Exp Brain Res (2016) 234:3037–3047
DOI 10.1007/s00221-016-4705-6

RESEARCH ARTICLE

Musicians have enhanced audiovisual multisensory binding: experience-dependent effects in the double-flash illusion

Gavin M. Bidelman[1,2]

Received: 9 February 2016 / Accepted: 14 June 2016 / Published online: 22 June 2016

Abstract Musical training is associated with behavioral and neurophysiological enhancements in auditory processing for both musical and nonmusical sounds (e.g., speech). Yet, whether the benefits of musicianship extend beyond enhancements to auditory-specific skills and impact multisensory (e.g., audiovisual) processing has yet to be fully validated. Here, we investigated multisensory integration of auditory and visual information in musicians and nonmusicians using a double-flash illusion, whereby the presentation of multiple auditory stimuli (beeps) concurrent with a *single* visual object (flash) induces an illusory perception of multiple flashes. We parametrically varied the onset asynchrony between auditory and visual events (leads and lags of ±300 ms) to quantify participants' "temporal window" of integration, i.e., stimuli in which auditory and visual cues were fused into a single percept. Results show that musically trained individuals were both faster and more accurate at processing concurrent audiovisual cues than their nonmusician peers; nonmusicians had a higher susceptibility for responding to audiovisual illusions and perceived double flashes over an extended range of onset asynchronies compared to trained musicians. Moreover, temporal window estimates indicated that musicians' windows (<100 ms) were ~2–3× shorter than nonmusicians' (~200 ms), suggesting more refined multisensory integration and audiovisual binding. Collectively, findings indicate a more refined binding of auditory and visual cues in musically trained individuals. We conclude that experience-dependent plasticity of intensive musical experience extends beyond simple listening skills, improving multimodal processing and the integration of multiple sensory systems in a domain-general manner.

Keywords Audiovisual integration · Experience-dependent plasticity · Multisensory facilitation · Musical training · Temporal binding window

Introduction

Our perception and interactions with the external world consist not of isolated sensory events, but rather, a rich combination of multisensory experiences. Indeed, individual sensory systems such as audition regularly integrate and interact with the other modalities (e.g., vision) in service of enhancing perceptual processing. The utility of combining auditory and visual cues is perhaps best exemplified in the case of speech perception where auditory recognition for normal (Sumby and Pollack 1954) and even noise-degraded speech (Erber 1975; Vatikiotis-Bateson et al. 1998) is improved when listeners are provided concurrent visual cues of the talker. While visual information can enhance listening experiences, it can also interact with the auditory modality. Multisensory integration in the audiovisual domain is illustrated by the well-known McGurk effect, where visual speech cues (e.g., seeing a talker's lips) influence the auditory input to create illusory speech percepts (McGurk and MacDonald 1976). Needless to say, multisensory processing represents a ubiquitous operation in everyday human behaviors that shapes our perception for communicative and noncommunicative signals alike.

✉ Gavin M. Bidelman
g.bidelman@memphis.edu

1 Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA

2 School of Communication Sciences and Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN 38152, USA

Given the pervasive nature of multisensory processing in synthesizing the perceptual world, there is growing interest to examine how different disorders and experiential factors might alter this fundamental process. Emerging evidence from behavioral and neurophysiological studies suggests the inability to assimilate information from more than one sense may underlie a series of neurodevelopmental disorders including autism and dyslexia (for review, see Wallace and Stevenson 2014). Under this proposition, the brain's "temporal window" for integrating multiple sensory cues is extended, producing an aberrant binding of multisensory features and deficits in creating single unified percepts (Foss-Feig et al. 2010; Kaganovich et al. 2014; Wallace and Stevenson 2014). Similarly, synaesthetes who display multimodal percepts (e.g., tones inducing color percepts) show altered multisensory integration windows, consistent with their hypersensitive perception and cross-pairing of the senses (Neufeld et al. 2012). While temporal binding is prolonged in cases of neurodevelopmental disorders and rare cases of perceptual phenomena (e.g., synesthesia), whether or not it can be shortened with certain listening experience(s) has yet to be established (e.g., for short-term perceptual learning effects, see Powers et al. 2009).

The current study investigated the hypothesis that a salient form of multimodal experience, musical training, can sharpen audiovisual processing and the temporal binding window for combining multisensory cues. To date, musicians have represented an ideal model to study *auditory* plasticity (Kraus and Chandrasekaran 2010; Herholz and Zatorre 2012; Moreno and Bidelman 2014) given music's intensive demands on listening skills and its ability to generalize (i.e., transfer) to benefit auditory processing in nonmusical domains (e.g., speech perception). However, music production recruits a rich array of brain networks subserving, among other things, an interplay of auditory-, motor-, memory- and visual-related processes (Zatorre and McGill 2005). Musical training not only engages the auditory system, but also involves auditory-motor coordination, audio–visual integration, and requires learning implicit and explicit "rules" that govern musical systems (for a review see, Herholz and Zatorre 2012). Nevertheless, while studies have clearly demonstrated musician advantages in nearly all aspects of auditory processing, evidence that musical experience can enhance multimodal processing (e.g., audiovisual integration) is scarce or conflicting. For example, some reports show improvements in verbal (i.e., auditory), but not visual working memory following formal music training (Brandler and Rammsayer 2003; Ho et al. 2003; Tierney et al. 2008; Parbery-Clark et al. 2009b; Strait et al. 2010), while others show increased working memory performance for musicians independent of modality (George and Coch 2011; Bidelman et al. 2013). In fact, our recent studies have revealed enhanced visuospatial processing in musicians relative to musically naïve individuals (Bidelman et al. 2013). This opens the possibility that the functional benefits of musicianship may not be exclusively auditory in nature.

A growing number of behavioral and neurophysiological studies are beginning to recognize that musical experience might confer multisensory processing advantages in addition to those described in the domain of hearing. In this vein, recent studies have revealed that musicians show more refined multisensory integration for the neural encoding and processing of audiovisual signals for both *speech* and *musical* stimuli (Musacchia et al. 2007; Lee and Noppeney 2011; Paraskevopoulos et al. 2012; Lee and Noppeney 2014). For example, using stimuli in which auditory (speech utterances) and visual (speaker's lips) cues are delivered in a temporally asynchronous manner, several studies have reported higher sensitivity for detecting audiovisual coherence and hence more restricted integration windows in trained musicians (Lee and Noppeney 2011; Paraskevopoulos et al. 2012; Lee and Noppeney 2014). Yet, with few exceptions (Lu et al. 2014), these studies have employed speech and musical stimuli. Consequently, it remains unclear if musicians' alleged improvements in multisensory integration result from domain-general benefits in audiovisual processing, per se, or instead from musicians' well-known superiority in processing linguistically and musically relevant stimuli (Kraus and Chandrasekaran 2010; Bidelman 2013; Moreno and Bidelman 2014).

In the current study, we aimed to determine if musical experience enhances audiovisual processing and the temporal binding window for combining multisensory cues more broadly and in a domain-general manner. The stimulus paradigm consisted of the well-known double-flash illusion (Shams et al. 2000, 2002), whereby the presentation of multiple auditory stimuli (beeps) concurrent with a *single* visual object (flash) induces an illusory perception of multiple flashes. These nonspeech/nonmusic stimuli are composed only of simple visual flashes and auditory beeps and thus contain no lexical–semantic meaning or relation to familiar musical stimuli. By parametrically varying the onset asynchrony between auditory and visual events (leads and lags) we quantified group differences in the "temporal window" for fusing audiovisual perceptual objects. Findings show that musically trained individuals are both faster and more accurate at processing concurrent audiovisual cues than their nonmusician peers and have more refined multisensory temporal binding windows for integrating the auditory and visual senses.

Methods

Participants

Twenty young adults participated in the experiment: 10 musicians (3 male; 7 female) and 10 nonmusicians (1 male; 9 female). All were monolingual speakers of American English. Each participant completed a questionnaire to assess musical background (Wong and Perrachione 2007). Musicians (M) were defined as amateur instrumentalists who had received ≥7 years of continuous private instruction on their principal instrument (mean ± *SD*; 10.6 ± 2.7 yrs), beginning prior to age 14 (10.0 ± 2.3 yrs) (Table 1). Beyond formal private or group lessons, each was currently active in music practice or ensemble engagement. These inclusion criteria are consistent with similar definitions for "musicians" used in previous studies from our lab and others examining the neuroplastic effects of musicianship on perceptual processing (Wong et al. 2007; Parbery-Clark et al. 2009a; Bidelman et al. 2011a, 2014b). Nonmusicians (NM) had no more than 2 years of self-directed music training (0.40 ± 0.84 yrs) and had not received instruction within the past 5 years. All participants showed normal audiometric sensitivity (i.e., pure tone thresholds <25 dB HL at octave frequencies between 500 and 8000 Hz), normal or corrected-to-normal vision and no previous history of neuropsychiatric illnesses. The two groups were also closely matched in age (M: 25.3 ± 4.1 yrs, NM: 23.0 ± 2.3 yrs; $t_{18} = 1.54$, $p = 0.14$), years of formal education (M: 17.1 ± 2.1 yrs, NM: 18.5 ± 1.9 yrs; $t_{18} = 1.52$, $p = 0.14$) and gender balanced ($p = 0.58$, Fisher's exact test). All participants were paid for their time and gave informed consent in compliance with a protocol approved by the Institutional Review Board at the University of Memphis.

Table 1 Musical demographics of participants

Participant	Instrument(s)	Years of music training	Age of onset
Musicians			
M1	Trumpet	8	10
M2	Trombone	9	11
M3	Piano	10	8
M4	Viola	16	6
M5	Trumpet/guitar/piano	12	11
M6	French horn	12	12
M7	Voice/piano	12	10
M8	Voice	8	14
M9	Voice/piano	7	8
M10	Viola	12	10
Mean (SD)		10.6 (2.7)	10.0 (2.3)
Nonmusicians			
NM1	Violin	2	8
NM2	Clarinet	2	7
NM3	–	0	–
NM4	–	0	–
NM5	–	0	–
NM6	–	0	–
NM7	–	0	–
NM8	–	0	–
NM9	–	0	–
NM10	–	0	–
Mean (SD)		0.4 (0.8)	7.5 (0.7)*

* Age of onset statistics for nonmusicians were computed from the two participants with minimal musical training

Stimuli

Stimuli were constructed to replicate the sound-induced double-flash illusion (Shams et al. 2000, 2002; Foss-Feig et al. 2010). In this paradigm, the presentation of multiple auditory stimuli (beeps) concurrent with a *single* visual object (flash) induces an illusory perception of multiple flashes (Shams et al. 2000) (for examples, see: https://shamslab.psych.ucla.edu/demos/). Stimulus onset asynchrony (SOA) between the auditory and visual stimulus pairing can be parametrically varied to either promote or deny the illusory percept. The illusion (i.e., erroneously perceiving two flashes) is higher at shorter SOAs when beeps occur in closer proximity to the flash. The illusion is less likely (i.e., individuals perceive only a single flash) at long SOAs when the auditory and visual objects are well separated in time. A schematic of the stimulus time course is shown in Fig. 1.

On each trial, participants reported the number of flashes they perceived. Each trial was initiated with a fixation cross on the screen. The visual stimulus was a brief (13.33 ms; a single screen refresh) uniform white disk displayed on the center of the screen on a black background, subtending ~4.5° visual angle. In illusory trials, a single flash was accompanied by a pair of auditory beeps, whereas nonillusory trials actually contained two flashes and two beeps. The auditory stimulus consisted of a 3.5 kHz pure tone of 7-ms duration including 3 ms of onset/offset ramping (Shams et al. 2002). In illusory (single flash) trials, two beeps were presented with varying SOA relative to the single flash. We parametrically varied the SOA between beeps and the single flash from −300 and +300 ms (cf. Foss-Feig et al. 2010) (see Fig. 1). This allowed us to quantify the temporal spacing by which listeners bind auditory and visual cues (i.e., report the illusory percept) and compare the temporal window for audiovisual integration between groups. The onset of one beep always coincided with the onset of the single flash.

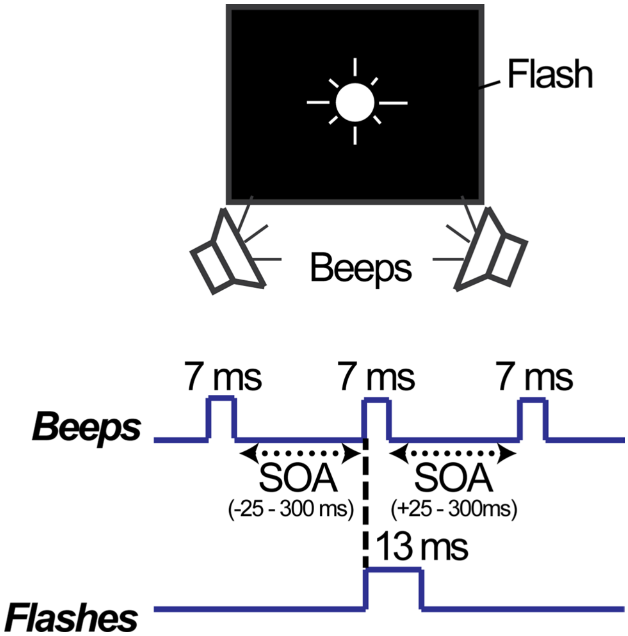

Fig. 1 Task schematic for the double-flash illusion. *Flashes* (13.33 ms *white disks*) were presented on the computer screen concurrent with auditory beeps (7 ms, 3.5 kHz tone) delivered via headphones (*top*). Single trial time course (*bottom*). A single beep was always presented simultaneous with the onset of the flash. A second beep was then presented either before (negative SOAs) or after (positive SOAs) the first. SOAs ranged from ±300 ms relative to the single flash. Despite seeing only a single flash, listeners report perceiving two visual flashes indicating that auditory cues modulate the visual percept. The strength of this double-flash illusion varies with the proximity of the second beep (i.e., SOA)

However, the second beep was either delayed (+300, +200, +150, +100, +50, +25 ms) or advanced (−300, −200, −100, −50, −25 ms) relative to the flash. Both illusory (1F/2B) and nonillusory (2F/2B) trials used these same SOAs (randomly ordered). A total of 30 trials were run for each of these SOA conditions, spread across three blocks. Thus, in aggregate, there were a total of 330 illusory (1F/2B) and 330 illusory (2F/2B) SOA trials. Interleaving illusory and nonillusory conditions also helps to minimize response bias effects in the flash–beep task (Mishra et al. 2007). In addition, trials containing only a single flash and one beep (i.e., 1F/1B) were intermixed with the SOA trials. 1F/1B trials were included as control catch trials and were dispersed randomly throughout the task. Nonillusory trials allowed us to estimate participants' response bias as these trials do not evoke a perceptual illusion and are clearly perceived as having one (1F/1B) or two (2F/2B) flashes, respectively. Illusory (1F/2B) and nonillusory (2F/2B or 1F/1B) conditions were interleaved and trial order was randomized throughout each block. In total, participants performed 690 trials of the task (= 23 stimuli*30 trials).

Procedure

Listeners were seated in a double-walled sound attenuating chamber (Industrial Acoustics, Inc.) ~90 cm from a computer monitor. Stimulus delivery and responses data collection was controlled by E-prime® (Psychological Software Tools, Inc.). Visual stimuli were presented as white flashes on a black background via computer monitor (Samsung SyncMaster S24B350HL; nominal 75 Hz refresh rate). Auditory stimuli were presented binaurally using high-fidelity circumaural headphones (Sennheiser HD 280 Pro) at comfortable level (80 dB SPL). On each trial of the task, listeners indicated via button press whether they perceived "1" or "2" *flashes*. Participants were aware that trials would also contain auditory stimuli but were instructed to make their response based solely on their perception of the visual stimulus. They were encouraged to respond as accurately and quickly as possible. Both response accuracy and reaction time (RT) were recorded for each stimulus condition. Participants were provided a break after each of the three blocks to avoid fatigue.

Data analysis

Behavioral data (% and RT)

For each SOA per subject, we computed the mean percentage of trials two flashes reported. For 1F/2B presentations (illusory trials), higher percentages indicate that listeners erroneously perceived two flashes when only one was presented (i.e., the illusion). Tracing the presence of the double-flash illusion across SOAs allowed us to examine the temporal characteristics of multisensory integration and the audiovisual synchrony needed to bind auditory and visual cues. RTs were also computed per condition for each participant, calculated as the median response time between the end of stimulus presentation and execution of the response button press.

Dependent measures (% two flashes reported; RTs) were analyzed using a two-way mixed model ANOVA with fixed effects of group as the between-subjects factor and SOA as the within-subjects factor. Subjects were modeled as a random effect. Following this omnibus analysis, post hoc multiple comparisons were employed; pairwise contrasts were adjusted using Tukey–Kramer corrections to control type I error inflation. Unless otherwise noted, the alpha level was set at $\alpha = 0.05$ for all statistical tests.

Temporal window quantification

We aimed to characterize the temporal extent required to perceive the double-flash illusion—the temporal binding window for audiovisual integration. To this end, we

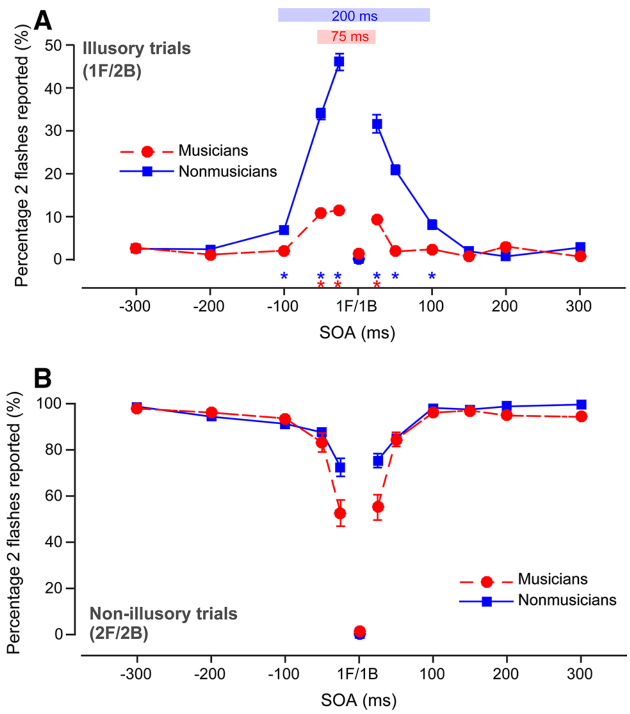

Fig. 2 Musicians show lower susceptibility for perceiving the illusory double-flash percept. **a** Illusory trials; **b** nonillusory trials. **a** Regardless of group, psychometric functions reveal the illusion was strongest for short SOAs and progressively weakened with increasing asynchrony. Musicians show far less susceptibility to the illusion, demonstrating lower incidence of perceiving two illusory flashes, particularly in the shortest SOAs. *Stars* denote SOAs which show significant ($p < 0.001$) increase in two-flash responses relative to the 1F/1B control condition. *Solid bars* indicate the breadth of each group's temporal window estimated from their psychometric functions (see text for details). **b** As in **a** but for nonillusory trials. *Error bars* = ±1 s.e.m.; *$p < 0.001$

measured the width of each participant's temporal window via two methods. In the first approach following Foss-Feig et al. (2010), we quantified the window as the contiguous span of consecutive 1F/2B SOAs where two-flash reports were significantly greater than the 1F/1B nonillusory condition (Foss-Feig et al. 2010). Significance was assessed via paired samples *t* tests (i.e., each SOA verses the 1F/1B condition) using a stringent criterion ($p = 0.001$). Limiting the temporal window to *contiguous* SOAs that were highly significant relative to the 1F/1B control condition further controls the family-wise error rate (Foss-Feig et al. 2010).

In a complementary approach, the width of the temporal window was quantified via the psychometric functions (i.e., Fig. 2a) by computing its integral (between −300 and +300 ms (Lee and Noppeney 2011, 2014). Considering area under the curve accounts for possible differences not only in the width of the psychometric functions but also in their height, reflecting the propensity to perceive double-flash illusion. Group differences were then assessed via an independent samples *t* test contrasting the temporal window between groups. As there is no standard for measuring the temporal window, employing both measures allowed us to generalize our findings more broadly and ensure that group differences in audiovisual processing were not idiosyncratic to the specific choice of temporal window metric.

Results

Behavioral data (%)

The proportion of two-flash reports for each SOA and group is shown for illusory and non illusory trials in Fig. 2, panels A and B, respectively. Higher proportions of reporting the presence of two flashes are indicative of a greater strength or susceptibility to the illusion. Consistent with previous reports (Foss-Feig et al. 2010; Neufeld et al. 2012), both groups showed a similar pattern of responses where the illusion was strong for short SOAs (±25 ms), progressively weakened with increasing asynchrony, and was absent for the longest intervals (±300 ms). Yet across groups, musicians showed far less susceptibility to the illusion, demonstrating lower incidence of perceiving two illusory flashes, particularly in the shortest SOAs where the effect is generally strongest. These observations were confirmed with a two-way ANOVA, which revealed a significant group × SOA interaction [$F_{11, 198} = 113.86$, $p < 0.001$]. Follow-up Tukey–Kramer contrasts revealed that musicians reported fewer illusory double flashes for short SOAs (±100, ±50, ±25 ms; all *p*s < 0.05). These findings reveal that musicians showed a lower propensity (i.e., susceptibility) for the double-flash illusion and more accurately parsed audiovisual cues.

Differences between musicians and nonmusicians could result from group-specific response biases, e.g., if nonmusicians had a higher tendency to report "two flashes." To rule out this possibility, we analyzed performance on the 1F1/B control trials, which should be perceived as a single flash. Higher percentages in this condition would indicate increased responses bias for reporting two visual objects. We found that response bias was minimal (<2 %) for both groups (M = 1.4 ± 1.1 %; NM = 0 ± 0 %), meaning that listeners rarely reported the illusion and perceived only a single flash during 1F/1B trials. Furthermore, while there was a group × SOA interaction for nonillusory trials which contained two flashes and two beeps [$F_{1, 198} = 8.04$, $p < 0.001$] (Fig. 2b), this effect was driven by nonmusicians having more accurate identification only in the ±25 ms condition; all other nonillusory SOA conditions failed to show a group effect. Moreover, pooling the nonillusory SOAs (i.e., ±300, ±200, ±150, ±100, ±50,

±25 ms), we found musicians and nonmusicians on the whole, did not differ in their behavioral report of veridical 2F/2B trials ($t_{18} = -0.85$, $p = 0.41$). Collectively, these analyses help confirm that the observed group effects (Fig. 2a) do not result from one group being more likely to respond "two flashes" per se, but rather, from differences in the perception of the illusion between musicians and nonmusicians.

Group differences in the temporal binding window

The width of participants' multisensory binding window was first quantified from each groups' psychometric function (i.e., identification curves shown in Fig. 2a) as a significant increase in the proportion of trials in which two flashes were perceived (i.e., illusory percepts) in the 1F/2B conditions compared to the nonillusory single flash–beep condition (Foss-Feig et al. 2010). Significance was determined for each group using paired samples *t* test contrasting each groups' responses against the 1F/1B control condition. Applying this procedure, the breadth of the temporal window was defined for each group as the extent of contiguous SOAs showing a reliable difference (i.e., illusory percept). Group differences in the width of the temporal window are shown in Fig. 2a as stars, representing significant SOAs, and by horizontal bars (top). Results showed that nonmusicians' temporal window was ~2–3× longer than musicians (M: 75 ms, NM: 200 ms), indicating that the binding of auditory and visual cues is considerably more refined in musically trained individuals.

To further validate these findings and conduct direct comparisons of audiovisual binding between groups, we measured the temporal window using a second method which considers the area (integral) under the psychometric functions (Lee and Noppeney 2011, 2014). The area under the curve accounts for possible differences in not only the width of the psychometric functions but also their height and thus the propensity to perceive double-flash illusion. We found that the area of musicians' temporal window was considerably (3–4×) narrower than in nonmusicians [$t_{18} = -16.09$, $p < 0.0001$] (Fig. 3). Smaller breadth and height of the psychometric curve further indicates that musicians not only had a more selective temporal window but also showed less susceptibility to the illusion than their nonmusician peers. These findings further bolster the notion that musicians have enhanced multisensory integration and are better able to parse rapid and visual auditory cues compared to nonmusicians. The fact that we find high consistency between the two measures also suggests that the observed group differences in audiovisual processing are not idiosyncratic to the specific choice of metric to quantify the temporal window.

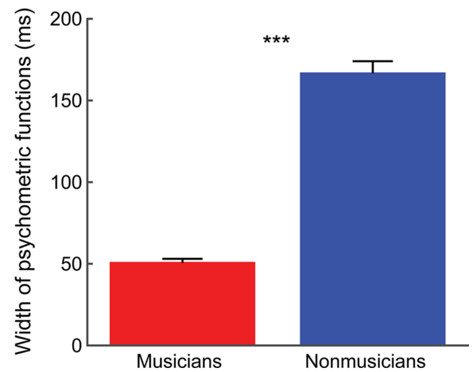

Fig. 3 Musicians have more acute temporal windows for multisensory integration. Temporal windows were estimated as the area under the psychometric curves (i.e., identification functions shown in Fig. 2a) between ± 300 ms (Lee and Noppeney 2014). This integrand reveals differences not only in the width of the psychometric functions but also in their height and thus the propensity to perceive double-flash illusion. Musicians' temporal binding window is 2–3× narrower than nonmusicians indicating a more refined binding of auditory and visual information. *Error bars* = ±1 s.e.m.; ***$p < 0.001$

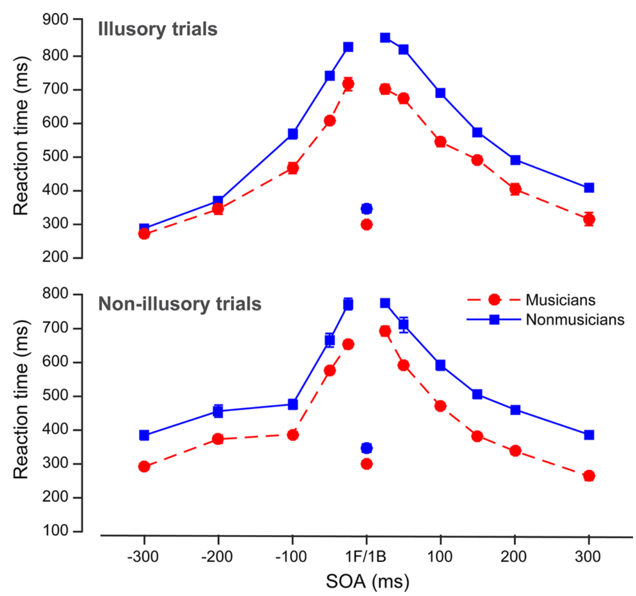

Fig. 4 Reaction times by group. Across the board for both illusory (**a**) and nonillusory (**b**) trials, musicians show faster decisions than nonmusicians when judging audiovisual stimuli. Musicians are not only more accurate at processing concurrent audiovisual cues (e.g., Fig. 2) but on average, respond ~20 % faster than nonmusicians. *Error bars* = ±1 s.e.m

Reaction times (RTs)

Group reaction times across SOAs are shown in Fig. 4 for illusory and nonillusory trials, panels A and B, respectively. Pooled across conditions, musicians RTs were ~20 % faster than nonmusicians' responses (M: 467 ± 26 ms; NM:

565 ± 24 ms). This was confirmed by an ANOVA which revealed a significant group × SOA interaction on behavioral RTs [$F_{11, 198} = 18.31$, $p < 0.001$]. Follow-up contrasts revealed that musicians were faster at making their response than nonmusicians for the majority of SOAs (all but −300 and −200 ms). A nearly identical pattern of results was found for nonillusory trials (Fig. 4b) [group × SOA interaction: $F_{11, 198} = 6.85$, $p < 0.001$], where musicians showed faster behavioral responses across the board. Collectively, these findings indicate that musically trained participants were not only more accurate at processing concurrent audiovisual cues than nonmusicians but considerably faster at judging the composition of audiovisual stimuli.

Discussion

In the present study, we measured multisensory integration in musicians and nonmusicians via the double-flash illusion (Shams et al. 2000; Foss et al. 2007), a task requiring the perceptual binding of temporally offset auditory and visual cues. Collectively, these findings indicate that musically trained individuals (1) are faster and more accurate at processing concurrent audiovisual objects than their nonmusician peers and (2) show more refined (~2–3× shorter) temporal windows for multisensory integration and audiovisual binding. These findings reveal that experience-dependent plasticity of intensive musical training extends beyond the auditory domain and improves the integration of information from multiple sensory systems (audition and vision).

Domain-general benefits of music-related plasticity

The present data reveal that a salient form of auditory experience, musical training, extends beyond simple auditory processing benefits to enhance multisensory integration. These results extend prior work demonstrating auditory-specific enhancements in musicians (Bidelman and Krishnan 2010; Kraus and Chandrasekaran 2010; Strait et al. 2010; Bidelman et al. 2014a) by revealing a multimodal component to music's benefits on brain function. They further extend recent work on musicianship and multisensory integration for *speech* and *musical* stimuli (Musacchia et al. 2007; Lee and Noppeney 2011, 2014; Paraskevopoulos et al. 2012) to domain-general stimuli.

To date, studies have established that protracted musical training improves the neurobiological processing (Shahin et al. 2003; Bidelman et al. 2014b) and behavioral control of linguistic stimuli including providing enrichments to speech perception (Chartrand and Belin 2006; Bidelman and Krishnan 2010; Bidelman et al. 2014b; Kraus et al. 2014; Bidelman and Alain 2015), phonological awareness (Anvari et al. 2002; Slevc and Miyake 2006) and second language learning proficiency (Slevc and Miyake 2006; Cooper and Wang 2012). Similarly, musical training is known to improve the neural encoding and mental control of musically relevant stimuli (Shahin et al. 2003; Foster and Zatorre 2010; Bidelman et al. 2011b; Paraskevopoulos et al. 2012). Hence, musicians' more refined multisensory integration observed previously for speech and musical stimuli (Musacchia et al. 2007; Lee and Noppeney 2011, 2014; Paraskevopoulos et al. 2012) may have resulted not from their improved audiovisual processing, *per se*, but rather, from musicians' superiority in speech–language (for reviews, see Besson et al. 2011; Moreno and Bidelman 2014) and musically relevant tasks (e.g., Pantev et al. 2001; Bidelman et al. 2011b). Here, we show that musicians have enhanced audiovisual processing for *nonspeech* and *nonmusical* stimuli. Moreover, musicians' temporal window for binding auditory and visual percepts was ~3–5× (several hundred milliseconds) faster than their nonmusician peers (Fig. 3). These findings therefore extend previous results by revealing a musician enhancement in audiovisual integration in the absence of lexical–semantic meaning or musical familiarly of the stimulus. These data suggest that musical training sharpens the overall tracking and binding of perceptually relevant information from multiple sensory systems, and does so for speech and nonspeech stimuli alike (cf. Lu et al. 2014). This notion is supported by other recent studies which have similarly shown that intensive musical training improves aspects of working memory (Bidelman et al. 2013) and general temporal acuity (Rammsayer et al. 2012) irrespective of sensory modality (i.e., both auditory and visual enhancements).

The current data corroborate neuroimaging studies demonstrating enhancements in nonauditory brain regions (e.g., motor cortex; Elbert et al. 1996), increased auditory-motor coupling (Zatorre et al. 2007) and shorter temporal auditory-motor integration windows in trained musicians (van Vugt and Tillmann 2014). Conceivably, long-term musical rehearsal and production might also act to improve the prediction of both when and what auditory and visual events are likely to occur as it does in the auditory-motor system (Novembre and Keller 2014; van Vugt and Tillmann 2014). As conceived by Novembre and Keller (2014), experience-dependent coupling of perception and actions (or other percepts) might help "*scaffold the human ability to represent complex (structured) actions and entrain to multiple agents* (Novembre and Keller 2014; p.1)." Structured actions and multiagent entrainment is essential in joint musical tasks such as ensemble performance. Coupling an additional sensory modality (e.g., motor activity) during auditory learning has also been shown to increase cortical responses to behaviorally relevant sounds (Lappe et al. 2008; Paraskevopoulos et al. 2012). Thus, in the case of the auditory-motor system, higher multisensory coupling would be advantageous in musical performance as coordination with the motor system could act to increase sensory encoding within the auditory modality. Given that music production also involves intensive visual engagement (e.g., reading notes from a score), it stands to reason that intensive musical experience might similarly tighten coupling between auditory and visual cortices and account for the data herein.

Alternatively, it is plausible that the more refined audiovisual binding seen here in musicians might instead result from an augmentation of more general cognitive mechanisms (e.g., attention or executive control; Moreno and Bidelman 2014) that are known to differ in musically trained individuals (Pallesen et al. 2010; Strait et al. 2010; Strait and Kraus 2011; Zuk et al. 2014). The notion of top-down, attentional/executive regulation of sensory processing has also been highlighted in recent animal (Fritz et al. 2003) and human studies (Myers and Swan 2012), where increased "feedback" can act to enhance or inhibit the activity in stimulus-selective sensory cortices, driven by the engagement of prefrontal control regions. Distributing attention across a wider variety of sensory modalities has also been shown to enhance performance in complex audio–visual tasks (Mishra and Gazzaley 2012). Thus, it is conceivable that if musical training increases and/or enables one to deploy attentional resources more effectively (e.g., Strait et al. 2010; Strait and Kraus 2011)—and possibly across modalities—this could lead to musicians' cross-modal enhancements observed in the present study. On the other hand, short-term improvements in the temporal binding window with (short-term) training are remarkability similar despite drastic alternations in task structure (Powers et al. 2009). This implies that experience-dependent effects in the double-flash illusion are driven by changes in lower level, sensory–perceptual representations rather than cognitive processing (cf. attention) (Powers et al. 2009).

The double-flash illusion requires a behavioral decision on the visual stimulus that must be informed by the perception of a concurrent auditory event. As such, it is often considered a measure of multisensory integration (Mishra et al. 2007; Powers et al. 2009; Foss-Feig et al. 2010). While it is clear that musicians show enhanced audiovisual processing, it remains possible that group differences in the double-flash susceptibility result, at least in part, from enhanced unisensory processing rather than multisensory integration, *per se*. For example, if musicians have an enhanced ability to attend to the auditory modality (cf. "auditory dominant individuals"; Giard and Peronnet 1999), this may allow them to more effectively parse sound from visual stimuli. Under this interpretation, musician's lower susceptibility to the illusion may result from unimodal enhancements in audition (e.g., Bidelman et al. 2011b, 2013). Alternatively, rather than leading to improvements in multisensory processing per se, musical training might enhance temporal processing in each modality separately (e.g., Rammsayer et al. 2012). However, if this were the case, we might have expected more pervasive group differences across the board, resulting in parallel psychometric functions. Instead, we find an interaction in the behavioral pattern (e.g., Fig. 2a) with group differences circumscribed to conditions with only the most rapid SOAs (<100 ms). Moreover, while neuroimaging studies of the double-flash illusion have shown engagement both unisensory (auditory, visual) and polysensory brain areas (Mishra et al. 2007, 2008), it is the latter (i.e., cross-modal interactions) which drive the illusory percept. Nevertheless, future neuroimaging studies are warranted to assess the relative contribution of uni- and multisensory brain mechanisms in musician's shorter temporal windows observed here at the behavioral level.

Experience-dependent changes in multisensory perception

Several previous studies have investigated whether the temporal window can be enlarged or contracted with disorders or learning. For example, using the double-flash illusion, Wallace and colleagues have shown that children with autism spectrum disorder perceive illusory flashes over an wider range of SOAs, suggesting an extended multisensory binding window compared to typically developing children (Foss-Feig et al. 2010; Stevenson et al. 2014; Wallace and Stevenson 2014). Similarly, normal aging seems to increase multisensory integration (Diederich et al. 2008; DeLoss et al. 2013), as evidenced by broader temporal binding window (Laurienti et al. 2006). Presumably, age-related declines in multisensory binding result from a slowing in peripheral sensory processing, resulting in a decrease in the ability to parse multiple sensory representations (Diederich et al. 2008).

While certain disorders might elongate the temporal window (reviewed by Wallace and Stevenson 2014), it is useful to determine whether certain experiences or learning can improve multisensory integration. Indeed, short-term perceptual training (<5 days) in either a combined or unisensory regimen has been shown to narrow the temporal window of multisensory binding (Powers et al. 2009; Stevenson et al. 2013). However, other studies have reported that the sound-induced flash illusion is largely resistance to feedback training (Rosenthal et al. 2009). Equivocal findings across studies might be explained by differences in the longevity of short-term training effects and/or the inclusion feedback during learning (Powers et al. 2009). Indeed, training-related effects in audiovisual integration are not observed with mere passive stimulus exposure and have only been assessed for longevity after one week post-training (Powers et al. 2009). Consequently, the long-term "sticking power" of short-term audiovisual training remains unknown.

The current data demonstrate that certain long-term perceptual–cognitive experiences (i.e., musical training) can refine the audiovisual temporal integration window. It is conceivable that other intense audiovisual experiences might also enhance multisensory processing. With respect to auditory perception and working memory, we have recently shown that certain forms of intensive language experience (i.e., tone-language speaking bilinguals) confer similar enhancements as musical training (Bidelman et al. 2011a, 2013). In this regard, it would be of interest in future studies to directly contrast the degree to which different forms of expertise that draw on audiovisual processing engender multisensory benefits (e.g., musicianship vs. bilingualism).

How then do we account for musicians' higher sensitivity to audiovisual processing? It is conceivable that musicianship changes functional brain organization so as to enhance connectivity between sensory systems that are highly engaged by music rehearsal (i.e., audition, vision, motor). Indeed, enhanced connectivity in musicians has been observed between auditory and motor cortices (Grahn and Rowe 2009), suggesting the potential for increased and/or faster access between sensory modalities with training. Unfortunately, we are aware of no study to date which has examined potential music-induced differences in brain connectivity between primary auditory and visual brain regions that would presumably mediate the double-flash illusion (cf. Paraskevopoulos et al. 2015). However, in musically lay individuals, prior studies have indicated that the likelihood of perceiving the double-flash illusion is highly correlated with white matter connectivity between occipito-parietal regions, the putative ventral/dorsal streams comprising the "what/where" pathways (Kaposvari et al. 2015). This suggests that parallel visual channels play an important role in audiovisual interactions and the temporal binding of disparate cues as required by double-flash percepts (Shams et al. 2000, 2002). It is possible that musicians might show more refined temporal binding of auditory and visual events as we observe behaviorally due to increased functional connectivity between the auditory and visual systems. Future neuroimaging experiments are warranted to test this possibility.

Does musical experience causally relate to more refined audiovisual integration?

To date, experience-dependent benefits of musical training to auditory perceptual and cognitive functions have largely been identified through cross-sectional studies. The current study is no different with regard to this limitation. It remains possible, for example, that musicians self-select to pursue music activities in early life, perhaps due to superiorities in auditory processing or listening skills before they commence training. While self-selection might account for musicians' benefits observed within the *auditory* modality (e.g., Parbery-Clark et al. 2009b; Zendel and Alain 2009; Bidelman and Krishnan 2010; Bidelman et al. 2011a; Moreno and Bidelman 2014), such an explanation seems unlikely to account for the *multisensory (visual) processing* benefits observed here. For example, it would be doubtful that musicians choose to pursue long-term musical training (a largely auditory experience) due to some preexisting enhancement in their *visual* capacity. We argue that protracted musical experience is likely a causal factor in yielding the observed enhancements in audiovisual binding skills bore out of the repeated exposure to and experience with combining auditory (instrument sound) and visual cues (music notation) during music engagement. Nevertheless, future training studies are needed to fully explore the causal relation between musical training, audiovisual processing and other nonauditory perceptual–cognitive abilities.

Broader implications and directions for future studies

Several neuroimaging studies have investigated the less well-studied "fusion illusion," the complement of the double-flash illusion in which a single flash is perceived when two brief flashes are accompanied by a single beep (e.g., Mishra et al. 2007, 2008). Interesting, these studies have suggested that the two illusions (i.e., "fission" vs. "fusion") are supported by different underlying mechanisms; the "fission" (double-flash) illusion examined here seems to depend on an early (90–150 ms) propagation of neural activity in auditory, visual and superior temporal cortices, which occurs prior to the first modulations signaling the "fusion" percept (~180 ms) (Mishra et al. 2007, 2008). In light of potential mechanistic differences between these two audiovisual effects, future studies might compare both illusions between musicians and nonmusicians in order to shed further light on the neural mechanisms underlying experience-dependent changes in audiovisual processing.

Interestingly, emerging evidence suggests that the temporal, multisensory binding window might be prolonged in a handful of diffuse disorders including autism (Foss-Feig et al. 2010), dyslexia and schizophrenia (Wallace and Stevenson 2014). Deficits in audiovisual integration and reduced sensitively to audiovisual asynchrony have also been observed in children with a history of language learning disorders (Kaganovich et al. 2014). In contrast, musical training has been shown to actually enhance both of these perceptual–cognitive traits, i.e., speech–language function (Kraus and Chandrasekaran 2010; Besson et al. 2011; Moreno and Bidelman 2014) and multisensory processing (present study; Musacchia et al. 2007; Lee and Noppeney 2011; Paraskevopoulos et al. 2012; Lee and Noppeney

2014). If multisensory dysregulation proves to be at the core of these higher-level conditions (Wallace and Stevenson 2014), it stands to reason that musical training—which we find to improve multisensory processing—might be used as a rehabilitation strategy in the treatment and management of certain disorders. Future studies could explore the impact of musical engagement in remediating audiovisual deficits characteristic of various neurodevelopmental disorders.

Acknowledgments The author thanks Haley Sanders for assistance in data collection and Amy Fehrenbach for comments on earlier versions of this manuscript. This work was supported in part by a grant from the GRAMMY® Foundation awarded to G.M.B.

References

Anvari SH, Trainor LJ, Woodside J, Levy BA (2002) Relations among musical skills, phonological processing and early reading ability in preschool children. J Exp Child Psychol 83:111–130

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

Bidelman GM (2013) The role of the auditory brainstem in processing musically-relevant pitch. Front Psychol 4:1–13

Bidelman GM, Alain C (2015) Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. J Neurosci 35:1240–1249

Bidelman GM, Krishnan A (2010) Effects of reverberation on brainstem representation of speech in musicians and non-musicians. Brain Res 1355:112–125

Bidelman GM, Gandour JT, Krishnan A (2011a) Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. J Cogn Neurosci 23:425–434

Bidelman GM, Krishnan A, Gandour JT (2011b) Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. Eur J Neurosci 33:530–538

Bidelman GM, Hutka S, Moreno S (2013) Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. PLoS ONE 8:e60676

Bidelman GM, Schug JM, Jennings SG, Bhagat SP (2014a) Psychophysical auditory filter estimates reveal sharper cochlear tuning in musicians. J Acoust Soc Am 136:EL33–EL39

Bidelman GM, Weiss MW, Moreno S, Alain C (2014b) Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. Eur J Neurosci 40:2662–2673

Brandler S, Rammsayer TH (2003) Differences in mental abilities between musicians and nonmusicians. Psychol Music 31:123–138

Chartrand JP, Belin P (2006) Superior voice timbre processing in musicians. Neurosci Lett 405:164–167

Cooper A, Wang Y (2012) The influence of linguistic and musical experience on Cantonese word learning. J Acoust Soc Am 131:4756–4769

DeLoss DJ, Pierce RS, Andersen GJ (2013) Multisensory integration, aging, and the sound-induced flash illusion. Psychol Aging 28:802–812

Diederich A, Colonius H, Schomburg A (2008) Assessing age-related multisensory enhancement with the time-window-of-integration model. Neuropsychologia 46:2556–2562

Elbert T, Pantev C, Wienbruch C, Rockstroh B, Taub E (1996) Increased cortical representation of the fingers of the left hand in string players. Science 270:305–307

Erber NP (1975) Auditory-visual perception of speech. J Speech Hear Disord 40:481–492

Foss AH, Altschuler EL, James KH (2007) Neural correlates of the Pythagorean ratio rules. NeuroReport 18:1521–1525

Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT (2010) An extended multisensory temporal binding window in autism spectrum disorders. Exp Brain Res 203:381–389

Foster NE, Zatorre RJ (2010) A role for the intraparietal sulcus in transforming musical pitch information. Cereb Cortex 20:1350–1359

Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat Neurosci 6:1216–1223

George EM, Coch D (2011) Music training and working memory: an ERP study. Neuropsychologia 49:1083–1094

Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11:473–490

Grahn JA, Rowe JB (2009) Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. J Neurosci 29:7540–7548

Herholz SC, Zatorre RJ (2012) Musical training as a framework for brain plasticity: behavior, function, and structure. Neuron 76:486–502

Ho Y, Cheung M, Chan A (2003) Music training improves verbal but not visual memory: cross sectional and longitudinal explorations in children. Neuropsychology 17:439–450

Kaganovich N, Schumaker J, Leonard LB, Gustafson D, Macias D (2014) Children with a history of SLI show reduced sensitivity to audiovisual temporal asynchrony: an ERP study. J Speech Lang Hear Res 57:1480–1502

Kaposvari P, Csete G, Bognar A, Csibri P, Toth E, Szabo N, Vecsei L, Sary G, Kincses ZT (2015) Audio-visual integration through the parallel visual pathways. Brain Res 1624:71–77

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

Kraus N, Slater J, Thompson EC, Hornickel J, Strait DL, Nicol T, White-Schwoch T (2014) Music enrichment programs improve the neural encoding of speech in at-risk children. J Neurosci 34:11913–11918

Lappe C, Herholz SC, Trainor LJ, Pantev C (2008) Cortical plasticity induced by short-term unimodal and multimodal musical training. J Neurosci 28:9632–9639

Laurienti PJ, Burdette JH, Maldjian JAW, Wallace MT (2006) Enhanced multisensory integration in older adults. Neurobiol Aging 27:1155–1163

Lee H, Noppeney U (2011) Long-term music training tunes how the brain temporally binds signals from multiple senses. Proc Natl Acad Sci USA 108:E1441–E1450

Lee HL, Noppeney U (2014) Music expertise shapes audiovisual temporal integration windows for speech, sinewave speech and music. Front Psychol 5:1–9

Lu Y, Paraskevopoulos E, Herholz SC, Kuchenbuch A, Pantev C (2014) Temporal processing of audiovisual stimuli is enhanced in musicians: evidence from magnetoencephalography (MEG). PLoS ONE 9:e90686

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748

Mishra J, Gazzaley A (2012) Attention distributed across sensory modalities enhances perceptual performance. J Neurosci 32:12294–12302

Mishra J, Martinez A, Sejnowski TJ, Hillyard SA (2007) Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. J Neurosci 27:4120–4131

Mishra J, Martinez A, Hillyard SA (2008) Cortical processes underlying sound-induced flash fusion. Brain Res 1242:102–115

Moreno S, Bidelman GM (2014) Understanding neural plasticity and cognitive benefit through the unique lens of musical training. Hear Res 308:84–97

Musacchia G, Sams M, Skoe E, Kraus N (2007) Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc Natl Acad Sci USA 104:15894–15898

Myers EB, Swan K (2012) Effects of category learning on neural sensitivity to non-native phonetic categories. J Cogn Neurosci 24:1695–1708

Neufeld J, Sinke C, Zedler M, Emrich HM, Szycik GR (2012) Reduced audio–visual integration in synaesthetes indicated by the double-flash illusion. Brain Res 1473:78–86

Novembre G, Keller PE (2014) A conceptual review on action-perception coupling in the musicians' brain: what is it good for? Front Human Neurosci 8:603

Pallesen KJ, Brattico E, Bailey CJ, Korvenoja A, Koivisto J, Gjedde A, Carlson S (2010) Cognitive control in auditory working memory is enhanced in musicians. PLoS ONE 5:e11120

Pantev C, Roberts LE, Schulz M, Engelien A, Ross B (2001) Timbre-specific enhancement of auditory cortical representations in musicians. NeuroReport 12:169–174

Paraskevopoulos E, Kuchenbuch A, Herholz SC, Pantev C (2012) Evidence for training-induced plasticity in multisensory brain structures: an MEG study. PLoS ONE 7:e36534

Paraskevopoulos E, Kraneburg A, Herholz SC, Bamidis PD, Pantev C (2015) Musical expertise is related to altered functional connectivity during audiovisual integration. Proc Natl Acad Sci USA 112:12522–12527

Parbery-Clark A, Skoe E, Kraus N (2009a) Musical experience limits the degradative effects of background noise on the neural processing of sound. J Neurosci 29:14100–14107

Parbery-Clark A, Skoe E, Lam C, Kraus N (2009b) Musician enhancement for speech-in-noise. Ear Hear 30:653–661

Powers AR, Hillock AR, Wallace MT (2009) Perceptual training narrows the temporal window of multisensory binding. J Neurosci 29:12265–12274

Rammsayer T, Buttkus F, Altenmüller E (2012) Musicians do better than non-musicians in both auditory and visual timing tasks. Music Percept 30:85–96

Rosenthal O, Shimojo S, Shams L (2009) Sound-induced flash illusion is resistant to feedback training. Brain Topogr 21:185–192

Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE (2003) Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J Neurosci 23:5545–5552

Shams L, Kamitani Y, Shimojo S (2000) What you see is what you hear. Nature 408:788

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. Cogn Brain Res 14:147–152

Slevc RL, Miyake A (2006) Individual differences in second-language proficiency: does musical ability matter? Psychol Sci 17:675–681

Stevenson RA, Wilson MM, Powers AR, Wallace MT (2013) The effects of visual training on multisensory temporal processing. Exp Brain Res 225:479–489

Stevenson RA, Siemann JK, Schneider BC, Eberly HE, Woynaroski TG, Camarata SM, Wallace MT (2014) Multisensory temporal integration in autism spectrum disorders. J Neurosci 34:691–697

Strait DL, Kraus N (2011) Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. Front Psychol 2:113

Strait DL, Kraus N, Parbery-Clark A, Ashley R (2010) Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. Hear Res 261:22–29

Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. J Acoust Soc Am 26:212–215

Tierney AT, Bergeson TR, Pisoni DB (2008) Effects of early musical experience on auditory sequence memory. Empir Musicol Rev 3:178–186

van Vugt FT, Tillmann B (2014) Thresholds of auditory-motor coupling measured with a simple task in musicians and non-musicians: was the sound simultaneous to the key press? PLoS ONE 9:e87176

Vatikiotis-Bateson E, Eigsti I-M, Yano S, Munhall KG (1998) Eye movement of perceivers during audiovisual speech perception. Percept Psychophys 60:926–940

Wallace MT, Stevenson RA (2014) The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. Neuropsychologia 64C:105–123

Wong PC, Perrachione TK (2007) Learning pitch patterns in lexical identification by native English-speaking adults. Appl Psycholinguist 28:565–585

Wong PC, Skoe E, Russo NM, Dees T, Kraus N (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci 10:420–422

Zatorre R, McGill J (2005) Music, the food of neuroscience? Nature 434:312–315

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

Zendel BR, Alain C (2009) Concurrent sound segregation is enhanced in musicians. J Cogn Neurosci 21:1488–1498

Zuk J, Benjamin C, Kenyon A, Gaab N (2014) Behavioral and neural correlates of executive functioning in musicians and non-musicians. PLoS ONE 9:e99868