

Sounds are perceived as louder when accompanied by visual movement

Article (Accepted Version)

Maniglia, Marcello, Grassi, Massimo and Ward, Jamie (2017) Sounds are perceived as louder when accompanied by visual movement. *Multisensory Research*, 30 (2). pp. 159-177. ISSN 2213-4794

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/68076/>

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.

Sounds Are Perceived as Louder when Accompanied by Visual Movement

Marcello Maniglia¹, Massimo Grassi² and Jamie Ward^{3,4,*}

¹ University of California, Riverside, CA, USA

² Department of Psychology, University of Padova, Italy

³ School of Psychology, University of Sussex, Falmer, Brighton, BN1 9QH, UK

⁴ Sackler Centre for Consciousness Science, University of Sussex, Brighton, UK

Received 5 September 2016; accepted 4 April 2017

* To whom correspondence should be addressed. E-mail: jamiew@sussex.ac.uk

Abstract

In this study, we present three experiments investigating the influence of visual movement on auditory judgments. In Experiments 1 and 2, two bursts of noise were presented and participants were required to judge which was louder using a forced-choice task. One of the two bursts was accompanied by a moving disc. The other burst either was accompanied by no visual stimulus (Experiment 1) or by a static disc (Experiment 2). When the two sounds were of identical intensity participants judged the sound accompanied by the moving disc as louder. The effect was greater when auditory stimuli were of the same intensity but it was still present for mid-to-high intensities. In a third, control, experiment participants judged the pitch (and not the loudness) of a pair of tones. Here the pattern was different: there was no effect of visual motion for sounds of the same pitch, with a reversed effect for mid-to-high pitch differences (the effect of motion lowered the pitch). This showed no shift of response towards the interval accompanied by the moving disc. In contrast, the effect on pitch was reversed in comparison to what observed for loudness, with mid-to-high frequency sound accompanied by motion rated as lower in pitch respect to the static intervals.

The natural tendency for moving objects to elicit sounds may lead to an automatic perceptual influence of vision over sound particularly when the latter is ambiguous. This is the first account of this novel audio-visual interaction.

Keywords

Cross-modal correspondence, loudness judgment, visual movement, auditory perception

1. Introduction

Structured correspondences exist between different dimensions of perception (such as luminance, weight, size, pitch, loudness) both within and across sensory modalities. For instance, darker objects look heavier (but feel lighter — Walker et al., 2010), a sound may be perceived as louder when accompanied by a bright light (Odgaard et al., 2004), and higher pitch and smaller size tend to be perceived as perceptually grouped (Grassi et al., 2013; Parise and Spence, 2009). These cross-modal correspondences do have real world consequences in cognition (e.g. in sound symbolism — Monaghan et al., 2012) and perception (Parise and Spence, 2009). This study documents a novel cross-modal correspondence between visual motion and loudness judgments: sounds accompanied by visual motion are judged as louder.

The movement of objects can be perceived through multiple senses particularly vision, hearing and touch (Soto-Faraco et al., 2003). The brain processes motion in both unimodal sensory cortices and polymodal areas, located in the parietal and frontal regions, that gather inputs from different senses, as revealed by fMRI (Bremmer et al., 2001). The magnitude of this polymodal activity is greatest when auditory and visual motion are coherent (Baumann and Greenlee, 2007). In the audio visual bouncing effect (ABE), the the perception of bouncing is induced by the presence of an acoustic stimulus at the moment of two moving stimuli intercepting in a display otherwise perceived as streaming (Sekuler, Sekuler and Lau, 1997). This is associated with higher activation in multimodal areas and concomitant reduction of

activation in unimodal regions, hinting at a competitive interaction between unimodal and multimodal areas, and that this process may be disrupted through brain stimulation (Grassi and Casco, 2012; Maniglia et al., 2012; Sekuler et al., 1997). Certain static sounds, such as pitch modulations (which imply ascending/descending), can activate regions involved in visual motion that have been traditionally classed as unimodal (Sadaghiani et al., 2009). In summary, these studies demonstrate how motion perception is constructed from multiple sensory signals both static and moving.

In terms of behaviour, the timing of a static sound (Freeman and Driver, 2008) or the spatial position of two static sounds (Teramoto et al., 2012) can drive the direction of apparent visual motion. Moreover, the use of congruent auditory movement facilitates learning of visual motion detection (but see Alais and Burr, 2004; Seitz et al., 2006). The movement of our own bodies can also affect the way that we structure our percepts. Phillips-Silver and Trainor (2007) showed that participants perceived unaccented beats as waltz or march according to their previous motor movements to the same rhythms. Similarly, observed lip movements can change what is heard when the two streams are incongruent (McGurk and MacDonald, 1976).

While it is well known that visual motion information can bias auditory and tactile perception of motion (and the other way around, with sound biasing the perceived number of visual stimuli or the perceived trajectory of moving objects, see Shams et al., 2000, and Sekuler et al., 1997, respectively) however, no studies — to date — investigated the influence of visual motion on the perception of other acoustic parameters such as loudness or pitch (but see Grassi and Pavan, 2012, for a modulatory action of a fronto-parallel visual motion on the subjective duration of a sound). The preferential pairing of certain auditory and visual dimensions (e.g., brightness and loudness) has been investigated in a number of ways. Stein et al. (1996) reported that observers rated the brightness of a weak visual stimulus as higher when it was accompanied by a brief burst of white noise relative to when it was presented alone. The auditory stimulus

produced brightness enhancement regardless of the relative location of the auditory cue source. Odgaard et al. (2003) attempted to replicate this finding and tested whether the reported enhancement reflected an early-stage sensory interaction or a later-stage response bias effect. They argued that single stimulus ratings are more susceptible to criterion shifts or response bias (Green and Swets, 1966; Tanner et al., 1967) than two interval forced choice procedures (2IFC). In their procedure, two visual stimuli were presented consecutively (with one accompanied by a sound) and the task was to determine which was brighter. Under these test conditions there was no tendency to perceive the visual stimulus that was accompanied by a sound as brighter (when the two visual stimuli were equated). They concluded that the influence of sound on brightness is not a perceptual phenomenon but rather a post-perceptual decision. In a subsequent study, using the same 2IFC procedure, Odgaard et al. (2004) demonstrated that when two sounds are presented in succession the presence of a visual stimulus during one of the sounds enhanced the perceived loudness (when the intensity of the sounds was in fact identical). Thus, there appears to be an asymmetry in cross-modal correspondences: a visual stimulus can make a sound appear louder, but an auditory stimulus does not make a visual stimulus appear brighter (Note 1).

This basic paradigm has been adapted by others to explore, for instance, auditory-vibrotactile interactions (Yarrow et al., 2008) and auditory influences on contrast detection (Lippert et al., 2007). Jaekl and Soto-Faraco (2010), in particular, showed that audiovisual enhancement induced by multisensory integration is more effective when visual stimuli present features compatible with the magnocellular pathway. In the present study we extend it to consider the influence of visual movement on loudness judgments. Our initial motivation for exploring this was from unpublished observations of individuals with a visual-to-auditory synaesthesia (Saenz and Koch, 2008) for whom faster moving visual stimuli were reported as louder. In natural environments there also tends to be a correlation between whether an object

moves and whether it makes a sound that may form the basis of such a cross-modal correspondence. Sounds originate because of a vibration (i.e., a motion). Therefore, still objects tend to be silent whereas moving ones often produce sounds.

The current study includes three experiments. In Experiments 1 and 2, we investigated whether visual motion enhances loudness perception. Participants were asked to judge which of two bursts of noise was louder in a two interval forced choice paradigm. In each trial, one interval (i.e., one noise burst) was accompanied by a moving visual stimulus whereas the other was accompanied by no visual stimulus or a static stimulus. In Experiment 3, the experiment was replicated to investigate whether visual motion perception enhances pitch perception. Participants were asked to judge which of two tone intervals was higher in pitch.

2. Experiment 1

2.1. Method

2.1.1. Participants

A group of eight participants (five females) were recruited from the University of Sussex (mean age = 21.23 years, S.D. = 8.19). They all reported normal hearing and normal (or corrected to normal) vision. Each participant was tested for up to one hour as described below. They were refunded with credits or paid £5 for participating.

2.1.2. Materials

The experiment was conducted in a dark, sound-attenuating booth (thereby occluding background luminance and ambient noise), with a chinrest mounted in place to keep the

participant's head positioned 57 cm from the computer's monitor. The monitor used was a 19 inch, Dell 1908FP, ATI Radeon HD 2400 pro. The maximum luminance was 70 c/d^2 as measured with a Minolta- LS100. Visual stimuli were generated using the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). The auditory stimulus was a broadband noise, 20 Hz–20 kHz, produced using the Matlab PSYCHOACOUSTICS toolbox (Soranzo and Grassi, 2014). For the auditory stimuli, we selected five intensities spaced by approximately by 0.4 log units, similarly to Odgaard et al. (2003, 2004). Intensities were 33, 37, 41, 45, and 49 dB (A) and were measured through headphones with an Bruel and Kjaer 2610 Measuring Amplifier. Auditory stimuli were delivered through Sennheiser HD 497 Headphones. The duration of both the sound and the moving stimulus were 80 ms. Onset and offset of the noise bursts were modulated in amplitude with a 5 ms raised cosine ramp. Stimuli presentation and data collection were carried on in a Matlab program, operating on a Dell Optiplex 755 Intel Core Duo CPU 2.33 GHz.

2.1.3. Procedure

The design followed Odgaard et al. (2003), except that the target was the auditory stimulus. The participants' first dark adapted for 15 minutes in the dark room before entering the booth and received further instructions. On each trial, the participants heard two noises in succession, separated by a brief delay (temporal two interval forced-choice); the task was to decide whether the first or second of them was louder by pressing key 'z' or 'm' on a keyboard. The participants were instructed to fixate a central point of the screen and also informed that sometimes a moving stimulus (a white disc) would appear simultaneously with one of the two noises in each pair, but that they should ignore it when performing the task.

Each trial began with a 1 s fixation. Both sounds in the first and second interval were presented for 80 ms, with a 500 ms Inter Stimulus Interval. Movement started at the onset of

the sound, either in the first or in the second interval, and lasted for 80 msec. The disc's diameter was 2 degrees of visual angle. The speed of the disc was 12.6 deg/sec, with the disc moving randomly towards the left or right, starting from the centre of the screen. After the presentation of the second stimulus participants were cued to enter their responses (without time constraint) before the next trial began.

The combination of each possible intensity level (5) in the first and second interval produced 25 possible intensity pairs. This created 50 unique trials when considering that the visual movement could appear in the first or in the second interval. We added a further condition containing 'catch trials' with sound presented without movement. In total there were 350 experimental trials (7 repetitions of 50) plus 75 catch trials (3 repetitions of 25), bringing the total number of trials to 425. Each participant conducted a brief session of approximately 20 trials to familiarize with the task. The participant performed the task in two separate blocks of 20 minutes each.

2.2. Results

First, we compared the overall accuracy when participants judged the intensities of the two bursts without visual stimuli and when they judged them with it. This comparison was performed by excluding the trials in which the intensity of the auditory stimuli was identical. The mean accuracy in the audio-visual trials was 78.1% (S.D. = 11.4) and in the audio only trials was 78.5% (S.D. = 10.6). The accuracy did not differ in the two conditions ($t_7 = 0.16$, $p = 0.87$).

Then, following Odgaard et al. (2004), trials were divided into two categories: trials in which the intensity of the two bursts was identical (the 'diagonal' of the factorial design — the 'same intensity' condition), and the remaining trials (the 'different intensity' condition). We

then considered for each participant the proportion of trials in which the interval accompanied by the visual movement was selected as louder. The average proportion by condition across participants is plotted in Fig. 1. We compared the proportion of answers with the chance level (50%), in order to verify the effect of motion on loudness judgment. A one-sample t -test showed that when the intensity of the auditory stimuli were identical ('same intensity' condition), participants were significantly more likely to choose the interval with movement as the loudest one ($t_7 = 3.44$, $p = 0.011$; Cohen's $d = 2.60$), while in the 'different intensity' condition, the difference between the observed proportion of answers and chance is not different ($t_7 = 1.94$, $p = 0.093$; Cohen's $d = 1.46$). Furthermore, a comparison between the two conditions showed a greater influence of visual movement for the 'same intensity' condition relative to the 'different intensity' condition ($t_7 = 2.57$, $p = 0.036$).

As a further analysis, we calculated the proportion of 'louder' responses for each stimulus intensity separately for motion and non-motion condition, in order to test for effects of movement at different stimulus intensities. We fitted a logistic psychometric function on the average data (see Fig. 2) via maximum likelihood estimation. Successively, we conducted a two-way ANOVA on the 'louder' proportions with Intensity (33, 37, 41, 44, 49 dB) and Presence of movement (present vs absent) as within factors. Results of the two way ANOVA showed a main effect of intensity ($F_{4,28} = 44.76$, $p < 0.0001$, partial- $\eta^2 = 0.865$) and the interaction between the two factors ($F_{4,28} = 9.46$, $p < 0.0001$, partial- $\eta^2 = 0.575$). Moreover, post hoc comparisons (Bonferroni-corrected t -tests) showed a significant difference between the intervals with movement and those without movement for the second highest intensity ($p = 0.034$ for 45 dB), where participants rated the interval with the movement as louder.

2.3. Discussion

The results show an effect of cross-modal interaction between visual movement and loudness judgments. Moreover, the procedure we adopted allowed us to reduce the effect of response bias (Odgaard et al., 2003; 2004), suggesting an early-stage, perceptual origin for this cross-modal interaction. Observers tended to choose the interval with the movement as being louder, although intensity was identical in both sound intervals. Additionally, we compute the overall proportion of intervals accompanied by movement chosen as louder than those without movement, observing the tendency to rate as louder the intervals with movement when presented at the second highest intensity (45 dB). This result seems to reflect an interaction between stimulus intensity and crossmodal modulation, i.e. the effect emerges for high stimulus intensities.

3. Experiment 2

In the first experiment we showed that the presence of a moving stimulus in one interval during a two interval forced choice procedure enhances the loudness of a simultaneous acoustic stimulus. However, since we presented a moving visual stimulus in one interval and just the fixation point in the other one, it could be argued that the mere presence of a visual stimulus per se (or its brightness) can account for this enhancing effect rather than the movement (as in Odgaard et al., 2004). In order to disentangle the role of brightness from the role of movement, we created a new version of the experiment in which the white disc was always presented during the experiment (in this way brightness cannot account for loudness enhancement).

3.1. Methods

3.1.1. Participants

A new group of twelve participants (six females) was recruited from the University of Sussex (age = 20.25 SD = 2.76). They all reported normal (or corrected to normal) vision and normal hearing. Each participant completed two 30-min sessions as described above. They were refunded with credits or paid £5 for participating.

3.1.2. Materials and Procedure

Materials and procedures were identical to Experiment 1. Here, however, the sound interval that did not contain the visual motion was presented simultaneously with a static disc.

3.2. Results

As for Experiment 1, we calculated the accuracy for the intensity judgments with moving stimuli in one of the intervals (audio-visual) and for the condition without moving stimuli (audio only). The accuracy was not different in the two conditions ($t_{11} = 0.49, p = 0.62$) with a mean accuracy of 79.5% (S.D. = 7.6%) in audio-visual trials and 80.3% (S.D. = 10.7) in audio-only trials.

Successively, we divided the trials into two categories (same intensity and different intensity). The average proportion of trials in which the burst accompanied by visual movement was reported as louder is plotted in Fig . 3. A one-sample t -test showed that in the ‘same intensity’ condition participants significantly rated the interval with movement as louder ($t_{11} = 4.04, p = 0.002$; Cohen’s $d = 2.43$) with respect to the chance (50%). In the ‘different intensity’ condition, the difference between the observed proportion of answers and chance is not different but shows the same trend ($t_{11} = 2.08, p = 0.061$; Cohen’s $d = 1.25$). Furthermore, a comparison between the two conditions showed a greater influence of visual movement for the ‘same intensity’ condition over the ‘different intensity’ condition ($t_{11} = 3.86, p = 0.003$).

As for Experiment 1, we calculated the proportion of ‘louder’ responses for each stimulus intensity separately for motion and non-motion condition. We fitted a logistic psychometric function on the average data (see Fig. 4) via maximum likelihood estimation. Successively, we conducted a two-way ANOVA on the ‘louder’ proportions with Intensity (33, 37, 41, 44, 49 dB) and Presence of movement (present vs absent) as within factors. Results of the two way ANOVA showed a main effect of Intensity ($F_{4,44} = 182.43, p < 0.0001, \text{partial-}\eta^2 = 0.943$) and interaction between the two factors ($F_{4,44} = 4.2, p < 0.0001, \text{partial-}\eta^2 = 0.276$). Post hoc comparisons (Bonferroni-corrected t -tests) showed no significant difference between the intervals with movement and those without movement for any of the stimuli intensities.

3.3. Discussion

The results of the current experiment replicated those of Experiment 1 and extend them by ruling out a rival interpretation, namely that it is the presence of a visual stimulus per se (e.g., its brightness) rather than motion that affects loudness perception. However, differently from Experiment 1, in which participants showed a tendency to rate as overall louder the stimuli with movement when presented at mid-high intensities, here we did not find intensity-specific effects. A reason for this difference might lie in the paradigm used in Experiment 2, in which, unlike Experiment 1, we presented a visual stimulus, although static, in the no movement interval.

4. Experiment 3

Experiments 1 and 2 show that if we present a sound together with a moving stimulus, a perceptual characteristics of the sound (i.e., its loudness) can be enhanced. However, they do

not enable to understand whether visual motion enhances loudness only or any perceptual characteristic of sound (e.g., its pitch).

4.1. Methods

4.1.1. Participants

Fourteen participants (seven females) were recruited from the University of Padova, Italy (age = 21.4 SD = 3.55). They all reported normal (or corrected to normal) vision and normal hearing. Each participant completed two 30-min sessions as described above.

4.1.2. Materials and Procedure

The procedure was the same as in Experiment 2, with the difference that participants had to judge which one of two consecutively presented sounds had the highest pitch. The sounds accompanying the visual stimulus were pure tones of five different frequencies: 872, 936, 1000, 1063 and 1127 Hz. The experiment was administered using an ASUS computer (Cpu Intel i5 650 3.20 GHz; Motherboard Asus P7H55-V; RAM 4 GB; Graphic Card AMD Radeon HD 5700 Series; OS Windows 7 Professional 64 bit). The computer was connected to a monitor NEC MultiSync FE950+ and to a M-AUDIO FastTrack Pro sound card. The output of the sound card was presented dichotically through Sennheiser HD 580 headphones at 60 dB SPL. Sounds were synthesized in real time at 44.1 kHz sample rate and 24 bits resolution. The participant took the experiment inside a single walled IAC sound-proof booth.

4.2. Results

We computed the accuracy for pitch judgment for the moving and non-moving condition for the trials in which the frequency was actually different in the two intervals (i.e., by excluding the diagonal trial in the factorial design matrix). Accuracy in the moving trials was 83.8% (S.D. = 6.1) and 85.1% (S.D. = 8.75 %) in the non-moving trials. The accuracy did not differ statistically between the two conditions ($t_{13} = 1.052, p = 0.312$).

Then we compared trials in the two categories ‘same frequency’ and ‘different frequency’, by plotting the proportion of trials in which participants rated as higher in pitch the interval containing the moving stimulus (Fig. 5). A one-sample t -test showed that in the ‘same frequency’ condition participants did not rate the interval with movement as higher (or lower) in pitch ($t_{13} = 0.663, p = 0.52$) with respect to the chance (50%). Interestingly, in the ‘different frequency’ condition, there is a bias towards reporting the interval with the moving stimulus as lower in pitch ($t_{13} = 2.28, p = 0.044$). Finally, a paired t -test between the two conditions showed no difference in the proportion of trials with movement rated as higher (or lower) in pitch ($t_{13} = 1.86, p = 0.085$).

Similarly to Experiment 1 and Experiment 2, we calculated the proportion of ‘higher in pitch’ responses for each stimulus frequency separately for motion and no motion condition. We fitted a logistic psychometric function on the average data (see Fig. 6) via maximum likelihood estimation. Successively, we conducted a two-way ANOVA on the ‘higher in pitch’ proportions with Frequency (872, 936, 1000, 1063 and 1127 Hz) and Presence of movement (present vs absent) as within factors. Results of the two way ANOVA showed a main effect of Frequency ($F_{4,52} = 415.97, p < 0.0001, \text{partial-}\eta^2 = 0.97$) and Movement ($F_{1,13} = 9.75, p = 0.008, \text{partial-}\eta^2 = 0.428$) and interaction between the two factors ($F_{4,52} = 4.79, p = 0.002, \text{partial-}\eta^2 = 0.269$) Post hoc comparisons (Bonferroni-corrected t -tests) showed that participants rated the interval with the movement significantly lower in pitch than the ones without movement when presented with the third highest Frequency (1000 Hz, $p = 0.005$). As

a further analysis, we calculated the overall accuracy of the participants in the three Experiments, reported in the Appendix. In general, participants were accurate in their loudness and pitch judgements (77% to 84% among the three experiments).

4.3. Discussion

The results of Experiment 3 showed that when asked to judge the pitch of two consecutively presented tones, participants do not perceive as ‘higher in pitch’ the interval containing the moving stimulus more often than the one containing the static stimulus. However, there seems to be a little effect, overall, of movement: In fact, when presented with movement, participants rated the sound as being lower in pitch for the middle frequency (1000 Hz). As for loudness, it seems that the amount of the crossmodal interaction depends on the level of the sound feature that we manipulated, being intensity in Experiment 1 and 2 and frequency in Experiment 3. However, taking into account only the trials in which participants were asked to rate the sounds when they were of the same frequency, we did not observe this effect (however the means show a trend in the same direction). Overall, the result of Experiment 3 supports the idea that the effect observed in Experiment 1 and Experiment 2 is specific for loudness and it is not a response bias.

5. General Discussion

In recent years, a number of studies showed that auditory stimuli can alter visual perception in different ways: sounds can aid detection, improve temporal resolution, guide attention, and affect motion perception in visual processing (Bolognini et al., 2005; Frassinetti et al., 2002; Lippert et al., 2007). In this study we add to the literature showing that the reverse is also true: namely, visual stimuli can affect auditory perception. Odgaard et al. (2004) had previously

demonstrated that visual brightness can affect judgments of loudness. Using the same methodology we established that visual motion can also affect judgments of loudness (even when visual brightness is identical). The effect is greater when auditory stimuli are the same intensity, similarly to what reported by Odgaard et al. (2004), but the effect is still present when the intensities are higher, as showed by the psychometric fit in Experiment 1 and 2. The method and procedure used here are less contaminated by response bias than, say, perceptual ratings of individual stimuli and — hence — is consistent with a pre-decision, perceptual locus (Odgaard et al., 2004). Moreover, a control experiment in which, presented with the same experimental configuration, participants had to judge the pitch of the sound showed no effect of movement on the task in case of physically identical stimuli (although in the trials in which the sound were physically different in pitch, we observed a slight bias towards judging the sound accompanying the interval with movement as lower in pitch). We speculate that the origins of this correspondence may lie in the natural tendency for moving (relative to static) objects to produce sound. This may affect the way that the auditory and visual pathways communicate leading to an automatic tendency for purely auditory judgments, such as loudness discrimination, to be affected by potentially relevant (although in this case, wholly irrelevant) information concerning visual motion.

Although it is perhaps unsurprising that auditory motion and visual motion should interact (Baumann and Greenlee, 2007), as has been shown, for example, in a number of brain stimulation studies targeting parietal cortex (Maniglia et al., 2012) it is more surprising that visual motion should interact with other qualities of audition such as loudness or pitch. Some previous research has examined cross-modal correspondences between visual motion and pitch. For instance, ascending and descending pitch modulations are linked to preferential looking towards upwards and downwards visual movement in children (Nava et al, 2016) and infants (Walker et al., 2010; but see Lewkowicz and Minar, 2014 who did not observe a

difference). However, this probably reflects spatial position rather than movement per se because the same tendency is found for static visual stimuli occupying high and low positions (Ben-Artzi and Marks, 1995). The effect reported here of visual motion on loudness is unlikely to affect spatial position as the visual motion was collapsed across leftwards and rightwards motion. Using a similar audiovisual paradigm, in which participants were asked to judge the brightness of a visual stimulus while concomitantly exposed to auditory stimulation, Odgaard et al. (2003) demonstrated that the enhancing effect of noise observed by Stein et al. (1996), and by these authors interpreted as a product of multisensory interaction, was likely the result of late stage response bias (however see Jaekl and Soto-Faraco, 2010 for a different take on the possible multisensory nature of this effect). Odgaard et al. (2003) reported two manipulations of the original experimental paradigm that removed the effect, first, asking participants to compare two intervals rather than rate the brightness for each trial; second, reducing the probability of co-occurrence of brightness and noise from 0.5 to 0.25. While in the present study we kept the same probability of co-occurrence of audio and movement stimulation (~80% of the trials), we did test a two-interval paradigm, that, unlike Odgaard et al. (2003), showed that the multisensory modulation was present. Consequently, while we cannot exclude that a change in the probability of co-occurrence might modulate the effect reported in the present paper, the two-interval paradigm seems to offer a certain degree of certainty about the genuine multisensory origin of the effect rather than it being the product of a response bias.

Of note, Experiment 1 showed an intensity specific interaction, with participants rating as overall louder the stimuli with movement when presented at mid-high intensities, while in Experiment 2 such tendency was not observed. A possible reason for this discrepancy might be that while in Experiment 1 the static interval contained a blank display, in Experiment 2 it had a (static) visual stimulus. Hence, the brightness quality of the interval without movement might have induced an enhancement of perceived loudness, as in Odgaard et al. (2004).

Future research is needed to clarify the correspondence further. For instance, it is unclear whether a faster speed would produce a stronger effect than a slower speed. That is, is the effect due to visual movement per se or to speed of movement? Also it is unclear whether the effect is asymmetric. Using the same procedure as used here, Odgaard et al. (2003, 2004) demonstrated that brightness affects loudness perception, but loudness does not affect brightness perception. In the present context it is unknown whether a loud sound influences perceptual judgments of the speed of movement. It would also be interesting to explore receding and approaching motion. As far as we know, only subjective duration has been investigated with audiovisual looming stimuli (Grassi and Pavan, 2012) whereas no investigation studies the effect on perceived speed. Nonetheless, an approaching (unseen) sound source is judged as louder (Neuhoff, 2001), so we would expect an approaching visual stimulus (paired with a static sound source) to elicit a comparable effect.

In summary, we established that visual motion affects the perception of auditory loudness and this adds to the growing literature on normal cross-modal correspondences between different sensory dimensions.

Note

1 However, it is worth noticing that some crossmodal effects might emerge from a general perception that two stimuli ‘go well together’.

References

- Alais, D. and Burr, D. (2004). No direction-specific bimodal facilitation for audiovisual motion detection, *Cogn. Brain Res.* **19**, 185–194.
- Baumann, O. and Greenlee, M. W. (2007). Neural correlates of coherent audiovisual motion perception, *Cereb. Cortex* **17**, 1433–1443.

- Ben-Artzi, E. and Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference, *Percept. Psychophys.* **57**, 1151–1162.
- Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). “Acoustical vision” of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, **160**, 273-282.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial vision* , **10**, 433–436.
- Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K. P., . . . and 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.
- Frassinetti F., Bolognini N. and Làdavas E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction, *Exp. Brain Res.* **147**, 332–343.
- Freeman, E. and Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound, *Curr. Biol.* *18*(16), 1262–1266.
- Grassi, M. and Casco, C. (2012). Revealing the origin of the audiovisual bounce-inducing effect, *See. Perceiv.* **25**, 223–233.
- Grassi, M. and Pavan, A. (2012). The subjective duration of audiovisual looming and receding stimuli, *Attent. Percept. Psychophys.* **74**, 1321–1333.
- Grassi, M., Pastore, M. and Lemaitre, G. (2013). Looking at the world with your ears: How do we get the size of an object from its sound? *Acta Psychol.* **143**, 96–104.
- Green, D.M. and Swets J.A. (1966). Signal detection theory and psychophysics. Wiley, New York., NY, USA.
- Jaekl P.M. and Soto-Faraco S. (2010). Audiovisual contrast enhancement is articulated primarily via the M-pathway, *Brain Res.* **1366**, 85–92.
- Lippert, M., Logothetis, N. K. and Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound, *Brain Res.* **1173**, 102–109.
- Lewkowicz, D. J. and Minar, N. J. (2014). Infants are not sensitive to synesthetic cross-modality correspondences. A comment to Walker et al. (2010), *Psychol. Sci.* **25**, 832–834.
- Maniglia, M., Grassi, M., Casco, C. and Campana, G. (2012). The origin of the audiovisual bounce inducing effect: A TMS study, *Neuropsychologia* **50**, 1478–1482.
- McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices, *Nature* **264**(5588), 746–748.

- Monaghan, P., Mattock, K. and Walker, P. (2012). The role of sound symbolism in language learning, *J. Exp. Psycho. Learn. Mem. Cogn.* **38**, 1152–1164.
- Nava, E., Grassi, M. and Turati, C. (2016). Audio-visual, visuo-tactile and audio-tactile correspondences in preschoolers, *Multisens. Res.* **29**, 93–111.
- Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecol. Psychol.* **13**, 87–110.
- Odgaard, E. C., Arieh, Y. and Marks, L. E. (2003). Cross-modal enhancement of perceived brightness; Sensory interaction versus response bias, *Percept. Psychophys.* **65**, 123–132.
- Odgaard, E. C., Arieh, Y. and Marks, L. E. (2004). Brighter noise: Sensory enhancement of perceived loudness by concurrent visual stimulation, *Cogn. Affect. Behav. Neurosci.* **4**, 127–132.
- Parise, C. V. and Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes, *PLoS One* **4**. doi: 10.1371/journal.pone.0005664
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial vision*, **10**, 437–442.
- Phillips-Silver, J. and Trainor, L. J. (2007). Hearing what the body feels: auditory encoding of rhythmic movement, *Cognition* **105**, 533_546. doi: 10.1016/j.cognition.2006.11.006
- Sadaghiani S, Maier JX and Noppeney U (2009). Natural, metaphoric and linguistic auditory direction signals have distinct influences on visual motion processing, *J. Neurosci.* **29**, 6490–6499.
- Saenz, M. and Koch, C. (2008). The sound of change: Visually-induced auditory synesthesia, *Curr. Biol.* **18**, R650–R651.
- Seitz, A. R., Kim, R. and Shams, L. (2006). Sound facilitates visual learning, *Curr. Biol.* **16**, 1422–1427. doi: 10.1016/j.cub.2006.05.048
- Sekuler, R., Sekuler, A. B., and Lau, R. (1997). Sound alters visual motion perception. *Nature* **385** (6614), 308
- Shams, L., Kamitani, Y. and Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature* **408**(6814), 788–788.
- Soranzo, A. and Grassi, M. (2014). PSYCHOACOUSTICS: a comprehensive MATLAB toolbox for auditory testing, *Front. Psychol.* **5**, 712. doi: 10.3389/fpsyg.2014.00712
- Soto-Faraco, S., Kingstone, A. and Spence, C. (2003). Multisensory contributions to the perception of motion, *Neuropsychologia* **41**, 1847–1862.

- Stein, B. E., London, N., Wilkinson, L. K. and Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis, *J. Cogn. Neurosci.* **8**, 497–506.
- Tanner, T. A. J., Haller, R. W., & Atkinson, R. C. (1967). Signal recognition as influenced by presentation schedules. *Perception & Psychophysics*, **2**, 349–358.
- Teramoto, W., Hidaka, S., Sugita, Y., Sakamoto, S., Gyoba, J., Iwaya, Y. and Suzuki, Y. (2012). Sounds can alter the perceived direction of a moving visual object, *J. Vis.* **12**. doi: 10.1167/12.3.11
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A. and Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences, *Psychol. Sci.* **21**, 21–25.
- Walker, P., Francis, B. J. and Walker, L. (2010). The brightness-weight illusion darker objects look heavier but feel lighter, *Exp. Psychol.* **57**, 462–469.
- Yarrow, K., Haggard, P. and Rothwell, J. C. (2008). Vibrotactile-auditory interactions are post-perceptual, *Perception* **37**, 1114–1130. doi: 10.1068/p5824

Appendix

As a further analysis, we calculated the overall accuracy of the participants in the three Experiments. For Experiment 1, the overall accuracy was 0.77 and 0.79 for the moving stimulus in the first and in the second interval, respectively (Table 1). For Experiment 2, the overall accuracy was 0.78 and 0.81 for the moving stimulus in the first and in the second interval, respectively (Table 2). For Experiment 3, the accuracy was 0.84 for both conditions (Table 3).

Table 1.

Experiment 1. Mean proportion of louder responses for the interval presented with the moving stimulus. (A) First interval presented with the moving stimulus only; (B) second interval presented with the moving stimulus only. The gray boxes indicate conditions in which the interval with the moving stimulus was physically the loudest.

(A)	Second interval (dB)				
First interval (dB)	33.00	37.00	41.00	45.00	49.00
33.00	x	0.46	0.23	0.24	0.18
37.00	0.79	x	0.50	0.25	0.18
41.00	0.84	0.82	x	0.36	0.21
45.00	0.88	0.95	0.86	x	0.50
49.00	0.86	0.93	0.88	0.79	x

(B)	Second interval (dB)				
First interval (dB)	33.00	37.00	41.00	45.00	49.00
33.00	x	0.55	0.80	0.71	0.77
37.00	0.20	x	0.79	0.91	0.89
41.00	0.30	0.23	x	0.91	0.95
45.00	0.16	0.20	0.34	x	0.91
49.00	0.11	0.24	0.16	0.52	x

Table 2.

Experiment 2. Mean proportion of louder responses for the interval presented with the moving stimulus. (A) First interval presented with the moving stimulus only; (B) Second interval presented with the moving stimulus only. The gray boxes indicate conditions in which the interval with the moving stimulus was physically the loudest.

(A)	Second interval (dB)				
First interval (dB)	33.00	37.00	41.00	45.00	49.00
33.00	x	0.56	0.39	0.29	0.18
37.00	0.80	x	0.56	0.15	0.10
41.00	0.87	0.76	x	0.36	0.13
45.00	0.85	0.89	0.89	x	0.27
49.00	0.93	0.82	0.85	0.85	x

(B)	Second interval (dB)				
First interval (dB)	33.00	37.00	41.00	45.00	49,00
33.00	x	0.71	0.73	0.71	0,74
37.00	0.26	x	0.63	0.85	0,88
41.00	0.13	0.23	x	0.86	0,87
45.00	0.11	0.10	0.26	x	0,85
49.00	0.06	0.07	0.11	0.23	x

Table 3.

Experiment 3. Mean proportion of ‘higher in pitch’ responses for the interval presented with the moving stimulus. (A) First interval presented with the moving stimulus only; (B) Second interval presented with the moving stimulus only. The gray boxes indicate conditions in which the interval with the moving stimulus was physically higher in pitch.

(A)		Second interval (Hz)				
First interval (Hz)	872.29	936.14	1000.00	1063.85	1127.70	
872.29	x	0.62	0.19	0.08	0.03	
936.14	0.90	x	0.39	0.12	0.03	
1000.00	0.95	0.85	x	0.16	0.05	
1063.85	0.95	0.90	0.74	x	0.12	
1127.70	0.98	0.93	0.82	0.60	x	

(B)		Second interval (Hz)				
First interval (Hz)	872.29	936.14	1000.00	1063.85	1127.70	
872.29	x	0.49	0.72	0.89	0.93	
936.14	0.06	x	0.55	0.88	0.92	
1000.00	0.08	0.08	x	0.80	0.87	
1063.85	0.02	0.06	0.22	x	0.90	
1127.70	0.04	0.07	0.15	0.41	x	

Figure 1. Experiment 1. Proportion of louder responses for intervals presented with the moving stimulus. On the left, responses to trials where the intensity of the two sounds intervals was identical, on the right responses to trials where the intensity of two sounds was different. In each boxplot, the central mark is the median. The edges of the box are the 25th and 75th percentiles. The whiskers are the interquartile range (i.e., Q3–Q1) augmented by 50% and symbols are outliers. The black circles represent the mean and the vertical bars are +/- SEM.

Figure 2. Proportion of louder responses as a function of stimulus intensity for Experiment 1. Dotted and dashed lines (maximum likelihood fit) represent fit on aggregated data for No movement and Movement intervals, respectively.

Figure 3. Experiment 2. Proportion of louder responses for intervals presented with the moving stimulus. On the left, responses to trials where the intensity of the two sounds intervals was identical, on the right responses to trials where the intensity of two sounds was different. In each boxplot, the central mark is the median. The edges of the box are the 25th and 75th percentiles. The whiskers are the interquartile range (i.e., Q3–Q1) augmented by 50% and symbols are outliers. The black circles represent the mean and the vertical bars are +/- SEM.

Figure 4. Proportion of louder responses as a function of stimulus intensity for Experiment 2. Dotted and dashed lines (maximum likelihood fit) represent fit on aggregated data for No movement and Movement intervals, respectively.

Figure 5. Experiment 3. Proportion of ‘higher in pitch’ responses for intervals presented with the moving stimulus. On the left, responses to trials where the frequency of the two sounds intervals was identical, on the right responses to trials where the frequency of two sounds was different. In each boxplot, the central mark is the median. The edges of the box are the 25th and 75th percentiles. The whiskers are the interquartile range (i.e., Q3–Q1) augmented by 50% and symbols are outliers. The black circles represent the mean and the vertical bars are +/- SEM.

Figure 6. Proportion of louder responses as a function of stimulus frequency for Experiment 3. Dotted and dashed lines (maximum likelihood fit) represent fit on aggregated data for No movement and Movement intervals, respectively.