



*Citation for published version:*

Jicol, C, Proulx, M, Pollick, F & Petrini, K 2018, 'Long-term music training modulates the recalibration of audiovisual simultaneity', *Experimental Brain Research*, vol. 236, no. 7, pp. 1869–1880.  
<https://doi.org/10.1007/s00221-018-5269-4>

*DOI:*

[10.1007/s00221-018-5269-4](https://doi.org/10.1007/s00221-018-5269-4)

*Publication date:*

2018

*Document Version*

Peer reviewed version

[Link to publication](#)

This is a post-peer-review, pre-copyedit version of an article published in *Experimental Brain Research*. The final authenticated version is available online at: <https://doi.org/10.1007/s00221-018-5269-4>.

**University of Bath**

**Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## **Long-term music training modulates the recalibration of audiovisual simultaneity**

Crescent Jicol <sup>1</sup> , Michael J. Proulx<sup>1</sup>, Frank E Pollick <sup>2</sup> , Karin Petrini <sup>1</sup>

<sup>1</sup> Department of Psychology, University of Bath, United Kingdom

<sup>2</sup> 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.

26 Due to the difference between the speed of light and that of sound, there exist distance-  
27 dependent changes in the times at which visual and auditory stimuli reach the respective  
28 sensory receptors (Arnold, Johnston & Nishida, 2005; Noel, Łukowska, Wallace, & Serino,  
29 2016; Spence & Squire, 2003). Moreover, there are differences between the neural  
30 processing times of these cues (Alais & Charlile, 2005; King, 2005; Schroeder & Foxe,  
31 2004). Nevertheless, for relatively small temporal differences humans are seldom aware of  
32 the asynchrony between these sensory cues thanks to the brain's capacity to shift  
33 (recalibrate) the point at which a person perceives their simultaneity (e.g. Di Luca, Machulla,  
34 & Ernst, 2009; Harrar & Harris, 2008; Keetels & Vroomen, 2007; Van der Burg, Orchard-  
35 Mills, & Alais, 2015; Vatakis, Navarra, Soto-Faraco, & Spence, 2007).

36 Even short exposure times to audiovisual asynchronous stimuli (circa three minutes) can  
37 affect the perceived synchrony of subsequent similar audiovisual stimuli (Fujisaki et al.,  
38 2004; Vroomen, Keetels, De Gelder, & Bertelson, 2004). In fact, a study by Van der Burg,  
39 Alais, and Cass (2013) showed that recalibration to asynchronous stimuli can occur almost  
40 instantaneously, following a single exposure to an asynchronous multisensory event (Simon,  
41 Noel, & Wallace, 2017). This suggests that recalibration could be a fast sensory process,  
42 rather than a higher-level cognitive process (Van der Burg et al., 2013). However, Rohde  
43 and Ernst (2016) showed that asynchronies in visuo-motor tasks, such as delays between a  
44 button press and a visual flash (Rohde & Ernst, 2013), can be compensated with training  
45 and increased perceptual accuracy (higher ability to detect asynchrony), and are subject to  
46 perceived agency (i.e. the prior knowledge that pressing the button is causing the flash to  
47 appear, and thus the flash should follow the button press), suggesting that higher-level  
48 cognitive processes might actually affect recalibration.

49 The effect of multisensory training and perceptual accuracy on recalibration has emerged  
50 from studies focusing on actively sensed modalities, such as those involving motor action in  
51 the recalibration task (Rohde & Ernst, 2013, Rohde, Scheller & Ernst 2014, Rohde & Ernst,

52 2016), rather than on passively sensed audiovisual modalities (Desantis & Haggard, 2016;  
53 Roach, Heron, Whitaker, & McGraw 2010; Vroomen et al., 2004) and on short periods of  
54 exposure rather than long and naturally occurring periods of multisensory training (Noel,  
55 Niew, Van der Burg, & Wallace, 2017; Simon et al., 2017; Van der Burg et al., 2013). Hence,  
56 we do not know whether the changes in recalibration and perceptual accuracy are specific to  
57 sensorimotor tasks or if they are a general multisensory mechanism, and whether it can be  
58 facilitated by long-term multisensory practice (known to affect brain plasticity as well as  
59 perceptual accuracy; Lee & Noppeney, 2011; Petrini et al., 2011).

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

79 recalibration will also allow us to discern whether these processes are performed by the  
80 same mechanism, as Rohde and Ernst (2013) suggest; or if there are two different cognitive  
81 processes which are unequally impacted by long-term expertise with multisensory stimuli.

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

95 Several studies have reported that prolonged and rapid recalibration are two different  
96 processes and independent of each other (Bruns & Röder, 2015; De Nier, Noel, & Wallace,  
97 2017; Van der Burg, Alais, & Cass, 2015; Van der Burg & Goodbourn, 2015; Van der Burg,  
98 Orchard-Mills, & Alais, 2015), suggesting that rapid recalibration is an early sensory effect,  
99 whereas the prolonged recalibration reflects a more cognitive process, here we focused on  
100 prolonged recalibration. Hence, we asked whether long-term music training affects the  
101 higher-cognitive recalibration process (e.g. Desantis & Haggard, 2016; Fujisaki et al., 2004;  
102 Vroomen et al., 2004).

103 We hypothesised that musicians would show a reduced effect of recalibration due to their  
104 increased perceptual accuracy when compared to non-musicians, and that this reduction in  
105 recalibration would be greater after adaptation with a music clip (for which musicians have

106 prior knowledge and sense of agency) than a flash and beep clip. Secondly, we expected  
107 drummers to show an even weaker effect of recalibration with drumming displays, due to  
108 their added motor experience and sense of agency with the stimulus.

109

110

## Method

### 111 *Participants*

112 A total sample size of 24 was calculated for a Cohen's F effect size equal to 0.25 (for a  
113 medium effect size) through a priori type of power analysis for an ANOVA repeated  
114 measures within-between interaction. We used G\*Power 3.1 (Faul, Erdfelder, Lang, &  
115 Buchner, 2007) and assumed a level of power of 0.80, 3 groups, 6 measurements, and an  
116 alpha level of 0.05. We aimed to test more than 8 participants per group and have an equal  
117 number of participants in the three groups. We tested 42 participants in total, but had to  
118 exclude the data for 4 non-musicians and 3 musicians because their performance was at  
119 chance level in at least one of the six testing blocks. . We also had to exclude the data for  
120 another musician because of a technical problem and for another non-musician because he  
121 listened to music for more than six hours every day. This decision was taken based on  
122 evidence that untrained music listeners can at times show similar capabilities to trained  
123 musicians (Bigand & Poulin-Charronnat, 2006). No data for the tested drummers had to be  
124 excluded. The data for eleven drummers (Mean age = 24.45, SD = 1.65, two females), 11  
125 musicians (Mean age = 24.91, SD = 2.32, five females), and 11 non-musicians (Mean age =  
126 21.91, SD = 1.42, eight females) were included in the study. The number of participants is  
127 similar or higher than other studies investigating recalibration effects (e.g. Fujisaki et al.,  
128 2004; Navarra, García-Morera, & Spence, 2012; Noel et al., 2016; Roach et al., 2010;  
129 Vroomen et al., 2004; Petrini et al., 2011). All participants reported normal or corrected-to-  
130 normal vision and hearing. Non-musicians had no experience with playing any instrument.  
131 Musicians and drummers were selected to have at least four years of active music

132 training/practice and have played their instrument for at least 1h per week over the period of  
133 training (e.g. Lee and Noppeney, 2011; Vines et al., 2006). We defined musicians as those  
134 who played any musical instrument besides the drums (Mean = 8.73, SD = 3.58). Drummers  
135 had to have significantly more experience in drumming than any other instrument (at least 2  
136 years more) and preferably to only have played the drums (Mean = 10.64 years, SD = 5.26).  
137 Participants gave informed consent to participate, and the study received ethical approval  
138 from the research ethics board at University of Bath. All subjects gave informed consent to  
139 participate and received cash for their participation.

#### 140 *Apparatus and Stimuli*

141 The flash-beep displays consisted of a pure tone at 2000 Hz and 84 dB mean intensity and a  
142 white dot (luminance: 85 cd/m<sup>2</sup>). These were presented on a black background (luminance:  
143 0.12 cd/m<sup>2</sup>) and were 460 ms in duration. Detailed description of the creation and  
144 characteristics of the drumming point-light displays has been published elsewhere (Petrini et  
145 al., 2009a; Petrini, Russell & Pollick, 2009b; Petrini, Holt & Pollick, 2010). The drumming  
146 displays consisted of a point-light display of a professional jazz drummer playing a simple  
147 swing groove at 120 BPM and accent on the second beat (see examples of clips online).  
148 The 3D motion coordinates were transformed into point-light displays using a Matlab script  
149 with PsychToolbox routines (Brainard, 1997; Pelli, 1997). The matching synthetic sounds  
150 were created using a simulation of the first 25 modes of a circular membrane (Fontana,  
151 Avanzini, & Rocchesso, 2004). This takes as input the time and impact velocity of an impact  
152 and provides the audio signal. The 60Hz movies (AVI) and audio (WAV) were combined in  
153 Adobe Premiere 1.5 to produce the audiovisual displays. The audiovisual displays containing  
154 asynchronous audio and video were generated by either delaying the video with respect to  
155 the audio, or the audio respect to the video, by 67, 133, 200 and 267 ms. The resulting  
156 audiovisual clips were three seconds in duration. All displays were presented in focus and  
157 were preceded by a fixation point. We used a point light display rather than a full clip  
158 because we wanted to avoid possible effects of context as we were interested in the action



159 and kept the low-level information as similar as possible between the flash-beep and the  
160 drumming display.

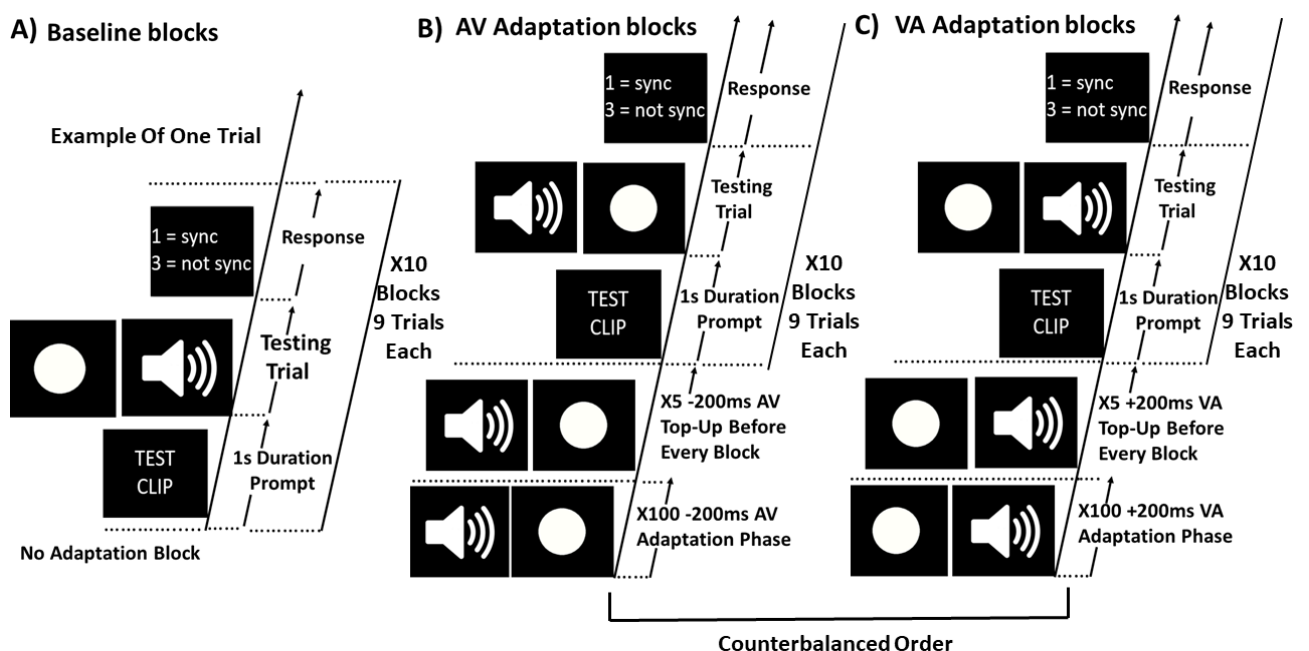
161 All displays were presented via an Apple Macintosh MacPro with Retina display (60 Hz  
162 refresh rate) laptop running OS X 10.9 and an AMD Radeon R9 M370X graphics card with  
163 2GB of GDDR5 memory. The visual cues were displayed on a HannsG HP222 monitor,  
164 which was placed approximately 50 cm from the observer. Auditory cues were presented  
165 through high quality Sennheiser HD 380 pro headphones and the volume at the sound  
166 source was 50 dB intensity for the drumming displays and 55 dB for the flash-beep. The  
167 experiment was controlled using MATLAB 2013b (MATHWORKS Inc., Natick, MA) and the  
168 PsychToolbox (Brainard, 1997; Pelli, 1997).

#### 169 *Procedure*

170 Participants completed a 90-minute experiment composed of six blocks (two baseline blocks  
171 and four adaptation blocks). The first two blocks were aimed at measuring participants' point  
172 of subjective simultaneity before adaptation (i.e., individual baseline). One block presented  
173 the audiovisual drumming displays and the other block the flash-beep display (see clip  
174 examples online). The presentation of these two blocks was counterbalanced across  
175 participants. The displays varied in the level of asynchrony between the visual and the  
176 auditory cue (-266.67, -200, -133.33, -66.67, 0, 66.67, 133.33, 200, 266.67ms; where  
177 negative offsets indicate the audio stream preceded the video stream). For both display  
178 types, each level of asynchrony was repeated 10 times at random for a total number of 90  
179 trials in each block and an overall total number of 180 trials (2 display types X 9 audiovisual  
180 asynchronies X 10 repetitions) for the full study. Participants had to indicate for each trial  
181 whether the audio and video were in synchrony or not by pressing one of two keys on the  
182 computer keyboard (see Fig. 1A). Each one of the subsequent four adaptation blocks (flash-  
183 beep -200ms block (Fig. 1B), flash-beep +200ms block (Fig. 1C), drumming -200ms block,  
184 and drumming +200ms block), started with an adaptation phase and their presentation was  
185 counterbalanced across participants. At the beginning of each block the adaptation phase

186 was conducted by repeating 100 times either the display for which the auditory cue led the  
 187 visual of 200ms (-200) or the display for which the visual cue led the auditory of 200ms  
 188 (200). This duration of the adaptation asynchrony was selected based on previous literature  
 189 (e.g. Fujisaki et al., 2004; Vroomen et al., 2004). During the adaptation phase, participants  
 190 were instructed to carefully watch the repeated displays until the end. To make sure  
 191 participants paid attention to the display during the adaptation phase, they were asked to  
 192 count how many animal pictures were presented during this phase. These images were  
 193 flashed randomly between the SJ trials throughout each testing block. The number of  
 194 pictures changed in each block and participants had to report the number at the end of the  
 195 adaptation phase. After the adaptation phase ended participants were asked, similar to the  
 196 initial two blocks (baseline), to judge the synchrony between audio and video in the 9 clips  
 197 10 times. To ensure adaptation was maintained, before each set of 9 randomly presented  
 198 displays the adaptation display (either -200 or 200ms) was repeated 5 more times (see Fig.  
 199 1B and 1C).

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



203

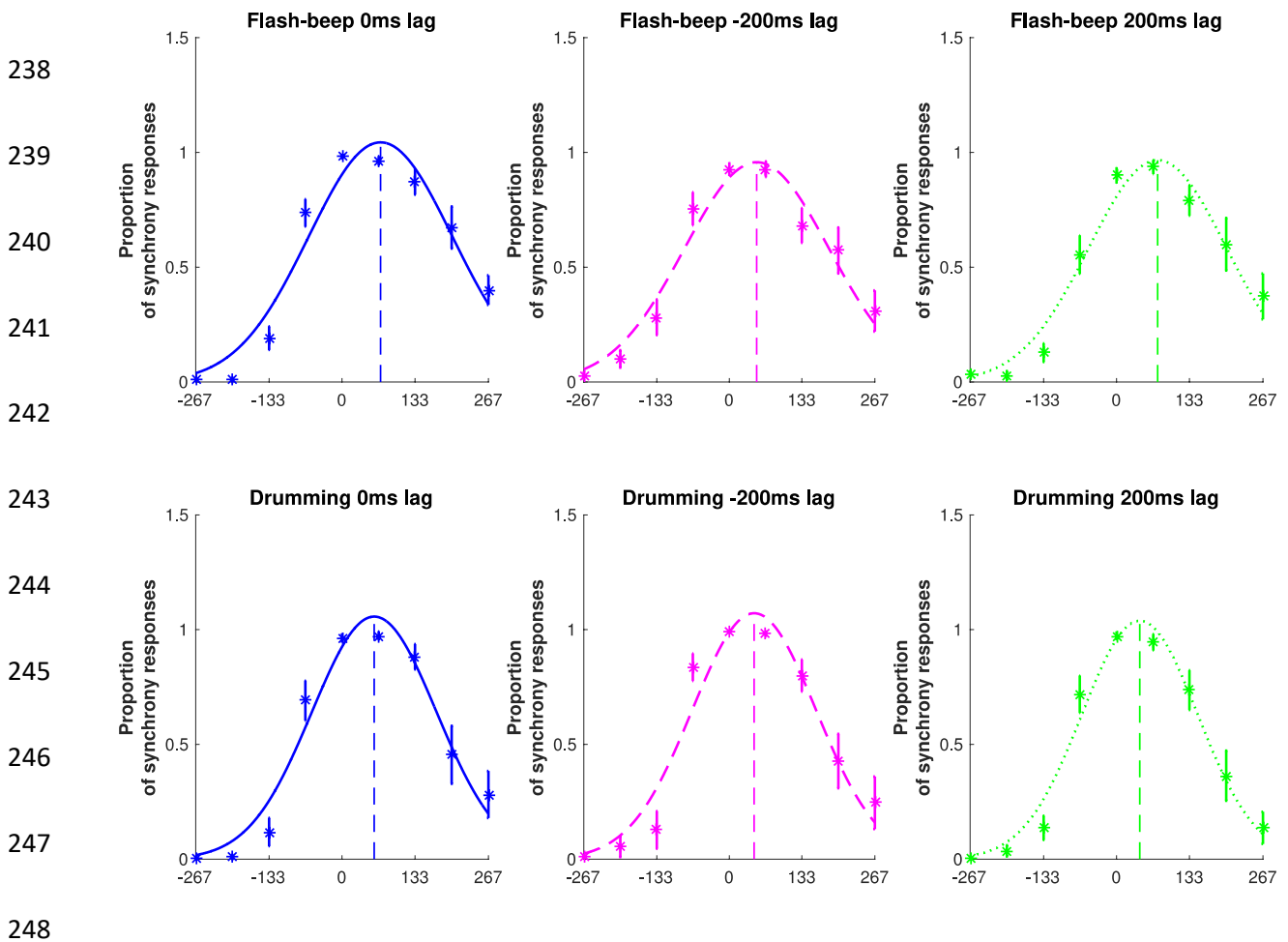
204 Fig. 1 Schematic of display conditions and experimental design for the baseline (left) and for  
205 the adaptation blocks (middle and right). (A) Participants were presented with 10 blocks of 9  
206 trials (corresponding to the 9 levels of audiovisual asynchrony) and were asked to judge if  
207 the sound and video in each trial were in synchrony or not. Prior to the display a prompt was  
208 flashed on the screen for one second. This was done for both flash-beep and drumming  
209 displays, but for simplicity here we show the flash-beep display. (B) Adaptation blocks with  
210 audiovisual drumming and flash-beep displays. Participants were exposed to 100 repetitions  
211 of -200ms AV asynchrony of one display after which 10 blocks of 9 testing trials were again  
212 presented. Before each block of testing trials an adaptation top-up consisting of another 5 -  
213 200ms AV repetitions was also presented, in order to maintain the adaptation throughout the  
214 study duration. For simplicity here we show this procedure for the flash-beep display only, as  
215 it was identical for the drumming display. (C) Adaptation blocks with the visual-audio  
216 drumming and flash-beep displays. Participants were exposed to 100 repetitions of +200ms  
217 VA asynchrony of one display after which 10 blocks of 9 testing trials were again presented.  
218 Before each block of testing trials an adaptation top-up consisting of another 5 +200ms VA  
219 repetitions was also presented, in order to maintain the adaptation throughout the study  
220 duration. For simplicity here we show this procedure for the flash-beep display only, as it  
221 was identical for the drumming display.

### 222 *Analysis Procedure*

223 For both types of displays, the proportion of synchronous responses for each level of  
224 stimulus onset asynchrony (SOA) was fit with a Gaussian probability density function  
225 similarly to several studies that examined audiovisual recalibration effects (e.g., Fujisaki et  
226 al., 2004; Van der burg et al., 2013). From these fits, two parameters of interest were  
227 derived: the point of subjective simultaneity (PSS) and the temporal integration window  
228 (TIW). The PSS represented the level of SOA at which the highest perceived simultaneity  
229 between video and audio was perceived by the individual and it was taken as the peak of the  
230 Gaussian curve. The TIW represents the range of cue onset asynchronies, where

231 participants were not able to reliably identify the physical asynchrony between the cues. We  
 232 estimated participants' TIW using the standard deviation (SD) of the Gaussian fit (e.g. Love  
 233 et al., 2013; Desantis & Haggard, 2016). This procedure was followed for both non-adaption  
 234 and adaption conditions. Please see Fig. 2 for an example of the drummers group and also  
 235 Fig. S1 and S2 in the supplemental material for the average fitting of the musician and non-  
 236 musician groups.

237



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

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

260

261

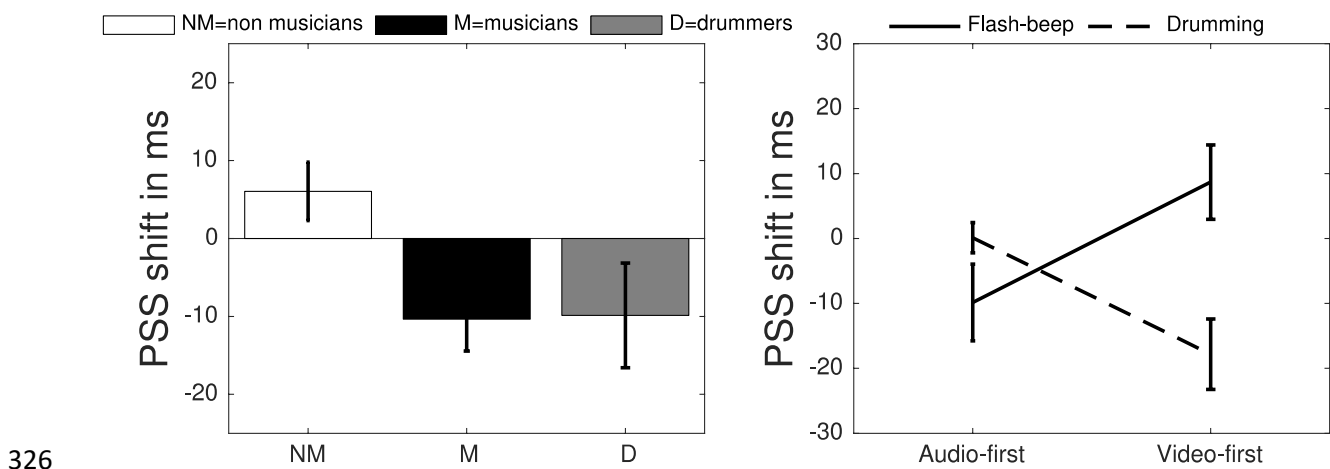
## Results

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

279 To examine the effect of long-term musical training on audiovisual recalibration we first  
280 calculated how much the point of subjective simultaneity (PSS) in the drumming and the  
281 flash-beep display conditions shifted after adaptation by subtracting the value of each

282 individual PSS after adaptation from that before adaptation (PSS from baseline phase). We  
283 calculated the effect of recalibration this way, rather than as a difference in PSS shift  
284 between the two adaptation conditions (e.g. Desantis & Haggard, 2016; Fujisaki et al., 2004;  
285 Vroomen et al., 2004), as we wanted to account for differences in the individuals' initial  
286 ability to detect asynchrony between audio and video (as we know musicians and drummers  
287 have an enhanced ability to detect asynchrony compared to non-musicians; Lee &  
288 Noppeney, 2011; Petrini et al., 2011). However, if we had calculated the recalibration effect  
289 in terms of the difference in PSS shift under the two adaptation conditions, we would have  
290 found very similar values to previous studies (Desantis & Haggard, 2016; Fujisaki et al.,  
291 2004; Navarra et al., 2012; Vroomen et al., 2004). The obtained data were then analysed  
292 with a mixed factorial ANOVA with group (drummers, musicians, and non-musicians) as  
293 between-subjects factor and display type (drumming and flash-beep) and adaptation  
294 asynchrony (-200 and +200ms) as within-subjects factors. We found a main effect of group,  
295  $F(2,30)= 3.440$ ,  $p=.045$ ,  $\eta_p^2 = .187$ , and a significant interaction between display type and  
296 adaptation asynchrony  $F(1,30)=17.986$ ,  $p<.001$ ,  $\eta_p^2 = .375$ . All other effects did not reach  
297 significance level ( $F\leq 1.889$ ,  $p\geq .180$ ). Planned simple contrasts returned no significant  
298 difference between the effect of adaptation for the drummers and the musicians group ( $p =$   
299  $.947$ ; 95% CI [-15, 14.05] based on 1000 bootstrap samples), but showed a significant  
300 difference between the effect of adaptation for drummers and non-musicians ( $p = .033$ , 95%  
301 CI [1.39, 30.45] based on 1000 bootstrap samples) and musicians and non-musicians ( $p =$   
302  $.028$ , 95% CI [1.86, 30.92] based on 1000 bootstrap samples). Fig. 3, left panel, shows that  
303 the effect of recalibration was very similar for drummers and musicians whose PSS shifted to  
304 an audio-leading asynchrony irrespective of the display type and of the adaptation  
305 asynchrony. The recalibration effect of non-musicians, however, was very different with their  
306 PSS shifting towards video-leading asynchrony irrespective of the display type and of the  
307 adaptation asynchrony.

308 Fig. 3, right panel, shows that the interaction between type of display and adaptation was  
 309 due to the flash-beep display inducing a PSS shift in the direction of the adapted  
 310 asynchrony; that is, towards visual-leading asynchrony if the asynchrony used during  
 311 adaptation had the video leading the auditory or towards audio-leading asynchrony if the  
 312 asynchrony used during adaptation had the audio leading the video. This result is in line with  
 313 the previous studies where a simple flash-beep type of display was used and only non-  
 314 musicians (that we know of) were tested (e.g., Desantis & Haggard, 2016; Fujisaki et al.,  
 315 2004; Foss-Feig et al., 2010; Garcia-Perez & Alcala-Quintana, 2012; Shams, Kamitani, &  
 316 Shimojo, 2000; Vroomen et al., 2004). In contrast, for the drumming display the PSS shifted  
 317 towards audio-leading asynchrony when the visual-leading asynchrony was used during the  
 318 adaptation phase. Post hoc paired-samples t-test analyses, Bonferroni corrected, showed  
 319 that there was a significant difference between the effect of visual-leading adaptation for the  
 320 flash-beep and the drumming display ( $t(32)= 3.934, p= .002, 95\% \text{ CI } [14.21, 38.79]$  based on  
 321 1000 bootstrap samples). No difference, in contrast, was found between the effect of audio-  
 322 leading adaptation for the flash-beep and the drumming display ( $t(32)= -1.310, p= .208, 95\%$   
 323  $\text{CI } [-24.30, 4.67]$  based on 1000 bootstrap samples). These results were replicated by  
 324 running the analysis for the male only sample which had the larger number of participants  
 325 (see supplementary material for these additional analyses and figure).



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

334

335 A directional Pearson's correlation was run to test whether the PSS shift towards audio-  
336 leading asynchronies for musicians and drummers increased with years of music training  
337 (Fig. 4, left panel). The results showed that the PSS shift towards audio-leading asynchrony  
338 increased significantly with years of music training when musicians and drummers were  
339 adapted to the visual-leading asynchrony ( $r = -.378$ ,  $p = .042$ ). For the audio-leading  
340 asynchrony however, no such effect was found ( $r = -.144$ ,  $p = .261$ ).

341 Hence, the PSS shift towards audio-leading asynchronies for drummers and musicians was  
342 driven by a change in the recalibration process specific to the adaptation with the visual-  
343 leading asynchrony. Similarly, we examined whether the size and sign of the recalibration  
344 effect decreased with a decrease in the size of the TIW by running a directional Pearson's  
345 correlation separately for the audio-leading and the visual-leading asynchrony (Fig. 4, right  
346 panel). The results showed that the size of the TIW and the recalibration correlated positively  
347 for the adaptation with the audio-leading asynchrony, in that the smaller the TIW the smaller  
348 and more negative was the recalibration effect ( $r = .443$ ,  $p = .005$ ). The same correlation for  
349 the adaptation with the visual-leading asynchrony did not reach significance despite showing  
350 a similar trend ( $r = .264$ ,  $p = .069$ ).

351

352

353

354



355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382

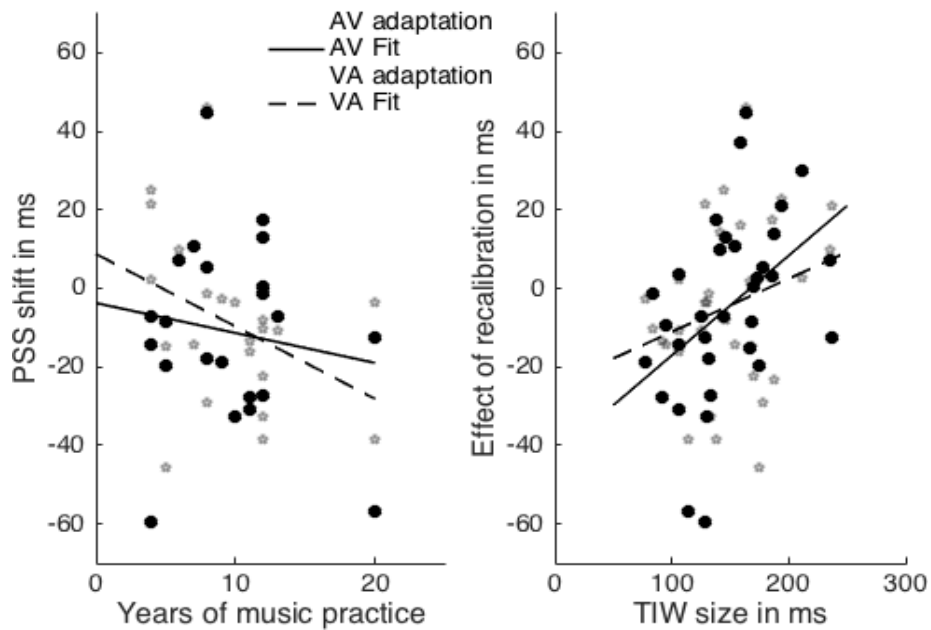


Fig. 4 *Left*. The shift in milliseconds of the point of subjective simultaneity (PSS) plotted against years of music training for the audio-leading adaptation (AV) and visual-leading asynchrony (VA) conditions. Data shown are together for drummers and musicians and drumming and flash-beep displays but 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.

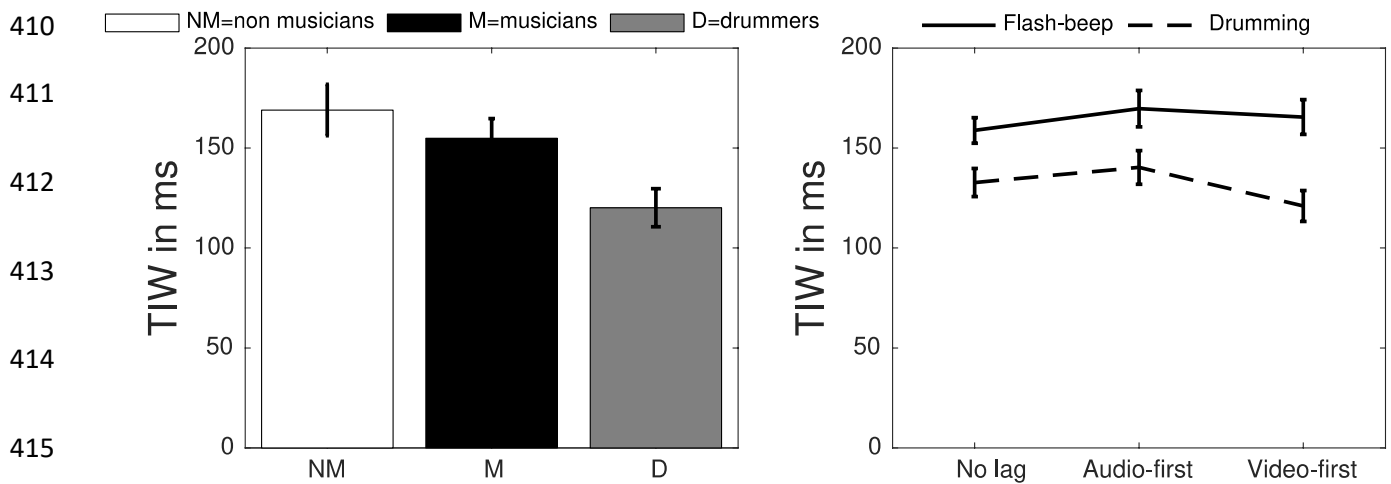
These results further show that the effect of long-term music training on the recalibration process is driven by drummers and musicians shifting their PSS towards audio-leading asynchronies and that this effect is linked to a narrowing of the TIW (see Fig. 5).

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,  $F(2,30)= 5.394$ ,  $p= .010$ ,

383  $\eta_p^2 = .264$ , a significant effect of display type,  $F(1,30)= 21.908$ ,  $p < .001$ ,  $\eta_p^2 = .422$ , a  
384 significant effect of adaptation lag,  $F(2,60)= 3.834$ ,  $p = .027$ ,  $\eta_p^2 = .113$ , and a significant  
385 interaction between display type and adaptation lag  $F(2,60)= 4.135$ ,  $p = .021$ ,  $\eta_p^2 = .121$ . All  
386 other effects did not reach significance level ( $F < 1.299$ ,  $p > .80$ ). Planned simple contrasts  
387 returned a significant difference between the size of the TIW for the drummers and the  
388 musicians group ( $p = .031$ , 95% CI [3.43, 66.02] based on 1000 bootstrap samples), with  
389 drummers showing a smaller TIW, and thus greater ability to detect asynchrony, than the  
390 other musicians, and a significant difference between the size of TIW for drummers and non-  
391 musicians ( $p = .003$ , 95% CI [17.58, 80.13] based on 1000 bootstrap samples), with  
392 drummers showing a far smaller TIW than non-musicians. Fig. 5 shows the decrease in TIW  
393 width (or increase in asynchrony detection ability) when going from non-musicians to  
394 musicians and then to drummers. The significant effect of display type was due to drumming  
395 displays leading overall to a smaller TIW ( $M = 131.34$  and  $SD = 39.54$ ) than flash-beep  
396 ( $M = 164.64$  and  $SD = 56.90$ ). Post hoc paired-samples t-test analyses, Bonferroni corrected,  
397 showed that the effect of adaptation lag was a consequence of the audio-leading  
398 asynchrony widening participants' TIW (decreasing their asynchrony detection ability) when  
399 compared to the video-leading lag ( $t(32) = 3.330$ ,  $p = .006$ , 95% CI [4.56, 18.93] based on  
400 1000 bootstrap samples). The significant interaction between display type and adaptation lag  
401 was due to visual-leading asynchrony resulting in the smaller TIW with respect to no lag  
402 ( $t(32) = 2.876$ ,  $p = .042$ , 95% CI [3.42, 20.07] based on 1000 bootstrap samples) and audio-  
403 leading adaptation ( $t(32) = 4.44$ ,  $p < .001$ , 95% CI [10.45, 28.18] based on 1000 bootstrap  
404 samples) for the drumming display but not for the flash-beep display ( $t(32) = -1.323$ ,  $p = .195$ ).  
405 This result is similar to that found for the recalibration effect, in that the adaptation with a  
406 visual-leading asynchrony has a strong effect on the drumming display but not on the flash-  
407 beep display condition.

408

409



416 Fig. 5. *Left.* TIW for the non-musician (NM), musicians (M) and drummer group (D).

417 Drummers showed the narrowest TIW, followed by musicians, whereas non-musicians  
 418 showed the widest TIW (please see Fig. S4 for a breakdown of the Figure into the different  
 419 conditions). *Right.* TIW width for the flash-beep and drumming displays before adaptation,  
 420 and after adaptation with audio-leading and visual-leading asynchrony. Error bars show  
 421 standard error of the mean.

422

### 423 Discussion

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

436 (Lee & Noppeney, 2011; Petrini et al., 2011) but also modulates the adaptive recalibration  
437 process. Additionally, because musicians and drummers showed the narrowest TIW but not  
438 the weakest adaptation, this suggests that greater perceptual accuracy cannot fully explain  
439 changes in the examined recalibration process, as it has been suggested before (Noel et al.,  
440 2016; Rohde & Ernst, 2013; Van De Burg et al., 2013). Furthermore, while musicians and  
441 drummers showed the same recalibration effect, they did not show the same perceptual  
442 accuracy (i.e. drummers were significantly more accurate). Hence, our results suggest that  
443 whilst these processes might have overlapping mechanisms, they are also independent.

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

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

460 Van der Burg et al. (2013) showed that for rapid recalibration with audiovisual stimuli, the  
461 size of the TIW and the recalibration effect are directly proportional (see also Noel et al.,  
462 2016). In their 2013 study, Rohde and colleagues also showed that this correlation between

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

490 a strong conclusion on whether it is the music training that directly affects the brain  
491 recalibration mechanism or whether it is the refinement of perceptual abilities following music  
492 training that affects this mechanism. Future studies could tackle this question by examining,  
493 for example, performance on judgements of simultaneity by musicians with different levels of  
494 perceptual accuracy but similar training.

495 The reason why musicians and drummers consistently recalibrated their perceived  
496 synchrony between sound and vision towards audio-leading asynchronies after adaptation is  
497 unclear. It has been shown that having predictable targets and training increases motor  
498 anticipation and recalibration in sensorimotor tasks (Rohde, van Dam, & Ernst, 2014). In the  
499 present study no active motor task was used, however, musicians have been shown to have  
500 a higher ability to predict the arrival of auditory information by filling in missing visual  
501 information with their acquired motor repertoire (Petrini et al., 2009b). For example,  
502 drummers can predict when a drumming impact will occur and judge the asynchrony  
503 between visual information and sound even if the visual movement of the drummer is  
504 missing/occluded. In other words, musicians have enhanced abilities to predict when a  
505 sound should occur based on their long-term sensorimotor training (Lee and Noppeney,  
506 2011; Petrini et al., 2011). Interestingly, when predicting the time of impact based on missing  
507 visual information the perceived synchrony of drummers shifted from visual-leading to audio-  
508 leading asynchrony (Petrini et al., 2009b), similarly to our present results. The explanation of  
509 why in musicians the sound needs to lead the video to perceive simultaneity after adaptation  
510 with visual-leading asynchrony can reside in their ability to map the sound occurrence based  
511 on the learnt action (Lee & Noppeney, 2011; Petrini et al., 2009b; Desantis & Haggard,  
512 2016). That is, musicians may not rely on vision (as in Petrini et al., 2009b) and may predict  
513 and anticipate the arrival of the sound based on their audio-motor mapping process (Lee and  
514 Noppeney, 2011; Petrini et al., 2009b) as suggested by tapping studies showing that touch  
515 needs to precede the other stimuli to perceive synchrony (Aschersleben & Prinz, 1995;  
516 Miyake, Onishi, & Pöppel, 2004; Repp & Su, 2013). If musicians were using motor

517 simulation/mapping in place of visual information to decide whether visual and auditory  
518 information were synchronised they would anticipate the sound occurrence with respect to  
519 the visual stimulus (to coincide with their anticipated motor event) and report synchronization  
520 when the sound preceded the visual information. Non-musicians in turn might not use this  
521 sensorimotor mapping (Lee & Noppeney, 2011) and consequently show overall the usual  
522 bias found in synchrony perception towards visual-leading asynchronies (e.g. Love, Petrini,  
523 Cheng, & Pollick, 2013). Hence, the adaptation to fixed audiovisual lags could exacerbate  
524 these existing differences in synchrony perception between musicians and non-musicians.

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

543 Drummers showed the highest perceptual accuracy, followed by the other musicians and the  
544 non-musicians. This result replicates and extends previous findings showing that long-term  
545 music training strongly enhances perceptual accuracy (e.g. Lee & Noppeney, 2011; Petrini  
546 et al., 2009a; Petrini et al., 2011), and showing that the type of instrument played affects this  
547 enhancement process at least in the temporal domain. This result was again general for  
548 flash-beep and drumming clips rather than stronger for the drumming clips (for which  
549 drummers have increased sense of agency and motor repertoire; Calvo-Merino et al.,  
550 2005,2006; Rohde & Ernst, 2016). Numerous studies have emphasised the role of rhythm  
551 maintenance when playing a percussion instrument, such as the drums (Botella, 2008;  
552 Flatischler, 1992; Nichols, 2012). This aspect is not as important in the large majority of  
553 instruments which can produce melody (e.g. piano, flute, guitar etc.). Drummers are also  
554 responsible with maintaining the rhythm and synchronicity between instruments in a band  
555 (Nichols, 2012), which may explain why drummers perform better than both other musicians  
556 and non-musicians in multisensory simultaneity judgement tasks (Bishop & Goebel, 2014;  
557 Hodges et al., 2005; Petrini, Dahl et al., 2009; Petrini, Russell et al., 2009; Vatakis &  
558 Spence, 2006; Lee and Noppeney, 2011).

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



570 Petrini et al., 2010) we showed that with cyclic stimuli temporal order judgements (TOJ)  
571 become really difficult and at times impossible for both drummers and non-musicians (Petrini  
572 et al., 2010). Indeed, in Petrini et al 2010 we showed that although drummers were more  
573 precise than non-musicians in both SJ and TOJ tasks when using the drumming displays  
574 used here, there were still drummers and non-musicians unable to perform the TOJ task.  
575 This means that the shift of the participants' PSS towards auditory-leading asynchronies for  
576 drumming displays could have been a consequence of their inability to discriminate what  
577 sense was coming first during adaptation. In other words, the adaptation might not have  
578 been effective with the drumming displays because the sensory order of the asynchrony  
579 used during adaptation was unclear for that stimulus condition. We would also expect based  
580 on our previous findings (e.g. Love et al., 2013; Petrini et al., 2010) that this uncertainty  
581 during the adaptation phase would affect more the +200ms than the -200ms adaptation  
582 condition as we know that participants even for cyclic and complex stimuli are quite good at  
583 judging the temporal order for large auditory-first asynchronies while for large vision-leading  
584 asynchronies participants are not as good (Petrini et al., 2010). When looking at Figure 3  
585 right panel and at its breakdown in the supplementary material the shift towards auditory first  
586 was indeed stronger and more common across the groups for visual leading adaptation  
587 (+200ms) than auditory-leading adaptation (-200ms). Also visual-leading adaptation did have  
588 a stronger effect than auditory-leading adaptation on TIW for drumming displays but not  
589 flash-beep thus supporting this argument. That said what is still unclear is why participants  
590 recalibrated to auditory leading PSS more when they were less sure of the sensory order in  
591 the visual-leading adaptation condition than when they were more sure of it in the auditory-  
592 leading adaptation condition. That is, we would expect that for auditory-leading adaptation all  
593 groups will show a PSS shift towards auditory-leading asynchronies if in this condition the  
594 adaptation with the drumming displays was generally more efficient, but the only group that  
595 showed this trend was the drummers group while the non-musicians PSS shifted towards  
596 vision-leading asynchronies. Future studies could run both TOJ and SJ tasks with similar  
597 recalibration tasks and groups to the present study to examine the contribution of the cyclic

598 nature of the stimulus to the recalibration process. This would help to understand when the  
599 recalibration process is disrupted, especially considering that cyclic stimuli are very common  
600 in everyday life.

601 Interestingly, under both adaptation conditions (although more for auditory-leading  
602 adaptation), a shift towards an audio-leading perceived synchrony was accompanied by an  
603 increased perceptual accuracy or narrowing of the TIW (Noel et al., 2016; Rohde & Ernst,  
604 2013). Furthermore, in general the average TIW of the drumming displays was narrower  
605 than that of the flash-beep. We do not know as yet why participants had a higher level of  
606 perceptual accuracy for the drumming display when compared to the flash-beep display,  
607 what we do know is that this is not the first time this result was found with the same stimuli  
608 (Love et al., 2013) when using simultaneity judgements. We assume that this is due to  
609 differences in level of complexity and amount of information between the two stimuli,  
610 however, because this is the first study examining the recalibration effect for stimuli with very  
611 different levels of complexity, future studies could further examine how the brain uses these  
612 features to flexibly recalibrate to audiovisual asynchrony, by including a higher number of  
613 natural and complex stimuli.

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

622

623

624 Acknowledgements

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

627 Conflict of interest

628  
629 The authors declare no competing financial interests.

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

## References

649

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

653 Arnold, D. H., Johnston, A., & Nishida, S. (2005). Timing sight and sound. *Vision Research*,  
654 45(10), 1275-1284.

655 Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the  
656 musical capacities that do not depend on formal musical training. *Cognition*, 100(1),  
657 100-130.

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

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

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

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

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

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

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

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

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

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

681 Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\*Power 3: a flexible statistical  
682 power analysis program for the social, behavioral, and biomedical sciences. *Behav*  
683 *Res Methods*, 39(2), 175-191.

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

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

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

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

695

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

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

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

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

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

709

710 King, A. J. (2005). Multisensory integration: strategies for synchronization. *Current Biology*,  
711 15(9), R339-R341.

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

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

718 Navarra, J., García-Morera, J., & Spence, C. (2012). Temporal adaptation to audiovisual  
719 asynchrony generalizes across different sound frequencies. *Frontiers in Psychology*,  
720 3, 152.

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

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

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

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

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

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

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

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

742 Petrini, K., Pollick, F. E., Dahl, S., McAleer, P., McKay, L., Rocchesso, D., ... & Puce, A.  
743 (2011). Action expertise reduces brain activity for audiovisual matching actions: an  
744 fMRI study with expert drummers. *Neuroimage, 56*(3), 1480-1492.

- 745 Petrini, K., Russell, M., & Pollick, F. (2009). When knowing can replace seeing in audiovisual  
746 integration of actions. *Cognition*, 110(3), 432-439.
- 747 Proverbio, A. M., Attardo, L., Cozzi, M., & Zani, A. (2015). The effect of musical practice on  
748 gesture/sound pairing. *Frontiers in Psychology*, 6.
- 749 Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: a review of recent research  
750 (2006–2012). *Psychonomic Bulletin & Review*, 20(3), 403-452.
- 751 Roach, N. W., Heron, J., Whitaker, D., & McGraw, P. V. (2010). Asynchrony adaptation  
752 reveals neural population code for audio-visual timing. *Proceedings of the Royal  
753 Society of London B: Biological Sciences*, rspb20101737.
- 754 Rohde, M., & Ernst, M. O. (2013). To lead and to lag—forward and backward recalibration of  
755 perceived visuo-motor simultaneity. *Frontiers in Psychology*, 3, 599.
- 756 Rohde, M., & Ernst, M. O. (2016). Time, agency, and sensory feedback delays during action.  
757 *Current Opinion in Behavioral Sciences*, 8, 193-199.
- 758 Rohde, M., Scheller, M., & Ernst, M. O. (2014). Effects can precede their cause in the sense  
759 of agency. *Neuropsychologia*, 65, 191-196.
- 760 Rohde, M., van Dam, L. C., & Ernst, M. O. (2014). Predictability is necessary for closed-loop  
761 visual feedback delay adaptation. *Journal of Vision*, 14(3), 4-4.
- 762 Roseboom, W., & Arnold, D. H. (2011). Twice upon a time. Multiple concurrent temporal  
763 recalibrations of audiovisual speech. *Psychological Science*, 22(7), 872-877.
- 764 Schroeder, C. E., & Foxe, J. J. (2004). 18 Multisensory Convergence in Early Cortical  
765 Processing. *The Handbook of Multisensory Processes*, 295.
- 766 Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you  
767 hear. *Nature*, 408(6814), 788.



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

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

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

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

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

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

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

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

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

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

793 Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Temporal recalibration during  
794 asynchronous audiovisual speech perception. *Experimental Brain Research*, 181(1),  
795 173-181.

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

798 Vroomen, J., Keetels, M., De Gelder, B., & Bertelson, P. (2004). Recalibration of temporal  
799 order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*,  
800 22(1), 32-35.

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820 **Long-term music training modulates the recalibration of audiovisual simultaneity**

821

822 Crescent Jicol<sup>1</sup>, Michael J. Proulx<sup>1</sup>, Frank E Pollick<sup>2</sup>, Karin Petrini<sup>1</sup>

823

824 <sup>1</sup> Department of Psychology, University of Bath, United Kingdom

825 <sup>2</sup> School of Psychology, University of Glasgow, United Kingdom

826

827

828

829

830

831 **Corresponding Author:**

832 Crescent Jicol

833 Department of Computer Science

834 University of Bath

835 Claverton Down

836 Bath

837 BA2 7AY

838 United Kingdom

839 Email: C.Jicol@bath.ac.uk

840

841

842

843

844

845

846

847

848

849

850

851

852

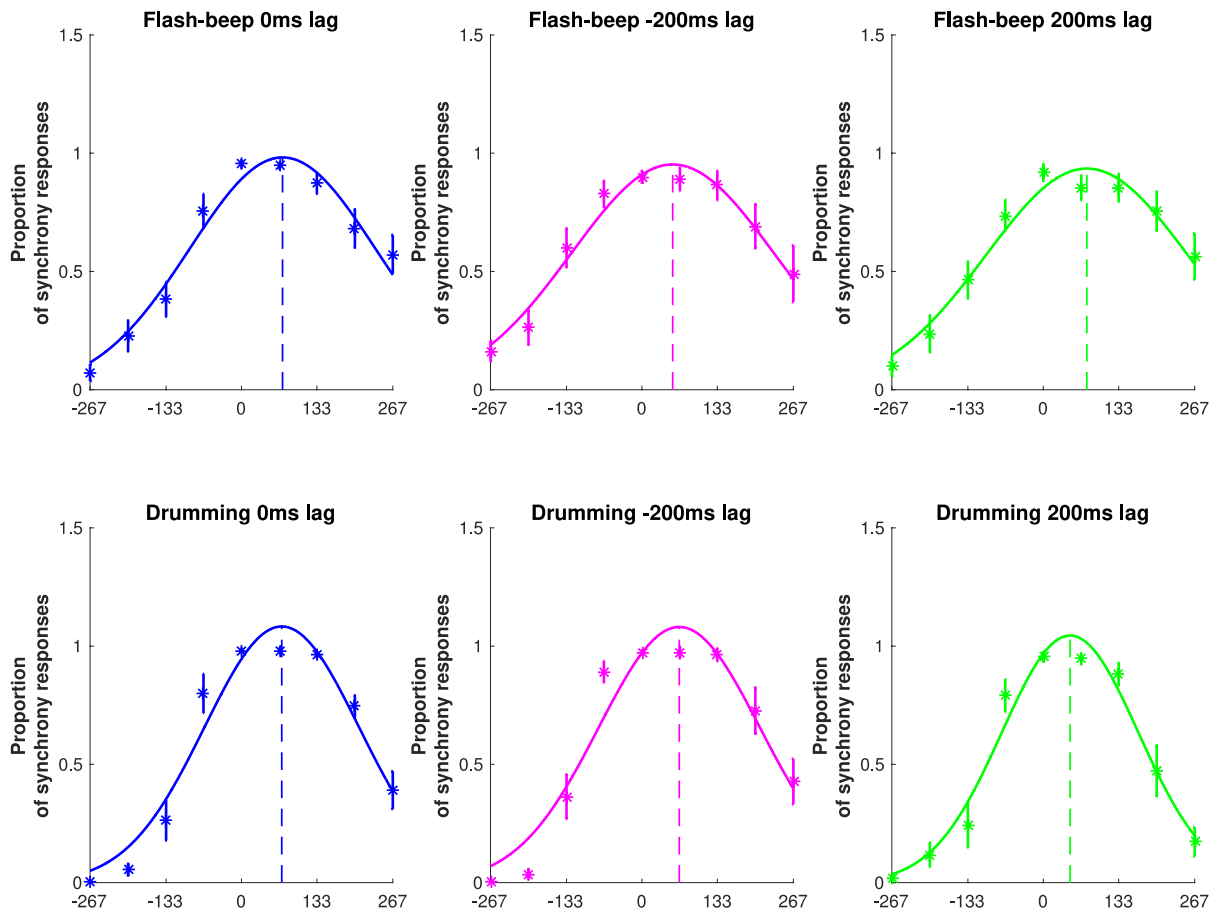
853

854

855

856 Supplemental Results

857



858

859 **Fig S1.** Fit to average proportion of “synchrony” responses as a function of audiovisual  
 860 SOAs for the musicians (non drummers) group shown separately for no adaptation (blue), -  
 861 200 (magenta) and 200ms (green) adaptation conditions. Results for drumming displays are  
 862 shown in the bottom panels and flash-beep displays in the top panels. Solid lines represent  
 863 the best-fitting Gaussian curves while the asterisks represent the average data at each  
 864 audiovisual SOA. The peak of the Gaussian curves provides an estimate of the PSS (point  
 865 of subjective simultaneity), marked by the dashed vertical lines, while the width of the  
 866 Gaussian represents the TIW (temporal Integration window). The error bars represent the  
 867 standard error of the mean.

868

869

870

871

872

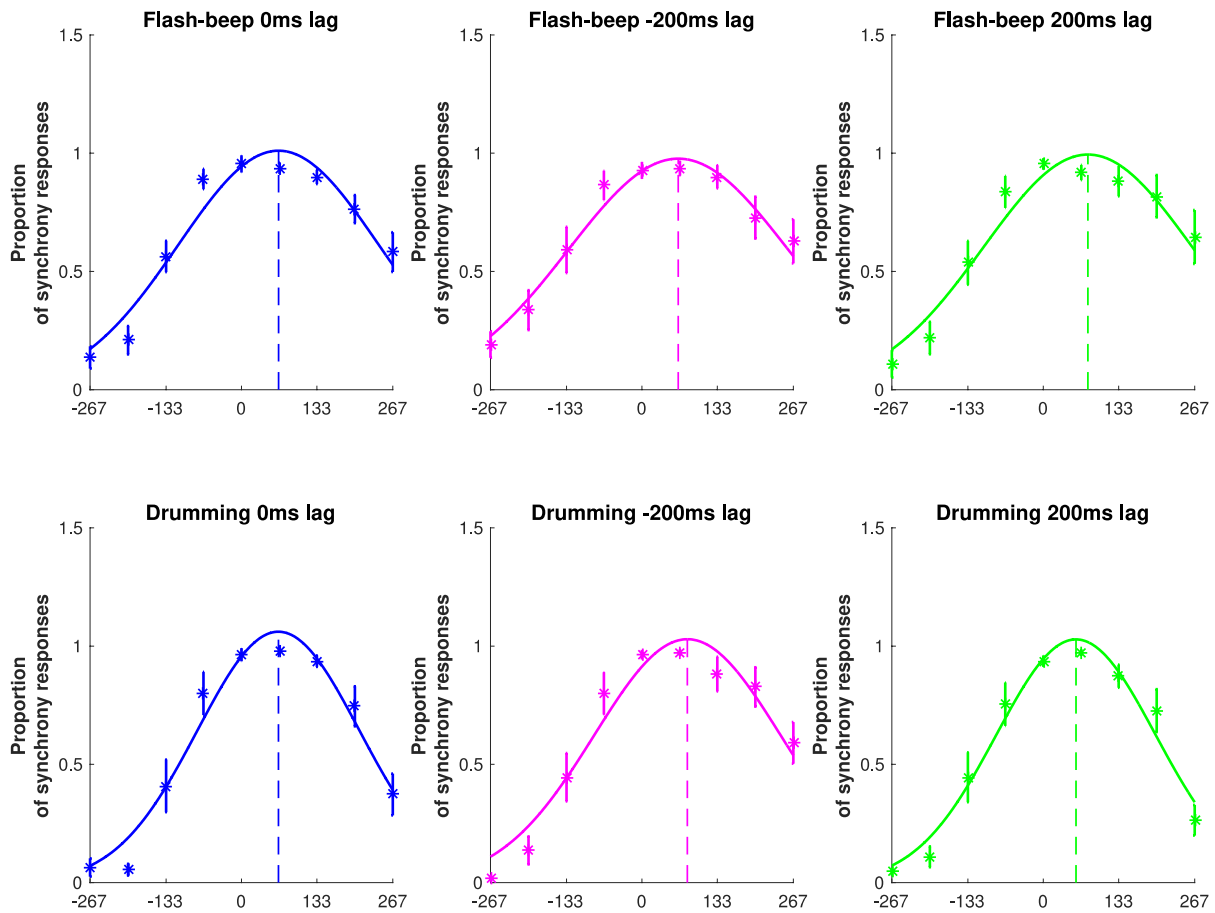
873

874

875

876

877



878

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

888

889

890

891

892

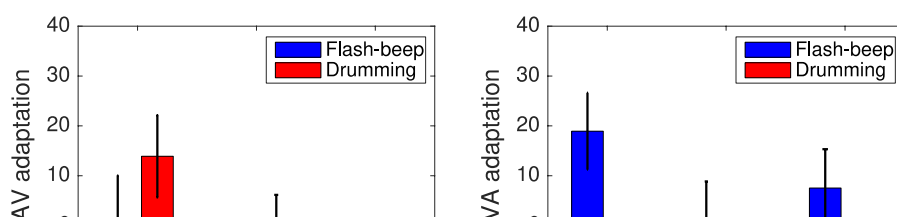
893

894

895

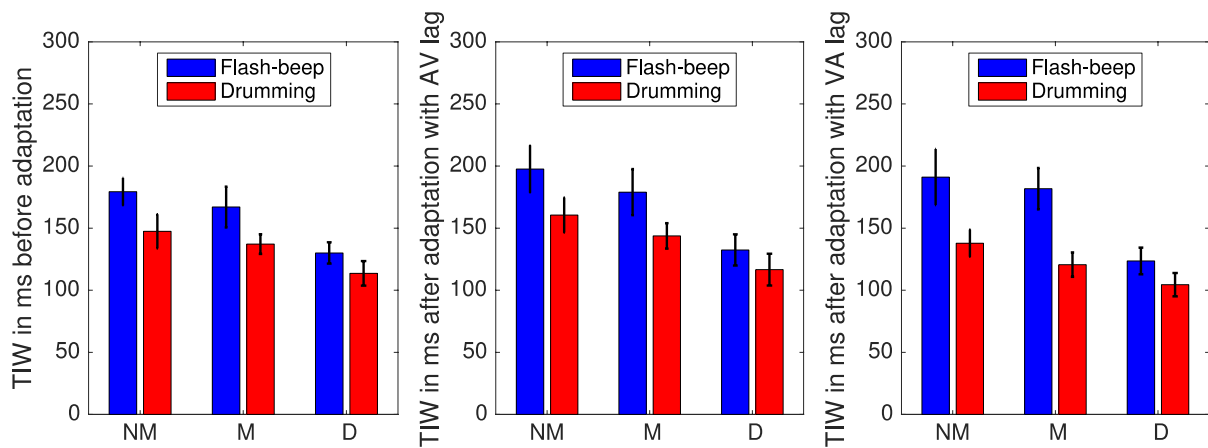
896

897



898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919

**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.



920  
921  
922  
923  
924  
925  
926  
927  
928

**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.

929 Recalibration analyses for male sample only

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

947

948

949

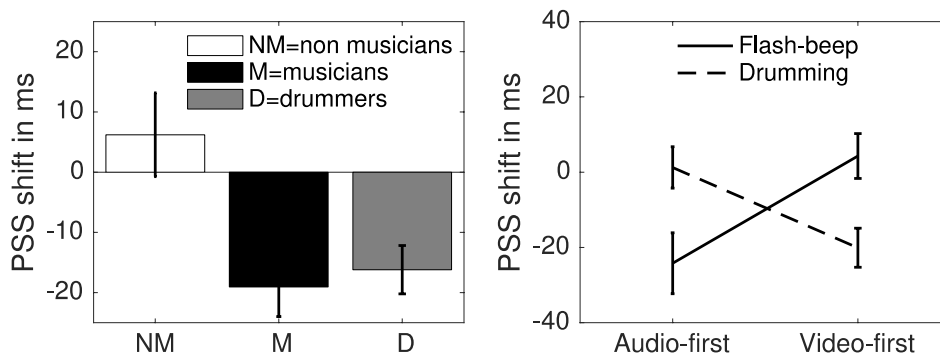
950

951

952

953

954



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

962

963