: entrainment to English and French speech rhythms. Psychon Bull Rev 18:1035–1041

London J (2012) Hearing in time. Psychological aspects of musical meter. New York, Oxford University

Mégevand P, Molholm S, Nayak A, Foxe JJ (2013) Recalibration of the multisensory temporal window of integration results from changing task demands. PLoS One 8(8):e71608

Nozaradan S, Zerouali Y, Peretz I, Mouraux A (2013) Capturing with EEG the neural entrainment and coupling underlying sensorimotor synchronization to the beat. Cereb Cortex 25(3):736–747

Occelli V, Spence C, Zampini M (2011) Audiotactile interactions in temporal perception. Psychon Bull Rev 18:429–454

Patel AD (2008) Music, language, and the brain. Oxford University, New York

Patel AD (2011) Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. Front Psychol 2(142):142

Patel AD, Daniele JR (2003) An empirical comparison of rhythm in language and music. Cognition 87:B35–B45

Pike KN (1945) The intonation of American English. University of Michigan, Ann Arbor

Repp BH (1998) A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. J Acoust Soc Am 104:1085–1100

Repp BH (2000) Compensation for subliminal timing perturbations in perceptual-motor synchronization. Psychol Res 63:106–128

Repp BH (2001) Phase correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization. J Exp Psych Hum Percept Perform 27:600–621

Repp BH (2002) Automaticity and voluntary control of phase correction following event onset shifts in sensorimotor synchronization. J Exp Psych Hum Percept Perform 28:410–430

Repp BH (2003) Phase attraction in sensorimotor synchronization with auditory sequences: effects of single and periodic distractors on synchronization accuracy. J Exp Psychol Hum Percept Perform 29:290–309

Repp BH (2004) On the nature of phase attraction in sensorimotor synchronization with interleaved auditory sequences. Hum Mov Sci 23:389–413

Repp BH (2005) Sensorimotor synchronization: a review of the tapping literature. Psychon Bull Rev 12(6):969–992

Repp BH (2006) Rate limits of sensorimotor synchronization. Adv Cog Psych 2:163–181

Repp BH (2010) Sensorimotor synchronization and perception of timing: effect of music training and task experience. Hum Mov Sci 29:200–213

Repp BH, Doggett R (2007) Tapping to a very slow beat: a comparison of musicians and non-musicians. Music Percept 24:367–376

Repp BH, Keller PE (2004) Adaptation to tempo changes in sensorimotor synchronization: effects of intention, attention, and awareness. Q J Exp Psychol 57A:499–521

Repp BH, Penel A (2004) Rhythmic movement is attracted more strongly to auditory than to visual rhythms. Psychol Res 68:252–270

Repp BH, Su Y-H (2013) Sensorimotor synchronization: a review of recent research (2006–2012). Psychon Bull Rev 20:403–452

Roy C, Dalla Bella S, Lagarde J (2016) To bridge or not to bridge the multisensory time gap: bimanual coordination to sound and touch with temporal lags. Exp Brain Res. doi:10.1007/s00221-016-4776-4

Selkirk EO (1984) Phonology and syntax: the relation between sound and structure. MIT Press, Cambridge, MA

Slater J, Kraus N (2015) The role of rhythm in perceiving speech in noise: a comparison -of percussionists, vocalists and non-musicians. Cogn Process. doi:10.1007/s10339-015-0740-7

Stevens C, Byron T (2009) Universals in musical processing. In: Hallam S, Cross I, Thaut M (eds) Oxford handbook of music psychology. Oxford University, New York, pp 14–123

Thaut MH, Miller RA, Schauer LM (1998) Multiple synchronization strategies in rhythmic sensorimotor tasks: phase vs. period correction. Biol Cybern 79:241–250

Thompson EC, White-Schwoch T, Tierney A, Kraus N (2015) Beat synchronization across the lifespan: intersection of development and musical experience. PLoS One 10(6):e0128839

Tierney A, Kraus N (2014) The ability to move to a beat is linked to the consistency of neural responses to sound. J Neurosci 33(38):14981–14988

Tillmann B, Bigand E (2001) Global context effect in normal and scrambled musical sequences. J Exp Psychol Hum Percept Perform 27(5):1185–1196

Trainor LJ, Shahin AJ, Roberts LE (2009) Understanding the benefits of musical training: effects on oscillatory brain activity. Ann NY Acad Sci 1169:133–142

Tuwim J (1980) Wiersze dla dzieci. Warszawa, Nasza Księgarnia

Villing RC, Repp BH, Ward TE, Timoney JM (2011) Measuring perceptual centers using the phase correction response. Atten Percept Psychophys 73:1614–1629

Woodruff Carr K, White-Schwoch T, Tierney AT, Straita DL, Kraus N (2014) Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. PNAS 111(40):14559–14564

Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: auditory–motor interactions in music perception and production. Nat Rev Neurosci 8:547–558