BRILL

Multisensory Research 29 (2016) 113–132

MULTISENSORY RESEARCH

brill.com/msr

Which Direction Is up for a High Pitch?

Michael J. Carnevale and Laurence R. Harris *

Centre for Vision Research, York University, Toronto, Canada

Received 12 December 2014; accepted 23 August 2015

Abstract

Low- and high-pitched sounds are perceptually associated with low and high visuospatial elevations, respectively. The spatial properties of this association are not well understood. Here we report two experiments that investigated whether low and high tones can be used as spatial cues to upright for self-orientation and identified the spatial frame(s) of reference used in perceptually binding auditory pitch to visuospatial 'up' and 'down'. In experiment 1, participants' perceptual upright (PU) was measured while lying on their right side with and without high- and low-pitched sounds played through speakers above their left ear and below their right ear. The sounds were ineffective in moving the perceived upright from a direction intermediate between the body and gravity towards the direction indicated by the sounds. In experiment 2, we measured the biasing effects of ascending and descending tones played through headphones on ambiguous vertical or horizontal visual motion created by combining gratings drifting in opposite directions while participants either sat upright or laid on their right side. Ascending and descending tones biased the interpretation of ambiguous motion along both the gravitational vertical and the long-axis of the body with the strongest effect along the body axis. The combination of these two effects showed that axis of maximum effect of sound corresponded approximately to the direction of the perceptual upright, compatible with the idea that 'high' and 'low' sounds are defined along this axis.

Keywords

Cross modal correspondences, pitch, elevation, frequency, spatial orientation, auditory cues to orientation

1. Introduction

The crossmodal association between auditory pitch and visuospatial height, in which high-pitched sounds are perceived as being located higher in space, is one of the best-known and perceptually robust correspondences (Spence, 2011). The correspondence appears to have structural, statistical and semantic

* To whom correspondence should be addressed. E-mail: harris@yorku.ca

DOI:10.1163/22134808-00002516

origins: structural because of the filtering properties of the pinna, statistical because of the distribution of sounds in the natural world, and semantic as evidenced by the use of the word ‘high’ to describe both dimensions. However, what is meant by ‘high’ has never been definitively identified. The frequency–elevation mapping related to the filtering properties of the external ear would suggest high relative to the body (as the ears are fixed on the head), whereas the frequency–elevation mapping associated with the natural statistics of the physical environment would suggest high relative to gravity. Here we assess whether sounds (either with real spatial locations or with the ‘high’ correspondence resulting from their frequency content) can contribute to our sense of orientation, and whether any such effects are defined primarily with respect to the body axis or to gravity.

The association between auditory pitch and spatial elevation has been recognized since the 19th century and can be traced to cross-cultural studies of language, long before the term ‘crossmodal correspondence’ was even coined (Stumpf, 1883). In the first experimental quantification of this phenomenon, Pratt (1930) found that the ability to localize sounds presented from loudspeaker in the vertical plane was highly biased by frequency content. These mislocalization effects have been replicated and extended many times (e.g., Mudd, 1963; Parise *et al.*, 2014; Roffler and Butler, 1968), and the effect has even been found in blind individuals. The perceptual effects related to the pitch–height correspondence also extend to visual motion (either real or apparent), where ascending tones are associated with upward visual motion and descending tones are associated with downward visual motion (Maeda *et al.*, 2004; Sadaghiaini *et al.*, 2009). This correspondence is consequential to perception and cognition as shown in a wide-variety of paradigms, including the speeded-classification task (i.e., where reaction times are faster when relative pitch and elevation match) and experiments on attention (Ben-Artzi and Marks, 1995; Bernstein and Edelstein, 1971; Chiou and Rich, 2012; Evans and Treisman, 2010; Melara and Marks, 1990; Melara and O’Brien, 1987; Mossbridge *et al.*, 2011; Patching and Quinlan, 2002). Where might these sensory effects come from?

One possibility is that higher frequency sounds might actually tend to come from higher up and that we have either learned, or are innately programmed to recognize this fact. Parise *et al.* (2014) recorded the distribution of sounds in the natural environment by having participants walk around with directional microphones attached to their heads with one pointing up and the other pointing down. They found that indeed higher-frequency sounds tend to come from above and lower-frequency sounds from below. They also showed that the frequency filtering properties of the external ear tend to accentuate this distinction. So sounds, counter-intuitively, carry orientation information in their very makeup. Can this information then contribute to our perception of orientation in a way similar to the orienting cues provided by vision? In almost any scene there are many visual cues that indicate the direction of gravity (Harris *et al.*, 2011). Can we use a 'high sound' to indicate the direction of gravity also?

The role of sound in defining orientation is unclear. Moving auditory cues appear to play only a minor role in evoking self-motion (Väljamäe, 2009). Some studies show potential for sounds with sufficient spatial cues to enhance balance and postural stability (Easton *et al.*, 1998; Petersen *et al.*, 1995, 1996) and in this sense sounds can be used as a reference cue to orientation relative to the environment. However, there appears to be no work in the literature using high- and low-pitched sounds as orientation cues to define high and low space despite their strong perceptual associations to spatial 'high' and 'low'.

When tilted sideways, participants mislocalize sounds in a way consistent with using both head-based and gravity-based cues (Roffler and Butler, 1968). Parise *et al.* (2014) explained this as resulting from a combination of distal (i.e., with respect to the external environment, here represented by the axis of gravity) and proximal (i.e., with respect to the self, here represented by the head axis) reference frame priors for relating auditory pitch to visuospatial height. The individual priors refer to the perceptual system's expected mapping between auditory frequency and visual elevation as defined both in terms of the body and gravitational axes.

As it turns out our perceived direction of upright also has both head- and gravity-based components. Traditionally, our perception of 'up' has been measured using the subjective visual vertical (SVV — Aubert, 1861; Howard, 1982) in which subjects judge or set the orientation of a visual line with respect to the perceived direction of gravity (i.e., the direction a physical object would fall). The 'perceptual upright' (PU) is a complementary concept of 'up' (Dyde *et al.*, 2006), defined as the orientation at which an object or character is most easily recognized. Both SVV and PU representations of 'up' are influenced by various perceptual cues. These cues include the idiotropic (or body) vector, the gravity vector (which is built up from vestibular and other cues, Lackner and DiZio, 2005), and the visual vector (based on visual cues to upright such as the direction of light, the physical relationships between objects, and the horizon, Harris, 2009). The relative influence of these cues on a participant's SVV and PU can be modeled as a weighted vector sum (Dyde *et al.*, 2006; Mittelstaedt, 1983).

The SVV typically shows a very large influence of the gravitational vector and little else, whereas the PU shows a more evenly distributed set of weighted influences. In experiment 1 we therefore chose the PU as our measure of perceived upright and attempted to influence it by presenting both static and dynamic sounds that had both physical and spectral cues to up. Participants lay on their side, a posture that causes the perceptual upright to lie between the upright indicated by the long axis of the body and gravity. By positioning sounds above and below the participant, aligned with gravity, we attempted to strengthen the gravity component and swing the perceptual upright towards this direction. In experiment 2, non-spatial frequency sweeps (high-to-low or low-to high) were played through headphones to try and influence the perceived direction of ambiguous visual motion with this auditory–spatial correspondence in a way similar to Maeda *et al.*, 2004. In their experiment, participants judged the perceived direction of ambiguous vertical visual motion while sitting upright. Gravity and body frames were thus always aligned. Here we separated these frames by testing our subjects upright and while lying on their side to identify the reference frame(s) in which this purely crossmodal correspondence operated.

There were several possible outcomes for experiment 2. One possibility is that our crossmodal correspondence defined non-spatial sounds could have no effect on disambiguating ambiguous visual motion. Another possibility is that the sounds could have an influence only along the axis of the body (indicating that higher frequencies were associated with the top of the body). Conversely, the sounds could have an influence only along the gravity axis (indicating that higher frequencies were associated with the gravitationally defined up). Finally, the sounds could have an influence along both the body and the gravity axes (indicating that they were associated with both reference frames).

2. Experiment 1. Methods

2.1. Participants

Twelve graduate students (six male, six female, mean age 25) participated in the experiment with static sounds. For the experiments with dynamic sounds there were thirteen participants (five male, eight female, mean age 24) who were either graduate students or undergraduate students who volunteered from the Undergraduate Registered Participant Pool (URPP) and who were awarded class credit for participation. All participants reported having normal or corrected-to-normal vision, normal hearing, and no known vestibular or self-orientation issues. All experiments were approved by the ethics board of York University and followed the guidelines of Helsinki.

2.2. Apparatus and General Setup

Participants laid on their right side on a massage table in a large dark room with their head off the edge of the massage table supported in a way that ensured that their ears were unoccluded (Fig. 1A). A computer monitor (ViewSonic VG732M-LED) was aligned orthogonal to the participants' line of sight. Participants viewed the monitor through a circular black shroud (diameter 35°), which kept their head at a constant distance of 20 cm from the display and

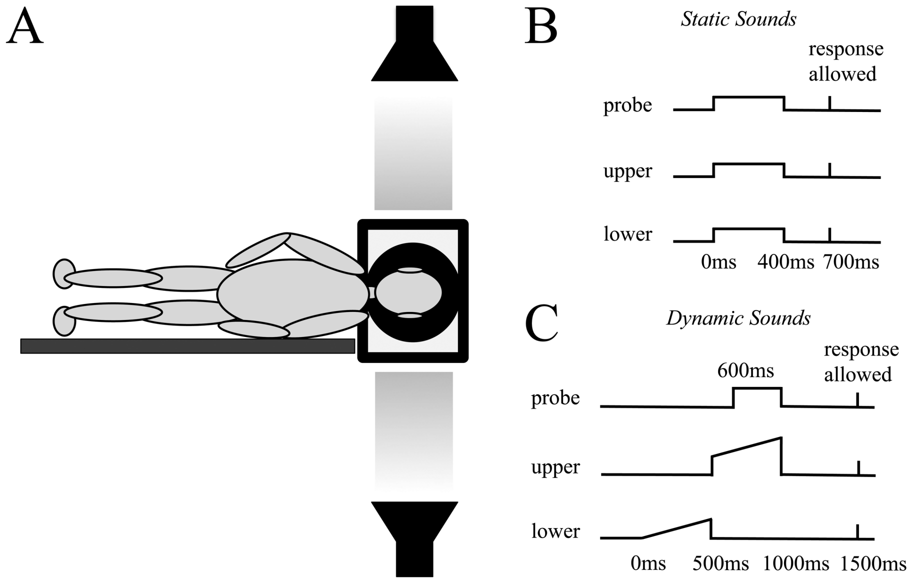

Figure 1. Setup and stimulus timings (ms) for experiment 1. (A) Participants lay on their right side with their ears unoccluded. Participants viewed the screen through a circular shroud. Speakers were positioned above and below the participant's head. (B) Timing for static sounds showing the 'p/d' probe stimulus, and auditory stimuli from the upper and lower speakers. Responses could not be made until after a delay period. (C) Timing for the dynamic sounds.

blocked out external visual stimuli. Speakers were positioned 54 cm above and below the participant's head with the top speaker aimed down to the participant's left ear and the bottom speaker aimed up at the participant's right ear. Participants held a computer mouse in their right hand with which they made responses. The monitor, speakers, and mouse were all connected to a MacBook Pro and all inputs and outputs were controlled using MATLAB. All visual stimuli for all experiments were generated and presented using the Psychtoolbox package of functions (Kleiner *et al.*, 2007). Anti-aliasing features were utilized for visual smoothing of all visual objects. All auditory stimuli were presented using the Psychportaudio MATLAB package functions.

2.3. Visual Stimulus and Adaptive Staircase

The perceptual upright was measured using the Oriented Character Recognition Test (OCHART) (Dyde *et al.*, 2006). The perceptual upright is defined as the orientation at which a character is most easily recognized. Participants were presented with the ambiguous visual character 'p' which can be interpreted as a 'p' or a 'd' depending on its perceived orientation. The character was presented as a capital 'P' in a sans serif Calibri font and subtended 5.75° × 4°. To obtain the participant's perceptual upright, the two orientations where the character was most ambiguous were determined. To determine these two orientations, two QUEST adaptive staircases were used (Watson and Pelli, 1983) to adjust the orientation of the presented characters to find the orientations at which participants reported a ‘p’ or ‘d’ 50% of the time. The PU was calculated by finding the point midway between the two orientations at which the p/d character was most ambiguous (i.e., the orientation at which the character was most unambiguous). The two QUESTs began with the character at opposing orientations of $-90°$ and $+90°$ (positive angles representing rotation in the clockwise direction and counterclockwise rotation for negative angles where 0° is defined as the top of the participant’s head). The two QUESTs were programmed such that when participants responded that they saw a ‘p’, the presented character that was initially tilted over at $-90°$ was tilted further counter-clockwise on the next presentation while the character initially presented at 90° would be tilted in the clockwise direction. The QUEST procedures used a cumulative Gaussian function with an initial slope estimate of 5 for both the static- and dynamic-sounds experiments.

2.4. Auditory Stimulus

2.4.1. Static Sounds

The auditory stimulus for the static-sounds experimental condition was composed of pure tones where the speaker located above the participant’s head played a pure tone of 1200 Hz and the speaker below simultaneously played a pure tone of 200 Hz (Note 1), lasting for 400 ms (Fig. 1B). The sounds had a volume of 83 dB, measured at the ear. During control trials, no auditory tone was presented. For all experiments a sampling frequency of 44 100 Hz was used to generate and play all auditory stimuli.

2.4.2. Dynamic Sounds

The auditory stimulus for the dynamic-sounds experimental condition was composed of a tone that began with the bottom speaker and ended with the above speaker and changed its frequency and volume as it did so. The sound from the bottom speaker swept from 0 Hz to 600 Hz over a period of 500 ms and immediately following this the above speaker played a tone sweep from 600 Hz to 1200 Hz over a period of 500 ms (Fig. 1C). This lead to the perception of an auditory ‘object’ (see Lewald and Ehrenstein, 1998) rising both physically (from low space to high space, travelling through the head) and in frequency (from 0 Hz to 1200 Hz) over a period of 1 s. The whole auditory sweep (i.e., the two sweeps combined as one array) was multiplied by a ramp from 0–1 leading to a linear increase in volume as the sound ‘travelled’ from low to high physical space. This was done to make the sound appear more as an object perceptually and add to the sense that it was travelling from low to high space (Eitan *et al.*, 2008).

2.5. Experimental Paradigm

The paradigm was to measure the PU in participants lying on their side. Previous studies (Dyde *et al.*, 2006) have shown that the PU lies between the gravity and body axes. Thus, if sound were to have an effect on the PU then sounds aligned with the gravity axis would pull the PU more towards the gravity axis. The PU was therefore measured in participants lying on their side in the presence and absence of sounds. To obtain the PU, each participant completed two QUEST adaptive staircases (one to home in on the p-to-d transition and one for the d-to-p transition) for the with- and without-sound conditions. For the static-sound experiment each QUEST was set to run for 60 trials leading to a total of 240 trials (2 conditions $\times$ 2 QUEST adaptive staircases $\times$ 60 trials each). For the dynamic-sound experiment each QUEST was set to run for 50 trials leading to a total of 200 trials (2 conditions $\times$ 2 QUEST adaptive staircases $\times$ 50 trials each). The QUEST procedures in the dynamic-sound experiment had fewer trials because the results of the static-sounds experiment showed that convergence on the final estimate was reached in less than 50 trials. The presentations of all four QUEST conditions were randomly interleaved in the experimental paradigm.

For the static-sound experiment, every trial began with the ‘p/d’ probe that was presented for 400 ms in a particular orientation chosen by the QUEST adaptive staircase depending on the participant’s previous responses. The auditory stimulus was presented with the visual stimulus for the duration of the probe stimulus while in control trials no sounds were presented. After the probe disappeared participants responded after an enforced 300 ms delay. A left mouse-click denoted that they perceived a ‘d’ and right mouse-click denoted a ‘p’. After their response there was a 400 ms delay before the next trial. The timings are shown graphically in Fig. 1B.

For the dynamic-sound experiment each trial began with the auditory stimulus. After 600 ms, the visual probe was presented for 400 ms after which both the visual probe and auditory stimulus were switched off and the screen went grey. During no-sound control trials there was a 600 ms silent delay before the character probe was presented for 400 ms followed by grey screen. There was then a 500 ms delay before the participants could register their response, and the next trial would begin. The timings are shown graphically in Fig. 1C.

2.6. Data Analysis

The PSE values from the four adaptive staircases (QUEST procedures starting at $-90°$ and $+90°$ for each of the sound and no-sound conditions) were obtained using the ‘QUESTMean’ function from PsychToolbox running within MATLAB (Watson and Pelli, 1983). The QUESTMean function provides the final estimate of the routine calculated as the mean of the posterior distribution function and represents the best estimate of PSE. The PSEs represent the orientations at which the character was most ambiguous. The perceptual upright (PU) is defined as the orientation midway between the two PSE values. To determine if there were any effects of sound on participants' PUs, within-subjects t-tests were performed comparing the sound and no-sound conditions.

3. Experiment 1. Results

The orientations of the PSEs and PUs for each individual participant and their means are shown in Fig. 2 for both the static and dynamic experiments. Within-subjects t-tests revealed no significant difference in PU values between the sound and no-sound conditions for either the static sounds $t(11) = -1.57$, $p = 0.14$ or the dynamic sounds $t(12) = -0.72$, $p = 0.48$.

4. Experiment 1. Discussion

There were no significant differences between the sound and no-sound conditions using any of the sound cues we presented. The mean of all the PUs collapsed across both experiments was $-27.3°$ (standard error = 2.6). The PU was thus between the 'up' defined by the body (0°) and the 'up' defined by gravity ($-90°$). Dyde *et al.* (2006) reported a body:gravity ratio of 2.5 which would correspond to a tilt of $-21.8°$ but acknowledged a large inter-subject variability as can be seen in Fig. 2. The sounds in the frequency range we used produced no tendency to strengthen the gravity vector and swing the PU towards the gravitational vertical ($-90°$) or to strengthen the body vector and swing the PU towards the body (0°). This suggests that the crossmodal correspondence between frequency and spatial elevation, even when aided by physically locating sounds above and below the participant and adding a loudness cue, does not provide an independent cue that contributes to perceived self-orientation. While the potential spatial cues of the loudness–height and pitch–height crossmodal correspondences were combined and thus confounded in the dynamic-sounds experiment, this only strengthens the claim that their effect on perceived upright is negligible. This finding agrees with the extremely minor contribution that auditory cues seem to play in evoking self-motion (Väljamäe, 2009) and stabilizing posture (Easton *et al.*, 1998; Sakamoto *et al.*, 2004; Sakellari and Soames, 1996).

The results of experiment 1 showed that spatial sound (real or *via* the pitch–height crossmodal correspondence) had no effect on a person's perceived orientation and did not appear to strengthen the influence of either the body or gravitational cues. It may be the case that sound simply does not influence perceived orientation, but it may also be the case that these particular sounds were not optimized to elicit an effect. Ernst and Banks (2002) revealed that

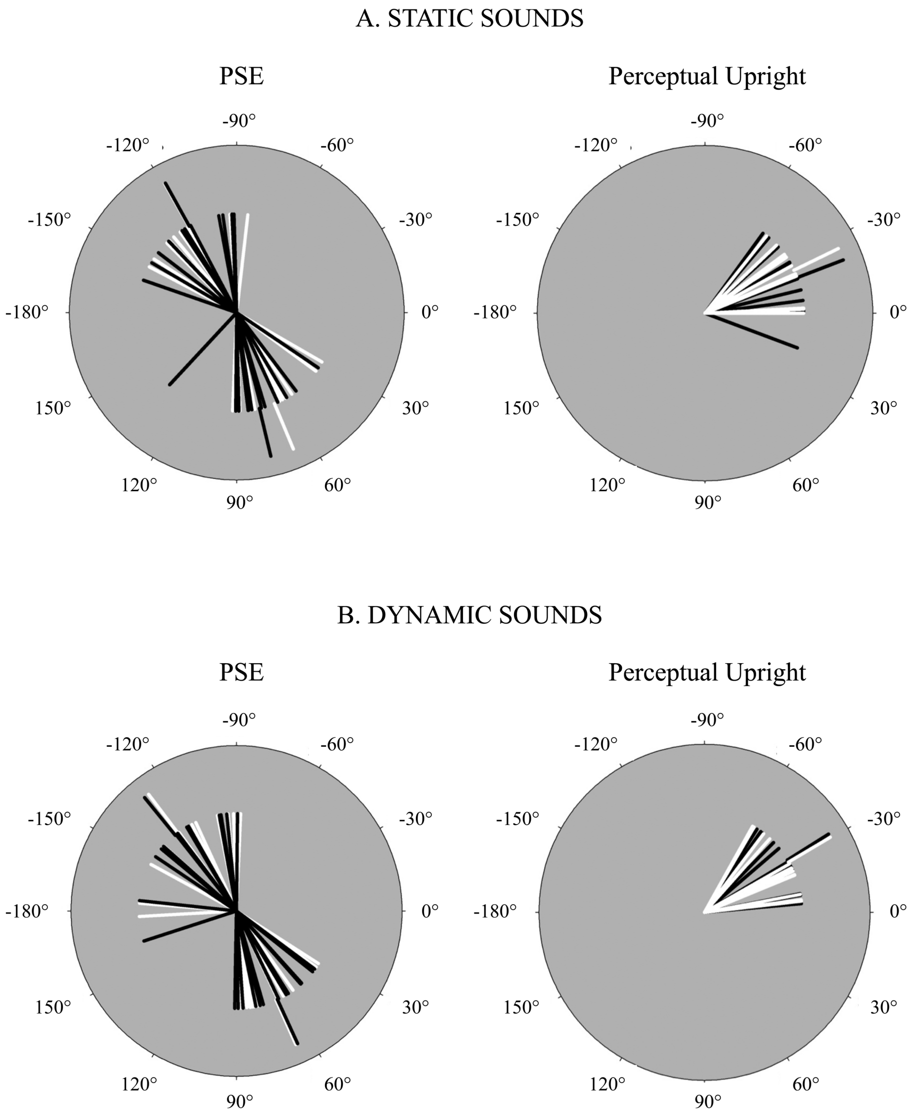

Figure 2. Polar plots showing PSEs (left panels) and perceptual upright (PU, right panels) for the sound (black) vs. no-sound (white) conditions in the static-sound experiment (A) and dynamic-sound experiment (B). 0° corresponds to the top of the participant's head and −90° to gravitational up. Inner lines show the orientations for each individual subject while the outer lines show mean values.

the relative influence of combined perceptual cues depends on the reliability of those cues, and it is possible that our spatially localized sound stimuli were simply not reliable enough (i.e., the precise spatial source locations were not entirely clear) to influence the PU. Our results do not negate the possibility that if sounds were made to be more reliable than other contributing orientation cues (e.g., vestibular, visual, tactile, or proprioceptive) there might be a detectable influence of sound on perceived orientation.

To potentially increase the spatial reliability and effectiveness of auditory spatial stimuli used as orientation cues some suggestions are as follows. Rather than using simple tones, which are relatively difficult to localize (Roffler and Butler, 1968), the auditory stimulus could include a broader spectral range (e.g., band-passed filtered noise). Another consideration is the possibility that the particular frequencies chosen in the static and dynamic sounds experiment were not in the most effective range. Parise *et al.* (2014) showed that the frequency–elevation mapping for auditory localization was more pronounced in some frequency ranges than others. Our auditory stimuli were less than 1200 Hz while their stimuli ranged to above 8000 Hz and showed variable effects at different frequencies. Another methodological change that could have made our experimental design more sensitive was that we compared a sound condition to a no-sound condition rather than comparing a congruent sound (i.e., high sound above and low sound below) condition to an incongruent sound condition (i.e., low sound above and high sound below). Further experiments are warranted.

In experiment 2 we replicated a study by Maeda *et al.* (2004), which demonstrated that non-spatial sounds of ascending and descending pitch could bias the perception of ambiguous visual motion, and asked the question: in which reference frame does this effect operate?

5. Experiment 2. Methods

5.1. Participants

Nine participants (five male, four female, mean age 25) volunteered to participate in the four within-subject experimental conditions. All participants reported having normal or corrected-to-normal vision, normal hearing, and no known vestibular or self-orientation issues. All experiments were approved by the ethics board of York University and followed the guidelines of Helsinki.

5.2. Apparatus and General Setup

Participants performed the experiment either upright or laying on their right side in a dark room. In the upright conditions, participants sat comfortably on a chair at a table and looked at a monitor through a black circular shroud (diameter 35°), which kept their head at a constant distance of 20 cm from the display and blocked out external visual stimuli. During the on-side conditions participants laid on their right side on a padded massage table and looked at an identical monitor through an identical shroud, both of which were tilted with the participant. Padding was offered to participants to make them more comfortable. Auditory stimuli were presented *via* noise-cancelling headphones (Maxell NC-11) in both the upright and on-side orientations. Participants held a computer mouse in their right hand, which was used to make responses. The monitor, headphones, and mouse were all connected to a MacBook Pro and all inputs and outputs were controlled using MATLAB.

5.3. Visual Stimuli

The visual stimuli were composed of two, superimposed, spatially enveloped sinusoidal luminance gratings with a Michelson contrast of 0.05, spatial frequency 2 cycles/degree, and temporal frequency 6.25 Hz moving in opposite directions. The grating filled the 35° circular aperture. To create ambiguous motion, two component gratings, drifting in opposite directions, with contrast ratios of 1.0/0.0, 0.7/0.3, 0.6/0.4, 0.5/0.5, 0.4/0.6, 0.3/0.7, and 0.0/1.0 were added together. The 0.5/0.5 contrast ratio produced completely ambiguous motion. Two visual motion conditions (left/right and upward/downward relative to the head) and two body orientations (upright and right side down) were run in four separate blocks in counterbalanced order. The arrangement is shown diagrammatically in Fig. 3A.

5.4. Auditory Stimuli

There were three distinct sounds presented to participants in this study. Two of the auditory stimuli were tones with a constant rate of change, either ascending from 0.3 to 2.0 kHz, or descending from 2.0 to 0.3 kHz over a period of 1 s. The other auditory stimulus was broadband pink noise (i.e., noise with frequency spectrum such that the energy per Hz is inversely proportional to the frequency of the signal; Bak *et al.*, 1987) presented for a period of 1 s. Sounds were delivered through headphones. Presentation was the same in each ear and volume was constant throughout the experiment. The volume was selected to be loud enough to hear clearly but to not cause discomfort.

5.5. Experimental Paradigm

Each trial began with a centered fixation cross (2° in each direction) on a grey background. The cross was replaced after a random time delay of between 1 and 2 s with a compound moving grating (duration 400 ms) with one of the seven contrast ratios chosen randomly. One of the three sounds (i.e., ascending, descending, or pink noise, duration 200 ms) was played starting 50 ms after onset of the visual gratings. The visual motion stimulus was followed by a grey screen, which signaled participants to make a forced choice of the direction of motion they saw (i.e., either upwards or downwards, or leftwards or rightwards, depending on the direction of visual motion condition). If they saw

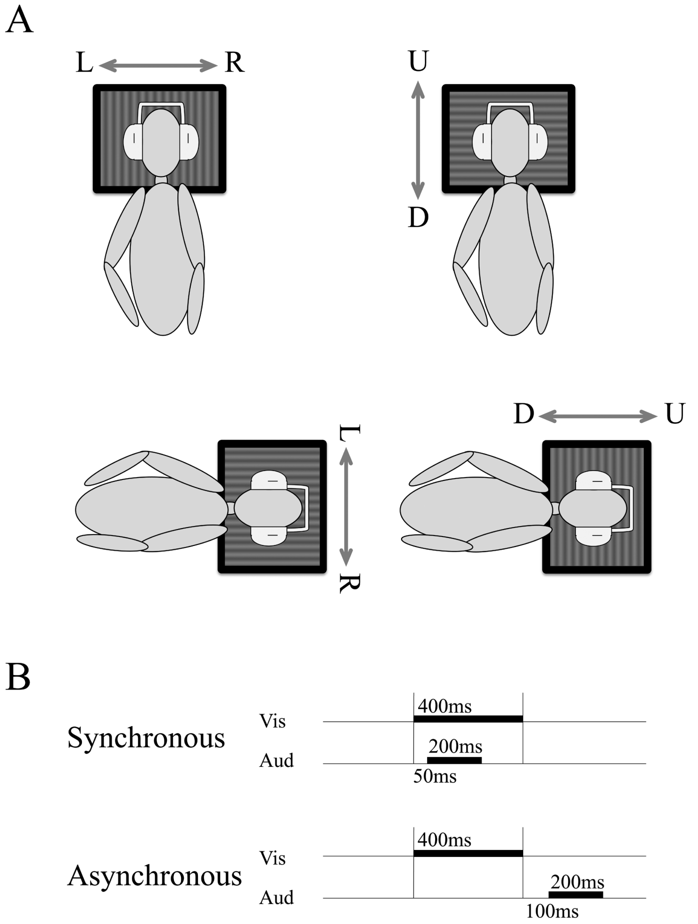

Figure 3. Conditions for experiment 2. (A) The four orientation and visual motion combinations. In the top row, body orientation is upright whereas in the bottom row they are laying on their right side. In the left column visual motion direction is leftward/rightward relative to head, and in the right column visual motion direction is upward/downward relative to the head. (B) The timings of stimulus presentations in ms for the synchronous and asynchronous trials.

downward or leftward visual motion relative to their head they responded with a left-click of the mouse, and a right-click for upward or rightward motion.

5.6. Response Bias

Mixed into the experimental design were a series of trials designed to detect possible response bias. In these trials only the three most ambiguous contrast ratios for the gratings were presented (contrast ratios: 0.4/0.6, 0.5/0.5, 0.6/0.4). One of two sounds (ascending or descending) was played but was desynchronized from visual display by being played 100 ms after the visual gratings stimulus presentation ended (Fig. 3B). Subjects responded about the perceived direction of the grating after the sound had ended.

In each of the four orientation and visual motion direction combinations there were 20 synchronous trials for each of the three sound conditions at each of the seven contrast ratios, leading to 420 trials. There were also 20 asynchronous trials for each of two sounds at three contrast ratios leading to a further 120 trials for each of the four direction of visual motion and body orientation combinations. The synchronous trials and the asynchronous trials were randomly interleaved, leading to a total of 540 trials in each condition.

5.7. Data Analysis

Data were plotted as percentage of times one particular direction was chosen as a function of the relative contrast of the visual gratings. Logistic functions:

$$\%\text{ up or }\%\text{ left} = y_0 + a/(1 + \exp[-\{x - x_0\}/b]) \qquad (1)$$

were fit through each participant's data to obtain the points of subjective equality (PSE x_0). The curve-fitting algorithm searched for four parameter values (PSE x_0, slope b, the amplitude of the sigmoid a, and the lower lapse rate y_0) using a maximum likelihood optimization routine (Myung, 2003), which is the preferred method when dealing with generalized non-linear models such as this one (Fesselier and Knoblauch, 2006). PSE here is defined as halfway between the lower and upper lapse rates, where upper lapse rate is given by $100 - a - y_0\%$.

To explore the effects of the three independent variables (head orientation, direction of visual motion, and influence of sound) on participants' responses, the PSEs from the descending and ascending sound conditions were subjected to a $2 \times 2 \times 2$ within-subjects mixed-models analysis of variance (ANOVA) using an unstructured covariance structure (this structure had the best measure of fit) to control sphericity (Field *et al.*, 2012).

6. Experiment 2. Results

Figure 4 shows illustrative psychometric functions fitted through the mean data from all the participants. A $2 \times 2 \times 2$ ANOVA was performed on the PSEs obtained from each subject individually to explore the PSE data for the factors of body orientation (on-side or upright), direction of visual motion (upward/downward or leftward/rightward), and sound (ascending and descending). The ANOVA revealed a significant interaction between the effects of sound and visual direction with $F(1, 56) = 18.60$, $p < 0.001$. The interaction between sound and visual direction revealed that, across orientations, the effect of sound in the up-down tasks was greater than the effect of sound in the left-right tasks. There were no interactions between orientation and visual direction $F(1, 56) = 0.36$, $p > 0.05$, or between sound and

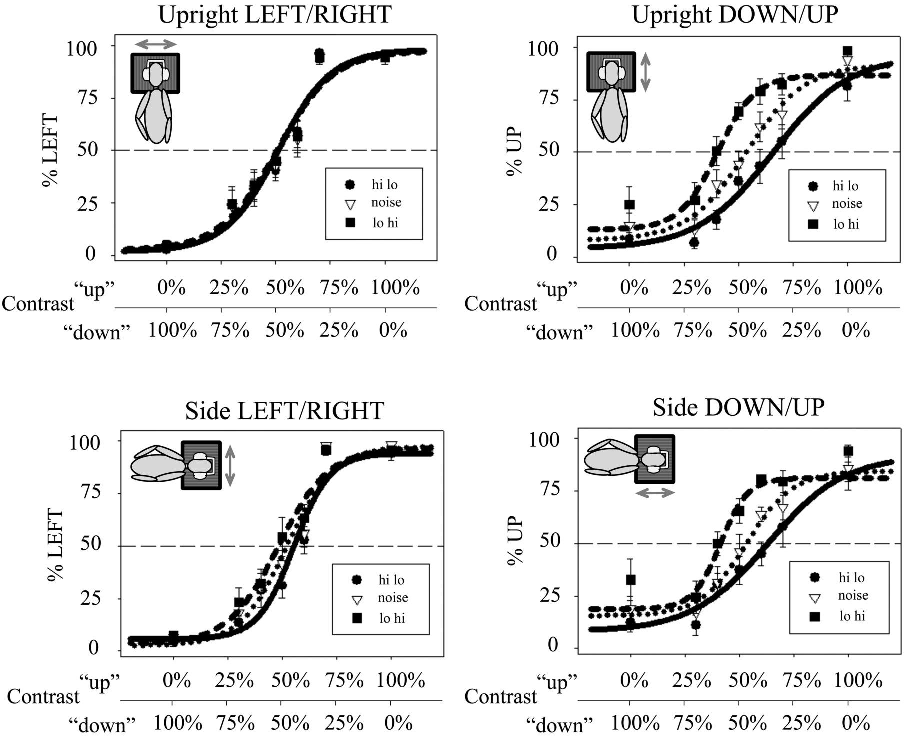

Figure 4. Psychometric functions for each of the four orientation and visual motion direction combinations. These illustrative psychometric functions were fit to the means of all participants. The y-axis represents the percent of leftward (left panels) or upward responses (right panels) while the x-axis represents the relative contrast of the oppositely moving gratings that made up each stimulus. For the on-side leftward/rightward panels, leftward visual motion corresponds to gravitational upwards. The three curves in each graph represent best-fit logistics to the mean participant responses for each of the three sound conditions (solid for descending, dotted for noise, dashed for ascending). Error bars show standard errors between participants.

orientation $F(1,56) = 0.88$, $p > 0.05$. There was a main effect of sound $F(1,56) = 25.44$, $p < 0.001$ showing a difference between the ascending and descending sound conditions. There were no main effects of orientation $F(1,56) = 0.07$, $p > 0.05$ or visual direction $F(1,56) = 0.25$, $p > 0.05$. There were no three-way interactions $F(1,56) = 0.04$, $p = 0.83$.

T-tests were then performed to compare the PSEs between the ascending and descending conditions in each of the orientation by visual motion direction conditions. The PSEs measured with the ascending sound stimulus were all significantly different from those measured with the descending sound stimulus except for the condition where participants were upright performing the leftward/rightward visual motion task, which gave results of $t(8) = 0.39$, $p = 0.70$. The results for the upright upward/downward condition were $t(8) = -3.30$, $p < 0.05$, on-side upward/downward $t(8) = -3.81$, $p < 0.05$, and the on-side leftward/rightward $t(8) = -3.23$, $p < 0.05$. These values were error corrected as a family using the false discovery rate method (Benjamini and Hochberg, 1995). Mean differences between the PSEs were calculated as effect sizes for the upright leftward/rightward (0.01), upright upward/downward (0.33), on-side leftward/rightward (0.07), and on-side upward/downward conditions (0.39).

Comparing the upper and lower lapse rate values from the ascending and descending sound conditions showed no significant differences in any of the body orientation/visual direction combinations (Note 2), thus confirming that PSE differences reflect the influence of sound condition rather than changes in the lapse rates.

6.1. Response Bias Control Trials

To confirm that the effect of sound biased the visual percept and was not due to response bias, the effects of sound in the synchronous and asynchronous trials were compared in the upright upward/downward condition. The raw response data (i.e., number of times participants responded upwards or rightwards out of the 20 trials) for the conditions where participants sat upright and made upward/downward judgments, as well as the on-side leftward/rightward judgments were tested to confirm Maeda *et al.*'s null findings and to verify the condition most crucial to the arguments presented in this paper.

To compare the effect of sound in the synchronous and asynchronous conditions four repeated measures, factorial ANOVAs were performed with contrast ratio (0.4/0.6, 0.5/0.5, 0.6, 0.4) and sound conditions (ascending tones, descending tones, and noise) as factors. For the asynchronous upright upward/downward condition, there was no main effect of sound $F(2, 16) = 3.98$, $p > 0.05$, a significant main effect of contrast ratio $F(1, 8) = 8.85$, $p < 0.05$, and no interaction $F(2, 16) = 0.37$, $p > 0.05$. For the asynchronous on-side leftward/rightward condition, there was no main effect of sound $F(2, 16) = 2.14$, $p > 0.05$, a significant main effect of contrast ratio $F(1, 8) = 8.76$, $p < 0.05$, and no interaction $F(2, 16) = 1.07$, $p > 0.05$. By contrast, confirming the results described above with the comparable subset of data, for the synchronous upright upward/downward condition, there was a main effect of sound with $F(2, 16) = 19.71$, $p < 0.05$, a main effect of contrast ratio $F(1, 8) = 8.38$, $p < 0.05$, and no interaction $F(2, 16) = 0.14$, $p > 0.05$. For the synchronous on-side leftward/rightward condition, there was a main effect of sound $F(2, 16) = 3.89$, $p < 0.05$, a main effect of contrast ratio $F(1, 8) = 8.47$, $p < 0.05$, and no interaction $F(2, 16) = 0.37$, $p > 0.05$.

7. Experiment 2. Discussion

Significant effects of synchronous sound were found in all conditions except the upright condition where participants made leftward/rightward judgments. Since significant effects were found both when visual motion was in line with the body axis (on-side upward/downward visual motion direction) and when it was in line with the gravitational axis (on-side leftward/rightward visual motion direction), the results suggest that the crossmodal correspondence binding auditory frequency to spatial elevation had components along *both* these axes.

The ascending and descending sound conditions shifted the PSE 7% along the gravitational axis (as seen in the on-side leftward/rightward judgments) and 39% along the body axis (as seen in the on-side up/down judgments), calculated as mean differences, corresponding to a body-to-gravity ratio of 5.6:1. Dyde *et al.* (2006) reported a ratio of 6.8:1 under comparable conditions although high inter-subject variability and low sample size in both studies makes direct comparison unreliable. Thus, both gravity and body reference frames appear to be used in binding auditory and visual orientation information, in line with Parise *et al.*'s (2014) and Roffler and Butler's (1968) conclusions from sound localization errors while lying on one side. Figure 5 shows the orientation of the resultant axis determined from the vector sum of the gravity and body components of the effect of sound on visual motion. For a person lying on their right side, this axis is tilted by 10.2° from the body axis, towards the direction of gravity.

7.1. No Response Bias

Analysis of the asynchronous response bias control trials revealed no main effects of sound when the auditory stimulus did not overlap temporally with the visual motion stimulus. In contrast, the synchronous on-side leftward/rightward and upright upward/downward conditions both showed a main effect of sound. This suggests that the influence of sound on visual motion (along both the gravitational and body axes) reflects a true perceptual ef-

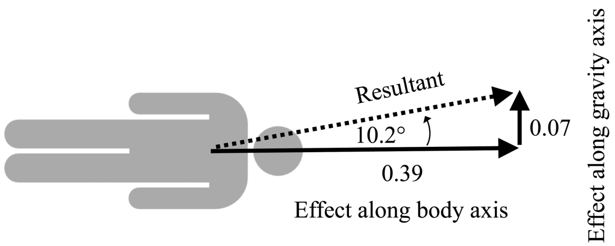

Figure 5. The relative sizes of the effects of sound along the long axis of the body and along the gravitational vertical are show here along with their vector sum. The angle of the resultant auditory pitch effect was tilted 10.2° towards gravitational vertical.

fect rather than introducing a response bias. These results are congruent with Maeda *et al.* (2004) who tested a comprehensive range of audiovisual temporal asynchronies and found no effect of the sound in disambiguating visual motion unless the sound was played synchronously with the visual stimulus.

7.2. Comparison of the Two Experiments

The results of experiment 1 showed no effect of sound, for either real or pitch-based auditory spatial cues, on the perceptual upright suggesting that the brain does not use the crossmodal correspondence between auditory pitch and spatial elevation as a cue for self-orientation. By contrast, in experiment 2 the ascending and descending tones biased ambiguous visual motion within both body and gravitational reference frames, suggesting that the crossmodal correspondence between auditory pitch and visuospatial height is integrated with visual motion systematically. How then can these results be reconciled? The 'optimal tilt' found in experiment 2 is comparable to the tilt of the perceptual upright. We speculate that the reason we did not find an effect in experiment 1, and the reason that sound has not been shown by anyone to have a quantitative effect on a person's perception of their orientation, may be because humans are not accurate at independently determining the three-dimensional direction of sound (Parise *et al.*, 2014 for a comprehensive description of the enormous errors that are made). Instead the perceived direction of a sound may be determined with reference to a person's perceptual upright, which is itself determined by a combination of visual, body and gravitational cues for a given angle of head or body tilt (see Barnett-Cowan *et al.*, 2013). This would explain how sound's effect could have been missed under all but the most sensitive conditions and is also discouraging to those looking to provide alternative orienting cues to people with balance problems.

Acknowledgements

The experiments were supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to LRH. MJC held a scholarship from the Visual Science and Applications CREATE grant.

Notes

1. In order to confirm that these sounds could be appropriately localized, an additional experiment was performed in which 5 participants identified whether the speaker above or below their right and left ear, respectively, was playing the high note. Performance was perfect over 80 trials.
2. We used simple t-tests with no correction for multiple tests. No tests were significant:

Lower lapse rates

side left/right: $t = 0.6738$, $\mathrm{df} = 8$, $p = 0.5194$

side up/down: $t = 1.6935$, $\mathrm{df} = 8$, $p = 0.1288$

upright left/right: $t = 0.6813$, $\mathrm{df} = 8$, $p = 0.5149$

upright up/down: $t = 0.5759$, $\mathrm{df} = 8$, $p = 0.5805$

Upper lapse rates

side left/right: $t = 1.4279$, $\mathrm{df} = 8$, $p = 0.1912$

side up/down: $t = 1.1078$, $\mathrm{df} = 8$, $p = 0.3001$

upright left/right: $t = -0.4583$, $\mathrm{df} = 8$, $p = 0.6589$

upright up/down: $t = 1.2709$, $\mathrm{df} = 8$, $p = 0.2395$

References

Aubert, H. (1861). Eine scheinbare Drehung von Objekten bei Neigung des Kopfes nach rechts oder links, *Virchows Arch.* **20**, 381–393.

Bak, P., Tang, C. and Wiesenfeld, K. (1987). Self-organized criticality: an explanation of $1/f$ noise, *Phys. Rev. Lett.* **59**, 381–384.

Barnett-Cowan, M., Jenkin, H. L., Jenkin, M. R. and Harris, L. R. (2013). Asymmetrical representation of body orientation, *J. Vis.* **13**, 1–9.

Ben-Artzi, E. and Marks, L. E. (1995). Visual–auditory interaction in speeded classification: role of stimulus difference, *Percept. Psychophys.* **57**, 1151–1162.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc.* **57**, 289–300.

Bernstein, I. H. and Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time, *J. Exp. Psychol.* **87**, 241–247.

Chiou, R. and Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting, *Perception* **41**, 339–353.

Dyde, R. T., Jenkin, M. R. and Harris, L. R. (2006). The subjective visual vertical and the perceptual upright, *Exp. Brain Res.* **173**, 612–622.

Easton, D. R., Greene, A. J., DiZio, P. and Lackner, J. R. (1998). Auditory cues for orientation and postural control in sighted and congenitally blind people, *Exp. Brain Res.* **118**, 541–550.

Eitan, Z., Schupak, A. and Marks, L. E. (2008). Louder is higher: Cross-modal interaction of loudness change and vertical motion in speeded classification, in: *Proceedings of the 10th International Conference on Music Perception and Cognition*, Sapporo, Japan.

Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion, *Nature* **415**, 429–433.

Evans, K. K. and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features, *J. Vis.* **10**, 1–12.

Fesselier, R. Y. and Knoblauch, K. (2006). Modeling psychometric functions in R, *Beh. Res. Methods.* **38**, 28–41.

Field, A., Miles, J. and Field, Z. (2012). *Discovering Statistics Using R*. Sage Publications, Thousand Oaks, CA, USA.

Harris, L. R. (2009). Visual–vestibular interactions, in: *The Encyclopedia of Neuroscience*, L. R. Squire (Ed.), Vol. 10, pp. 381–387. Academic Press, Oxford, UK.

Harris, L. R., Jenkin, M., Dyde, R. T. and Jenkin, H. L. (2011). Enhancing visual cues to orientation: suggestions for space travelers and the elderly, in: *Progress in Brain Research*, A. M. Green, C. E. Chapman, J. F. Kalaska and F. Lepore (Eds), Vol. 191, pp. 133–142. Elsevier, Amsterdam, Netherlands.

Howard, I. P. (1982). *Human Visual Orientation*. Wiley, New York, NY, USA.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R. and Broussard, C. (2007). What's new in Psychtoolbox-3, *Perception* **36**(14), 1.

Lackner, J. R. and DiZio, P. (2005). Vestibular, proprioceptive, and haptic contributions to spatial orientation, *Annu. Rev. Psychol.* **56**, 115–147.

Lewald, J. and Ehrenstein, W. H. (1998). Influence of head-to-trunk position on sound localization, *Exp. Brain Res.* **121**, 230–238.

Maeda, F., Kanai, R. and Shimojo, S. (2004). Changing pitch induced visual motion illusion, *Curr. Biol.* **14**, R990–R991.

Melara, R. D. and Marks, L. E. (1990). Interaction among auditory dimensions: timbre, pitch, and loudness, *Percept. Psychophys.* **48**, 169–178.

Melara, R. D. and O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions, *J. Exp. Psychol. Gen.* **116**, 323–336.

Mittelstaedt, H. (1983). A new solution to the problem of the subjective vertical, *Naturwissenschaften* **70**, 272–281.

Mossbridge, J. A., Grabowecky, M. and Suzuki, S. (2011). Changes in auditory frequency guide visual–spatial attention, *Cognition* **121**, 133–139.

Mudd, S. A. (1963). Spatial stereotypes of four dimensions of pure tone, *J. Exp. Psychol.* **66**, 347–352.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation, *J. Math. Psychol.* **47**, 90–100.

Parise, C. V., Knorre, K. and Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing, *Proc. Natl Acad. Sci.* **111**, 6104–6108.

Patching, G. R. and Quinlan, P. T. (2002). Garner and congruence effects in the speeded classification of bimodal signals, *J. Exp. Psychol. Human* **28**, 755–775.

Petersen, H., Magnusson, M., Johansson, R., Akesson, M. and Fransson, P. A. (1995). Acoustic cues and postural control, *Scand. J. Rehabil. Med.* **27**, 99–104.

Petersen, H., Magnusson, M., Johansson, R. and Fransson, P. A. (1996). Auditory feedback regulation of perturbed stance in stroke patients, *Scand. J. Rehabil. Med.* **28**, 217–223.

Pratt, C. C. (1930). The spatial character of high and low tones, *J. Exp. Psychol.* **13**, 278–285.

Roffler, S. K. and Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane, *J. Acoust. Soc. Am.* **43**, 1255–1259.

Sadaghiani, S., Joost, X. M. and Noppeney, U. (2009). Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing, *J. Neurosci.* **29**, 6490–6499.

Sakamoto, S., Osada, Y., Suzuki, Y. and Gyoba, J. (2004). The effects of linearly moving sound images on self-motion perception, *Acoust. Sci. Technol.* **25**, 100–102.

Sakellari, V. and Soames, R. W. (1996). Auditory and visual interactions in postural stabilization, *Ergonomics* **39**, 634–648.

Spence, C. (2011). Crossmodal correspondences: a tutorial review, *Atten. Percept. Psychophys.* **73**, 971–995.

Stumpf, K. (1883). *Tonpsychologie*. S. Hirzel, Leipzig, Germany.

Väljamäe, A. (2009). Auditorily-induced illusory self-motion: a review, *Brain Res. Rev.* **61**, 240–255.

Watson, A. B. and Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method, *Percept. Psychophys.* **33**, 113–120.