

## Electrophysiological correlates and psychoacoustic characteristics of hearing-motion synaesthesia

Article (Accepted Version)

Rothen, Nicolas, Bartl, Gergely Janos, Franklin, Anna and Ward, Jamie (2017)  
Electrophysiological correlates and psychoacoustic characteristics of hearing-motion  
synaesthesia. *Neuropsychologia*, 106. pp. 280-288. ISSN 0028-3932

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/70227/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

### **Copyright and reuse:**

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

## **Electrophysiological Correlates and Psychoacoustic Characteristics of Hearing-Motion Synaesthesia**

**Nicolas Rothen\*<sup>1,3</sup>, Gergely Bartl\*<sup>2,4</sup>, Anna Franklin<sup>2</sup>, and Jamie Ward<sup>2,3</sup>**

1) University of Bern, Switzerland

2) School of Psychology, University of Sussex, Brighton, UK

3) Sackler Centre for Consciousness Science, University of Sussex, Brighton, UK

4) Department of Psychology, University of Roehampton, London, UK

\* Equal contribution / shared first authorship

### **Running Head: Hearing-motion synaesthesia**

#### Address for correspondence:-

Prof. Jamie Ward,

School of Psychology,

University of Sussex,

Falmer, Brighton,

BN1 9QH, U.K.

Tel. : +44 (0)1273 876598

Fax. : +44 (0)1273 678058

E-mail : jamiew@sussex.ac.uk

**Abstract**

People with hearing-motion synaesthesia experience sounds from moving or changing (e.g. flickering) visual stimuli. This phenomenon may be one of the most common forms of synaesthesia but it has rarely been studied and there are no studies of its neural basis. We screened for this in a sample of 200+ individuals, and estimated a prevalence of 4.2%. We also document its characteristics: it tends to be induced by physically moving stimuli (more so than static stimuli which imply motion or trigger illusory motion); and the psychoacoustic features are simple (e.g. “whooshing”) with some systematic correspondences to vision (e.g. faster movement is higher pitch). We demonstrate using event-related potentials that it emerges from early perceptual processing of vision. The synaesthetes have a higher amplitude motion-evoked N2 (165-185 msec), with some evidence of group differences as early as 55-75 msec. We discuss similarities between hearing-motion synaesthesia and previous observations that visual motion triggers auditory activity in the congenitally deaf. It is possible that both conditions reflect the maintenance of multisensory pathways found in early development that most people lose but can be retained in certain people in response to sensory deprivation (in the deaf) or, in people with normal hearing, as a result of other differences (e.g. genes predisposing to synaesthesia).

**Keywords:**

Hearing; visual motion; EEG; synaesthesia/synesthesia; multisensory.

## **Introduction**

The movement of objects in the external world is processed through multiple sensory channels; notably vision, audition, and touch. This requires both specialised routines within each sensory system for detecting motion, as well as the ability to compare and contrast motion signals from different senses to determine whether they reflect a common event (i.e. resulting in a bound percept) or multiple events (Soto-Faraco, Spence, Lloyd, & Kingstone, 2004). As such, motion perception is a highly constructive process. This can lead to illusory perception of motion arising because the brain makes a 'best guess' from ambiguous or contradictory signals. In this paper, we also propose that this constructive nature of motion perception can, in certain individuals, lead to visual motion habitually giving rise to synaesthetic auditory experiences – a phenomenon termed hearing-motion synaesthesia (Saenz & Koch, 2008). We determine the prevalence and characteristics of this type of synaesthesia and show, using EEG, that it reflects early perceptual differences in the visual system.

To illustrate how motion perception is constructed from different multisensory signals, consider the stream-bounce illusion (Sekuler, Sekuler, & Lau, 1997). In this illusion, two moving lines are typically perceived to approach each other and then pass through each other ('streaming'), obeying the Gestalt law of good continuity. However, the presence of an auditory beep as the lines come together can alter the visual percept to one of 'bouncing' in which the lines reverse their direction of motion. In this instance, perceptual knowledge of the world (i.e., that collisions are often accompanied by sounds) alters the percept of visual motion.

There are multiple pathways in which visual and auditory motion-relevant information converge, and different stimuli might preferentially engage different pathways (Ursino, Cuppini, & Magosso, 2014). Connections between primary auditory

and primary visual cortex have been postulated to underlie the double-flash illusion in which a rapid double beep and a single flash give rise to an illusory visual experience (Watkins, Shams, Tanaka, Haynes, & Rees, 2006). Considering higher visual regions, sounds with ascending/descending pitch can activate, in fMRI, the region MT/V5 which is generally thought to code 'visual' motion (Sadaghiani, Maier, & Noppeney, 2009). In this instance, the sound source is not moving but movement is implied through other acoustic features. Certain static visual stimuli that induce a percept of motion have also been shown to activate this region including the Rotating Snakes illusion (Kuriki, Ashida, Murakami, & Kitaoka, 2008). Purely linguistic information describing motion does not activate this region but activates a 'higher' convergence zone in parietal cortex (Sadaghiani et al., 2009), and parietal regions appear to be essential for creating the bound percept in the stream-bounce illusion, as shown by TMS (Maniglia, Grassi, Casco, & Campana, 2012). Further still, regions involved in semantic memory may drive auditory and visual imagery (e.g. imagining the sound of a horse galloping) through top-down activation of sensory regions (Zvyagintsev et al., 2013).

The constructive nature of motion perception via multiple signals might give rise not only to various illusions (experienced by almost all people), but also to fundamentally different perceptual experiences (experienced by a few) as in the case of hearing-motion synaesthesia studied here. People with hearing-motion synaesthesia experience moving or changing (e.g. flickering) visual stimuli as sounds: in effect, a moving visual stimulus elicits a subjectively bound audio-visual percept. This was first documented by Saenz and Koch (2008) who discovered it when presenting an optic flow stimulus (dots moving inwards or outwards from a central

point) to students - one of whom claimed to 'hear' the movement<sup>1</sup>. Visual stimuli such as these have been previously shown to activate the auditory cortex of congenitally deaf people (Finney, Fine, & Dobkins, 2001). This is attributed to early multisensory plasticity such that pathways linking vision to audition were retained in these deaf individuals but eliminated (or reduced) in most people with normal hearing, a process that some researchers have likened to synaesthesia (Giraud & Lee, 2007). The 'neonatal synaesthesia hypothesis' argues that adult synaesthesia is a consequence of failing to remove multisensory pathways present in everyone during infancy (Maurer & Mondloch, 2006), or otherwise functionally suppressed (e.g. by inhibition; Grossenbacher and Lovelace, 2001). In the synaesthetic population the retention of these pathways (structural and/or functional) may arise due to genetic differences that affect brain maturation (Asher et al., 2009), whereas in the case of deaf people it arises in response to the absence of an appropriate sensory signal (Bavelier & Neville, 2002). An alternative scenario is that this form synaesthesia reflects a pattern of cross-wiring that is unique to some individuals and is not related to other groups (e.g. deaf adults, neurotypical infants) and is not related to the normal multisensory perception of motion. We return to this possibility in the discussion.

Whilst Saenz and Koch (2008) did not explore the neural basis, they did adapt a behavioural test (Guttman, Gilroy, & Blake, 2005; Guttman, Gilroy, & Blake, 2005; Guttman, Gilroy, & Blake, 2005). In general, people are better at discriminating rhythm in the auditory domain (e.g. from a signal of long and short beeps resembling Morse code) than the visual domain (e.g. a visual disc flickering on and off for short or long durations). Saenz and Koch (2008) reasoned that if a rhythmic visual stimulus

---

<sup>1</sup> The anecdote is reported here: <https://www.caltech.edu/news/caltech-neurobiologists-discover-individuals-who-hear-movement-1455>

induces an auditory percept then it would lead to auditory-like levels of performance for these visual stimuli, and this was observed in their N=4 hearing-motion synaesthetes.

A recent report by Fassnidge, Marcotti, and Freeman (2017) attempted to replicate the Saenz and Koch (2008) paradigm, and also introduced a new task in which participants had to detect the presence/absence of an auditory stimulus in either the presence/absence of visual motion. The latter should elicit an interfering auditory experience for the synaesthetes. They divided participants according to their subjective report of hearing-motion synaesthesia during the debrief ('Did you actually hear faint sounds when you saw flashes?') with 8/40 (22%) giving an affirmative response. They did not find a selective advantage for synaesthetes in visual rhythm perception, as reported by Saenz and Koch (2008), but rather a general advantage in both visual and auditory conditions. Nevertheless, across the whole group, there was a correlation between visual interference on auditory detection (which could arguably reflect the visual stimulus inducing a masking synaesthetic sound) and performance on the rhythm detection task.

The present research builds on prior research in several novel and important ways. We report the detailed characteristics of this phenomenon by screening a sample of over 200 people, documenting both the kinds of visual stimuli that elicit these sensations and the psychoacoustic characteristics of the sensations. We determine whether the association between visual and auditory features obeys certain 'rules', termed cross-modal correspondences (Spence, 2011). These may include a tendency to associate small objects with higher pitch (e.g. Mondloch & Maurer, 2004); to judge larger or looming objects as louder (Liu, Mercado, & Church, 2011); and for the presence of a visual movement to increase loudness but not pitch perception

(Maniglia, Grassi, & Ward, in press). We also report the electrophysiological correlates (EEG event-related potentials) of hearing-motion synaesthesia with **the aim of understanding at what stage in visual processing of the inducer group differences emerge (perceptual or post-perceptual). The assumption is that these are related to or concomitant with elicitation of the concurrent (i.e. audiovisual).** With regards to visual motion, several event-related components have been identified including the P1 and N2 (Kuba, Kubova, Kremlacek, & Langrova, 2007). The P1 tends to reflect motion onset, it peaks at around 100 msec and has been linked to activity in V1 (Schellart, Trindade, Reits, Verbunt, & Spekreijse, 2004). The N2 (or N2b) emerges around 160-200 msec (Kuba et al., 2007) and has been linked to motion processing per se, reflecting activity in V5/MT (Schellart et al., 2004). Our hypothesis is that there will be differences in these or other motion-related VEPs in hearing-motion synaesthetes that will enable us to infer specific underlying neural mechanisms. More generally, it will enable us to confirm that this is a perceptual phenomenon rather than post-perceptual (e.g. associative memory). For comparison we measure auditory evoked potentials (where we do not expect any differences) and audio-visual potentials. For the latter, we also expect group differences because the visual component of the stimulus should induce synaesthetic sound but the extent to which the synaesthetic sound and physical sound will interact is unknown.

## **Study 1: Prevalence and Characteristics of Visual Inducers and Synaesthetic**

### **Auditory Experiences**

#### **Methods**

##### **Participants**



221 participants (180 female, 192 right handed, with a mean age of 20.85, SD $\pm$ 6.25) were recruited. A sample of psychology undergraduates (N=189) took the study for course credits. As this sample was naïve as to the purposes of the study it enables an estimate of prevalence. A second sample (N=32) were recruited because they had indicated that they might have this type of synaesthesia. Self-referral was via our synaesthesia website ([www.sussex.ac.uk/synaesthesia](http://www.sussex.ac.uk/synaesthesia)) or via social media, and was based on indicating they may have this type of synaesthesia in response to the optic flow stimulus used by Saenz and Koch (2008). All participants gave consent according to the study protocol approved by the Psychology and Life Sciences Cluster-based Research Ethics Committee, University of Sussex.

### Stimuli

There were 12 silent movies, each lasting 20s, and 2 static images (see Supplementary Material). The silent movies consisted of moving single dots (N=8), a rotating checkboard (N=1), the optic flow stimulus (N=1) of Saenz and Koch (2008), and two real-world movies (a busy street scene, a flock of birds in flight). The static images consisted of an image implying motion (galloping horses), which are known to activate area V5/MT (Kourtzi & Kanwisher, 2000), and the Rotating Snakes stimulus (Kitaoka & Ashida, 2003) which induces an illusion of motion. The movies depicting moving single dots were created using Synfig Studio 0.64.1 and contrasted four visual dimensions: large v. small moving dot, high v. low moving dot, fast/slow speed of movement, and fast/slow flicker. All stimuli contained a moving or flickering white dot on a black background. The standard size was approximately 2.12 degrees, presented centrally, oscillating left and right at approximately 17.15 deg/s on average (based on a 15.6 inch monitor and viewing distance of 65cm). The size manipulation increased or decreased the standard diameter threefold. The speed manipulation increased or

decreased the speed by approximately 50%. The vertical position manipulation shifted the dot up/down by 25% of the screen height. In the flicker condition, the dot remained central and the dot appeared/disappeared with an average frequency of 10Hz or 5Hz.

### Procedure

Participants completed an online questionnaire using the Bristol Online Survey software. The stimuli were presented in a fixed order and participants answered a set of questions after viewing each one. They were first asked “Do you experience sound(s) whilst watching this video? [yes/no]” and, given an affirmative answer, they were asked five clarifying questions.

- 1) On a scale of 1 to 9, how loud did the sound feel? [1-silence; 2-just audible; 3-whisper; 4-quiet office; 5-ordinary conversation; 6-street noises; 7-machinery (vacuum cleaner, hairdryer); 8-loud radio; 9-jet plane taking off]
- 2) How low or high did you feel the pitch of the sound was? (1-very low, 9-very high)
- 3) On a scale of 1 to 9, how dynamic did the sound feel? (1-static, single tone, 9-changing, variable tone/sounds)
- 4) How internal or external did the sound feel? [1-as in my head/like listening through headphones 2-as from outside/like listening through speakers]
- 5) Did the sound resemble any of the following: [hissing, ringing, pulsing, buzzing, clicking, cracking, humming, popping, roaring, rushing, whistling, whoosing, like a tone, like a typewriter, other: specify\_\_\_\_]

The loudness measure (Q1) uses common environmental descriptors linked to the decibel scale, and the psychoacoustic descriptors (Q5) were taken from the Tinnitus Sample Case History Questionnaire (Langguth et al., 2007). Given that the stimuli were presented online, there was no experimental control over the absolute

size/speed but we can nevertheless assess how relative differences in visual features manifest themselves psychoacoustically.

Following presentation of the 12 movies and 2 static stimuli, they were asked a series of questions about their musical experience, whether synaesthetic sounds are affected/elicited by tiredness, touch and or body movements, and to describe any other relevant motion-sound experiences from everyday life (see Appendix for the full list of questions).

## **Results**

### **Prevalence and Characteristics of Hearing-Motion Synaesthetes**

For the 12 movies, the average number of stimuli inducing a sound was low for the overall group (1.49, SD=2.95, range=0-12, N=221) and was lower still for the naïve sample of undergraduates (average=0.87, SD=1.90). We classified people as having hearing-motion synaesthesia if at least half of these stimuli ( $\geq 6$ ) elicited sounds. Whilst the cut-off is somewhat arbitrary, and requires independent verification, we show that people above and below the cut-off have different characteristics (Table 1) and different electrophysiological correlates (Study 2). From the total sample, we classed 21 people as having hearing-motion synaesthesia, of which 8 were recruited from the undergraduate sample. This gives a prevalence estimate of 4.2% (8/189) based just on the naïve sample.

Table 1 shows the characteristics of the two groups in terms of the percentage of responses generated to different kinds of stimuli. The hearing-motion group reported an increased rate of auditory experiences (relative to the control group) across the whole range of stimuli. The first two rows show the previously reported data (for the 12 movies depicting physical motion) but broken down into abstract and real-world. The third and fourth rows show the data for illusory and implied motion.

The hearing-motion group were far more likely to report an auditory experience to physical motion than to illusory or implied motion. By contrast, controls showed the greatest proportion of responses to real-world movies. Rather than suggest 'strong' and 'weak' forms of synaesthesia (Martino & Marks, 2001), we suggest that the two groups draw preferentially from distinct mechanisms: one based on visual perception of motion/change (favouring physical motion over static, **irrespective of whether real-world or abstract**) versus auditory imagery (favouring real-world over abstract). These would perhaps map on to perceptual versus semantic mechanisms of multisensory motion processing that have been noted elsewhere (Sadaghiani et al., 2009).

Table 1 shows that the hearing motion group also report higher levels of musical engagement (albeit not from an earlier age). The complementary auditory-to-visual type of synaesthesia is also linked to increased musical engagement, although this is not found in other, non-auditory types of synaesthesia (Ward, Thompson-Lake, Ely, & Kaminski, 2008). The direction of cause and effect cannot be ascertained at present.

Almost half of the synaesthetes agreed that sounds are elicited by the movements of their own hands irrespective of whether the hands are seen. A smaller number of synaesthetes agreed with statements about whether touch elicits sounds. Tiredness was noted to intensify their hearing-motion synaesthesia in 47.6% of participants, but to reduce the synaesthesia in 14.3%.

INSERT TABLE 1 ABOUT HERE

Cross-Modal Correspondences and Psychoacoustic Characteristics

For the 21 hearing-motion synaesthetes, paired t-tests were conducted on the psychoacoustic properties (pitch, loudness, dynamism) that were induced by the matched pairs of visual stimuli varying in size, position, speed of movement, or rate of flicker (providing that a sound was elicited for both items in a pair). The results are summarised in Figure 1. Higher pitch was associated with faster motion (paired  $t(12)=4.454$ ,  $p=.001$ ; which survives a Bonferroni correction of  $\alpha<.05/4$ ) and smaller size ( $t(14)=2.219$ ,  $p=.044$ ; which would not survive correction). Higher pitch was not linked to higher spatial positions ( $t(14)=.739$ ,  $p=.472$ ) or faster flicker ( $t(16)=1.344$ ,  $p=.198$ ). None of the visual changes systematically affected loudness ( $p>.10$ ). The mean loudness across stimuli was  $\sim 3$ , corresponding to a subjective loudness of “whisper” ( $\sim 20\text{dB}$ ) on the scale used here. None of the pairs of visual changes systematically affected the rated dynamism of the sounds (all  $p$ 's  $>.07$ ), although the flickering dots were rated as having far less auditory dynamism than moving dots.

INSERT FIGURE 1 ABOUT HERE

For the hearing motion synaesthetes, across all movies, the majority of sounds (90%) are reported as being internalised, and the tendency to hear internal sounds was common across all the synaesthetes. For the qualitative descriptors, the most common ones were “pulsing” (N=23), “like a tone” (N=22), “whooshing” (N=20), “ringing” (N=13), “rushing” (N=10), “humming” (N=9), “hissing” (N=5) and “popping” (N=5) [N refers to the number of times a descriptor was chosen].

When asked to describe what the sounds are like in everyday experience, several synaesthetes gave detailed answers such as the following (all from different individuals):

The sounds are mainly like inhaled and exhaled humming noises/faint whooshing noises, sometimes blends of sounds that can't be phonetically pronounced, although other times they can. If they can, than I can see the letters or sounds they would make in my mind's eye. I also can feel physical corresponding pulses in my legs, which match up with the sounds.

Usually just walking outside and going to the shops my surroundings are more varied so I notice synaesthetic responses more. I enjoy watching dance for part of this reason maybe? I don't know, it's kinda normal so tricky to describe. Going to loud places like music clubs are hard because it gets a bit much.

The sounds are in my head but are hard to describe! They're not like normal sounds but fit the motion of objects well. Pitches are higher towards the right hand side and lower on the left.

This was not something I had given much thought to until watching these videos but I found the sounds I experienced to be very simple tones

I ride the train to work and I find I can "hear" the beat of the trees as they go by. Or I "hear" people walking as a rhythm which allows me to pick up on any slight limp or inconsistency in their walk. If I watch a machine or something moving in a repetitive motion, I create a rhythm in my head which is usually fairly complex and repetitive. However when I see lots of things moving all at once, like crowds of people or things, it sounds like a constant fuzzy tone. (No pure tone or beat, just like watching the "white snow" on a TV, but kind of muted & quieter)

The sounds are similar to comic animation sounds. When feeling a heartbeat, [I] hear the noise from it.

Blinking lights (including the onscreen cursor in a text field such as this) often hum softly, especially when I'm tired. Aside from that, sounds often only begin to appear when I pay singular attention to rhythmic movement, i.e. hardly ever.

## **Study 2: Event-Related Potentials for Visual Motion and Tones**

### **Method**

#### **Participants**

Nineteen participants, recruited from Study 1, with normal or corrected-to-normal vision completed the EEG task. Two control participants were excluded due to poor quality and noise in the EEG recording. Nine participants, perceived sound on at least 6/10 of the meaningless motion videos described above were classified as synaesthetes (mean age 24.25 with SD $\pm$ 5.04, 7 female, 6 right-handed), whilst 8 control participants (mean age 28.11 with SD $\pm$ 10.14, 7 female, 6 right-handed) reported no synaesthetic sound experience to any stimulus in Study 1.

#### **Materials**

Stimuli were presented in E-Prime 2.0, on a 22 inch CRT monitor with a 60Hz refresh rate, and Labtek speakers approximately 100cms from the ears bilaterally. Participants were shown high contrast checkerboards moving left or right that have previously been used to study motion-evoked potentials namely (Schellart et al., 2004). The checkerboard was 16.98 x 9.62 degrees in size and comprised a 10x6

grid of black and white squares – see Figure 2. Its centre point was aligned with a coloured fixation cross which always remained in the centre of the screen. On *visual trials*, the checkerboard pattern was displaced to the left/right in subsequent steps (1 step / refresh rate), creating a perception of movement at a constant speed to the left/right, returning to the same phase angle at end of the trial. *Audio-visual trials* consisted of the same stimuli, accompanied by a 400Hz pure tone (loudness equalized at 20 phons) synchronized with motion duration. In *auditory trials*, participants were played the 400Hz pure tone whilst the checkerboard pattern remained still. The duration of the stimuli was either 83ms or 183ms. Visual and audio-visual trials were presented at two different speeds, equivalent to movement of 24 and 48 deg/s of visual angle at the viewing distance of 100cm.

INSERT FIGURE 2 ABOUT HERE

### *Procedure*

Participants were introduced to the experimental environment and gave informed consent at the beginning of the session. The experimenter explained and set up the task, ensuring viewing distance of 100cm.

Blocks of stimuli (each 2-4 minutes in duration) were presented, with five blocks of visual, audio-visual and auditory blocks each presented in pseudo-randomized order, without consecutive repetition of the same block-type. Participants were free to break between blocks. Trials with different motion speed and duration were randomized within each block. This produced 240 trials per each condition. Participants were instructed to maintain fixation on a central cross throughout. In random and infrequent (16.7%) target trials the fixation cross changed its colour from



red to green for the duration of trial and participants were required to press a key. Interval lengths, trial/interval length-ratio and block lengths were selected to minimize motion-adaptation effects (Heinrich, 2007; Kuba et al., 2007) and ensure participant comfort. In addition, inter-stimulus intervals were selected randomly between 800ms and 1200ms, to prevent potential carry-over effects. The recording session lasted approximately 60-minutes excluding set-up and debrief.

### *EEG recording*

EEG data was recorded with ACQUIRE 4.3.1 software of the Neuroscan system using Ag/AgCl active electrodes positioned over the scalp (N=32) and mastoids (N=2). Four ocular electrodes were used to monitor horizontal and vertical eye-movement. After initial testing sessions, the upper vertical eye electrode was removed with the aim of estimating eye-movement data from FP1. The ground electrode was positioned at AFz and the reference electrode at CPz according to the 10-20 system. All electrodes were fitted on a 64/80-system compatible Easy-Cap cap, both manufactured by Easycap GmbH. Impedances were kept below 10k $\Omega$ . A/D acquisition rate was 500Hz, with an online low pass filter of 100Hz and high pass filter of 0.05Hz.

### *Data analysis*

EEG recordings were processed and analysed in Matlab using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) toolboxes. Data were visually screened for anomalies. All target trials were excluded (i.e., green fixation cross). The following channels were removed: FP1, FPZ, and FP2 because the noise in these channels did not survive any rejection criteria. Vertical eye-channel(s) were also removed because they were not recorded in all individuals. At the same time channels were re-referenced to linked mastoids (auditory evoked-

potentials) or Fz (visual evoked-potentials). First, a second order Butterworth high-pass filter with a half-amplitude cutoff of 0.1 Hz was applied and DC-bias was removed. Then a 50 Hz notch-filter was applied. Secondly, a second-order Butterworth low-pass filter with a half-amplitude cutoff of 25 Hz was applied. Thereafter, automatic artefact rejection methods were applied. Epochs where the signal exceeded -60/+60  $\mu\text{V}$  in any of the EEG channels were rejected. Next, a moving window with a width of 100ms and a step size of 50ms was applied to the HEOG (horizontal electro-oculogram; horizontal eye channel). Epochs containing saccades (where the signal changed by more than 20  $\mu\text{V}$  in any of the moving windows) were rejected.

EEG epochs were averaged using baseline correction of 100ms preceding stimuli onset. For visual-evoked potentials (VEPs), an occipital cluster of electrodes (O1, Oz, O2) was averaged and referenced to Fz. The same occipital electrode clusters, referenced to Fz, were considered in the audio-visual stimulus condition. Auditory-evoked potentials (AEPs) are maximal over fronto-central sites, and a cluster of three electrodes (FC1, Fz, Cz, FC2) were averaged and referenced to linked mastoids as is conventional for AEPs. The peak latency was extracted from the grand mean average (collapsing across speed, stimulus duration, and group) and the mean amplitude was calculated for each participant in a 20 msec time window centred on P1 and N2 peaks identified based on previous research (Schellart, Trindade, Reits, Verbunt, & Spekrijse, 2004).

## **Results**

Participants performed well overall in the oddball task and identified the fixation-cross colour change reliably, indicated by near-ceiling hit rates for both synaesthetes ( $M=.98$ ,  $SE=.01$ ) and controls ( $M=.96$ ,  $SE=.01$ ). Mean overall hit rates in all conditions

were between .95 and .99. Mean false alarm rates for both groups were 0.7% and 0.5% respectively.

The visual condition is primarily of interest as this induces a difference in phenomenology across the groups. We may also expect some differences in the audio-visual condition too, although this would depend on the extent to which the physical sounds masked the synaesthetic sounds. We do not predict group differences in the auditory condition. **If differences in the auditory condition were found this might reflect greater excitability (or less inhibition) of auditory areas in the synaesthete group.**

The VEPs are shown in Figure 3. Three visual motion components were analysed: the P1 (90-110 msec) which has been linked to visual transients rather than motion per se, the N2 (165-185 msec) that has previously been linked to motion processing (Schellart et al., 2004), and a short latency component (SLC; 55-75 msec). The latter has not been routinely studied in the motion VEP literature but was identified by visual inspection of the waveforms and may be related to the N75 (Odom et al., 2010) that is found for checkerboard pattern-reversal (black and white squares reversed in contrast) and linked to activity in V1 (di Russo et al., 2005). A later component, the visual P2, is related to motion offset and was not analysed here (but is clearly visible for the short duration stimuli within the time window viewed). The results were analysed as a 2x2x2 mixed ANOVA contrasting group, speed and duration. For the visual P1, there was a main effect of speed with faster stimuli eliciting a greater amplitude ( $F(1,15) = 7.30, p = .016, \eta^2 = .33$ ), but no other main effects (group:  $F(1,15) = 0.17, p = .685, \eta^2 = .01$ ; duration  $F(1,15) = 2.23, p = .156, \eta^2 = .13$ ) and no interactions (group X duration;  $F(1,15) = 1.74, p = .207, \eta^2 = .10$ ; group X speed  $F(1,15) = 2.05, p = .172, \eta^2 = .12$ ; duration X speed  $F(1,15) = 2.89, p =$

.110,  $np2 = .16$ ; group X duration X speed  $F(1,15) = 1.57$ ,  $p = .229$ ,  $np2 = .10$ ). For the visual N2, there was a main effect of group with synaesthetes showing a larger peak than the controls ( $F(1,15) = 4.72$ ,  $p = .046$ ,  $np2 = .24$ ). There were also main effects of speed (fast > slow;  $F(1,15) = 20.72$ ,  $p < .001$ ,  $np2 = .58$ ) and duration (long>short;  $F(1,15) = 17.75$ ,  $p = .001$ ,  $np2 = .54$ ) and an interaction between speed and duration ( $F(1,15) = 5.45$ ,  $p = .034$ ,  $np2 = .27$ ). No other interactions were significant (group X duration  $F(1,15) = 1.36$ ,  $p = .262$ ,  $np2 = .08$ ; group X speed  $F(1,15) = 0.52$ ,  $p = .482$ ,  $np2 = .03$ ; group X duration X speed  $F(1,15) = 0.14$ ,  $p = .710$ ,  $np2 = .01$ ). For the SLC (55-75 msec), the effect of group approached significance ( $F(1,15) = 4.23$ ,  $p = .057$ ,  $np2 = .22$ ) due to greater negativity for the synaesthetes, but there were no other effects or interactions (duration  $F(1,15) = 1.06$ ,  $p = .319$ ,  $np2 = .07$ ; speed  $F(1,15) = 0.90$ ,  $p = .359$ ,  $np2 = .06$ ; group X duration  $F(1,15) = 0.02$ ,  $p = .900$ ,  $np2 < .01$ ; group X speed  $F(1,15) = 2.84$ ,  $p = .112$ ,  $np2 = .16$ ; duration X speed  $F(1,15) = 1.44$ ,  $p = .248$ ,  $np2 = .09$ ; group X duration X speed  $F(1,15) = 1.09$ ,  $p = .312$ ,  $np2 = .07$ ).

INSERT FIGURE 3 ABOUT HERE

The audio-visual condition is summarised in Figure 4, based on the same occipital electrode cluster and Fz reference as the VEP analysis above. In this analysis, the SLC window (55-75 msec) did reveal a significant group differences with synaesthetes showing a small negative-going peak that was not visible at all in the controls (group  $F(1,15) = 5.07$ ,  $p = .040$ ,  $np2 = .25$ ), and this was more pronounced for longer duration stimuli (group X duration  $F(1,15) = 8.32$ ,  $p = .011$ ,  $np2 = .36$ ). Neither the visual P1 nor N2 showed a group difference (P1:  $F(1,15) = 0.71$ ,  $p = .414$ ,

$np2 = .04$ ; N2:  $F(1,15) = 1.73$ ,  $p = .208$ ,  $np2 = .10$ ). As such the pattern is somewhat different to that reported for VEPs. There is evidence for an early difference between the groups (55-75 msec), which is significant in the audio-visual condition and borderline significant in the visual condition. There is a significant later difference, in the N2, that is present in the visual condition that is not present in the audio-visual condition. This may suggest an auditory masking of synaesthetic sound in this later time window. In the audio-visual condition, both the P1 and N2 showed main effects of speed (fast > slower; P1:  $F(1,15) = 5.78$ ,  $p = .030$ ,  $np2 = .28$ ; N2: speed  $F(1,15) = 18.20$ ,  $p = .001$ ,  $np2 = .55$ ) and the N2 showed an interaction between speed and duration ( $F(1,15) = 8.16$ ,  $p = .012$ ,  $np2 = .35$ ) as observed in the unimodal visual condition. No other interactions or effects approached significance (all  $p$ 's > .10).

INSERT FIGURE 4 ABOUT HERE

For auditory stimuli, a 2x2 mixed ANOVA was carried out contrasting group and stimulus duration using the same time windows identified previously and fronto-central electrodes (note that fast and slow was not a feature of the auditory stimuli). The results are shown in Figure 5. The only effect of significance was an effect of duration for the 165-185 msec positive component, such that the shorter duration was linked to a higher peak amplitude ( $F(1,15) = 16.58$ ,  $p = .001$ ,  $np2 = .52$ ). No other main effects or interactions approached significance (all  $p$ 's > .10). As such, the group differences that were apparent in visual and audio-visual evoked potentials were not found for unimodal auditory stimuli – consistent with the former, but not the latter, eliciting anomalous experiences.

INSERT FIGURE 5 ABOUT HERE

Unfortunately, EEG does not typically have the spatial resolution to determine whether the electrophysiological components originate from visual cortex, auditory cortex, or elsewhere. We can, however, examine the effects of our visual stimuli over more anterior (fronto-central) sites that show the greatest effect to auditory stimuli. Considering the visual N2 which showed a significant group difference over occipital sites, this component also shows a significant group difference over fronto-central sites re-referenced to linked mastoids rather than Fz, ( $F(1,15) = 6.09$ ,  $p = .026$ ,  $\eta^2 = .29$ . SYN mean = 2.877, SE = 0.705 and CON mean = 0.293, SE = 0.778). One-sample t-tests, against a reference of zero, showed that whilst controls showed no electrophysiological signature over these electrodes ( $t(7) = 0.38$ ,  $p = .718$ ,  $d = 0.19$ ) the synaesthetes did ( $t(8) = 4.08$ ,  $p = .004$ ,  $d = 1.92$ ).

## **Discussion**

This study examined the characteristics of hearing-motion synaesthesia. Our primary conclusion is that it is a perceptual phenomenon that originates at some of the earliest stages of cortical visual processing. With regards to the phenomenology (Study 1), we show that it is elicited by simple visual stimuli (e.g. a single moving dot) rather than elaborated stimuli (e.g. that involve object recognition or semantics, as in the case of grapheme-colour synaesthesia). Static images that imply or induce motion tended not to trigger sounds even though these stimuli are known to activate V5/MT (Kourtzi & Kanwisher, 2000; Kuriki et al., 2008) Non-synaesthetes, insofar as they reported any auditory experiences, tended to do so for meaningful stimuli (e.g. a silent movie clip of a bustling street scene). We suggest that synaesthetic auditory experiences are elicited by perceptual processing of physical visual movement, whereas the rarer auditory experiences reported by non-synaesthetes reflect other

processes (e.g. semantically induced auditory imagery). Just as the inducing visual stimuli can be described as 'simple', so were the associated auditory experiences (e.g. "hissing", "whooshing"). The motion VEPs (Study 2) indicated a group difference in the N2 component at 165-185 msec. This was centered on posterior electrodes (for both groups) but with an associated fronto-central component for synaesthetes only. The N2 is a perceptual component that is linked to motion processing in V5/MT (Schellart et al., 2004), although in hearing-motion synaesthetes other regions (e.g. auditory) could be co-activated. There was also some evidence of an earlier difference (55-75 msec) that was significant in the audio-visual condition, and close to significant in the visual condition. This resembles the N75 component that has been observed in checkerboard pattern-reversal (in which black and white squares are reversed in contrast) and linked to the processing of visual transients in V1 (di Russo et al., 2005).

Other researchers have observed that fMRI activation of auditory cortex by moving visual stimuli in early deaf people resembles synaesthesia (Giraud & Lee, 2007). It is important to note that we have no reason to believe that our synaesthetes have impaired hearing. This is backed up by our findings of normal auditory evoked potentials and the fact that the synaesthetes report being more musical than our control sample. However, it is possible that the synaesthetes are using some of the same neural pathways as those observed in congenitally deaf groups that are typically absent (or reduced) in most others. This is consistent with the neonatal synaesthesia hypothesis (Maurer & Mondloch, 2006). Campbell and Sharma (2016) contrasted motion-related VEPs in hearing children and early deaf children fitted with cochlear implants. Similar to our study of synaesthetes, they also noted increased amplitude of VEPs in the previously deaf children, and source localisation revealed an involvement of right auditory cortex in this VEP enhancement. Congenitally deaf

adults also show enhanced VEPs to briefly flashed stimuli, particularly in the peripheral vision (Neville, Schmidt, & Kutas, 1983). More recently it has been shown, using the Saenz and Koch (2008) visual rhythm paradigm, that this task does activate auditory cortex in the congenitally deaf whereas only the auditory version of the task does so in participants with normal hearing (Bola et al., 2017). Whilst we believe that our results are consistent with this interpretation, we cannot rule out the alternative hypothesis that hearing-motion synaesthetes have specialised neural pathways that are unrelated to other adults (hearing or deaf). This will require research on the structural anatomy and direct comparisons between these groups.

Our study also provides preliminary evidence for the prevalence of hearing-motion synaesthesia. Our estimate of 4.2% is comparable to the prevalence estimates for types of synaesthesia involving visual experiences (Simner et al., 2006). Saenz and Koch (2008) found their first case of hearing-motion by chance, but recruited a further 3 cases “after querying a few hundred individuals”, i.e. a somewhat lower prevalence than our study. Fassnidge et al. (2017), however, report a prevalence rate much higher than us (22%) but this was based on a single question at debrief. We suggest that further research needs to combine both the more detailed phenomenological report from our Study 1 with the objective measures used by Saenz and Koch (2008) and Fassnidge et al. (2017). Cluster analysis **avoids arbitrary cut-offs** and offers a bottom-up approach for defining groups based on multiple dimensions (e.g. behavioural, phenomenological) and has been applied to synaesthesia-like conditions (Grice-Jackson, Critchley, Banissy & Ward, in press). It is important to note that our sample was not demographically representative (containing mainly younger females), although other research suggests that synaesthesia is not strongly linked to either age or gender (Simner et al., 2006).



Historically, the nature of synaesthetic associations has been described as 'idiosyncratic' (i.e. random or unprincipled). However, a large body of research has subsequently shown that whilst synaesthetes vary greatly between each other they are nevertheless constrained by cross-modal correspondences (Sagiv & Ward, 2006). We found evidence that cross-modal correspondences are also implicated in hearing-motion synaesthesia in terms of smaller moving objects generating higher pitch (as noted elsewhere, Bien, ten Oever, Goebel, & Sack, 2012) and faster moving objects linked to higher pitch. The latter has not been previously noted in the literature on cross-modal correspondences (to our knowledge), but may derive from properties of the physical world. Faster vibrations of an object do generate higher acoustic frequencies. Other correspondences that we anticipated were not found. Higher elevation of a moving object was not linked to higher pitch, despite pitch-height correspondences being widespread (Parise, Knorre, & Ernst, 2014). It is possible that the height of a computer screen does not generate a large enough spatial comparison, or that a pitch modulation would be found for ascending-descending objects (i.e. for changes in elevation, rather the level of elevation per se). Similarly, judgments of synaesthetic loudness were not related to any properties of the visual stimuli tested here despite evidence elsewhere of, for instance, size-loudness correspondences (Liu et al., 2011).

In summary, we establish that hearing-motion synaesthesia arises from early visual processing of motion-related signals and draws on some cross-modal correspondences. It is likely to be at least as common as other forms of synaesthesia, although our prevalence estimate is preliminary. We propose a neurodevelopmental model that is consistent with the neonatal synaesthesia hypothesis, and consistent

with other lines of converging evidence such as cross-modal plasticity in early deafness.

### **Acknowledgements**

We thank Angela Gurney who conducted pilot research leading up to Study 1. NR and GB contributed equally to the overall project in terms of design, implantation and analyses. The study was designed by all authors and the manuscript was written by JW with key contributions from all authors. NR is supported by the Swiss National Science Foundation (Grant: PZ00P1\_154954)

### **References**

- Asher, J. E., Lamb, J. A., Brocklebank, D., Cazier, J. B., Maestrini, E., Addis, L., . . . Monaco, A. P. (2009). A Whole-Genome Scan and Fine-Mapping Linkage Study of Auditory-Visual Synesthesia Reveals Evidence of Linkage to Chromosomes 2q24, 5q33, 6p12, and 12p12. *American Journal of Human Genetics*, *84*(2), 279-285. doi: 10.1016/j.ajhg.2009.01.012
- Bavelier, D., & Neville, H. J. (2002). Cross-modal plasticity: Where and how? *Nature Reviews Neuroscience*, *3*, 443-452.
- Bien, N., ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study. *NeuroImage*, *59*(1), 663-672. doi: 10.1016/j.neuroimage.2011.06.095
- Bola, L., Zimmermann, M., Mostowski, P., Jednoróg, K., Marchewka, A., Rutkowski, P., & Szwed, M. (2017). Task-specific reorganization of the auditory cortex in deaf humans. *Proceedings of the National Academy of Sciences of the USA*, e600-e609.
- Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K. P., . . . Fink, G. R. (2001). Polymodal motion processing in posterior parietal and premotor cortex: A human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron*, *29*, 287-296.
- Campbell, J., & Sharma, A. (2016). Visual Cross-Modal Re-Organization in Children with Cochlear Implants. *PLoS One*, *11*(1). doi: 10.1371/journal.pone.0147793
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

- Di Russo, F., Pitzalis, S., Spitoni, G., Aprile, T., Patria, F., Spinelli, D. & Hillyard, S.A. (2005). Identification of the neural sources of the pattern-reversal VEP. *Neuroimage*, *24*, 874-886.
- Fassnidge, C., Marcotti, C. C., & Freeman, E. (2017). A deafening flash! Visual interference of auditory signal detection. *Consciousness & Cognition*, *49*, 15-24.
- Finney, E. M., Fine, I., & Dobkins, K. R. (2001). Visual stimuli activate auditory cortex in the deaf. *Nature Neuroscience*, *4*(12), 1171-1173. doi: 10.1038/nn763
- Giraud, A. L., & Lee, H. J. (2007). Predicting cochlear implant outcome from brain organisation in the deaf. *Restorative Neurology and Neuroscience*, *25*(3-4), 381-390.
- Grice-Jackson, T., Critchley, H.D., Banissy, M.J. & Ward, J. (in press). Common and distinct neural mechanisms associated with the conscious experience of vicarious pain. *Cortex*.
- Grossenbacher, P, & Lovelace, C. (2001). Mechanisms of synesthesia: cognitive and physiological constraints. *Trends in Cognitive Sciences*, *5*, 36-41.
- Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see - Auditory encoding of visual temporal sequences. *Psychological Science*, *16*(3), 228-235. doi: 10.1111/j.0956-7976.2005.00808.x
- Heinrich, S. P. (2007). A primer on motion visual evoked potentials. *Documenta Ophthalmologica*, *114*(2), 83-105. doi: 10.1007/s10633-006-9043-8
- Kitaoka, A., & Ashida, H. (2003). Phenomenal characteristics of the peripheral drift illusion. *Vision*, *15*, 261–262.
- Kourtzi, Z., & Kanwisher, N. (2000). Activation of human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience*, *12*, 48-55.
- Kuba, M., Kubova, Z., Kremlacek, J., & Langrova, J. (2007). Motion-onset VEPs: Characteristics, methods, and diagnostic use. *Vision Research*, *47*(2), 189-202. doi: 10.1016/j.visres.2006.09.020
- Kuriki, I., Ashida, H., Murakami, I., & Kitaoka, A. (2008). Functional brain imaging of the Rotating Snakes illusion by fMRI. *Journal of Vision*, *8*(10). doi: 10.1167/8.10.16
- Langguth, B., Goodey, R., Azevedo, A., Bjorne, A., Cacace, A., Crocetti, A., . . . Vergara, R. (2007). Consensus for tinnitus patient assessment and treatment outcome measurement: Tinnitus Research Initiative meeting, Regensburg, July 2006. *Tinnitus: Pathophysiology and Treatment*, *166*, 525-536. doi: 10.1016/s0079-6123(07)66050-6
- Liu, E. H., Mercado, E., & Church, B. A. (2011). Multidimensional processing of dynamic sounds: more than meets the ear. *Attention Perception & Psychophysics*, *73*(8), 2624-2638. doi: 10.3758/s13414-011-0190-y
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213.
- Maniglia, M., Grassi, M., Casco, C., & Campana, G. (2012). The origin of the audiovisual bounce inducing effect: A TMS study. *Neuropsychologia*, *50*(7), 1478-1482. doi: 10.1016/j.neuropsychologia.2012.02.033
- Maniglia, M., Grassi, M., & Ward, J. (in press). Sounds are Perceived as Louder when Accompanied by Visual Movement. *Multisensory Research*.
- Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science*, *10*, 61-65.
- Maurer, D., & Mondloch, C. J. (2006). The infant as synesthete? *Attention and Performance XXI*, 449-471.

- Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective and Behavioral Neuroscience*, *4*, 133-136.
- Neville, H. J., Schmidt, A., & Kutas, M. (1983). Altered visual evoked potentials in congenitally deaf adults. *Brain Research*, *266*, 127-132.
- Odom, J.V., Bach, M., Brigell, M., Holder, G.E., McCulloch, D.L., & Tormene, A.P. (2010). ISCEV standard for clinical visual evoked potentials. *Documenta Ophthalmologica*, *120*, 111–119
- Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, *111*(16), 6104-6108. doi: 10.1073/pnas.1322705111
- Sadaghiani, S., Maier, J. X., & Noppeney, U. (2009). Natural, Metaphoric, and Linguistic Auditory Direction Signals Have Distinct Influences on Visual Motion Processing. *Journal of Neuroscience*, *29*(20), 6490-6499. doi: 10.1523/jneurosci.5437-08.2009
- Saenz, M., & Koch, C. (2008). The sound of change: Visually-induced auditory synesthesia. *Current Biology*, *18*, R650-R651.
- Sagiv, N., & Ward, J. (2006). Cross-modal interactions: Lessons from synesthesia. *Progress in Brain Research*, *155*, 259-271.
- Schellart, N. A. M., Trindade, M. J. G., Reits, D., Verbunt, J. P. A., & Spekreijse, H. (2004). Temporal and spatial congruence of components of motion-onset evoked responses investigated by whole-head magneto-electroencephalography. *Vision Research*, *44*(2), 119-134. doi: 10.1016/j.visres.2003.09.016
- Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*(6614), 308-308. doi: 10.1038/385308a0
- Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., . . . Ward, J. (2006). Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception*, *35*, 1024-1033.
- Soto-Faraco, S., Spence, C., Lloyd, D., & Kingstone, A. (2004). Moving multisensory research along: Motion perception across sensory modalities. *Current Directions in Psychological Science*, *13*(1), 29-32. doi: 10.1111/j.0963-7214.2004.01301008.x
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention Perception & Psychophysics*, *73*(4), 971-995. doi: 10.3758/s13414-010-0073-7
- Ursino, M., Cuppini, C., & Magosso, E. (2014). Neurocomputational approaches to modelling multisensory integration in the brain: A review. *Neural Networks*, *60*, 141-165. doi: 10.1016/j.neunet.2014.08.003
- Ward, J., Thompson-Lake, D., Ely, R., & Kaminski, F. (2008). Synaesthesia, creativity and art: What is the link? *British Journal of Psychology*, *99*, 127-141.
- Watkins, S., Shams, L., Tanaka, S., Haynes, J. D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, *31*, 1247-1256.
- Zvyagintsev, M., Clemens, B., Chechko, N., Mathiak, K. A., Sack, A. T., & Mathiak, K. (2013). Brain networks underlying mental imagery of auditory and visual information. *European Journal of Neuroscience*, *37*(9), 1421-1434. doi: 10.1111/ejn.12140



**Table 1.** The different characteristics of the hearing-motion synaesthetes and controls in terms of percentage of auditory experiences reported for different visual stimuli, and their background level of musicality. Fishers exact test is used where there is a single binary data point (0,1) for each participant. Parametric tests are used for likert scales (adjusting for unequal variances as necessary) and non-parametric tests used for other variables that were heavily skewed (e.g. hours of music per week). One control participant did not complete the debrief questionnaire.

	Hearing Motion	Controls	Difference
Abstract movies % and SEM (e.g. flickering dots)	81.9 (4.2)	4.5 (0.7)	Independent samples median test, $p < .001$
Real-world movies % and SEM (e.g. dynamic street scene)	71.4 (8.1)	9.5 (1.6)	Independent samples median test, $p < .001$
Implied motion (static image of galloping horses) %	33.3	4.0	Fishers exact test, $p < .001$
Illusory motion (Rotating Snakes illusion) %	23.8	0.5	Fishers exact test, $p < .001$
“I play musical instruments or sing often” 1-strongly disagree, 9- strongly agree, (mean, SEM)	5.81 (0.61)	3.29 (0.20)	$t(218)=3.93$ , $p < .001$
How many hours per week, on average [music playing]? (mean, SEM)	5.58 (2.13)	2.10 (0.30)	Independent samples median test, $p = .087$
“I read music sheets well/with ease.” 1-strongly disagree, 9- strongly agree, (mean, SEM)	4.05 (0.63)	2.84 (0.16)	$t(22.69)=1.86$ , $p = .076$
“At what age (years) did you receive musical training?” (mean in years, SEM)	9.31 (1.46)	7.79 (.23)	$t(15.76)=1.03$ , $p = .321$

When I move my hands in front of me with my eyes open, they produce a sound (e.g. wiggling fingers in front of your eyes). % agreed.	42.9	5.0	Fishers exact test, $p < .001$
“When I move my hands out of sight, they produce a sound (e.g. wiggling fingers with your eyes shut).” % agreed.	38.1	3.0	Fishers exact test, $p < .001$
“When I am being touched by someone I often hear a sound.”	9.5	1.5	Fishers exact test, $p = .072$
“When I touch my own body I often hear a sound.” % agreed.	19.0	4.5	Fishers exact test, $p = .025$

### Figure Captions

**Figure 1.** The perceived pitch (top), loudness (middle), and auditory dynamism (bottom) of contrasting pairs of stimuli that vary (from left to right) in size, position, speed of motion, and speed of flicker. The mean is shown (on 1-7 scale) and 1 SEM.

**Figure 2.** The checkerboard stimulus used in visual and audio-visual trials. Motion speed was varied by increasing the displacement per refresh cycle on faster trials

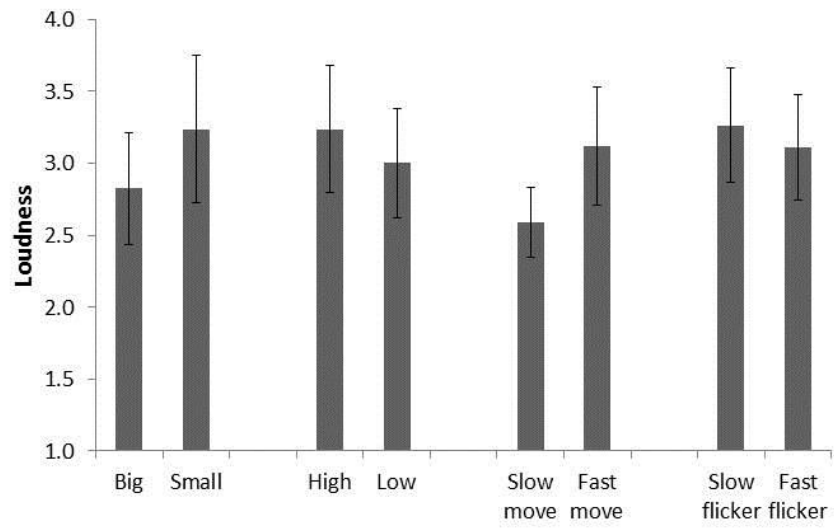
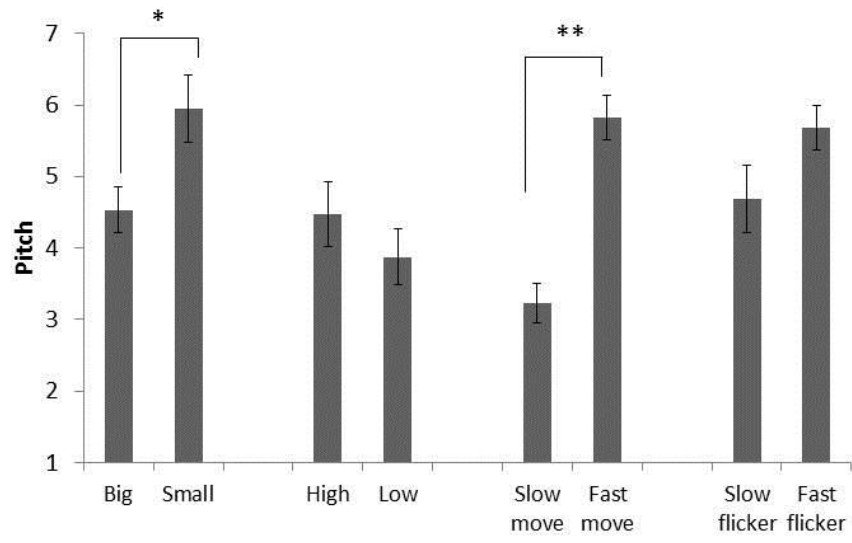
**Figure 3.** Left: Visual evoked potentials to moving checkerboards for synaesthetes (dashed) and controls (solid) for slow movement (left column), fast movement (right column), short duration (top row) and long duration (bottom row). The three shaded

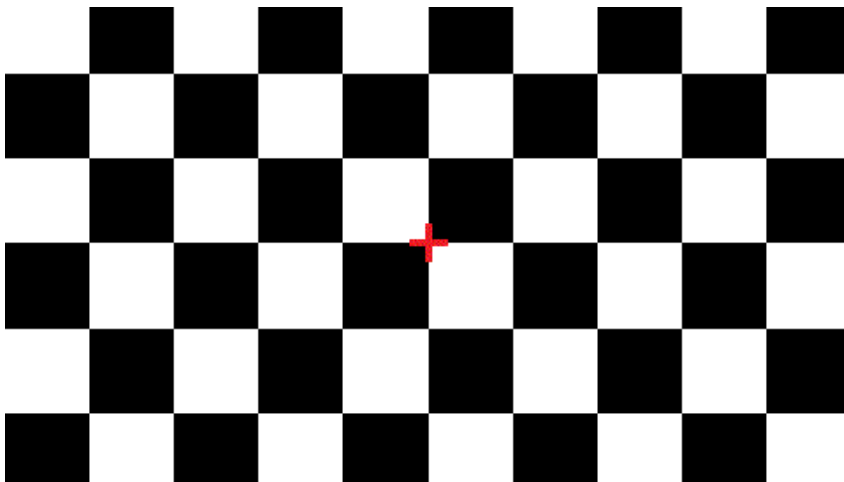
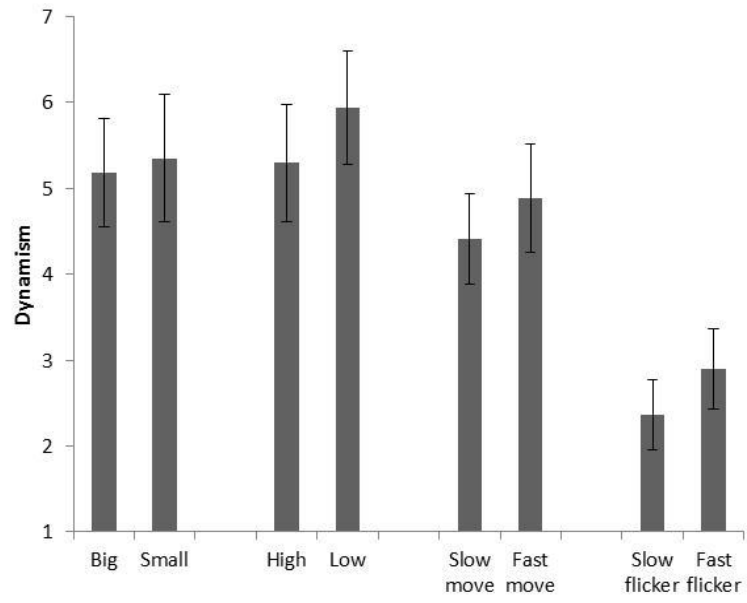


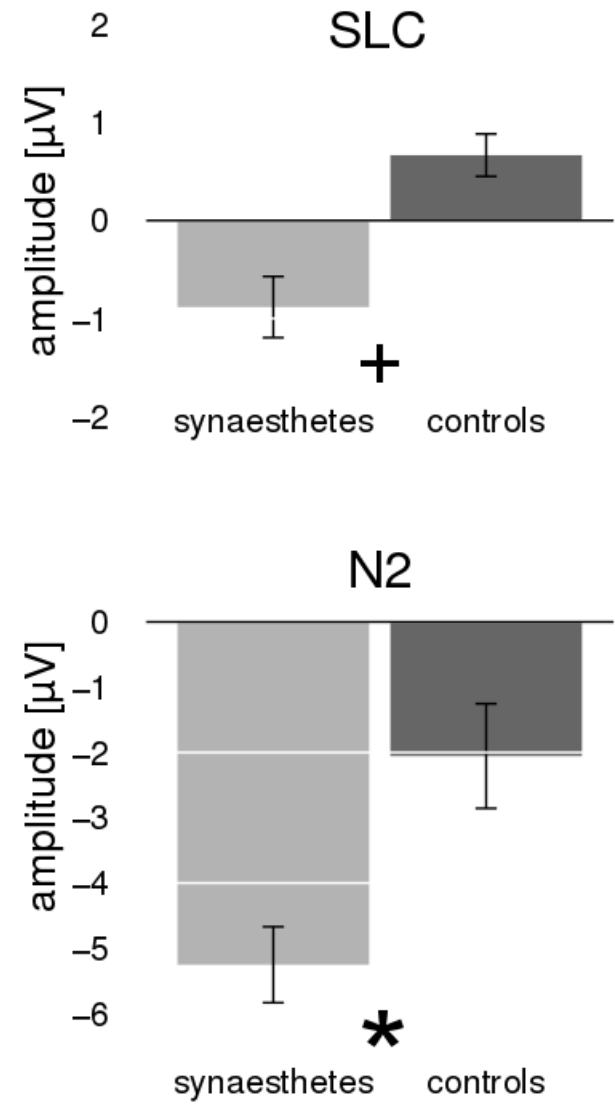
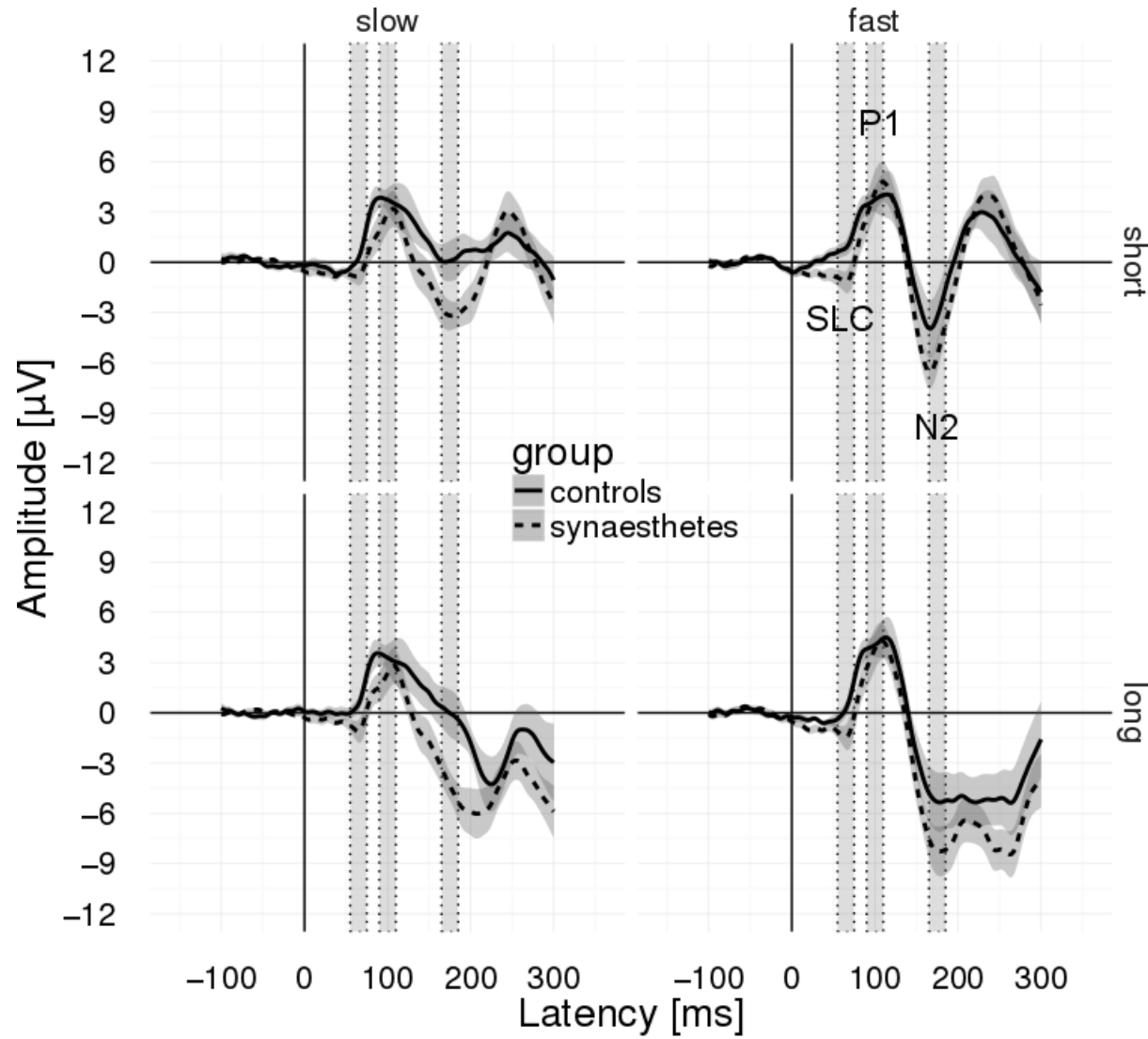
areas represent the three analysed time windows (i.e., SLC, P1, and N2). The shaded areas on the EEG signature represent standard errors. Top right: trending main effect Group for the SLC. Bottom right: significant main effect Group for the N2. Error bars represent standard errors.

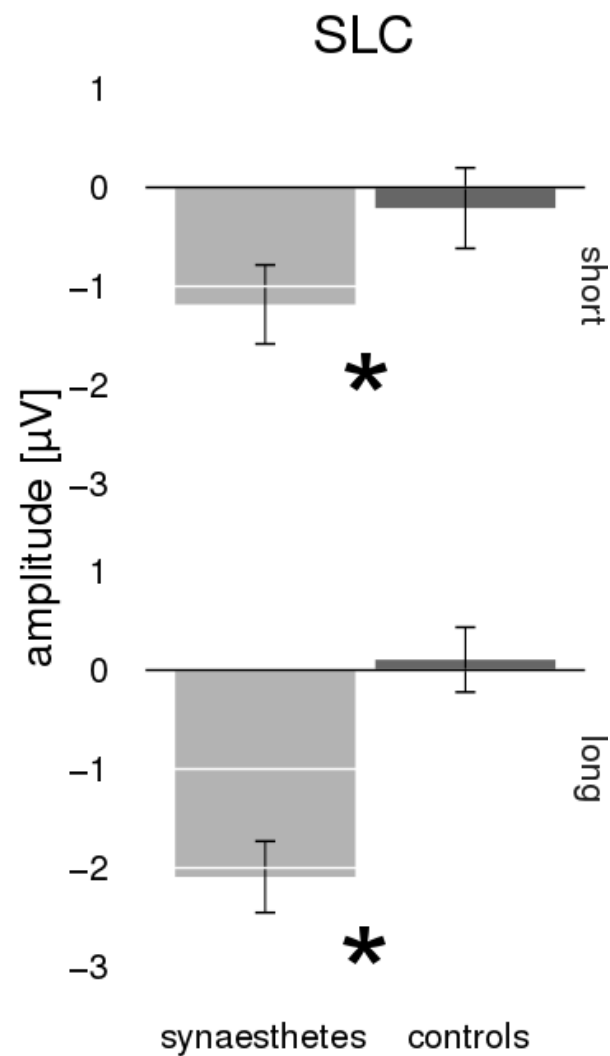
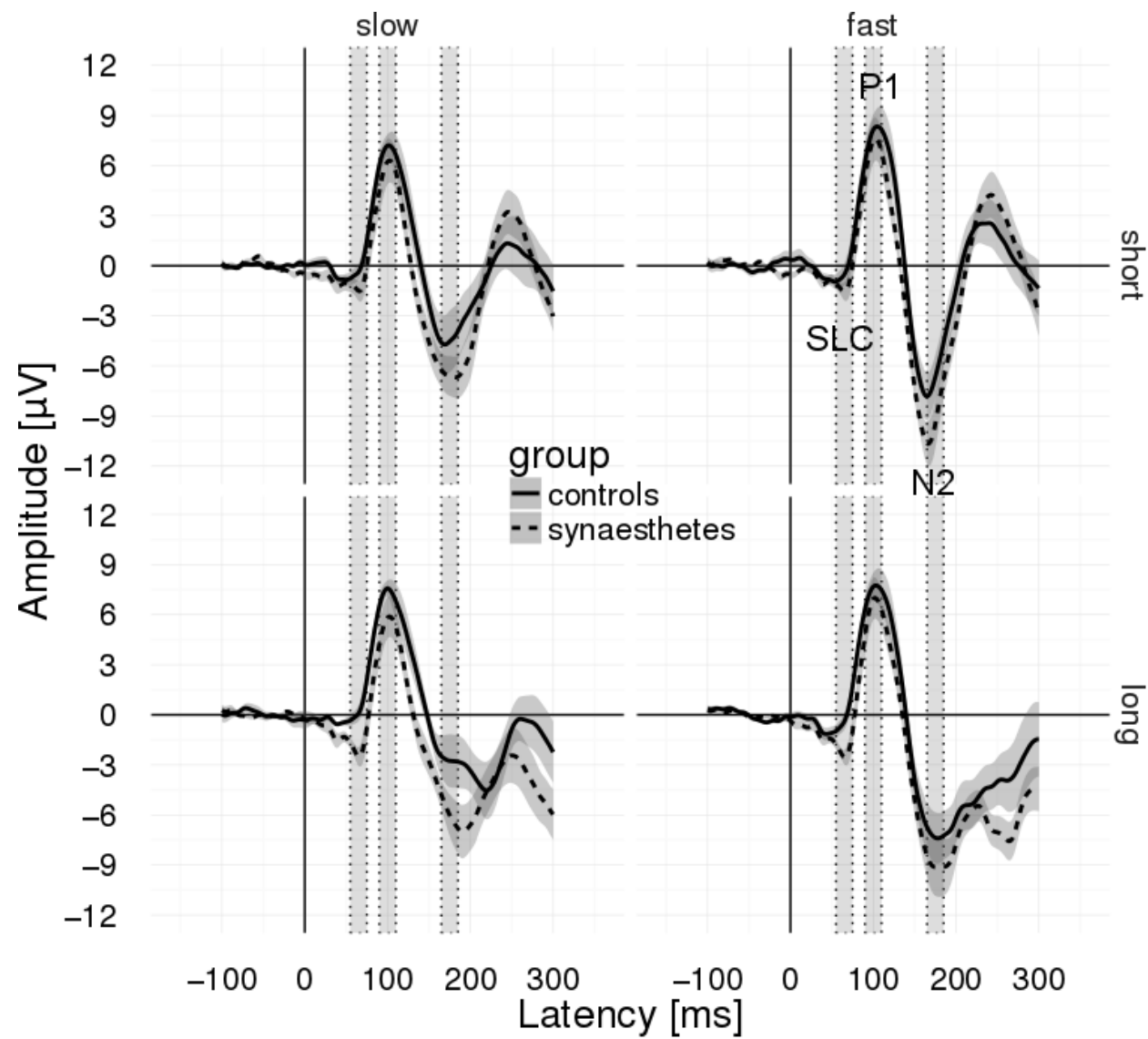
**Figure 4.** Left: Audio-visual evoked potentials to moving checkerboards accompanied by a temporally synchronous tone for synaesthetes (dashed) and controls (solid) for slow movement (left column), fast movement (right column), short duration (top row) and long duration (bottom row). The three shaded areas represent the three analysed time windows (i.e., SLC, P1, and N2). The shaded areas on the EEG signature represent standard errors. Right: significant Group x Duration interaction for the SLC. Error bars represent standard errors.

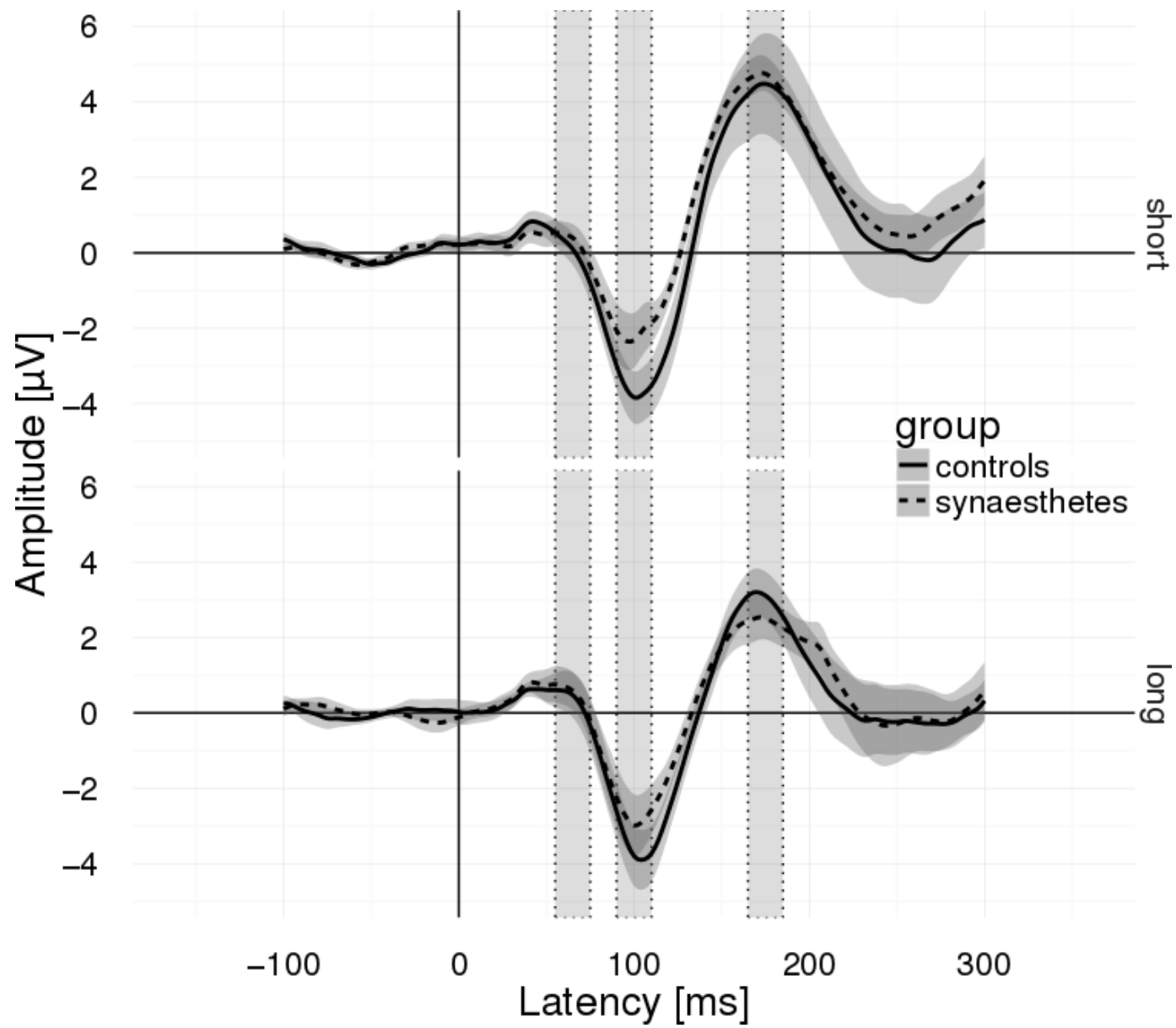
**Figure 5.** Auditory evoked potentials to a pure tone stimulus for synaesthetes (dashed) and controls (solid) for short duration (top) and long duration (bottom) stimuli. The three shaded areas represent the three analysed time windows (i.e., P1, N1, and P2). The shaded areas on the EEG signature represent standard errors.











## **Appendix**

The following statements relate to the experience of perceiving sounds when seeing motion. Please indicate how much you agree or disagree with them. (1-strongly disagree, 9-strongly agree)

I often find it difficult to distinguish between synaesthetic and real sound.

The synaesthetic sounds I perceive are simultaneous with the visual movement.

I perceive dynamic, rhythmic sounds when I see motion.

I perceive static tones when I see motion.

When I am tired I perceive sounds more when seeing something move.

When I am tired I perceive sounds less when seeing something move.

When I move my hands in front of me with my eyes open, they produce a sound (e.g. wiggling fingers in front of your eyes).

When I move my hands out of sight, they produce a sound (e.g. wiggling fingers with your eyes shut).

I play musical instruments or sing often.

(How many hours per week, on average \_\_\_\_)

(If you have been taught music, please indicate how old were you when first started: \_\_\_\_)

I read music sheets well/with ease.

When I am being touched by someone I often hear a sound.

When I touch my own body I often hear a sound.