

Visual motion perception induced by sounds in vertical plane

Wataru Teramoto^{a,b,*}, Yuko Manaka^c, Souta Hidaka^a, Yoichi Sugita^c, Ryota Miyauchi^b,
Shuichi Sakamoto^b, Jiro Gyoba^a, Yukio Iwaya^b, Yōiti Suzuki^b

^a Department of Psychology, Graduate School of Arts and Letters, Tohoku University, Sendai, Miyagi, Japan

^b Research Institute of Electrical Communication, Tohoku University, Sendai, Miyagi, Japan

^c Neuroscience Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

ARTICLE INFO

Article history:

Received 20 November 2009

Received in revised form 17 April 2010

Accepted 22 May 2010

Keywords:

Motion perception

Audio–visual interaction

Auditory spectral cue

Binaural cue

ABSTRACT

The alternation of sounds in the left and right ears induces motion perception of a static visual stimulus (SIVM: Sound-Induced Visual Motion). In this case, binaural cues were of considerable benefit in perceiving locations and movements of the sounds. The present study investigated how a spectral cue – another important cue for sound localization and motion perception – contributed to the SIVM. In experiments, two alternating sound sources aligned in the vertical plane were presented, synchronized with a static visual stimulus. We found that the proportion of the SIVM and the magnitude of the perceived movements of the static visual stimulus increased with an increase of retinal eccentricity (1.875–30°), indicating the influence of the spectral cue on the SIVM. These findings suggest that the SIVM can be generalized to the whole two dimensional audio–visual space, and strongly imply that there are common neural substrates for auditory and visual motion perception in the brain.

© 2010 Elsevier Ireland Ltd. All rights reserved.

Selecting signals from different sensory modalities and binding them appropriately as a single event are necessary to ensure a stable perceptual world and to guide appropriate actions to the external world [10]. Recent studies on multisensory perception have revealed that auditory information affects visual motion perception more than classical views had assumed. In a stream-bounce display, in which two visual stimuli moved towards and across each other, a bouncing perception occurs more frequently when a transient auditory stimulus was presented at the moment the visual stimuli were overlapped than when no sound was presented [24,30]. A transient auditory signal presented with an ambiguous visual motion display modulated the temporal positional information of a visual stimulus, resulting in the coherent visual motion perception and the visual motion aftereffects [11]. These studies revealed the modulatory, but not driving or inducing, effects of sounds on visual motion perception.

More recently, we clearly demonstrated that auditory signals could induce visual motion perception of a static stimulus [14]. A blinking visual stimulus at a fixed location was perceived to be laterally moving when its flash onset was synchronized to an alternating left–right sound source (Sound-Induced Visual Motion, SIVM). Together with a previous report, which showed that the

adaptation to a looming visual stimulus induced not only the motion aftereffect (MAE) in visual modality but also the aftereffect in auditory modality [15], our finding suggests that there are common neural substrates for auditory and visual motion perception [2,4,7,9,13,16–20,22,25–29].

It has been known that binaural cues as well as spectral cues can be used to localize sound sources and perceive shifts in sound location [5,3]. The relative importance of these cues is changed depending on a sound location or the direction of shifts in sound location. Sound localizations and the perception of shifts in sound location in the horizontal plane are mainly mediated by binaural cues such as interaural time differences and interaural level differences, while those in the vertical plane are mediated by spectral change cues produced by the directional filtering of the pinnae, head and shoulders. In our previous study [14], a sound was delivered alternately to the left and right ears via headphones, so that only binaural cues were likely to be used to localize each sound and/or to perceive shifts in sound location. Therefore, it is yet unclear how spectral cues contribute to the SIVM. Thus, we investigated how the SIVM occurred when a sound was delivered alternately from two loudspeakers aligned in a vertical meridian. If the SIVM is observed in the vertical plane as well as horizontal plane, it should mean that the SIVM can be generated by both binaural and spectral cues and occur in the whole two-dimensional audio–visual space.

In experiments, we also measured the SIVM in the horizontal plane with two loudspeakers. In our previous study, the sounds were presented via headphones [14]. Therefore, if we observe any

* Corresponding author at: Research Institute of Electrical Communication, Tohoku University, 2-1-1 Katahira Aoba-ku, Sendai, Miyagi 980-8577, Japan.
Tel.: +81 22 217 5461; fax: +81 22 217 5535.

E-mail address: teraw@ais.riec.tohoku.ac.jp (W. Teramoto).

difference in the SIVM between the vertical and horizontal planes in the current study, the difference could be attributed not only to the difference in cues for sound localization and/or shifts in sound location (binaural and spectral cues), but also to the difference in the origin of the sound source (external and internal). To control this confounding effect, we measured the SIVM in the horizontal as well as vertical planes, by using the same experimental setups. We call the SIVM in the vertical and horizontal planes as V-SIVM and H-SIVM, respectively, hereafter.

In Experiment 1, we measured how frequently the SIVM was observed, while varying with retinal eccentricity. Eight people including four of the authors (W.T., Y.M., S.H., and S.S.) participated in Experiment 1. All participants had normal or corrected-to-normal vision and normal hearing. Except for the authors, the participants were naïve as to the purpose of the experiment. Informed consent was obtained from each participant before the experiment. All procedures were approved by the local ethics committee of the Tohoku University. Visual stimuli were presented on a CRT monitor (Sony Trinitron GDM-F520, 21 inch, 800 × 600 pixels) with a refresh rate of 60 Hz and partly presented by red light emitted diodes (LEDs). Auditory stimuli were presented through two full-range loudspeakers (HOSIDEN, 0254-7N101, ϕ 30 mm) installed in a small plastics box (108 cm³) and positioned 76.4 cm above and below the CRT monitor for the measurement of the V-SIVM and 76.4 cm to the left and right of the monitor for the H-SIVM (i.e., 45° in elevation and in azimuth, respectively). The digital signals for the LEDs and auditory stimuli were converted to analog signals using the audio interfaces (M-Audio ProFire Lightbridge and BEHRINGER ADA8000 8ch AD-DA converter). The experiment was controlled by a customized PC (Dell-XPS 710) and MATLAB (The Mathworks, Inc.) with the Psychophysics Toolbox [6,21] and an open-source audio I/O library (Playrec, <http://www.playrec.co.uk/>). We confirmed that the onset of the visual and the auditory stimuli was almost synchronized (<10 ms) with a digital oscilloscope. The viewing distance was 76.4 cm and head movements were restrained by a chinrest. The experiment was conducted in a dark anechoic room.

The visual stimulus (5.0 cd/m², 400 ms duration) for the V-SIVM and the H-SIVM was a horizontal (0.2° × 3.0°) and vertical (3.0° × 0.2°) white bar, respectively. The white bar was presented in a black background (40° × 30°; 0.2 cd/m²) at the center of the monitor. The retinal eccentricity of the white bar varied by changing the position of a red fixation point. The fixation for the V-SIVM was placed 4.875° to the left of the white bar and at a retinal eccentricity along the vertical meridian of either ±1.875°, ±3.75°, ±7.5°, ±15° or ±30° (positive values indicate upper visual field, while negative values indicate lower visual field). The fixation for the H-SIVM was placed at a retinal eccentricity along the horizontal meridian of either 1.875°, 3.75°, 7.5°, 15° or 30° in the left visual field so that the white bar was presented in the right visual field (The right visual field was dominant for all the participants). The fixations at 15 and 30° were presented by the LEDs and the other fixations were presented on the monitor, irrespective of the tested visual field. The auditory stimulus was a 50 ms white noise burst with a cosine ramp of 5 ms at the onset and offset (sound pressure level: 74 dB, sampling frequency: 44.1 kHz). The white noise bursts were created from trial to trial.

There were two conditions for sound presentation. In the alternate condition, the sounds were presented alternately from two different loudspeakers – the upper and lower loudspeakers for the V-SIVM and the left and right loudspeakers for the H-SIVM. The onset of the sound was synchronized with that of the white bar. In the no-sound condition, any sound was not presented. The fixation point was presented from the beginning of each trial. 500 ms after the participant's pressing a start button, the white bar was presented six times (intersimulus interval: 100 ms, the duration of the

sequence: 2.5 s), accompanying the sounds in the alternate condition and without any sound in the no-sound condition. The retinal eccentricity of the white bar varied randomly from trial to trial. The participants' task was to report whether they perceived visual motion. The V-SIVM in the upper visual field, the V-SIVM in the lower visual field and the H-SIVM were measured in three different sessions. Each session consisted of one training and two experimental blocks. In the training block, the participants were asked to discriminate between static and moving visual stimuli without any sound for 100 trials: Visual stimuli (2: static/moving) × Eccentricity (5) × Repetition (10). This session was introduced in order for the participants to take firm fixation because our preliminary observation showed that eye movements also induce the illusory motion perception of static visual stimuli. The displacement angle for the moving visual stimuli was 0.2° upward and downward for the V-SIVM sessions and leftward and rightward for the H-SIVM sessions. The training session was repeated until the discrimination performance was reached above 75% for each retinal eccentricity. The main experimental blocks had 200 trials with static white bars: Eccentricity (5) × Auditory condition (2) × Repetitions (20). Additionally, 80 filler trials where the white bar was actually displaced by 0.2° were randomly introduced: Eccentricity (5) × Auditory condition (2) × Repetitions (8). For half of the trials in the alternate condition, the first sound originated from the upper loudspeaker for the V-SIVM sessions and from the left loudspeaker for the H-SIVM sessions, and for the other half, the first sound originated from the lower and the right loudspeakers, respectively. In the filler trials, the initial direction of displacement of the white bar was consistent with that of the sound in the alternate condition and randomly assigned in the no-sound condition. The presentation order of the conditions was randomized in each experimental block.

The proportion of the V-SIVM is shown in Fig. 1A. A blinking white bar appeared to be moving upward and downward when they were presented in conjunction with the alternating sounds in the vertical plane. This tendency was gradually strengthened with increasing retinal eccentricity, irrespective of the upper and lower visual fields. This illusory visual motion was not observed without sounds. We conducted a repeated analysis of variance (ANOVA) on data for each visual field with two within-participant factors; eccentricity (1.875°, 3.75°, 7.5°, 15°, and 30°) and two auditory conditions (alternate and no-sound). The ANOVA revealed significant main effects of auditory condition (lower visual field, $F_{1,7} = 64.78$, $p < .001$; upper visual field, $F_{1,7} = 18.77$, $p = .003$) and eccentricity (lower visual field, $F_{4,28} = 3.53$, $p = .019$; upper visual field, $F_{4,28} = 8.74$, $p < .001$). Additionally, an interaction effect between these factors was significant for the upper visual field ($F_{4,28} = 6.95$, $p < .001$). The illusory visual motion occurred more frequently in the alternate condition than in the no-sound condition at all the eccentricities for the lower visual field and at 7.5, 15, and 30° of retinal eccentricity for the upper visual field ($F_{s1,35} = 10.67$, $ps < .001$). These results indicate that the alternating two sounds in the vertical plane induce the visual motion perception of a static stimulus in the vertical direction, and the effect of the sounds becomes greater with the increment of the eccentricity of the visual stimuli.

The proportion of the H-SIVM is shown in Fig. 1B. A repeated two-way ANOVA revealed significant main effects of auditory condition ($F_{1,7} = 14.83$, $p = .006$) and eccentricity ($F_{4,28} = 3.76$, $p = .014$), and an interaction effect ($F_{4,28} = 3.42$, $p = .021$). The interaction revealed that the illusory visual motion occurred more frequently in the alternate condition than in the no-sound condition at all the eccentricities except for 1.875° ($F_{s1,35} > 8.39$, $ps < .001$). These results indicate that a visual stimulus blinking at a fixed location was more frequently perceived to be laterally moving when the sounds were presented alternately from the left and right loudspeakers, especially at larger retinal eccentricities. These findings were consistent with our previous findings, in which the alternat-

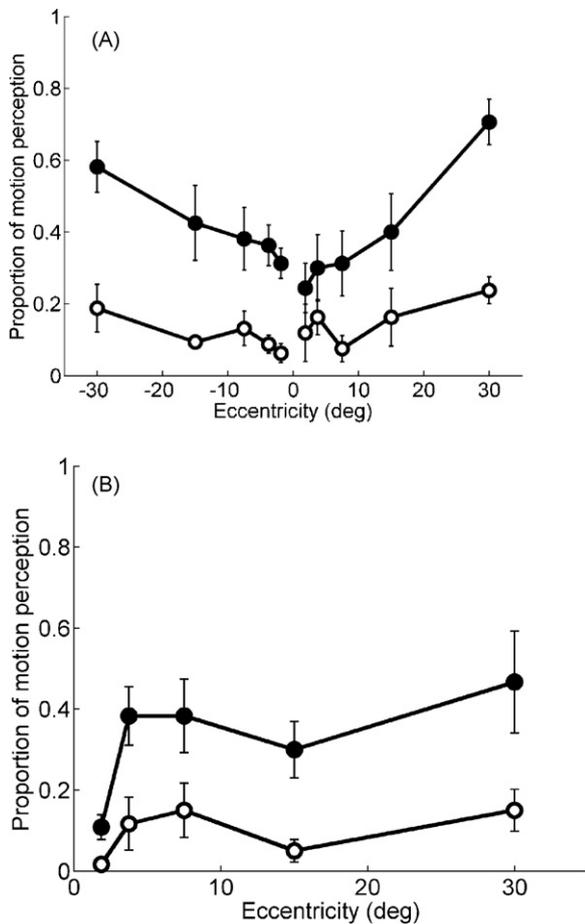


Fig. 1. The proportion of illusory visual motion perception as a function of retinal eccentricity. Sounds were presented alternately through two vertically-aligned loudspeakers (A) and horizontally-aligned loudspeakers (B). In figure A, negative and positive eccentricities indicate the lower and upper visual fields, respectively. The data in figure B was obtained from the right visual field. Filled and open circles indicate the alternate (sound) and no-sound conditions, respectively. Error bar denotes the standard error of the mean.

ing sounds were presented via headphones, instead of loudspeakers [14]. Thus, the internalized or artificial sound is not a crucial factor for the SIVM. In other words, the SIVM can occur under our everyday circumstances.

It could be argued that the participants simply judged, but not perceived, as motion when the visual stimulus was spatially uncertain at larger retinal eccentricities and the sound was clearly alternated between two loudspeakers. To rule out this possibility, we quantified the magnitude of the perceived displacement of a static visual stimulus for the V-SIVM and the H-SIVM in Experiment 2. If the participants' responses were related only to decisional processes, they would not be able to estimate the magnitude of the displacement of the visual stimulus accurately because they would not experience any perceptual displacement of the visual stimulus in such a case. Consequently, the magnitude of the displacement would not change as a function of retinal eccentricity. Those who served in Experiment 1 participated in the present experiment as well. The setup and the stimulus configuration in this experiment were identical with those of Experiment 1, except for the procedure described below. Only the alternate condition was tested. After the presentation of six pairs of the white bar and the sound, a probe bar, which was the same as the blinking bar, was presented. The participants' task was to adjust the width of the probe to the perceived displacements of the white bar. Each visual field (upper, lower and right) was tested in different sessions. The order of the sessions was

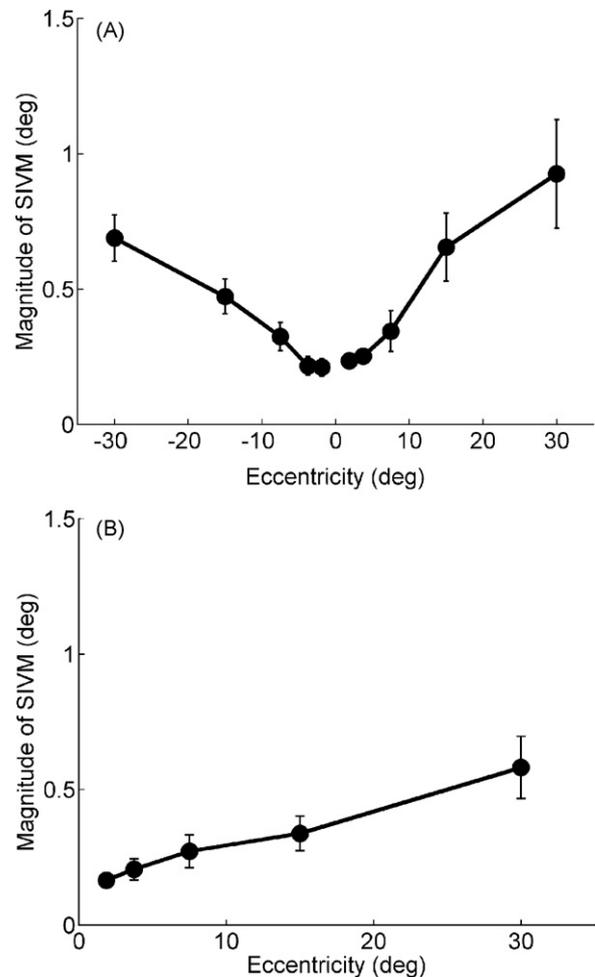


Fig. 2. The magnitude of the Sound-Induced Visual Motion (SIVM) as a function of retinal eccentricity. Sounds were presented alternately through two vertically-aligned loudspeakers (A) and horizontally-aligned loudspeakers (B). In figure A, negative and positive eccentricities indicate the lower and upper visual fields, respectively. The data in figure B was obtained from the right visual field. Filled and open circles indicate the alternate (sound) and no-sound conditions, respectively. Error bar denotes the standard error of the mean.

counterbalanced between the participants. A session consisted of a training block and two experimental blocks. The training block was the same as those in Experiment 1. Each experimental block had 25 trials: Eccentricity (5) × Repetitions (5). The first sound was delivered to one of the two loudspeakers for the half of the trials and the other loudspeaker for the other half.

The magnitude of the perceived displacements when the SIVM occurred was shown as a function of retinal eccentricity in Fig. 2. The trials in which the participants reported any displacements were considered as those in which the SIVM occurred. The proportions of motion perception increased with an increase of retinal eccentricity (see Supplemental Fig. 1), just as those in Experiment 1. We conducted a repeated ANOVA for the magnitude of the perceived displacements with two within-participant factors: tested visual field (upper, lower and right) and eccentricity (1.875, 3.75, 7.5, 15, and 30°). The ANOVA revealed a main effect of tested visual field ($F_{2,14} = 4.62$, $p = .029$) and eccentricity ($F_{4,28} = 19.18$, $p < .001$). The H-SIVM was significantly smaller than the V-SIVM for the upper visual field (Tukey's HSD test, $p < .05$). Furthermore, larger displacements were obtained with increasing retinal eccentricity, irrespective of the tested visual field.

To further single out the perceptual nature of the current phenomenon, the magnitude of the V-SIVM was measured using a motion-nulling procedure [19] with a method of constant stim-

uli. The results showed that the alternating sounds not only induce motion perception but also cancel the motion perception of visual stimuli that are displaced in the direction opposite to that of the sounds (see [Supplemental Fig. 2](#)). This finding clearly suggests that the participants experienced conspicuous displacements of visual stimuli when the sounds were alternated. These results suggest that the V-SIVM and H-SIVM are unattributable only to decisional and/or response processes.

In the present study, we demonstrated that a static visual stimulus blinking in one place was perceived to be moving up and down when it was accompanied with the alternation of sounds delivered from the upper and lower loudspeakers. The proportion of the SIVM (Experiment 1) and the magnitude of the perceived displacement of the static visual stimulus (Experiment 2) increased with increasing retinal eccentricity (1.875–30°). These characteristics of the SIVM were also observed when the sounds were presented in the horizontal plane. It has been reported that spectral cues are necessary for humans to perceive locations and shifts in sound location in the vertical plane, while binaural cues such as interaural time and level differences are of considerable benefit in perception of sounds in the horizontal plane [2,4]. Both of these localization cues have an essential role in perception of sounds in the diagonal plane (e.g., [12]). Therefore, our findings could be generalized to the whole two dimensional audio–visual space.

The current findings would be related to so-called cortical magnification [8,31]. Visual field is represented topographically in relatively lower cortical areas, but the scale of the map is changed depending on retinal location and cortical area. Representations in the central visual field are larger than those in the peripheral visual field; a visual area represented by a single neuron increases with increasing retinal eccentricity, resulting in low visual acuity and resolution. In our present study, the magnitude of the SIVM increased with increasing retinal eccentricity. We confirmed that this tendency was also observed for an actually moving visual stimulus (see [Supplemental Figure 3](#)). Furthermore, it is well known that the cortical magnification is changed depending on a given visual field [23]. The cortical magnification for the upper visual field decreases more sharply than that for the lower visual field and even more than that for the right (or temporal) visual field with increasing retinal eccentricity. Consistent with this fact, the SIVM was larger in the upper visual field than in the other visual fields. The difference in the magnitude of the SIVM among the visual fields would, therefore, be explained by the cortical magnification.

Recently, it was reported that auditory signals could modulate the perceived positions of visual inputs when the reliability or the intensity of visual stimuli decreased [1]. The lower reliability of visual information would lead to higher proportion and larger magnitude of the SIVM at larger eccentricities. This suggests that audio–visual interactions are regulated by a compensatory manner in motion processing in the brain depending on the reliability of the visual and auditory inputs. It can be assumed that the positional information originated from the auditory modality would alter the perceived spatial position of visual stimuli and, therefore, illusory visual apparent motion would be evoked. The second possibility is that the auditory apparent motion signal induced by the alternation of sounds would directly drive or trigger visual motion perception of a static stimulus. We think these assumptions about the underlying mechanisms are equally plausible at present. Further studies will clarify the underlying mechanisms of the SIVM.

In conclusion, the present study demonstrated that sounds aligned in the vertical plane as well as those in the horizontal plane could induce the visual motion perception of a static stimulus (SIVM) especially at larger retinal eccentricities. These strong

and general effects of sounds on visual motion perception imply that there are common neural substrates for auditory and visual motion perception in the brain.

Acknowledgements

This research was supported Grant-in-Aid for Specially Promoted Research (No. 19001004).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.neulet.2010.05.065](https://doi.org/10.1016/j.neulet.2010.05.065).

References

- [1] D. Alais, D. Burr, The ventriloquist effect results from near-optimal bimodal integration, *Curr. Biol.* 14 (2004) 257–262.
- [2] A. Alink, W. Singer, L. Muckli, Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex, *J. Neurosci.* 28 (2008) 2690–2697.
- [3] F. Asano, Y. Suzuki, T. Sone, Role of spectral cues in median plane localization, *J. Acoust. Soc. Am.* 88 (1990) 159–168.
- [4] O. Baumann, M.W. Greenlee, Neural correlates of coherent audiovisual motion perception, *Cereb. Cortex* 17 (2007) 1433–1443.
- [5] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA, 1983.
- [6] D.H. Brainard, The psychophysics toolbox, *Spat. Vis.* 10 (1997) 433–436.
- [7] F. Bremmer, A. Schlack, N.J. Shah, O. Zafiris, M. Kubischik, K. Hoffmann, K. Zilles, G.R. Fink, Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys, *Neuron* 29 (2001) 287–296.
- [8] P.M. Daniel, D. Whitteridge, The representation of the visual field on the cerebral cortex in monkeys, *J. Physiol.* 159 (1961) 203–221.
- [9] W.H. Ehrenstein, A.H. Reinhardt-Rutland, A cross-modal aftereffect: auditory displacement following adaptation to visual motion, *Percept. Mot. Skills* 82 (1996) 23–26.
- [10] M.O. Ernst, H.H. Bühlhoff, Merging the senses into a robust percept, *Trends Cogn. Sci.* 8 (2004) 162–169.
- [11] E. Freeman, J. Driver, Direction of visual apparent motion driven solely by timing of a static sound, *Curr. Biol.* 18 (2008) 1262–1266.
- [12] D.W. Grantham, B.W. Hornsby, E.A. Erpenbeck, Auditory spatial resolution in horizontal, vertical, and diagonal planes, *J. Acoust. Soc. Am.* 114 (2003) 1009–1022.
- [13] M.C. Hagen, O. Franzén, F. McGlone, G. Essick, C. Dancer, J.V. Pardo, Tactile motion activates the human middle temporal/V5 (MT/V5) complex, *Eur. J. Neurosci.* 16 (2002) 957–964.
- [14] S. Hidaka, Y. Manaka, W. Teramoto, Y. Sugita, R. Miyauchi, J. Gyoba, Y. Suzuki, Y. Iwaya, Alternation of sound location induces visual motion perception of a static object, *PLoS One* 4 (2009) e8188.
- [15] N. Kitagawa, S. Ichihara, Hearing visual motion in depth, *Nature* 416 (2002) 172–174.
- [16] J.W. Lewis, M.S. Beauchamp, E.A. DeYoe, A comparison of visual and auditory motion processing in human cerebral cortex, *Cereb. Cortex* 10 (2000) 873–888.
- [17] J. Lopez-Moliner, S. Soto-Faraco, Vision affects how fast we hear sounds move, *J. Vis.* 7 (2007) 1–7.
- [18] F. Maeda, R. Kanai, S. Shimojo, Changing pitch induced visual motion illusion, *Curr. Biol.* 14 (2004) R990–991.
- [19] S. Mateeff, J. Hohnsbein, T. Noack, Dynamic visual capture: apparent auditory motion induced by a moving visual target, *Perception* 14 (1985) 721–727.
- [20] G.F. Meyer, S.M. Wuerger, Cross-modal integration of auditory and visual motion signals, *Neuroreport* 12 (2001) 2557–2560.
- [21] D.G. Pelli, The videotoolbox software for visual psychophysics: transforming numbers into movies, *Spat. Vis.* 10 (1983) 437–442.
- [22] M. Radeau, P. Bertelson, The effect of a textured visual field on modality dominance in a ventriloquism situation, *Percept. Psychophys.* 20 (1976) 227–235.
- [23] J. Rovamo, V. Virsu, An estimation and application of the human cortical magnification factor, *Exp. Brain Res.* 37 (1979) 495–510.
- [24] R. Sekuler, A.B. Sekuler, R. Lau, Sound alters visual motion perception, *Nature* 385 (1997) 308.
- [25] S. Soto-Faraco, A. Kingstone, C. Spence, Multisensory contributions to the perception of motion, *Neuropsychologia* 41 (2003) 1847–1862.
- [26] S. Soto-Faraco, C. Spence, A. Kingstone, Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities, *J. Exp. Psychol. Hum. Percept. Perform.* 30 (2004) 330–345.

- [27] S. Soto-Faraco, C. Spence, D. Lloyd, A. Kingstone, Moving multisensory research along: Motion perception across sensory modalities, *Curr. Dir. Psychol. Sci.* 13 (2004) 29–32.
- [28] J.J. Stekelenburg, J. Vroomen, Neural correlates of audiovisual motion capture, *Exp. Brain Res.* 198 (2009) 383–390.
- [29] J. Vroomen, B de Gelder, Visual motion influences the contingent auditory motion aftereffect, *Psychol. Sci.* 14 (2003) 357–361.
- [30] K. Watanabe, S. Shimojo, When sound affects vision: effects of auditory grouping on visual motion perception, *Psychol. Sci.* 12 (2001) 109–116.
- [31] D. Whitteridge, P.M. Daniel, The representation of the visual field on the calcarine cortex, in: R. Jung, H. Kornhuber (Eds.), *The Visual System: Neurophysiology and Psychophysics*, Springer, Berlin, Göttingen, Heidelberg, 1961.