Acta Psychologica 129 (2008) 249–254

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Filling-in visual motion with sounds

A. Väljamäe [a,b,*], S. Soto-Faraco [c,d,e]

[a] Laboratory for Synthetic Perceptive, Emotive and Cognitive Systems (SPECS), Institute of Audiovisual Studies, Universitat Pompeu Fabra, Tànger, 135, Barcelona 08018, Spain
[b] Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden
[c] Parc Científic de Barcelona, Hospital Sant Joan de Déu Edifici Docent C/Santa Rosa,39-57, 08950 Esplugues - Barcelona - Spain
[d] Departament de Psicologia Bàsica, Universitat de Barcelona
[e] Institució Catalana de Recerca i Estudis Avançats (ICREA)

ARTICLE INFO

Article history:
Received 12 March 2008
Received in revised form 8 July 2008
Accepted 4 August 2008

PsyclNFO classification:
2320 (Sensory perception)
2560 (Psychophysiology)

Keywords:
Multisensory integration
Motion after-effect
Audition
Vision

ABSTRACT

Information about the motion of objects can be extracted by multiple sensory modalities, and, as a consequence, object motion perception typically involves the integration of multi-sensory information. Often, in naturalistic settings, the flow of such information can be rather discontinuous (e.g. a cat racing through the furniture in a cluttered room is partly seen and partly heard). This study addressed audio-visual interactions in the perception of time-sampled object motion by measuring adaptation after-effects. We found significant auditory after-effects following adaptation to unisensory auditory and visual motion in depth, sampled at 12.5 Hz. The visually induced (cross-modal) auditory motion after-effect was eliminated if visual adaptors flashed at half of the rate (6.25 Hz). Remarkably, the addition of the high-rate acoustic flutter (12.5 Hz) to this ineffective, sparsely time-sampled, visual adaptor restored the auditory after-effect to a level comparable to what was seen with high-rate bimodal adaptors (flashes and beeps). Our results suggest that this auditory-induced reinstatement of the motion after-effect from the poor visual signals resulted from the occurrence of sound-induced illusory flashes. This effect was found to be dependent both on the directional congruency between modalities and on the rate of auditory flutter. The auditory filling-in of time-sampled visual motion supports the feasibility of using reduced frame rate visual content in multisensory broadcasting and virtual reality applications.

1. Introduction

Many objects and events in our everyday environments are perceived concurrently through several sensory systems. It is now widely accepted that inter-sensory correlations can significantly sharpen our perceptual capabilities, for example, audio-visual speech perception in noisy environments (i.e. Massaro, 2004; Sumby & Pollack, 1954) or for difficult-to-perceive phonemes in the second language (Navarra & Soto-Faraco, 2007). In addition, these inter-sensory correlations also can be used in constructing unitary and coherent scenes from the fragmentary information available across different sensory channels. For example, one can use transient sounds and fleeting sightings when tracking a moving animal racing through the undergrowth of a cluttered forest. In order to accomplish perception under these less than optimal conditions, perceptual processes taking place in our brain often implement a filling-in strategy whereby sensory illusions help to complete missing external information, for example, restoring the shape of partly occluded objects (see Komatsu, 2006 for a recent review), the completion of lines and surfaces in Kanizsa figures (e.g. Kanizsa, 1976), or the filling-in of visual information missing in the blind spot of the eye (e.g. Ramachandran, 1992). However, while it is widely accepted that human perception is multisensory, there is barely any systematic research addressing the role of cross-modal interactions in the process of perceptual filling-in.

One avenue to start addressing cross-modal filling-in is to capitalize on the fact that temporal accuracy in perception is higher for the auditory than for the visual modality (Kohlrausch, Fassel, & Dau, 2000), and as a consequence the former usually dominates over the latter in temporal perception. For example, sound can significantly change the temporal perception of visual events, such as in the "auditory driving" effect (e.g. Welch, Duttenhurt, & Warren, 1986), the temporal ventriloquism (Morein-Zamir, Soto-Faraco, & Kingstone, 2003), or the freezing effect (e.g. Vroomen & de Gelder, 2000). In fact, outside the laboratory, in audio-visual media such as cinema and television, this auditory influence on vision has been extensively exploited, as sound has been traditionally used for highlighting the temporal structure of rapid visual events. Consider, for example, hitting sounds in Kung-Fu fighting scenes (Chion, 1994)

* Corresponding author. Address: Laboratory for Synthetic Perceptive, Emotive and Cognitive Systems (SPECS), Institute of Audiovisual Studies (IUA), Universitat Pompeu Fabra, Edifici La Nau, Tànger, 135, 08018 Barcelona, Spain. Tel.: +34 935421393; fax: +34 935422202.
E-mail addresses: aleksander.valjamae@iua.upf.edu (A. Väljamäe), salvador.soto@icrea.es (S. Soto-Faraco).

0001-6918/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2008.08.004

or Walt Disney's "Mickey Mousing" technique, whereby motion picture sounds are tightly synchronized with the character's movements (Thomas & Johnston, 1981). Interestingly, sound has also been used in cinema for creating an illusion of visual action continuity. In a classic example from George Lucas' film "The empire strikes back" (1980), the visual illusion of a spaceship door sliding open is created using two successive stills, a door closed and a door opened, plus a "swapping" sound effect (Chion, 1994).

Recent laboratory studies focusing on cross-modal interactions have found that sound can indeed induce the illusion of seeing a visual event when there is none. For example, Shams, Kamitani, and Shimojo (2000) reported an illusion of multiple visual flashes that were produced, when one single brief visual stimulus was coupled with the multiple auditory beeps. In these experiments, participants were asked to count the number of times a flickering white disk had flashed when presented with one or more task-irrelevant brief sounds. The number of flashes reported by observers increased with the number of beeps. In a later ERP study, Shams, Kamitani, Thompson, and Shimojo (2001) reported that the beeps modulated early visual evoked potentials originating from the occipital cortex. Interestingly, the electrophysiological activity corresponding to the illusory flashes was found to be very similar to the activity produced when a flash was physically presented. Later works confirmed that illusory flash is a perceptual effect with the psychophysically assessable characteristics (McCormick & Mamassian, 2008) and showed that human performance did not depend on whether visual stimuli in orientation-discrimination tasks were real or illusory (Berger, Martelli, & Pelli, 2003).

Here, we addressed the potential effects of sound-induced illusory flashes when embedded into time-sampled object motion. We used motion perception, as it has been shown to be a subject to strong multisensory interactions (e.g. Kitagawa & Ichihara, 2002; Meyer & Wuerger, 2001; Soto-Faraco, Lyons, Gazzaniga, Spence, & Kingstone, 2002; Soto-Faraco, Spence, & Kingstone, 2004; see Soto-Faraco, Kingstone, & Spence, 2003, for a review). In particular, we capitalized on the adaptation to motion in depth because it is known to produce consistent motion after-effects (MAE) both unimodally and cross-modally (Kitagawa & Ichihara, 2002). For example, after exposure to looming visual objects the viewer will perceive a steady visual stimulus as if it would be receding (Regan & Beverley, 1978). Importantly, as was shown by Kitagawa and Ichihara, the adaptation to visual continuous object motion in depth also results in an auditory changing-loudness after-effect thus highlighting the cross-modal nature of motion perception. Indeed, dynamic changes in sound intensity have been shown to be a strong cue to auditory motion in the horizontal plane (Lutfi & Wang, 1999), but this cue is even more important for the perception of auditory motion in depth, where binaural cues do not contribute to the localization in the median plane (Blauert, 1997). Therefore, keeping the same methodology as in Kitagawa and Ichihara (2002), in the rest of the text we will use changing-loudness after-effect as a correlate of auditory motion in the auditory motion after-effect (aMAE).

The goal of the present study was to measure the capacity of sounds to reinstate missing visual events. To this aim, we measured aMAE induced by sampled (discontinuous) visual events, and most critically, the interaction of acoustic and visual adaptors at different sampling rates. In Experiment 1, we compared the aMAE induced by high- and low-rate visual adaptors on their own with the same visual adaptors when combined with the high-rate acoustic events.

2. Experiment 1

Our first hypothesis was that (cross-modal) aMAE would depend on the frequency of the flashes representing discrete motion of the adapting visual object. Sparsely time-sampled visual motion should produce a lower aMAE compared to a higher rate (i.e. perceptually fused) moving stimuli. The second and critical hypothesis was that, if the combination of a slow train of flashes (flicker) with a rapid train of beeps (flutter) leads to the sound-induced illusory flash illusion (Shams, Kamitani, & Shimojo, 2002), then this high-rate flutter would help to fill-in sparsely sampled visual object motion. If this is the case, such audio-visual combination should produce aMAE comparable to aMAE induced by a real stimulus containing synchronized flashes and beeps at a high-rate.

2.1. Method

2.1.1. Participants

This study was conducted in accordance with the declaration of Helsinki and under approval of the local ethics committee at the University of Barcelona. Fifteen students from the University of Barcelona (mean age of 28.5 years, 10 females) took part in the experiment voluntarily and received a cinema ticket for their participation. All participants reported to have normal or corrected-to-normal vision and no hearing problems. Data from 2 additional participants were discarded because no aMAE was registered in more than half of the experimental conditions, and we thought it is possible that they had failed to pay full attention to the stimuli in the adapting phase. It has been shown that participants' attention during the adaptation phase can modulate the strength of the perceptual after-effects (Hong & Papathomas, 2006).

2.1.2. Stimuli

To simulate audio-visual motion in depth, we used the changing of visual stimuli size and intensity of sound (see Fig. 1). The auditory and visual adaptation stimuli were synthesized and rendered, using the psychophysics toolbox in Matlab (Brainard, 1997; Pelli, 1997). The factorial design was 2 (direction of adapting stimulus; approaching or receding) by 5 (stimulus type). Stimulus type defined the adapting motion modality and frequency rate: Ah (high-rate flutter at 12.5 beeps/s), Vh (high-rate flicker at 12.5 flashes/s), Vl (low-rate flicker, 6.25 flashes/s), AhVh (synchronized high-rate flicker and flutter), and AhVl (high-rate flutter combined with the low-rate visual flicker).

The adapting visual stimulus (2 s duration) consisted of a uniform white disk expanding (from 0 to 9 degrees of visual angle) or contracting (from 9 to 0 degrees of visual angle) at the center of the screen at ±4.5 deg/s velocity. Flash duration was 27 ms on and 53 ms off in the high-rate flicker condition, and 27 ms on

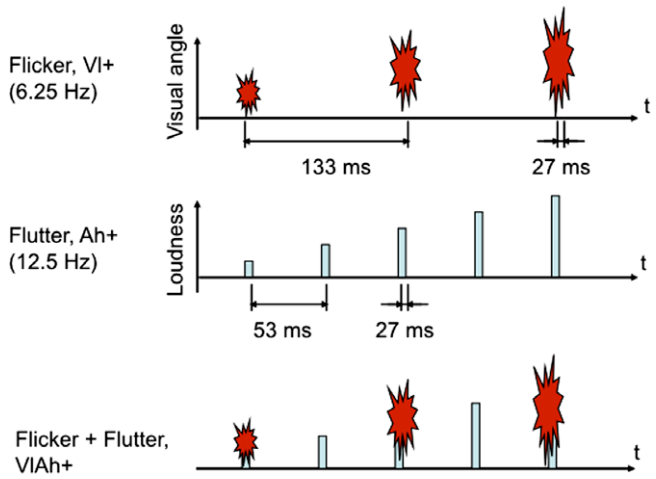

Fig. 1. Auditory, visual and audio-visual patterns representing approaching stimuli (see text for details).

and 133 ms off for the low-rate flicker. The adapting sound stimulus (2 s duration) consisted of a 400 Hz triangular waveform generated using Cool Edit Pro software (Syntrillium software corp.) at 48 kHz sampling rate. Similar tonal stimuli have been successfully used in the previous auditory looming experiments (e.g. Maier, Neuhoff, Logothetis, & Ghazanfar, 2004; Röhrbein, Schill, & Zetzsche, 2000). The auditory adaptor rose (from 40 to 80 dB) or fell (from 80 to 40 dB) in the sound pressure level (SPL) at ±20 dB/s. To create the high-rate flutter, auditory stimulus was windowed by intensity envelopes of 27 ms (8 ms ramps of Hann half-window) separated by 53-ms silent periods. The auditory aMAE was measured using a 1.5 s test sound of the same frequency as the adapting sound stimulus. The test sound pressure level was 60 dB at the onset, and it increased or decreased at a velocity calculated according to a psychophysical staircase PEST procedure (see the next section).

2.1.3. Apparatus and procedure

Participants sat in a dark, sound-attenuated, testing room 57 cm away from a computer monitor (Mitsubishi, mod. N0701, 75 Hz). Sound was rendered via two multimedia loudspeakers attached to either side of the monitor (thus, the auditory image was effectively located by the center of the screen).

A tone that is falling or rising in intensity at a certain rate (in dB/s) is perceived as being steady, when the auditory changing loudness after-effect occurs (e.g. Kitagawa & Ichihara, 2002). This point of subjective steadiness was assessed using a double-staircase method (Cornsweet, 1962), where the staircases were governed according to a PEST (parameter estimation by sequential testing) procedure (Taylor & Creelman, 1967). The adaptive staircases started with the test tone rising or falling at ±1.33 dB/s velocity and each of them was terminated when the adaptation step-size of the staircase was smaller than ±0.33 dB/s (the initial adaptation step-size was ±2.66 dB/s).

Participants were presented with 10 blocks of adapting stimuli. There were two blocks (one approaching and one receding) of motion direction for each of the five stimuli types. Each block was taking 5–6 min to complete on the average. Presentation order of these five block pairs followed a 5-item orthogonal Latin square design. Each block had the following temporal structure: pre-adaptation measurement of the point of subjective auditory steadiness using PEST; initial 60 repetitions of adaptation stimulus separated by 200 ms; post-adaptation measurement of the point of subjective auditory steadiness using PEST. During the staircase procedure in the post-adaption phase, each test stimulus presentation was preceded by five adaptation stimuli to preserve the after-effect.

Before running the first experiment, we tested our methodology using continuous audio-visual stimuli similar to the ones used by Kitagawa and Ichihara (2002). Data from two subjects showed comparable results, with congruent audio-visual stimuli pairs producing auditory aMAE in the range of 4–7 dB/s. The results from the first experiment, using time-sampled adaptors, show auditory aMAEs approximately twice smaller (see Fig. 2, left panel).

2.2. Results and discussion

The magnitude of motion after-effect values for each modality combination was obtained as the difference between the averaged points of subjective steadiness from the two pre-adaptation staircases and the two post-adaptation ones. Absolute values of individual aMAE were submitted to a within-subjects 2 (direction) × 5 (stimulus type) ANOVA with Greenhouse-Geisser correction applied whenever unequal variances had occurred.

In line with Kitagawa and Ichihara's (2002) results, neither the effect of direction of the adapting stimulus ($F(1,14) = 2.19$, $p = 0.16$) nor the interaction between direction and stimulus type ($F < 1$) reached significance. Critically, however, the main effect of stimulus type was significant $F(3,44) = 11.11$, $p < 0.001$, indicating the variation in the magnitude of the aMAE as a function of the particular combination of modalities used for adaptation (see Fig. 2, left panel). There were both unimodal and cross-modal aMAEs and we tested the difference between these using planned pair-wise comparisons (t-test, df = 14). There was no significant difference ($p = 0.82$) between the auditory after-effect produced by the

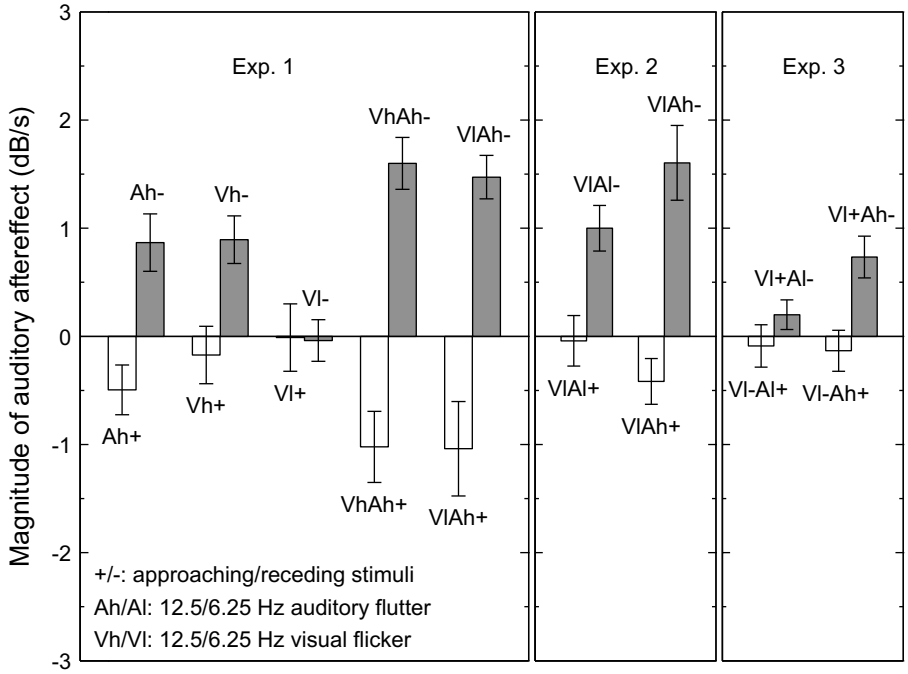

Fig. 2. Magnitude of the auditory after-effect (in dB/s) after adaptation to time-sampled approaching (+) or receding (-) motion in depth of the following types: auditory, visual and audio-visual stimuli.

high-rate flutter (Ah) and the high-rate flicker (Vh) adaptors. Furthermore, the combination of auditory and visual motion signals in the adapting stimulus resulted in a significant increase of the after-effect compared to the single modality conditions. Namely, the high-rate flicker synchronized with the high-rate flutter (VhAh) produced larger aMAE than the one by the high-rate flutter (Ah) alone ($p < 0.05$) or by the high-rate flicker (Vh) alone ($p < 0.01$). This pattern of results for time-sampled audio-visual object motion in depth is coherent with the data from Kitagawa and Ichihara (2002) using continuous motion.

In line with the first hypothesis outlined in the introduction, we found that sparsely time-sampled visual stimuli produced less aMAE than the visual motion sampled at a higher rate, as indicated by the significant difference ($p < 0.05$) between the high-rate flicker (Vh) and the low-rate flicker (Vl) conditions. In fact, the low-rate visual adaptor did not produce an after-effect significantly different from zero ($t < 1$ both for approaching and for receding). Crucially, our second hypothesis was validated since the addition of the high-rate beeps to the low-rate flashes (VlAh) restored the aMAE to the same levels as obtained with the high-rate flicker-flutter adaptor. In particular, the difference between the aMAE produced by the low-rate flicker (Vl) alone and the aMAE obtained for the low-rate flicker synchronized with the high-rate flutter (VlAh) was significant ($p < 0.001$). In fact, the aMAE produced by the combination VlAh was even larger ($p < 0.05$) than the aMAE of the high-rate flutter alone (Ah) and comparable in size ($p = 0.93$) with the aMAE obtained with the high-rate flashes combined with the high-rate flutter (VhAh). These results from Experiment 1 are consistent with the idea that the high-rate acoustic flutter might have filled-in the poorer visual motion signal. In Experiment 2, we addressed whether this indirect effect (of visual filling-in on aMAE) correlates with the subjective impression about continuity. In Experiment 3, we investigated the directional selectivity of the audio-visual adaptor.

3. Experiment 2

Audio-visually induced aMAEs reported in Experiment 1 show that the high-rate acoustic flutter might compensate the low-rate visual flicker perception. In Experiment 1 however, we did not assess the subjective sensation of the participants. In Experiment 2, we combined the aMAE measure with a questionnaire assessing the subjective sensation of smoothness regarding the visual motion in depth (the adapting stimulus). If our filling-in hypothesis is to hold, then the low-rate visual flicker should be perceived as more continuous when combined with the high-rate flutter than when presented by its own.

3.1. Method

3.1.1. Participants

Sixteen participants (mean age of 25 years, 8 females) took part in the experiment voluntarily at Chalmers University of Technology. All reported to have normal or corrected-to-normal vision and no hearing problems. They were compensated with one cinema ticket for their participation. The study was conducted under approval of the local ethics committee and in accordance with the declaration of Helsinki.

3.1.2. Stimuli

In the second experiment, we used 2 (direction of adaptors) by 2 (stimulus type) factorial design leading to four presentation blocks. All stimuli were audio-visual containing low-rate flicker combined with either low- or high-rate flutter (VlAl and VlAh).

3.1.3. Apparatus and procedure

The procedure was almost identical to the one used in Experiment 1 (Section 2.1), except for the aspects noted below. First, in addition to the task described for the first experiment, participants also had to rate the perceived continuity of the visual stimuli on the scale of 0 (clearly separate flashes) to 5 (continuous visual stream). The assessment was done at the end of each of two blocks with the same stimulus type. The presentation order of block pairs was balanced across participants. Second, participants sat in a dark testing room 57 cm away from a computer monitor (DELL, mod. D1226H, 75 Hz) and the sound was rendered via Beyerdynamic DT-990Pro circumaural headphones since a non-sound-attenuated room was used. In a dedicated study, we specifically confronted loudspeaker and headphone reproduction of auditory stimuli using same methodology based on the audio-visually induced aMAE. We did not find any significant difference in respect to aMAE size between these two reproduction methods. We can conclude, therefore, that in this particular methodology, spatial position of the auditory image plays a relatively weak role and does not affect aMAE values.

3.2. Results and discussion

The aMAE data in 2 (direction) $\times$ 2 (stimulus type) ANOVA confirmed the significant effects of flutter rate ($F(1,15) = 4.3$, $p < 0.05$), as in Experiment 1. In particular, combinations of the high-rate flutter with the low-rate flicker (VlAh) lead to significantly stronger aMAEs than the low-rate audio-visual adaptors (see Fig. 2, middle panel). More importantly, subjective ratings about motion continuity showed that the VlAh adaptors were perceived as more continuous ($M = 2.7$, SE = 0.3) than the VlAl adaptors ($M = 2.3$, SE = 0.2), $t(15) = -2.32$, $p < 0.05$. Out of the 16 participants, eight perceived the VlAh adaptors as more continuous than AlVl, six gave equivalent ratings for both, and only two rated the VlAl adaptors as being more continuous. In addition to the main pattern of results, a significant effect of direction on aMAE ($F(1,15) = 13.6$, $p < 0.005$) was observed, with the approaching AV motion producing larger after-effects. This disparity between looming and receding adaptors can be interesting in its own right because of the putative differences in the biological relevance between looming and receding objects (see Maier et al., 2004, and references therein), but it falls outside the scope of this study[1]. Suffice it to say that this trend is orthogonal to the effects under investigation here and that it was unreliable throughout the rest of experiments in this study (not significant in Experiments 1 and 3, nor, for instance, in Kitagawa & Ichihara, 2002).

The main conclusion of Experiment 2 is that, in accordance to what is seen in the aMAE measurements, participants' subjective ratings about the continuity of the low-rate visual events were higher when they were accompanied by the high-rate sounds. It is also worth noting that Experiment 2 has replicated the critical effect of sound on visual adaptors observed for the first time in Experiment 1.

4. Experiment 3

Results from the two previous experiments reveal that the combinations of low-rate visual flicker with high-rate auditory stimuli significantly increase the aMAE and the subjective ratings of visual smoothness. Yet, the present results do not speak directly about

[1] One possible explanation for the disparity between aMAEs produced by receding and approaching stimuli could be that short tones with a steady intensity level tend to be perceived as increasing in loudness (Reinhardt-Rutland, 1992; Small, 1977). Thus, for our PEST procedure, steeper test tones had to be applied for adaptors with falling intensity to compensate the increasing-loudness after-effects.

whether the observed effects are specific to motion per se or else they result just from the more continuous temporal signal provided by the higher rate flutter. This is especially relevant in light of the fact that the role of direction congruency in multisensory integration of motion has been somehow controversial (Alais & Burr, 2004; Meyer & Wuerger, 2001; Soto-Faraco et al., 2002, 2003). We tested the relevance of motion direction in this effect by using direction-incongruent audio-visual adaptors. If the effect of the audio-visual adaptor lacks direction specificity, then it should work equally well despite their respective directions of motion. If this was the case, then the adapting effects found in Experiments 1 and 2 would need to be explained by multisensory interactions in terms of temporal factors (i.e. change rate of the adaptor stimulus). If, on the contrary, direction incompatible adaptors fail to produce reliable aMAE, one would need to conclude that the interactions observed in Experiments 1 and 2, whereby aMAE is induced, occur at the level of motion signal processing. In order to test the direction specificity of the effect, we conducted a third experiment using incongruent combinations of auditory and visual adaptors.

4.1. Method

4.1.1. Participants

Fifteen participants (mean age of 24.3 years, 7 females) took part in the experiment voluntarily at Chalmers University of Technology. All reported to have normal or corrected-to-normal vision and no hearing problems. They were paid one cinema ticket for their participation. The study has been conducted under the approval of the local ethics committee.

4.1.2. Stimuli

In the third experiment, we used four types of adapting stimuli, consisting of combinations of low-rate flicker (expanding or contracting) paired with either low- or high-rate flutter always moving in the incongruent motion direction (e.g. contracting disk coupled with a rising tone, Ah+Vl-).

4.1.3. Apparatus and procedure

The same procedure as in Experiment 2 was used, except that no subjective ratings were assessed this time.

4.2. Results and discussion

Results from 2 (direction of sound) $\times$ 2 (stimulus type) ANOVA showed no significant effects of the main factors nor the interaction term, with $F(1,14) = 2.64$, $p = 0.13$ for the effect of stimulus type (VlAl vs. VlAh), as also can be seen in Fig. 2 (right panel). Adding to the argument, the between-subjects comparison between the present directionally incongruent VlAh adaptors in Experiment 3 with the congruent audio-visual VlAh motion adaptors used in Experiment 1 (using external loudspeakers) and in Experiment 2 (using headphones) resulted in significant differences: $t(28) = -3.9$, $p < 0.001$ and $t(22.6) = -2.4$, $p < 0.024$, respectively (averaged approaching and receding aMAEs). We can conclude that, similar to the results with the continuous audio-visual stimuli (Kitagawa & Ichihara, 2002), the aMAEs obtained with incongruent adaptors were weaker than the aMAEs seen with congruent ones, and not different in size and direction to the aMAE induced by unimodal acoustic adaptors. In particular, incongruent low-rate flicker-flutter failed to elicit a significant after-effect (as compared with zero; $p = 0.2$ for Vl+Al- and $p = 0.7$ for Vl-Al+). This direction specific audio-visual interaction is identical to the results from the second experiment in (Kitagawa & Ichihara, 2002), where a combination of continuous but directionally incongruent audio-visual adaptors produced aMAE similar to auditory only condition. This and our results show that directional consistency is an important factor for integrating multimodal information about object motion in depth.

5. General discussion

The present findings reveal that discrete audio-visual motion stimuli follow the same cross-modal interaction patterns as continuous stimuli in their capacity to produce aMAE. More importantly, however, our results provide empirical evidence supporting that sound can fill-in sparsely time-sampled visual motion, possibly arising from the occurrence of illusory visual events in the low-rate flicker / high-rate flutter condition (VlAh, Fig. 2). As was discussed earlier in the introduction, intensity change is the most salient cue for the perception of auditory motion in-depth. In the following sections, we discuss the implications of this finding.

We argue that the observed effects may be based on the sound-induced visual flash illusion (Shams et al., 2000), which demonstrates that irrelevant sounds can lead to the experience of extra visual flashes of a flickering disc (see also Kamitani and Shimojo (2001), for an apparent motion experiment). Note, for example, that the timing parameters of the present stimuli are similar to the ones producing the original visual flash illusion. Apart from aMAEs, in the second experiment we have also assessed the subjective ratings on visual motion "continuity-discontinuity" and, consistent with the filling-in hypothesis, the results showed that the low-rate flashes were perceived as more continuous when combined with the high-rate beeps. This can be interpreted as, similarly to Shams et al.'s (2000) findings, the sound beeps might induce illusory flashes which filled-in the low flash-rate stimulus. Such perceptual "upgrading" would explain the similarity between after-effect sizes for VlAh and VhAh conditions. Indeed, illusory visual stimuli can produce motion after-effects, as it has been shown in a study about perceptual filling-in of the blind spot area (Murakami, 1995).

Our results might be related to the "auditory driving" phenomenon (e.g. Welch et al., 1986), where perceived flicker frequency can be increased or decreased by the simultaneous presentation of auditory flutter. However, the results from our third experiment show that the results obtained were not solely dependent on the flutter rate, but also critically on the directional congruency between auditory and visual adaptors. This means that the potential of sounds to fill-in the visual series critically depends on some kind of compatibility, or congruence, between the motion signal being processed by hearing and sight. Thus, beyond the auditory driving, the present effect seems to belong to interactions between motion cues.

Interestingly, a recent study by Mastoropoulou, Debattista, Chalmers, and Troscianko (2005) investigated the influence of sound effects on the discrimination between 3 s video sequences sampled at 10, 12, 15, 20 and 24 frames per second (fps). These authors showed that, under audio-visual presentation conditions, naïve participants could reliably discriminate videos only when the displays differed by 14 fps (between 10 vs. 24 fps pair in their experiment). However, under visual-only presentation conditions participants could discriminate between displays differing by as little as 4 fps (all possible pairs of rates except for the minimally different ones; 10 vs. 12 fps and 12 vs. 15 fps). While Mastoropoulou et al. (2005) hypothesized that divided attention might be the cause of these effects, our results suggest that sound may exert a more direct impact on video by filling-in missing visual detail.

Finally, the present study provides a scientific grounding of the classic cinema techniques whereby rhythmic sound effects and music are used to enhance the visual action smoothness and

continuity (e.g. Walt Disney's "Mickey Mousing" technique; Thomas & Johnston, 1981). It has been shown that for purely visual information critical flicker-fusion frequency is typically around 60 Hz (Brown, 1965). However, the fusion of individual flashes into a perceptually continuous event might occur at lower stimulus repetition rates of approximately 20 Hz (Jewett et al., 2006 and references therein). Studies in virtual environments show a drastic decrease in task performance and users' subjective experience for frame rates below 15 Hz (Barfield & Hendrix, 1995; Reddy, 1997). In the light of the results of this study, the addition of specially designed sound might increase the deployment of media with low frame rates for applications, where data transmission rates are limited by bandwidth (e.g. mobile communication). Thus, our results open an avenue for the perceptual optimization of dynamic multisensory scenes (see Väljamäe & Tajadura-Jiménez, 2007). In such optimization strategy, amodal categories of scenes content (e.g. objects and their motion trajectories) may define the quality of individual sensory modalities used for media synthesis, delivery and reproduction.

Acknowledgements

AV work was supported by PRESENCCIA project (27731) under the IST programme and Swedish Science Council (VR); SS-F work was supported by grants of the Ministerio de Educación y Ciencia (SEJ2007-64103/PSIC and CDS2007-00012 Consolider-Ingenio programme).

References

Alais, D., & Burr, D. (2004). No direction-specific bimodal facilitation for audiovisual motion detection. *Cognitive Brain Research, 19*, 185–194.
Barfield, W., & Hendrix, C. (1995). The effect of update rate on the sense of presence within virtual environments. *Virtual Reality: The Journal of the Virtual Reality Society, 1*, 3–16.
Berger, T. D., Martelli, M., & Pelli, D. G. (2003). Flicker flutter: Is an illusory event as good as the real thing? *Journal of Vision, 3*(6), 406–412.
Blauert, J. (1997). *Spatial hearing* (revised ed.). Cambridge, MA: The MIT Press.
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.
Brown, J. L. (1965). Flicker and intermittent stimulation. In C. H. Graham (Ed.), *Vision and visual perception* (pp. 252–320). New York: Wiley.
Chion, M. (1994). *Audio-vision: Sound on screen*. New York: Columbia University Press.
Cornsweet, T. N. (1962). The staircase-method in psychophysics. *American Journal of Psychology, 75*, 485–491.
Hong, J., & Papathomas, T. V. (2006). Influences of attention on auditory aftereffects following purely visual adaptation. *Spatial Vision, 19*, 569–580.
Jewett, D. L., Hart, T., Larson-Prior, L. J., Baird, B., Olson, M., Trumpis, M., et al. (2006). Human sensory-evoked responses differ coincident with either "fusion-memory" or "flash-memory". *BMC Neuroscience, 1*(7), 18. doi:10.1186/1471-2202-7-1.
Kamitani, Y., & Shimojo, S. (2001). Sound-induced visual "rabbit". *Journal of Vision, 1*(3), 478a.
Kanizsa, G. (1976). Subjective contours. *Scientific American, 234*(4), 48–52.
Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature, 416*, 172–174.
Kohlrausch, A., Fassel, R., & Dau, T. (2000). The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. *Journal of the Acoustical Society of America, 108*, 723–734.
Komatsu, H. (2006). The neural mechanisms of perceptual filling-in. *Nature Review Neuroscience, 7*, 220–231.
Lutfi, R. A., & Wang, W. (1999). Correlated analysis of acoustic cues for the discrimination of auditory motion. *Journal of Acoustic Society of America, 106*(2), 919–928.
McCormick, D., & Mamassian, P. (2008). What does the illusory flash look like? *Vision Research, 48*(1), 63–69.
Maier, J. X., Neuhoff, J. G., Logothetis, N. K., & Ghazanfar, A. A. (2004). Multisensory integration of looming signals by rhesus monkeys. *Neuron, 43*, 177–181.
Massaro, D. M. (2004). From multisensory integration to talking heads and language learning. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of multisensory processes* (pp. 153–176). Cambridge: MIT Press.
Mastoropoulou, G., Debattista, K., Chalmers, A., & Troscianko, T. (2005). The influence of sound effects on the perceived smoothness of rendered animations. In Paper presented at APGV'05: Second symposium on applied perception in graphics and visualization. La Coruña, Spain (August).
Meyer, G. F., & Wuerger, S. M. (2001). Cross-modal integration of auditory and visual motion signals. *Neuroreport, 12*(11), 2557–2560.
Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17*, 154–163.
Murakami, I. (1995). Motion after-effect after monocular adaptation to filled-in motion at the blind spot. *Vision Research, 35*, 1041–1045.
Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of L2 sounds. *Psychological Research, 71*, 4–12.
Pelli, D. G. (1997). The video toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.
Ramachandran, V. S. (1992). Filling in gaps in perception: Part 1. *Current Directions in Psychological Science, 1*(6), 199–205.
Reddy, M. (1997). The effects of low frame rate on a measure for user performance in virtual environments (Technical Report ECS-CSG-36-97). University of Edinburg, Scotland, UK..
Regan, D., & Beverley, K. I. (1978). Illusory motion in depth: Aftereffect of adaptation to changing size. *Vision Research, 18*, 209–212.
Reinhardt-Rutland, A. H. (1992). Increasing- and decreasing-loudness aftereffects: Asymmetrical functions for absolute rate of sound level change in adapting stimulus. *The Journal of General Psychology, 122*(2), 187–193.
Röhrbein, F., Schill, K., & Zetzsche, C. (2000). Intermodal sensory interactions for ecologically valid intensity changes as caused by moving observers or moving objects. In H. H. Bülthoff, M. Fahle, K. Gegenfurthner, & H. Mallot (Eds.), *TWK 2000 - Beiträge zur 3. Tübinger Wahrnehmungskonferenz*. Kirchentellinsfurth: Knirsch Verlag. pp. 62.
Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature, 408*, 788.
Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research, 14*(1), 147–152.
Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potential in humans. *NeuroReport, 12*(17), 3849–3852.
Small, A. M. (1977). Loudness perception of signals of monotonically changing sound level. *Journal of the Acoustical Society of America, 61*, 1293–1297.
Soto-Faraco, S., Kingstone, A., & Spence, C. (2003). Multisensory contributions to the perception of motion. *Neuropsychologia, 41*(13), 1847–1862.
Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research, 14*, 139–146.
Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Crossmodal dynamic capture: Congruency effects of motion perception across sensory modalities. *Journal of Experimental Psychology: Human Perception & Performance, 30*, 330–345.
Sumby, W., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212–215.
Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America, 41*, 782–787.
Thomas, F., & Johnston, O. (1981). *Disney animation: The illusion of life*. New York: Abbeyville Press.
Väljamäe, A., & Tajadura-Jiménez, A. (2007). Perceptual optimization of audio-visual media: Moved by sound. In B. Anderson & J. Anderson (Eds.), *Narration and spectatorship in moving images*. Cambridge Scholars Press.
Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception & Performance, 26*, 1583–1590.
Welch, R. B., Duttenhurt, L. D., & Warren, D. H. (1986). Contributions of audition involved in the multimodal integration of perceptual and vision to temporal rate perception. *Perception & Psychophysics, 39*, 294–300.