Long-term music training modulates the recalibration of audiovisual simultaneity

Crescent Jicol [1] , Michael J. Proulx[1], Frank E Pollick [2] , Karin Petrini [1]

[1] Department of Psychology, University of Bath, United Kingdom
[2] School of Psychology, University of Glasgow, United Kingdom

Corresponding Author:

Crescent Jicol

Department of Computer Science

University of Bath

Claverton Down

Bath

BA2 7AY

United Kingdom

Email: C.Jicol@bath.ac.uk

Heading: Music training effect on audiovisual recalibration

Author contributions: KP designed the study, KP and FEP piloted the study, CJ conducted the experiment, KP and MJP supervised CJ during the experiment conduction, KP and CJ analyzed the data, KP, CJ, FEP and MJP wrote the manuscript.

Abstract

In order to overcome differences in physical transmission time and neural processing, the brain adaptively recalibrates the point of simultaneity between auditory and visual signals by adapting to audiovisual asynchronies. Here, we examine whether the prolonged recalibration process of passively sensed visual and auditory signals is affected by naturally occurring multisensory training known to enhance audiovisual perceptual accuracy. Hence we asked a group of drummers, of non-drummer musicians and of non-musicians to judge the audiovisual simultaneity of musical and non-musical audiovisual events, before and after adaptation with two fixed audiovisual asynchronies. We found that the recalibration for the musicians and drummers was in the opposite direction (sound leading vision) to that of non-musicians (vision leading sound), and change together with both increased music training and increased perceptual accuracy (i.e. ability to detect asynchrony). Our findings demonstrate that long-term musical training reshapes the way humans adaptively recalibrate simultaneity between auditory and visual signals.

Introduction

Due to the difference between the speed of light and that of sound, there exist distance-dependent changes in the times at which visual and auditory stimuli reach the respective sensory receptors (Arnold, Johnston & Nishida, 2005; Noel, Łukowska, Wallace, & Serino, 2016; Spence & Squire, 2003). Moreover, there are differences between the neural processing times of these cues (Alais & Charlile, 2005; King, 2005; Schroeder & Foxe, 2004). Nevertheless, for relatively small temporal differences humans are seldom aware of the asynchrony between these sensory cues thanks to the brain's capacity to shift (recalibrate) the point at which a person perceives their simultaneity (e.g. Di Luca, Machulla, & Ernst, 2009; Harrar & Harris, 2008; Keetels & Vroomen, 2007; Van der Burg, Orchard-Mills, & Alais, 2015; Vatakis, Navarra, Soto-Faraco, & Spence, 2007).

Even short exposure times to audiovisual asynchronous stimuli (circa three minutes) can affect the perceived synchrony of subsequent similar audiovisual stimuli (Fujisaki et al., 2004; Vroomen, Keetels, De Gelder, & Bertelson, 2004). In fact, a study by Van der Burg, Alais, and Cass (2013) showed that recalibration to asynchronous stimuli can occur almost instantaneously, following a single exposure to an asynchronous multisensory event (Simon, Noel, & Wallace, 2017). This suggests that recalibration could be a fast sensory process, rather than a higher-level cognitive process (Van der Burg et al., 2013). However, Rohde and Ernst (2016) showed that asynchronies in visuo-motor tasks, such as delays between a button press and a visual flash (Rohde & Ernst, 2013), can be compensated with training and increased perceptual accuracy (higher ability to detect asynchrony), and are subject to perceived agency (i.e. the prior knowledge that pressing the button is causing the flash to appear, and thus the flash should follow the button press), suggesting that higher-level cognitive processes might actually affect recalibration.

The effect of multisensory training and perceptual accuracy on recalibration has emerged from studies focusing on actively sensed modalities, such as those involving motor action in the recalibration task (Rohde & Ernst, 2013, Rohde, Scheller & Ernst 2014, Rohde & Ernst, 2016), rather than on passively sensed audiovisual modalities (Desantis & Haggard, 2016; Roach, Heron, Whitaker, & McGraw 2010; Vroomen et al., 2004) and on short periods of exposure rather than long and naturally occurring periods of multisensory training (Noel, Niear, Van der Burg, & Wallace, 2017; Simon et al., 2017; Van der Burg et al., 2013). Hence, we do not know whether the changes in recalibration and perceptual accuracy are specific to sensorimotor tasks or if they are a general multisensory mechanism, and whether it can be facilitated by long-term multisensory practice (known to affect brain plasticity as well as perceptual accuracy; Lee & Noppeney, 2011; Petrini et al., 2011).

Musical training is an example of such a rich naturally occurring multisensory activity because playing an instrument requires precise timing and synchronization among motor, visual and auditory information, as well as extensive practice with coordinating these modalities (Lee & Noppeney, 2011; Petrini et al., 2011). Indeed, a large body of research has shown that music expertise enhances audiovisual synchrony perception (Hodges, Hairston & Burdette, 2005; Petrini, Dahl et al., 2009; Proverbio, Attardo, Cozzi, & Zani, 2015; Vatakis & Spence, 2006). For example, studies by Lee and Noppeney (2011) and Petrini et al. (2011) showed that pianists and drummers are more precise than non-musicians when detecting audiovisual asynchrony between visual and auditory cues and differ from non-musicians in the associated neural mechanisms of audiovisual synchrony perception. Moreover, Rohde and Ernst (2013) found that the strength of recalibration depends on this perceptual accuracy, i.e. the more precisely a person can detect asynchrony the smaller their effect of recalibration would be (Van der Burg, Alais, & Cass, 2013; Noel et al., 2016). This could mean that judgements of simultaneity and adaptation to asynchronies are performed by the same mechanism. It is however still unknown whether naturally occurring multisensory training known to enhance audiovisual perceptual accuracy would also affect the recalibration process. If this were the case, then musicians, who have decreased tolerance to audiovisual asynchrony (i.e. have higher perceptual accuracy) should also show decreased recalibration to audiovisual asynchrony. Testing perceptual accuracy and recalibration will also allow us to discern whether these processes are performed by the same mechanism, as Rohde and Ernst (2013) suggest; or if there are two different cognitive processes which are unequally impacted by long-term expertise with multisensory stimuli.

Therefore, here we tested whether long-term music training affects the recalibration process by comparing how perception of simultaneity changes in musicians (drummers and other musicians) and non-musicians before and after adaptation with fixed audiovisual asynchrony. We examined both drummers and other musicians to test the effect of different types of sensory training (Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005; Calvo-Merino, Gresez, Glaser, Passingham, & Haggard, 2006) and sense of agency (Rohde & Ernst, 2016) on the brain recalibration process. Whereas drummers have long motor, auditory and visual experience with drumming actions, other musicians such as guitarists or pianists that play in bands have long auditory and visual experience with such actions, but do not have direct motor experience with it. Non-musicians, in contrast, have no other experience than that given by attending concerts or watching music videos. Besides the drumming display we used a simple flash-beep display for which none of the assessed groups should have a different level of experience.

Several studies have reported that prolonged and rapid recalibration are two different processes and independent of each other (Bruns & Röder, 2015; De Niear, Noel, & Wallace, 2017; Van der Burg, Alais, & Cass, 2015; Van der Burg & Goodbourn, 2015; Van der Burg, Orchard-Mills, & Alais, 2015), suggesting that rapid recalibration is an early sensory effect, whereas the prolonged recalibration reflects a more cognitive process, here we focused on prolonged recalibration. Hence, we asked whether long-term music training affects the higher-cognitive recalibration process (e.g. Desantis & Haggard, 2016; Fujisaki et al., 2004; Vroomen et al., 2004).

We hypothesised that musicians would show a reduced effect of recalibration due to their increased perceptual accuracy when compared to non-musicians, and that this reduction in recalibration would be greater after adaptation with a music clip (for which musicians have prior knowledge and sense of agency) than a flash and beep clip. Secondly, we expected drummers to show an even weaker effect of recalibration with drumming displays, due to their added motor experience and sense of agency with the stimulus.

Method

Participants

A total sample size of 24 was calculated for a Cohen's F effect size equal to 0.25 (for a medium effect size) through a priori type of power analysis for an ANOVA repeated measures within-between interaction. We used G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) and assumed a level of power of 0.80, 3 groups, 6 measurements, and an alpha level of 0.05. We aimed to test more than 8 participants per group and have an equal number of participants in the three groups. We tested 42 participants in total, but had to exclude the data for 4 non-musicians and 3 musicians because their performance was at chance level in at least one of the six testing blocks. . We also had to exclude the data for another musician because of a technical problem and for another non-musician because he listened to music for more than six hours every day. This decision was taken based on evidence that untrained music listeners can at times show similar capabilities to trained musicians (Bigand & Poulin-Charronnat, 2006). No data for the tested drummers had to be excluded. The data for eleven drummers (Mean age = 24.45, SD = 1.65, two females), 11 musicians (Mean age = 24.91, SD = 2.32, five females), and 11 non-musicians (Mean age = 21.91, SD = 1.42, eight females) were included in the study. The number of participants is similar or higher than other studies investigating recalibration effects (e.g. Fujisaki et al., 2004; Navarra, García-Morera, & Spence, 2012; Noel et al., 2016; Roach et al., 2010; Vroomen et al., 2004; Petrini et al., 2011). All participants reported normal or corrected-to-normal vision and hearing. Non-musicians had no experience with playing any instrument. Musicians and drummers were selected to have at least four years of active music training/practice and have played their instrument for at least 1h per week over the period of training (e.g. Lee and Noppeney, 2011; Vines et al., 2006). We defined musicians as those who played any musical instrument besides the drums (Mean = 8.73, SD = 3.58). Drummers had to have significantly more experience in drumming than any other instrument (at least 2 years more) and preferably to only have played the drums (Mean = 10.64 years, SD = 5.26). Participants gave informed consent to participate, and the study received ethical approval from the research ethics board at University of Bath. All subjects gave informed consent to participate and received cash for their participation.

Apparatus and Stimuli

The flash-beep displays consisted of a pure tone at 2000 Hz and 84 dB mean intensity and a white dot (luminance: 85 cd/m^2). These were presented on a black background (luminance: 0.12 cd/m^2) and were 460 ms in duration. Detailed description of the creation and characteristics of the drumming point-light displays has been published elsewhere (Petrini et al., 2009a; Petrini, Russell & Pollick, 2009b; Petrini, Holt & Pollick, 2010). The drumming displays consisted of a point-light display of a professional jazz drummer playing a simple swing groove at 120 BPM and accent on the second beat (see examples of clips online). The 3D motion coordinates were transformed into point-light displays using a Matlab script with PsychToolbox routines (Brainard, 1997; Pelli, 1997). The matching synthetic sounds were created using a simulation of the first 25 modes of a circular membrane (Fontana, Avanzini, & Rocchesso, 2004). This takes as input the time and impact velocity of an impact and provides the audio signal. The 60Hz movies (AVI) and audio (WAV) were combined in Adobe Premiere 1.5 to produce the audiovisual displays. The audiovisual displays containing asynchronous audio and video were generated by either delaying the video with respect to the audio, or the audio respect to the video, by 67, 133, 200 and 267 ms. The resulting audiovisual clips were three seconds in duration. All displays were presented in focus and were preceded by a fixation point. We used a point light display rather than a full clip because we wanted to avoid possible effects of context as we were interested in the action and kept the low-level information as similar as possible between the flash-beep and the drumming display.

All displays were presented via an Apple Macintosh MacPro with Retina display (60 Hz refresh rate) laptop running OS X 10.9 and an AMD Radeon R9 M370X graphics card with 2GB of GDDR5 memory. The visual cues were displayed on a HannsG HP222 monitor, which was placed approximately 50 cm from the observer. Auditory cues were presented through high quality Sennheiser HD 380 pro headphones and the volume at the sound source was 50 dB intensity for the drumming displays and 55 dB for the flash-beep. The experiment was controlled using MATLAB 2013b (MATHWORKS Inc., Natick, MA) and the PsychToolbox (Brainard, 1997; Pelli, 1997).

Procedure

Participants completed a 90-minute experiment composed of six blocks (two baseline blocks and four adaptation blocks). The first two blocks were aimed at measuring participants' point of subjective simultaneity before adaptation (i.e., individual baseline). One block presented the audiovisual drumming displays and the other block the flash-beep display (see clip examples online). The presentation of these two blocks was counterbalanced across participants. The displays varied in the level of asynchrony between the visual and the auditory cue (-266.67, -200, -133.33, -66.67, 0, 66.67, 133.33, 200, 266.67ms; where negative offsets indicate the audio stream preceded the video stream). For both display types, each level of asynchrony was repeated 10 times at random for a total number of 90 trials in each block and an overall total number of 180 trials (2 display types X 9 audiovisual asynchronies X 10 repetitions) for the full study. Participants had to indicate for each trial whether the audio and video were in synchrony or not by pressing one of two keys on the computer keyboard (see Fig. 1A). Each one of the subsequent four adaptation blocks (flash-beep -200ms block (Fig. 1B), flash-beep +200ms block (Fig. 1C), drumming -200ms block, and drumming +200ms block), started with an adaptation phase and their presentation was counterbalanced across participants. At the beginning of each block the adaptation phase was conducted by repeating 100 times either the display for which the auditory cue led the visual of 200ms (-200) or the display for which the visual cue led the auditory of 200ms (200). This duration of the adaptation asynchrony was selected based on previous literature (e.g. Fujisaki et al., 2004; Vroomen et al., 2004). During the adaptation phase, participants were instructed to carefully watch the repeated displays until the end. To make sure participants paid attention to the display during the adaptation phase, they were asked to count how many animal pictures were presented during this phase. These images were flashed randomly between the SJ trials throughout each testing block. The number of pictures changed in each block and participants had to report the number at the end of the adaptation phase. After the adaptation phase ended participants were asked, similar to the initial two blocks (baseline), to judge the synchrony between audio and video in the 9 clips 10 times. To ensure adaptation was maintained, before each set of 9 randomly presented displays the adaptation display (either -200 or 200ms) was repeated 5 more times (see Fig. 1B and 1C).

Participants had to take five-minute breaks after the baseline testing and then after both adaptation blocks. This served as relaxation time to prevent fatigue and also for the adaptation effects to wear off before adapting in the opposite direction.

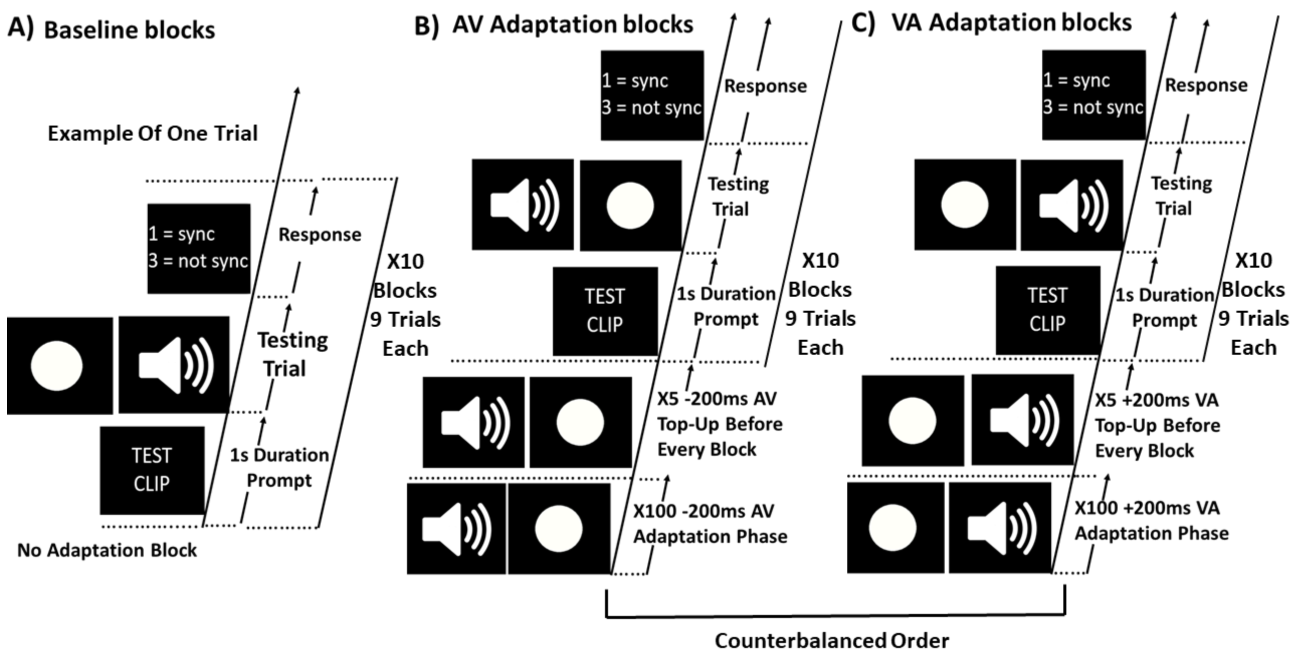

Fig. 1 Schematic of display conditions and experimental design for the baseline (left) and for the adaptation blocks (middle and right). (A) Participants were presented with 10 blocks of 9 trials (corresponding to the 9 levels of audiovisual asynchrony) and were asked to judge if the sound and video in each trial were in synch or not. Prior to the display a prompt was flashed on the screen for one second. This was done for both flash-beep and drumming displays, but for simplicity here we show the flash-beep display. (B) Adaptation blocks with audiovisual drumming and flash-beep displays. Participants were exposed to 100 repetitions of -200ms AV asynchrony of one display after which 10 blocks of 9 testing trials were again presented. Before each block of testing trials an adaptation top-up consisting of another 5 -200ms AV repetitions was also presented, in order to maintain the adaptation throughout the study duration. For simplicity here we show this procedure for the flash-beep display only, as it was identical for the drumming display. (C) Adaptation blocks with the visual-audio drumming and flash-beep displays. Participants were exposed to 100 repetitions of +200ms VA asynchrony of one display after which 10 blocks of 9 testing trials were again presented. Before each block of testing trials an adaptation top-up consisting of another 5 +200ms VA repetitions was also presented, in order to maintain the adaptation throughout the study duration. For simplicity here we show this procedure for the flash-beep display only, as it was identical for the drumming display.

Analysis Procedure

For both types of displays, the proportion of synchronous responses for each level of stimulus onset asynchrony (SOA) was fit with a Gaussian probability density function similarly to several studies that examined audiovisual recalibration effects (e.g., Fujisaki et al., 2004; Van der burg et al., 2013). From these fits, two parameters of interest were derived: the point of subjective simultaneity (PSS) and the temporal integration window (TIW). The PSS represented the level of SOA at which the highest perceived simultaneity between video and audio was perceived by the individual and it was taken as the peak of the Gaussian curve. The TIW represents the range of cue onset asynchronies, where participants were not able to reliably identify the physical asynchrony between the cues. We estimated participants' TIW using the standard deviation (SD) of the Gaussian fit (e.g. Love et al., 2013; Desantis & Haggard, 2016). This procedure was followed for both non-adaption and adaption conditions. Please see Fig. 2 for an example of the drummers group and also Fig. S1 and S2 in the supplemental material for the average fitting of the musician and non-musician groups.

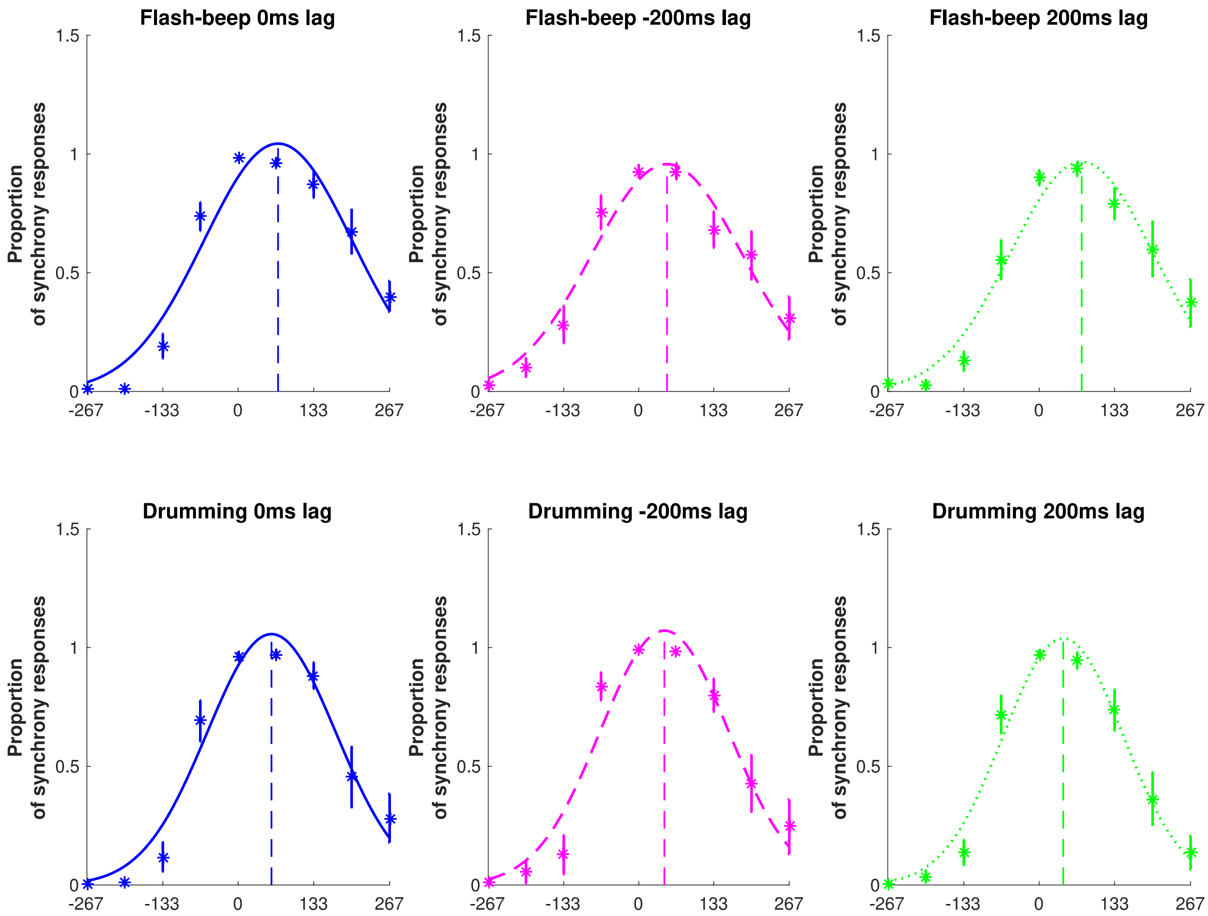

Fig. 2 Fit to average proportion of "synchrony" responses as a function of audiovisual SOAs (from -267ms audio leading asynchrony to 267ms visual leading asynchrony) for the drummers group shown separately for no adaptation (blue and solid line), -200ms (magenta and dashed line) and 200ms (green and dotted line) adaptation conditions and drumming (bottom panels) and flash-beep displays (top panels). Solid, dashed and dotted lines represent the best-fitting Gaussian curves while the asterisks represent the average data at

each audiovisual SOA. The peak of the Gaussian curves provides an estimate of the PSS (point of subjective simultaneity), marked by the dashed vertical lines, while the width of the Gaussian represents the TIW (temporal Integration window). The error bars represent the standard error of the mean. Please see Fig. S1 and S2 in the supplementary material for the fit to average data for the musician and non-musicians group.

Results

The r-square values for all three groups of participants were high indicating a good fit of the Gaussian to the data (drummers *Mean*=.91 and *SD*=.03, other musicians *Mean*=.89 and *SD*=.03, non-musicians *Mean*=.87 and *SD*=.05). Before examining the effect of music training on the prolonged recalibration process, we examine whether adaptation to the chosen fixed audiovisual asynchrony gave rise to a significant shift in PSS (when compared to the PSS before adaptation), irrespective of the shift direction, by comparing the absolute PSS shift separately for display conditions, adaptation conditions, and group. This was needed also to make sure that the two adaptation conditions (-200ms with auditory leading and +200ms with visual leading) were effective in shifting the participants PSS (i.e. had a significant aftereffect). Three one-sample t-tests showed that all the conditions and all the groups had a significant shift in PSS after adaptation when compared to 0 ($t \geq 2.829$, $p \leq .018$; when bootstrapped $p \leq .048$, 95% CI [10.55, 45.97] based on 1000 bootstrap samples). Furthermore, we examined whether there was any difference in PSS baseline (before any adaptation occurred) by analysing these data with a mixed factorial ANOVA with group (drummers, musicians, and non-musicians) as between-subjects factor and display type (drumming and flash-beep) as within-subjects factors. No significant effect was found ($F <= .312$, $p >= .697$).

To examine the effect of long-term musical training on audiovisual recalibration we first calculated how much the point of subjective simultaneity (PSS) in the drumming and the flash-beep display conditions shifted after adaptation by subtracting the value of each individual PSS after adaptation from that before adaptation (PSS from baseline phase). We calculated the effect of recalibration this way, rather than as a difference in PSS shift between the two adaptation conditions (e.g. Desantis & Haggard, 2016; Fujisaki et al., 2004; Vroomen et al., 2004), as we wanted to account for differences in the individuals' initial ability to detect asynchrony between audio and video (as we know musicians and drummers have an enhanced ability to detect asynchrony compared to non-musicians; Lee & Noppeney, 2011; Petrini et al., 2011). However, if we had calculated the recalibration effect in terms of the difference in PSS shift under the two adaptation conditions, we would have found very similar values to previous studies (Desantis & Haggard, 2016; Fujisaki et al., 2004; Navarra et al., 2012; Vroomen et al., 2004). The obtained data were then analysed with a mixed factorial ANOVA with group (drummers, musicians, and non-musicians) as between-subjects factor and display type (drumming and flash-beep) and adaptation asynchrony (-200 and +200ms) as within-subjects factors. We found a main effect of group, $F(2,30)= 3.440$, $p= .045$, $\eta_p^2 = .187$, and a significant interaction between display type and adaptation asynchrony $F(1,30)=17.986$, $p<.001$, $\eta_p^2 = .375$. All other effects did not reach significance level ($F \leq 1.889$, $p \geq .180$). Planned simple contrasts returned no significant difference between the effect of adaptation for the drummers and the musicians group (p = .947; 95% CI [-15, 14.05] based on 1000 bootstrap samples), but showed a significant difference between the effect of adaptation for drummers and non-musicians (p = .033, 95% CI [1.39, 30.45] based on 1000 bootstrap samples) and musicians and non-musicians (p = .028, 95% CI [1.86, 30.92] based on 1000 bootstrap samples). Fig. 3, left panel, shows that the effect of recalibration was very similar for drummers and musicians whose PSS shifted to an audio-leading asynchrony irrespective of the display type and of the adaptation asynchrony. The recalibration effect of non-musicians, however, was very different with their PSS shifting towards video-leading asynchrony irrespective of the display type and of the adaptation asynchrony.

Fig. 3, right panel, shows that the interaction between type of display and adaptation was due to the flash-beep display inducing a PSS shift in the direction of the adapted asynchrony; that is, towards visual-leading asynchrony if the asynchrony used during adaptation had the video leading the auditory or towards audio-leading asynchrony if the asynchrony used during adaptation had the audio leading the video. This result is in line with the previous studies where a simple flash-beep type of display was used and only non-musicians (that we know of) were tested (e.g., Desantis & Haggard, 2016; Fujisaki et al., 2004; Foss-Feig et al., 2010; Garcia-Perez & Alcala-Quintana, 2012; Shams, Kamitani, & Shimojo, 2000; Vroomen et al., 2004). In contrast, for the drumming display the PSS shifted towards audio-leading asynchrony when the visual-leading asynchrony was used during the adaptation phase. Post hoc paired-samples t-test analyses, Bonferroni corrected, showed that there was a significant difference between the effect of visual-leading adaptation for the flash-beep and the drumming display ($t(32)= 3.934$, $p= .002$, 95% CI [14.21, 38.79] based on 1000 bootstrap samples). No difference, in contrast, was found between the effect of audio-leading adaptation for the flash-beep and the drumming display ($t(32)= -1.310$, $p= .208$, 95% CI [-24.30, 4.67] based on 1000 bootstrap samples). These results were replicated by running the analysis for the male only sample which had the larger number of participants (see supplementary material for these additional analyses and figure).

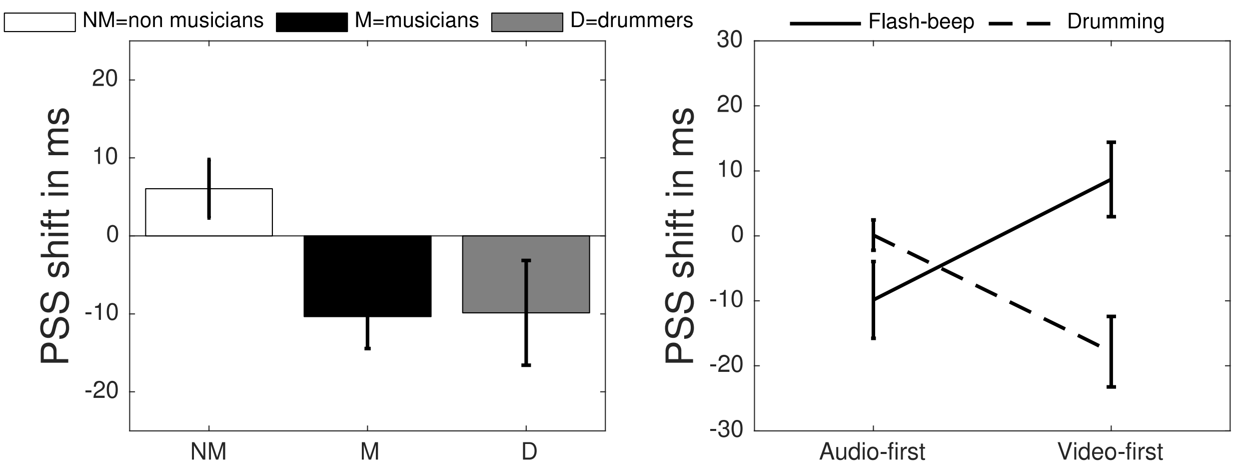

Fig. 3 *Left.* PSS shift for non-musicians (NM), musicians (M) and drummers (D). The PSS shift in ms was calculated by subtracting the value of each individual PSS after adaptation from that before adaptation (i.e. from the baseline or PSS before any adaptation took place). The adaptation for musicians and drummers was in the opposite direction to that of non-musicians (please see Figure S4 for a breakdown of the figure into the different conditions). *Right.* Overall PSS shift for flash-beep and drumming displays for the audio-leading and video-leading adaptations. Error bars show standard error of the mean.

A directional Pearson's correlation was run to test whether the PSS shift towards audio-leading asynchronies for musicians and drummers increased with years of music training (Fig. 4, left panel). The results showed that the PSS shift towards audio-leading asynchrony increased significantly with years of music training when musicians and drummers were adapted to the visual-leading asynchrony (r = -.378, p = .042). For the audio-leading asynchrony however, no such effect was found (r= -.144, p = .261).
Hence, the PSS shift towards audio-leading asynchronies for drummers and musicians was driven by a change in the recalibration process specific to the adaptation with the visual-leading asynchrony. Similarly, we examined whether the size and sign of the recalibration effect decreased with a decrease in the size of the TIW by running a directional Pearson's correlation separately for the audio-leading and the visual-leading asynchrony (Fig. 4, right panel). The results showed that the size of the TIW and the recalibration correlated positively for the adaptation with the audio-leading asynchrony, in that the smaller the TIW the smaller and more negative was the recalibration effect (r = .443, p = .005). The same correlation for the adaptation with the visual-leading asynchrony did not reach significance despite showing a similar trend (r = .264, p = .069).

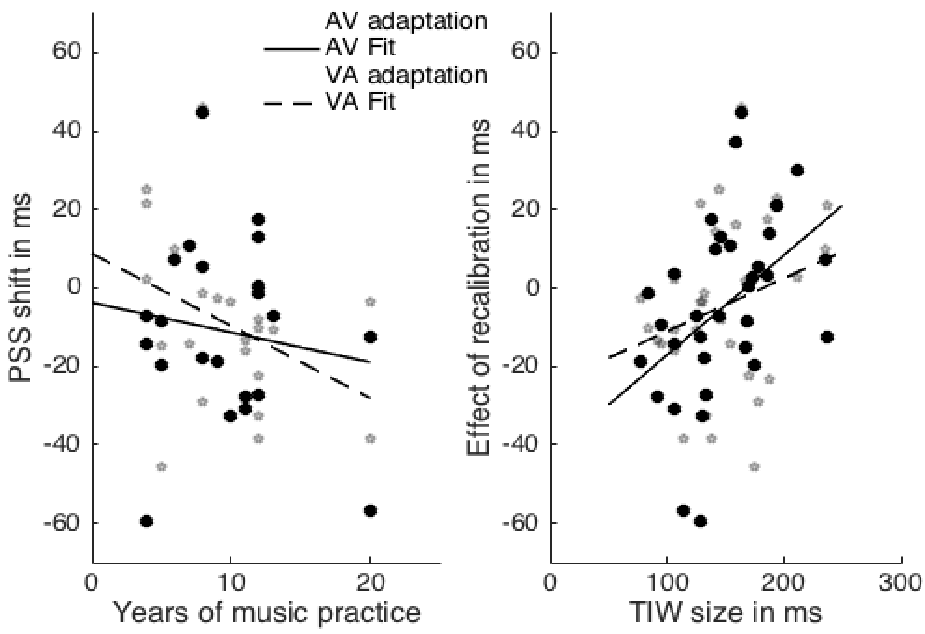

Fig. 4 *Left*. The shift in milliseconds of the point of subjective simultaneity (PSS) plotted against years of music training for the audio-leading adaptation (AV) and visual-leading asynchrony (VA) conditions. Data shown are together for drummers and musicians and drumming and flash-beep displays but separate for type of adaptation. *Right.* Relation between the temporal integration window (TIW) size and the shift in PSS for audio-leading adaptation (AV) and visual-leading asynchrony (VA) condition. Error bars show standard error of the mean.

These results further show that the effect of long-term music training on the recalibration process is driven by drummers and musicians shifting their PSS towards audio-leading asynchronies and that this effect is linked to a narrowing of the TIW (see Fig. 5).

Finally, we also examined the difference in perceptual accuracy due to long-term music training by analysis of the audiovisual temporal integration window (TIW) data with a mixed factorial ANOVA with group (drummers, musicians, and non-musicians) as between-subjects factor and display type (drumming and flash-beep) and adaptation lag (-200, 0ms and 200ms) as within-subjects factors. We found a main effect of group, $F(2,30)= 5.394$, $p= .010$,

η_p^2 = .264, a significant effect of display type, *F*(1,30)= 21.908, *p*< .001, η_p^2 = .422, a significant effect of adaptation lag, *F*(2,60)= 3.834, *p*= .027, η_p^2 = .113, and a significant interaction between display type and adaptation lag *F*(2,60)= 4.135, *p*=.021, η_p^2 = .121. All other effects did not reach significance level (F<=1.299, *p* >=.80). Planned simple contrasts returned a significant difference between the size of the TIW for the drummers and the musicians group (*p* = .031, 95% CI [3.43, 66.02] based on 1000 bootstrap samples), with drummers showing a smaller TIW, and thus greater ability to detect asynchrony, than the other musicians, and a significant difference between the size of TIW for drummers and non-musicians (*p* = .003, 95% CI [17.58, 80.13] based on 1000 bootstrap samples), with drummers showing a far smaller TIW than non-musicians. Fig. 5 shows the decrease in TIW width (or increase in asynchrony detection ability) when going from non-musicians to musicians and then to drummers. The significant effect of display type was due to drumming displays leading overall to a smaller TIW (*M* = 131.34 and *SD* = 39.54) than flash-beep (*M*=164.64 and *SD*= 56.90). Post hoc paired-samples t-test analyses, Bonferroni corrected, showed that the effect of adaptation lag was a consequence of the audio-leading asynchrony widening participants' TIW (decreasing their asynchrony detection ability) when compared to the video-leading lag (*t*(32)= 3.330, *p*= .006, 95% CI [4.56, 18.93] based on 1000 bootstrap samples). The significant interaction between display type and adaptation lag was due to visual-leading asynchrony resulting in the smaller TIW with respect to no lag (*t*(32)= 2.876, *p*= .042, 95% CI [3.42, 20.07] based on 1000 bootstrap samples) and audio-leading adaptation (*t*(32)= 4.44, *p*< .001, 95% CI [10.45, 28.18] based on 1000 bootstrap samples) for the drumming display but not for the flash-beep display (*t*(32) =-1.323, *p* =.195). This result is similar to that found for the recalibration effect, in that the adaptation with a visual-leading asynchrony has a strong effect on the drumming display but not on the flash-beep display condition.

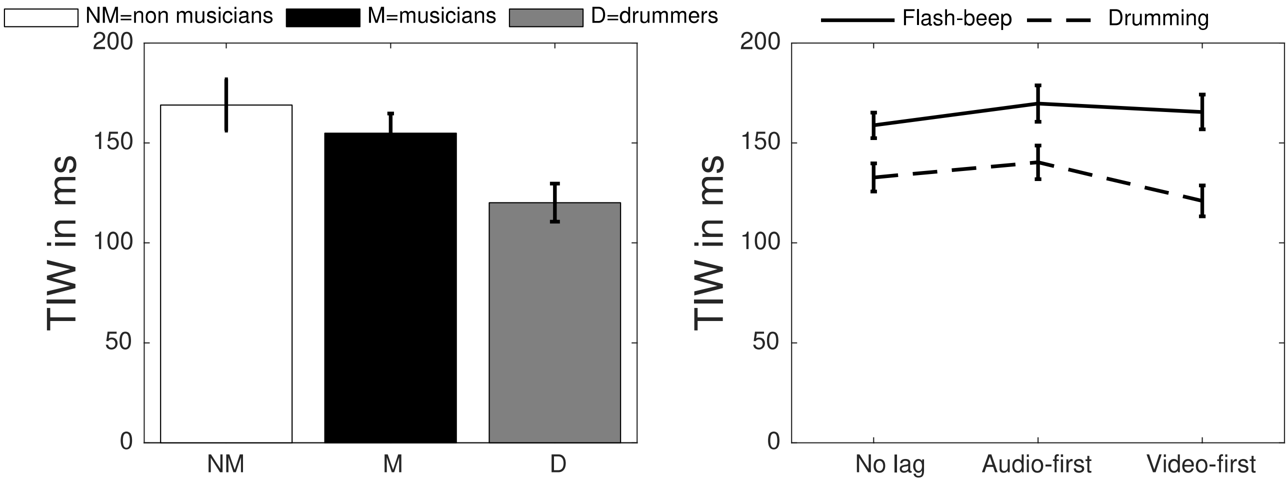

Fig. 5*. Left*. TIW for the non-musician (NM), musicians (M) and drummer group (D). Drummers showed the narrowest TIW, followed by musicians, whereas non-musicians showed the widest TIW (please see Fig. S4 for a breakdown of the Figure into the different conditions). *Right.* TIW width for the flash-beep and drumming displays before adaptation, and after adaptation with audio-leading and visual-leading asynchrony. Error bars show standard error of the mean.

Discussion

Long-term training with multisensory events affects the prolonged recalibration process for audiovisual integration. Our results show that both drummers and musicians had an opposite effect of recalibration (shift in PSS after adaptation) to non-musicians; that is while overall non-musicians recalibrated their perceived best synchrony towards visual-leading asynchronies, musicians and drummers recalibrated towards audio-leading asynchronies irrespective of the type of adaptation received. Interestingly, this shift towards audio-leading perceived synchrony increased with years of music practice and with an increase in perceptual accuracy (or decrease in the size of the TIW). However, the results for musicians and drummers were very similar, indicating that an added active motor experience tied to the stimulus (causing the sound) was not necessary for these changes to occur when recalibrating to passively sensed modalities (audiovisual displays). Our results show that long-term music training not only fine-tunes the binding process of visual and auditory cues

(Lee & Noppeney, 2011; Petrini et al., 2011) but also modulates the adaptive recalibration process. Additionally, because musicians and drummers showed the narrowest TIW but not the weakest adaptation, this suggests that greater perceptual accuracy cannot fully explain changes in the examined recalibration process, as it has been suggested before (Noel et al., 2016; Rohde & Ernst, 2013; Van De Burg et al., 2013). Furthermore, while musicians and drummers showed the same recalibration effect, they did not show the same perceptual accuracy (i.e. drummers were significantly more accurate). Hence, our results suggest that whilst these processes might have overlapping mechanisms, they are also independent.

Temporal correspondence is one of the factors that determine whether information from two senses will be perceived as belonging to the same event thus leading to multisensory integration (Spence & Squire, 2003; Stein et al., 1993). The extent to which we can tolerate a temporal misalignment between the cues and still bind them gives an estimate of how strongly they belong together. Because the extent of these cues relation depends also on environmental factors and the distance these cues have to travel the brain adaptively recalibrates their point of perceived simultaneity, which results in a perceptual realignment of these signals (Fujisaki et al., 2004; Vroomen et al., 2004) that otherwise would be perceived as asynchronous and separate. That is, the recalibration process determines a shift of the point of subjective simultaneity (PSS) in the direction of the leading sense after repeated exposure to an audiovisual asynchrony (i.e. shift towards auditory-leading PSS if overexposed to auditory-leading asynchronies).

Whilst it has been shown that training for a long-period with a music instrument, which is a rich multisensory activity, narrows the tolerance to the temporal misalignment between sound and vision cues (references), here we show that this long-term natural occurring multisensory training also affects the adaptive brain recalibration process.

Van der Burg et al. (2013) showed that for rapid recalibration with audiovisual stimuli, the size of the TIW and the recalibration effect are directly proportional (see also Noel et al., 2016). In their 2013 study, Rohde and colleagues also showed that this correlation between perceptual accuracy and strength of the recalibration was present for more prolonged adaptation (more than one asynchronous trial). These findings suggest that incoming multisensory information could be judged for simultaneity at every trial and individuals with lower tolerance to a misalignment between the cues could be able to dismiss this information as erroneous thus causing the recalibration to not occur or be weaker. In other words, individuals who have the narrowest TIW or highest perceptual accuracy should show either no or weaker recalibration. Here we show that even with prolonged periods of adaptation the extent of recalibration does correlate with the perceptual accuracy, indeed the smaller the TIW and the more the PSS shifts towards auditory-leading asynchronies after adaptation supporting the conclusion that these mechanisms may be intrinsically linked as the findings of Rohde et al. (2013) suggested. Nevertheless, we also showed that overall musicians and drummers did not differ in the extent of the recalibration to auditory-leading PSS while they did differ in their TIW size, thus suggesting that perceptual accuracy and recalibration might be subserved by separate cognitive processes, despite them correlating in the general population (Noel et al., 2016; Rohde & Ernst, 2013; Van der Burg et al., 2013). Whether long-term multisensory training as afforded by playing a musical instrument exacerbates the separation between these two multisensory mechanisms is still unclear, although our results do suggest that may be the case, since both musicians and drummers did recalibrate (although in the opposite direction to non-musicians) despite having smaller TIWs (higher perceptual accuracy). Future studies could examine how musicians and non-musicians perform in a rapid recalibration task to examine whether recalibration does or does not correlate with the level of perceptual accuracy in musicians, especially given that rapid and prolonged recalibration (the type of recalibration examined here) have been distinguished as two separate processes (Bruns & Röder, 2015; De Niear et al., 2017; Simon et al., 2017; Van der Burg et al., 2015a; Van der Burg et al., 2015b; Van der Burg & Goodbourn, 2015).

Furthermore, although our findings suggest that the effect of music training on audiovisual recalibration might be mediated by an enhancement in perceptual accuracy, we cannot draw a strong conclusion on whether it is the music training that directly affects the brain recalibration mechanism or whether it is the refinement of perceptual abilities following music training that affects this mechanism. Future studies could tackle this question by examining, for example, performance on judgements of simultaneity by musicians with different levels of perceptual accuracy but similar training.

The reason why musicians and drummers consistently recalibrated their perceived synchrony between sound and vision towards audio-leading asynchronies after adaptation is unclear. It has been shown that having predictable targets and training increases motor anticipation and recalibration in sensorimotor tasks (Rohde, van Dam, & Ernst, 2014). In the present study no active motor task was used, however, musicians have been shown to have a higher ability to predict the arrival of auditory information by filling in missing visual information with their acquired motor repertoire (Petrini et al., 2009b). For example, drummers can predict when a drumming impact will occur and judge the asynchrony between visual information and sound even if the visual movement of the drummer is missing/occluded. In other words, musicians have enhanced abilities to predict when a sound should occur based on their long-term sensorimotor training (Lee and Noppeney, 2011; Petrini et al., 2011). Interestingly, when predicting the time of impact based on missing visual information the perceived synchrony of drummers shifted from visual-leading to audio-leading asynchrony (Petrini et al., 2009b), similarly to our present results. The explanation of why in musicians the sound needs to lead the video to perceive simultaneity after adaptation with visual-leading asynchrony can reside in their ability to map the sound occurrence based on the learnt action (Lee & Noppeney, 2011; Petrini et al., 2009b; Desantis & Haggard, 2016). That is, musicians may not rely on vision (as in Petrini et al., 2009b) and may predict and anticipate the arrival of the sound based on their audio-motor mapping process (Lee and Noppeney, 2011; Petrini et al., 2009b) as suggested by tapping studies showing that touch needs to precede the other stimuli to perceive synchrony (Aschersleben & Prinz, 1995; Miyake, Onishi, & Pöppel, 2004; Repp & Su, 2013). If musicians were using motor simulation/mapping in place of visual information to decide whether visual and auditory information were synchronised they would anticipate the sound occurrence with respect to the visual stimulus (to coincide with their anticipated motor event) and report synchronization when the sound preceded the visual information. Non-musicians in turn might not use this sensorimotor mapping (Lee & Noppeney, 2011) and consequently show overall the usual bias found in synchrony perception towards visual-leading asynchronies (e.g. Love, Petrini, Cheng, & Pollick, 2013). Hence, the adaptation to fixed audiovisual lags could exacerbate these existing differences in synchrony perception between musicians and non-musicians.

In contrast to our predictions the effect of long-term music training extended to both displays (flash-beep and drumming clips) rather than being specific to or stronger for the music stimuli. This was the case not only for the recalibration effect but also for the TIW size. Indeed, musicians and drummers recalibrated towards audio-leading perceived synchrony for both displays and showed an increased perceptual ability, when compared to non-musicians, irrespective of the display used. Both drumming and flash-beep displays had auditory cues of short durations, and similarity in the visual information (white dots on a black background), although one was a cyclic event and the other was not. Hence, these displays might not have been different enough to affect simultaneity judgements. This, explanation is however unlikely as we did find an overall effect of type of display on simultaneity judgements. A more plausible explanation is that active experience with the motor action does not affect the recalibration of passively sensed modalities (for which active motion is not required) and rather both sensory (non-drummer musicians playing with drummers) and sensorimotor (drummers) experience affects the brain recalibration (Calvo-Merino et al., 2005; Calvo-Merino, et al., 2006). The group results seem to support this second possibility since as mentioned, the drummers and the other musicians showed a similar recalibration effect despite drummers having long-term active motor experience with the drumming display.

Drummers showed the highest perceptual accuracy, followed by the other musicians and the non-musicians. This result replicates and extends previous findings showing that long-term music training strongly enhances perceptual accuracy (e.g. Lee & Noppeney, 2011; Petrini et al., 2009a; Petrini et al., 2011), and showing that the type of instrument played affects this enhancement process at least in the temporal domain. This result was again general for flash-beep and drumming clips rather than stronger for the drumming clips (for which drummers have increased sense of agency and motor repertoire; Calvo-Merino et al., 2005,2006; Rohde & Ernst, 2016). Numerous studies have emphasised the role of rhythm maintenance when playing a percussion instrument, such as the drums (Botella, 2008; Flatischler, 1992; Nichols, 2012). This aspect is not as important in the large majority of instruments which can produce melody (e.g. piano, flute, guitar etc.). Drummers are also responsible with maintaining the rhythm and synchronicity between instruments in a band (Nichols, 2012), which may explain why drummers perform better than both other musicians and non-musicians in multisensory simultaneity judgement tasks (Bishop & Goebl, 2014; Hodges et al., 2005; Petrini, Dahl et al., 2009; Petrini, Russell et al., 2009; Vatakis & Spence, 2006; Lee and Noppeney, 2011).

Finally, we found that the effect of adaptation for the flash and beep displays was similar to previous studies (e.g. Navarra et al., 2012; Vroomen et al., 2004), in that overall the recalibration occurred in the direction of the adapted asynchrony. That is, participants usually perceived the synchrony when vision led the auditory cue if they were adapted with visual-leading asynchrony and perceived synchrony when the auditory cue led vision if they were adapted with audio-leading asynchrony. The drumming display, in contrast, did not have the same effect, and participants mostly perceived synchrony when the auditory cue led vision if adapted with visual-leading asynchrony. One evident difference between the flash-beep and drumming displays that could have contributed to the different results for these stimuli is that the drumming display is cyclical. In the present study we used a simultaneity judgement (SJ) task because in our previous studies (e.g. Love et al., 2013;

Petrini et al., 2010) we showed that with cyclic stimuli temporal order judgements (TOJ) become really difficult and at times impossible for both drummers and non-musicians (Petrini et al., 2010). Indeed, in Petrini et al 2010 we showed that although drummers were more precise than non-musicians in both SJ and TOJ tasks when using the drumming displays used here, there were still drummers and non-musicians unable to perform the TOJ task. This means that the shift of the participants' PSS towards auditory-leading asynchronies for drumming displays could have been a consequence of their inability to discriminate what sense was coming first during adaptation. In other words, the adaptation might not have been effective with the drumming displays because the sensory order of the asynchrony used during adaptation was unclear for that stimulus condition. We would also expect based on our previous findings (e.g. Love et al., 2013; Petrini et al., 2010) that this uncertainty during the adaptation phase would affect more the +200ms than the -200ms adaptation condition as we know that participants even for cyclic and complex stimuli are quite good at judging the temporal order for large auditory-first asynchronies while for large vision-leading asynchronies participants are not as good (Petrini et al., 2010). When looking at Figure 3 right panel and at its breakdown in the supplementary material the shift towards auditory first was indeed stronger and more common across the groups for visual leading adaptation (+200ms) than auditory-leading adaptation (-200ms). Also visual-leading adaptation did have a stronger effect than auditory-leading adaptation on TIW for drumming displays but not flash-beep thus supporting this argument. That said what is still unclear is why participants recalibrated to auditory leading PSS more when they were less sure of the sensory order in the visual-leading adaptation condition than when they were more sure of it in the auditory-leading adaptation condition. That is, we would expect that for auditory-leading adaptation all groups will show a PSS shift towards auditory-leading asynchronies if in this condition the adaptation with the drumming displays was generally more efficient, but the only group that showed this trend was the drummers group while the non-musicians PSS shifted towards vision-leading asynchronies. Future studies could run both TOJ and SJ tasks with similar recalibration tasks and groups to the present study to examine the contribution of the cyclic nature of the stimulus to the recalibration process. This would help to understand when the recalibration process is disrupted, especially considering that cyclic stimuli are very common in everyday life.

Interestingly, under both adaptation conditions (although more for auditory-leading adaptation), a shift towards an audio-leading perceived synchrony was accompanied by an increased perceptual accuracy or narrowing of the TIW (Noel et al., 2016; Rohde & Ernst, 2013). Furthermore, in general the average TIW of the drumming displays was narrower than that of the flash-beep. We do not know as yet why participants had a higher level of perceptual accuracy for the drumming display when compared to the flash-beep display, what we do know is that this is not the first time this result was found with the same stimuli (Love et al., 2013) when using simultaneity judgements. We assume that this is due to differences in level of complexity and amount of information between the two stimuli, however, because this is the first study examining the recalibration effect for stimuli with very different levels of complexity, future studies could further examine how the brain uses these features to flexibly recalibrate to audiovisual asynchrony, by including a higher number of natural and complex stimuli.

In conclusion, our results show that long-term music training affects both the perceived synchrony and the recalibration process of passively sensed modalities (audiovisual stimuli) indicating that both multisensory mechanisms can be shaped by naturally occurring multisensory training (Lee & Noppeney, 2011; Petrini et al., 2011). Such findings suggest that musical training could constitute a viable method of fine-tuning multisensory perception for those with deficits in this process, such as individuals with autism spectrum disorder (Foxe et al., 2013; Noel et al., 2017; Oberman & Ramachandran, 2008; Stevenson, Segers, Ferber, Barense, & Wallace 2015; Turi, Karaminis, Pellicano, & Burr, 2016).

Acknowledgements

We would like to thank Dr Chris Bevan and Eliot Farmer for engaging and useful discussion on what it means to be a drummer.

Conflict of interest

The authors declare no competing financial interests.

References

Alais, D., & Carlile, S. (2005). Synchronizing to real events: Subjective audiovisual alignment scales with perceived auditory depth and speed of sound. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(6), 2244-2247.

Arnold, D. H., Johnston, A., & Nishida, S. (2005). Timing sight and sound. *Vision Research*, *45*(10), 1275-1284.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100-130.

Bishop, L., & Goebl, W. (2014). Context-specific effects of musical expertise on audiovisual integration. *Frontiers in Psychology*, *5*, 1123.

Bosnyak, D. J., Eaton, R. A., & Roberts, L. E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cerebral Cortex*, *14*(10), 1088-1099.

Botella, L. (2008). Timekeeping is Everything 1: Rhythm and the Construction of Meaning. *Journal of Constructivist Psychology*, *21*(4), 309-320.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433-436.

Bruns, P., & Röder, B. (2015). Sensory recalibration integrates information from the immediate and the cumulative past. *Scientific Reports*, *5*, 12739.

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: an FMRI study with expert dancers. *Cerebral Cortex*, *15*(8), 1243-1249.

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, *16*(19), 1905-1910.

De Niear, M. A., Noel, J. P., & Wallace, M. T. (2017). The impact of feedback on the different time courses of multisensory temporal recalibration. *Neural Plasticity*, *2017*.

Desantis, A., & Haggard, P. (2016). Action-outcome learning and prediction shape the window of simultaneity of audiovisual outcomes. *Cognition*, *153*, 33-42.

Di Luca, M., Machulla, T. K., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision*, *9*(12), 7-7.

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 Methods, 39*(2), 175-191.

Flatischler, R. (1992). *The forgotten power of rhythm*. LifeRhythm.

Fontana, F., Avanzini, F., & Rocchesso, D. (2004, October). Computation of nonlinear filter networks containing delay-free paths. In *Proceedings of the 7th International Conference on Digital Audio Effects (DAFX-04), Naples, Italy* (pp. 113-118).

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research*, *203*(2), 381-389.

Foxe, J. J., Molholm, S., Del Bene, V. A., Frey, H. P., Russo, N. N., Blanco, D., ... & Ross, L. A. (2013). Severe multisensory speech integration deficits in high-functioning school-aged children with autism spectrum disorder (ASD) and their resolution during early adolescence. *Cerebral Cortex*, bht213.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. Y. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, 7(7), 773-778.

García-Pérez, M. A., & Alcalá-Quintana, R. (2012). On the discrepant results in synchrony judgment and temporal-order judgment tasks: a quantitative model. *Psychonomic Bulletin & Review*, *19*(5), 820-846.

Harrar, V., & Harris, L. R. (2008). The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Experimental Brain Research*, *186*(4), 517-524.

Hodges, D. A., Hairston, W. D., & Burdette, J. H. (2005). Aspects of multisensory perception: the integration of visual and auditory information in musical experiences. *Annals of the New York Academy of Sciences*, *1060*(1), 175-185.

Keetels, M., & Vroomen, J. (2007). No effect of auditory–visual spatial disparity on temporal recalibration. *Experimental Brain Research*, *182*(4), 559-565.

King, A. J. (2005). Multisensory integration: strategies for synchronization. *Current Biology*, *15*(9), R339-R341.

Lee, H., & Noppeney, U. (2011). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proceedings of The National Academy of Sciences*, *108*(51), E1441-E1450.

Love, S. A., Petrini, K., Cheng, A., & Pollick, F. E. (2013). A psychophysical investigation of differences between synchrony and temporal order judgments. *PloS One*, *8*(1), e54798.

Navarra, J., García-Morera, J., & Spence, C. (2012). Temporal adaptation to audiovisual asynchrony generalizes across different sound frequencies. *Frontiers in Psychology*, *3*, 152.

Nichols, J. (2012). Music education in homeschooling: Jamie's story. In *Narrative soundings: An anthology of narrative inquiry in music education* (pp. 115-125). Springer Netherlands.

Noel, J. P., De Niear, M. A., Stevenson, R., Alais, D., & Wallace, M. T. (2017). Atypical rapid audio-visual temporal recalibration in autism spectrum disorders. *Autism Research*, *10*(1), 121-129.

Noel, J. P., De Niear, M., Van der Burg, E., & Wallace, M. T. (2016). Audiovisual simultaneity judgment and rapid recalibration throughout the lifespan. *PloS one*, *11*(8), e0161698.

Noel, J. P., Łukowska, M., Wallace, M., & Serino, A. (2016). Multisensory simultaneity judgment and proximity to the body. *Journal of Vision*, *16*(3), 21-21.

Oberman, L. M., & Ramachandran, V. S. (2008). Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: The Mirror Neuron Hypothesis. *Social Neuroscience*, *3*(3-4), 348-355.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437-442.

Petrini, K., Dahl, S., Rocchesso, D., Waadeland, C. H., Avanzini, F., Puce, A., & Pollick, F. E. (2009). Multisensory integration of drumming actions: musical expertise affects perceived audiovisual asynchrony. *Experimental Brain Research, 198*(2-3), 339-352.

Petrini, K., Holt, S. P., & Pollick, F. (2010). Expertise with multisensory events eliminates the effect of biological motion rotation on audiovisual synchrony perception. *Journal of Vision*, *10*(5), 2-2.

Petrini, K., Pollick, F. E., Dahl, S., McAleer, P., McKay, L., Rocchesso, D., ... & Puce, A. (2011). Action expertise reduces brain activity for audiovisual matching actions: an fMRI study with expert drummers. *Neuroimage*, 56(3), 1480-1492.

Petrini, K., Russell, M., & Pollick, F. (2009). When knowing can replace seeing in audiovisual integration of actions. *Cognition*, 110(3), 432-439.

Proverbio, A. M., Attardo, L., Cozzi, M., & Zani, A. (2015). The effect of musical practice on gesture/sound pairing. *Frontiers in Psychology*, *6*.

Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). *Psychonomic Bulletin & Review*, *20*(3), 403-452.

Roach, N. W., Heron, J., Whitaker, D., & McGraw, P. V. (2010). Asynchrony adaptation reveals neural population code for audio-visual timing. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20101737.

Rohde, M., & Ernst, M. O. (2013). To lead and to lag–forward and backward recalibration of perceived visuo-motor simultaneity. *Frontiers in Psychology*, *3*, 599.

Rohde, M., & Ernst, M. O. (2016). Time, agency, and sensory feedback delays during action. *Current Opinion in Behavioral Sciences*, *8*, 193-199.

Rohde, M., Scheller, M., & Ernst, M. O. (2014). Effects can precede their cause in the sense of agency. *Neuropsychologia*, *65*, 191-196.

Rohde, M., van Dam, L. C., & Ernst, M. O. (2014). Predictability is necessary for closed-loop visual feedback delay adaptation. *Journal of Vision*, *14*(3), 4-4.

Roseboom, W., & Arnold, D. H. (2011). Twice upon a time. Multiple concurrent temporal recalibrations of audiovisual speech. *Psychological Science, 22*(7), 872-877.

Schroeder, C. E., & Foxe, J. J. (2004). 18 Multisensory Convergence in Early Cortical Processing. *The Handbook of Multisensory Processes*, 295.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, *408*(6814), 788.

Simon, D. M., Noel, J. P., & Wallace, M. T. (2017). Event Related Potentials Index Rapid Recalibration to Audiovisual Temporal Asynchrony. *Frontiers in Integrative Neuroscience*, *11*.

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology, 13*(13), R519-R521.

Stein, B. E., Meredith, M. A., & Wallace, M. T. (1993). The visually responsive neuron and beyond: multisensory integration in cat and monkey. In *Progress in Brain Research* (Vol. 95, pp. 79-90). Elsevier.

Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., & Wallace, M. T. (2015). The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. *Multisensory And Sensorimotor Interactions In Speech Perception, 249.* Studia Musicologica Norvegica, 32, 169–191.

Turi, M., Karaminis, T., Pellicano, E., & Burr, D. (2016). No rapid audiovisual recalibration in adults on the autism spectrum. *Scientific Reports*, *6*, 21756.

Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid recalibration to audiovisual asynchrony. *The Journal of Neuroscience*, *33*(37), 14633-14637.

Van der Burg, E., Alais, D., & Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales. *Scientific Reports*, *5*.

Van der Burg, E., & Goodbourn, P. T. (2015). Rapid, generalized adaptation to asynchronous audiovisual speech. *Proceedings of the Royal Society of London B: Biological Sciences*, *282*(1804), 20143083.

Van der Burg, E., Orchard-Mills, E., & Alais, D. (2015). Rapid temporal recalibration is unique to audiovisual stimuli. *Experimental Brain Research*, *233*(1), 53-59.

Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain Research*, *1111*(1), 134-142.

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Temporal recalibration during asynchronous audiovisual speech perception. *Experimental Brain Research*, *181*(1), 173-181.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*(1), 80-113.

Vroomen, J., Keetels, M., De Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, *22*(1), 32-35.

Long-term music training modulates the recalibration of audiovisual simultaneity

Crescent Jicol [1] , Michael J. Proulx[1], Frank E Pollick [2] , Karin Petrini [1]

[1] Department of Psychology, University of Bath, United Kingdom
[2] School of Psychology, University of Glasgow, United Kingdom

Corresponding Author:
Crescent Jicol
Department of Computer Science
University of Bath
Claverton Down
Bath
BA2 7AY
United Kingdom
Email: C.Jicol@bath.ac.uk

Supplemental Results

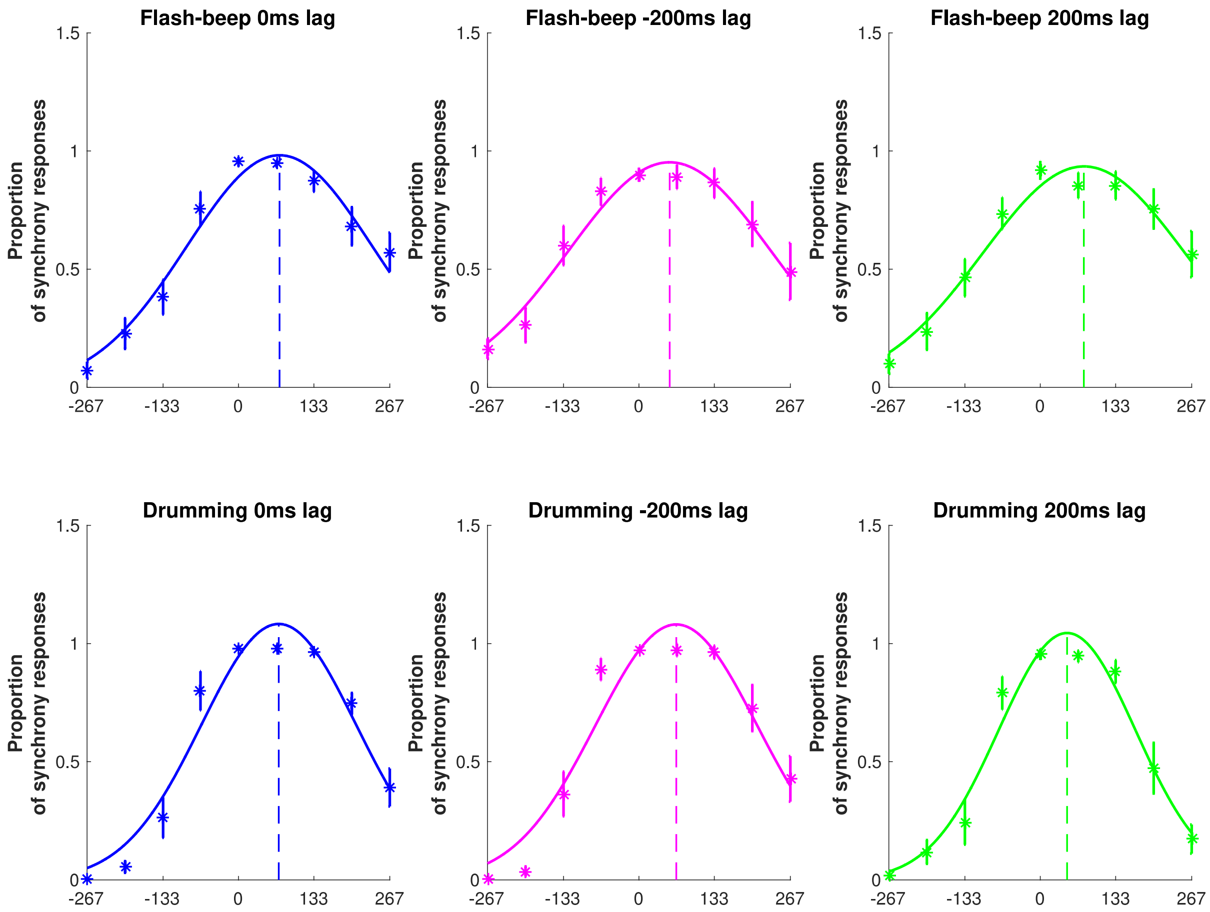

Fig S1. Fit to average proportion of "synchrony" responses as a function of audiovisual SOAs for the musicians (non drummers) group shown separately for no adaptation (blue), -200 (magenta) and 200ms (green) adaptation conditions. Results for drumming displays are shown in the bottom panels and flash-beep displays in the top panels. Solid lines represent the best-fitting Gaussian curves while the asterisks represent the average data at each audiovisual SOA. The peak of the Gaussian curves provides an estimate of the PSS (point of subjective simultaneity), marked by the dashed vertical lines, while the width of the Gaussian represents the TIW (temporal Integration window). The error bars represent the standard error of the mean.

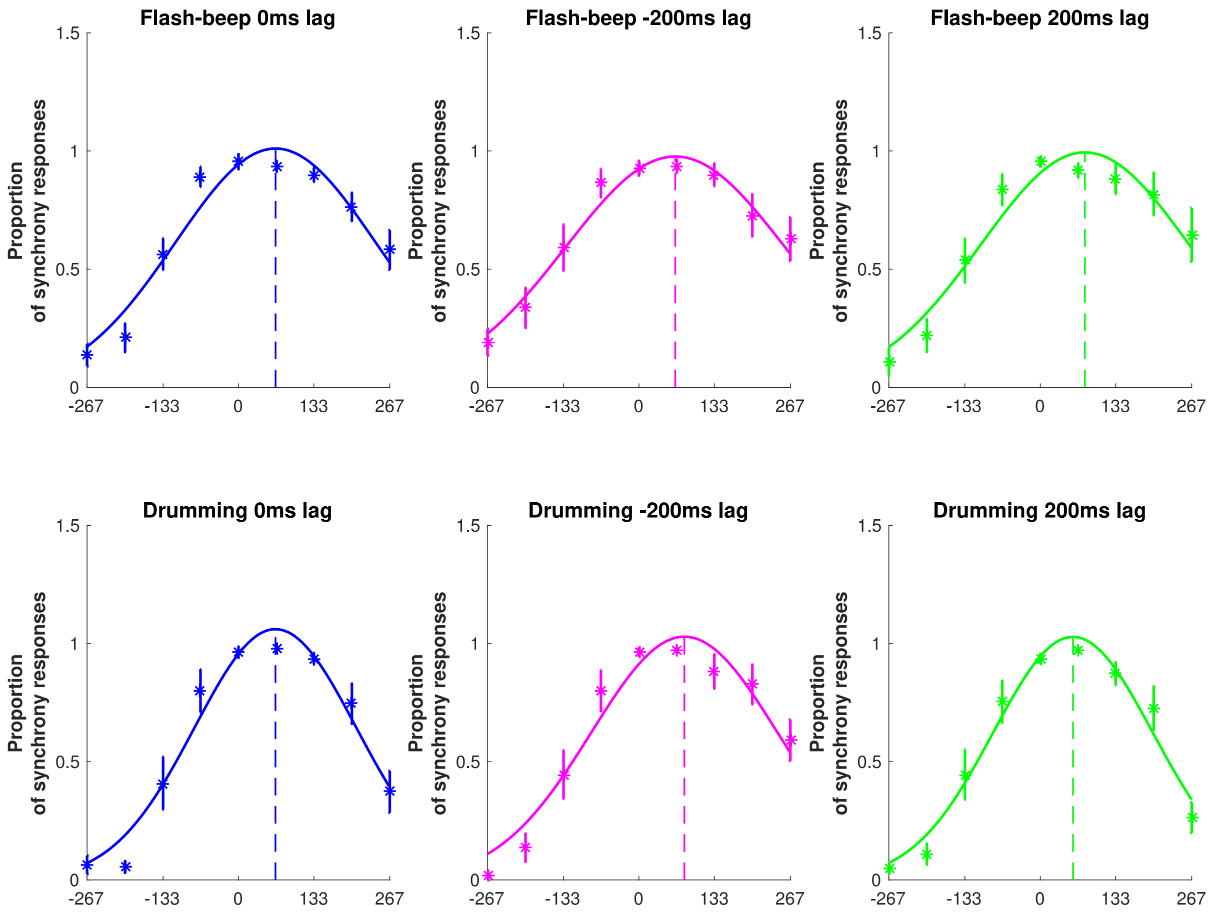

Fig S2. Fit to average proportion of “synchrony” responses as a function of audiovisual SOAs for the non-musicians group shown separately for no adaptation (blue), -200 (magenta) and 200ms (green) adaptation conditions and drumming (bottom panels). Results for drumming displays are shown in the bottom panels and flash-beep displays in the top panels. Solid lines represent the best-fitting Gaussian curves while the asterisks represent the average data at each audiovisual SOA. The peak of the Gaussian curves provides an estimate of the PSS (point of subjective simultaneity), marked by the dashed vertical lines, while the width of the Gaussian represents the TIW (temporal Integration window). The error bars represent the standard error of the mean.

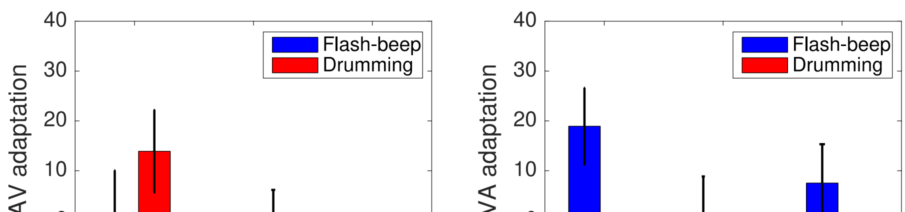

Fig S3. *Left:* PSS shift for non-musicians (NM), musicians (M) and drummers (D) after adaptation with the -200ms auditory leading fixed asynchrony lag (in blue for flash-beep displays and red for drumming displays). *Right:* PSS shift for non-musicians, musicians and drummers after adaptation with the +200ms visual leading fixed asynchrony lag (in blue for flash-beep displays and red for drumming displays). The PSS shift in milliseconds was calculated by subtracting the value of each individual PSS after adaptation from that before adaptation (i.e. from the baseline or PSS before any adaptation took place). The recalibration for musicians and drummers is mostly towards audio-leading asynchrony (negative values) for both adaptation conditions and both displays (drumming and flash-beep). For non-musicians recalibration is mostly towards visual-leading asynchrony (positive values) for both adaptation conditions and both displays (drumming and flash-beep). This trend is shown by the ANOVA results and by Fig. 1 in the manuscript. Error bars show standard error of the mean.

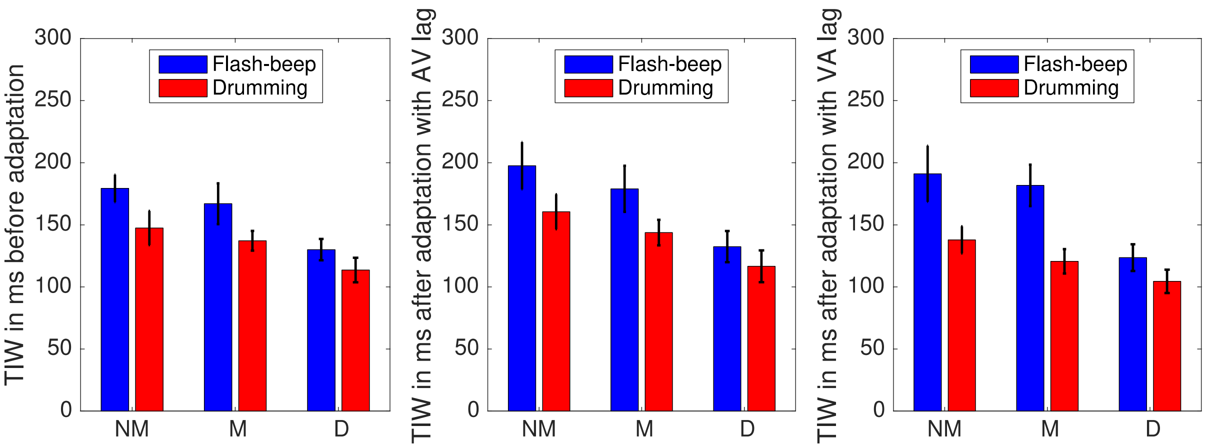

Fig S4. *Left*: TIW width for the non-musicians (NM), musicians (M) and drummers (D) in the baseline condition (before adaptation). *Middle*: TIW width for the non-musician, musicians and drummers after adaptation with the -200ms auditory leading fixed asynchrony lag (in blue for flash-beep displays and red for drumming displays). *Right:* TIW width for the non-musician, musicians and drummers after adaptation with the +200ms visual leading fixed asynchrony lag (in blue for flash-beep displays and red for drumming displays). Error bars show standard error of the mean.

Recalibration analyses for male sample only

The data for the male sample only were analysed with a mixed factorial ANOVA with group (drummers, musicians, and non-musicians) as between-subjects factor and display type (drumming and flash-beep) and adaptation asynchrony (-200 and +200ms) as within-subjects factors. We found a main effect of group, $F(2,15)= 4.860$, $p= .024$, $\eta_p^2 = .393$, a significant interaction between display type and adaptation asynchrony $F(1,15)=24.030$, $p<.001$, $\eta_p^2 = .616$, and a significant interaction of display type and group $F(2,15)=6.606$, $p=.009$, $\eta_p^2 = .468$. All other effects did not reach significance level ($F \leq .573$, $p \geq .071$). Fig. S5, left panel, shows that the effect of recalibration was very similar for drummers and musicians whose PSS shifted to an audio-leading asynchrony. The recalibration effect of non-musicians, however, was very different with their PSS shifting towards video-leading asynchrony. Fig. S5, right panel, shows that the interaction between type of display and adaptation was due to the flash-beep display inducing a PSS shift in the direction of the adapted asynchrony; that is, towards visual-leading asynchrony if the asynchrony used during adaptation had the video leading the auditory or towards audio-leading asynchrony if the asynchrony used during adaptation had the audio leading the video. In contrast, for the drumming display the PSS shifted towards audio-leading when the visual-leading asynchrony was used during the adaptation phase.

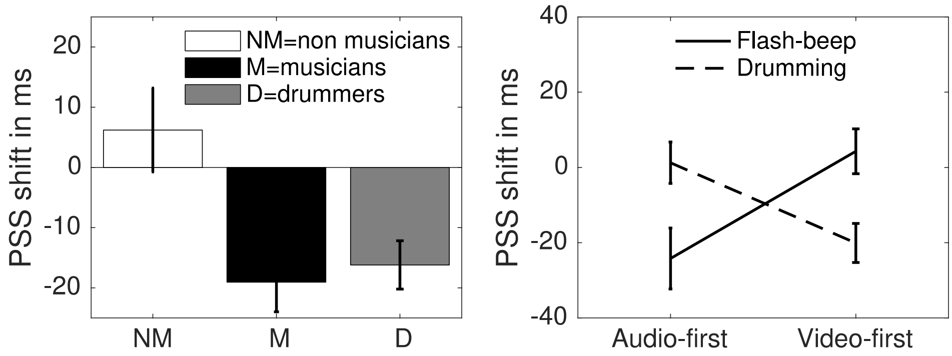

Fig S5. *Left:* PSS shift for non-musicians, musicians and drummers in the only male sample. The PSS shift in milliseconds was calculated by subtracting the value of each individual PSS after adaptation from that before adaptation (i.e. from the baseline or PSS before any adaptation took place). The adaptation for musicians and drummers was in the opposite direction to that of non-musicians. *Right:* Overall PSS shift for flash-beep and drumming displays for the audio-leading and video-leading adaptations. Error bars show standard error of the mean.