



## ORIGINAL ARTICLE

# Audio–visual interaction in visual motion detection: Synchrony versus Asynchrony



Stephanie Rosemann\*, Inga-Maria Wefel<sup>◇</sup>, Volkan Elis<sup>◇</sup>, Manfred Fahle

Department of Human-Neurobiology, University of Bremen, Hochschulring 18, 28359 Bremen, Germany

Received 31 May 2016; accepted 9 December 2016

Available online 23 February 2017

### KEYWORDS

Cross-modal integration;  
Audiovisual motion;  
Congruence;  
Vertical;  
Horizontal

### Abstract

**Objective:** Detection and identification of moving targets is of paramount importance in everyday life, even if it is not widely tested in optometric practice, mostly for technical reasons. There are clear indications in the literature that in perception of moving targets, vision and hearing interact, for example in noisy surrounds and in understanding speech. The main aim of visual perception, the ability that optometry aims to optimize, is the identification of objects, from everyday objects to letters, but also the spatial orientation of subjects in natural surrounds. To subserve this aim, corresponding visual and acoustic features from the rich spectrum of signals supplied by natural environments have to be combined.

**Methods:** Here, we investigated the influence of an auditory motion stimulus on visual motion detection, both with a concrete (left/right movement) and an abstract auditory motion (increase/decrease of pitch).

**Results:** We found that incongruent audiovisual stimuli led to significantly inferior detection compared to the visual only condition. Additionally, detection was significantly better in abstract congruent than incongruent trials. For the concrete stimuli the detection threshold was significantly better in asynchronous audiovisual conditions than in the unimodal visual condition.

**Conclusion:** We find a clear but complex pattern of partly synergistic and partly inhibitory audio–visual interactions. It seems that asynchrony plays only a positive role in audiovisual motion while incongruence mostly disturbs in simultaneous abstract configurations but not in concrete configurations. As in speech perception in hearing-impaired patients, patients suffering from visual deficits should be able to benefit from acoustic information.

© 2017 Spanish General Council of Optometry. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\* Corresponding author. Present address: Biological Psychology Lab, Department of Psychology, Faculty of Medicine and Health Sciences, European Medical School, Carl von Ossietzky University Oldenburg, 26111 Oldenburg, Germany.

E-mail address: [Stephanie.rosemann@uni-oldenburg.de](mailto:Stephanie.rosemann@uni-oldenburg.de) (S. Rosemann).

<sup>◇</sup> Both authors contributed equally to this work.

**PALABRAS CLAVE**

Integración intermodal;  
Movimiento audiovisual;  
Congruencia;  
Vertical;  
Horizontal;  
Concurrencia

**Interacción audio-visual en la detección del movimiento visual: sincronía frente a asincronía****Resumen**

**Objetivo:** La detección e identificación de los objetivos en movimiento es de extrema importancia en la vida diaria, aun cuando no se ha probado ampliamente en la práctica optométrica por motivos técnicos. La literatura incluye indicaciones claras acerca de la interacción entre la percepción de objetivos en movimiento, la visión y la audición, como por ejemplo en los ambientes ruidosos y en la comprensión del habla. La meta principal de la percepción visual, la capacidad que trata de optimizar la optometría, es la identificación de objetos, desde los cotidianos a las letras, y también la orientación espacial de los sujetos en entornos naturales. Para ayudar a lograr esta meta, deben combinarse las correspondientes características visuales y acústicas de entre el amplio espectro de señales que aportan los ambientes naturales.

**Métodos:** Investigamos la influencia de un estímulo de movimiento auditivo sobre la detección del movimiento visual, tanto en el movimiento auditivo concreto (movimiento izquierda/derecha) como abstracto (incremento/decremento de tono).

**Resultados:** Encontramos que los estímulos audiovisuales incongruentes originaban una detección significativamente inferior en comparación a la situación únicamente visual. Además, la detección fue considerablemente mejor en los campos congruentes abstractos que en los incongruentes. Para los estímulos concretos, el umbral de detección fue significativamente inferior en situaciones audiovisuales asíncronas que en la situación visual unimodal.

**Conclusión:** Encontramos un patrón claro aunque complejo de interacciones audio-visuales parcialmente sinérgicas y parcialmente inhibitorias. Parece ser que la asincronía juega únicamente un papel positivo en el movimiento audiovisual, mientras que la incongruencia se altera principalmente en las configuraciones abstractas simultáneas pero no en las configuraciones concretas. Como en la percepción del habla en pacientes con deficiencias auditivas, los pacientes que padecen déficits visuales deberían poder beneficiarse de la información acústica.

© 2017 Spanish General Council of Optometry. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Introduction**

We are surrounded by a loud, vivid, moving and changing environment where objects are perceived via different sensory systems. To achieve a unified perception, we have to localize, relate and integrate the different inputs to orient ourselves in the world and to interact with it – from cooking to driving a car. An important process subserving this goal is called cross-modal integration of visual and auditory signals and this interaction is more important than most people realize.

While the visual modality was long considered to be the dominant modality, it was shown that auditory stimuli can influence and modify visual perception.<sup>1–12</sup> Beneficial cross-modal interactions were reported, where an additional auditory stimulus facilitated visual perception compared to unimodal visual stimulation.<sup>3,9,13–26</sup> An advantage of cross-modal integration arises in this connection because our brain associates features of auditory and visual objects with one another, e.g. high tones are associated with small objects or bright surfaces while low tones are more often combined with large objects and dark surfaces.<sup>27,28</sup> In contrast, additional visual information during an auditory task shows only slight performance improvements.<sup>24,29–31</sup> In the context of speech perception, access to articular movement and gestures of the speaker enhances the perception of speech.<sup>32,33</sup>

This audiovisual integration in speech perception is of particular importance in older adults<sup>34</sup> or cochlear-implanted deaf patients.<sup>35</sup> Patient and animal studies show strong associations of vision and hearing impairments,<sup>36</sup> compensation and crossmodal plasticity of one modality when the other is impaired.<sup>37–39</sup> Patients with visual disorders (hemianopia or visuo-spatial neglect) benefit tremendously from audiovisual integration to improve or recover from visual deficits.<sup>40–42</sup> In addition, slight deficits in vision may be compensated by the auditory domain, leading to an increased reliance on auditory rather than visual cues. Therefore, this topic is highly relevant for optometry and optometric practice both concerning diagnostics and treatment of visual deficits by testing not only visual but also audio-visual abilities (analogous to hearing-impairments).

In relation to cross-modal processing the topic of motion perception has received much attention lately.<sup>1,43,44</sup> The connection between visual and auditory motion detection was confirmed to be beneficial if both stimuli move in the same direction.<sup>6,8,12,25,45–48</sup> If one of the inputs disturbs the perception of the others, detection of the target stimulus direction can be reduced.<sup>20</sup> These effects have been shown mostly for the horizontal plane using apparent motion stimuli and moving vertical bars, but also for random dot paradigms and moving gratings. Auditory stimuli were auditory tones fading from one ear to the

other or changing loudspeaker locations. Some studies also explored cross-modal integration for the vertical plane using abstract auditory stimuli like ascending or descending pitch tones.<sup>6,7,49</sup> The auditory input has an advantageous effect on visual motion detection as it gives additional temporal information to the already existing spatial information of the visual system.

Synchrony is an important topic in cross-modal binding and integration.<sup>50–52</sup> Greatest effects are observed if both stimuli are co-localized and simultaneously presented whereas a delayed presentation reduces the effect.<sup>7,13,21,53,54</sup> But also larger time windows can still lead to cross-modal integration, depending on context and complexity of the stimuli. In this case, the visual modality takes sound velocity into account.<sup>55–58</sup> Additionally, congruence is an essential factor in multisensory motion perception. If both stimuli move in the same direction, detection thresholds can be significantly reduced while incongruent stimuli lead to increased detection thresholds.<sup>7,13–15,18,45,49,53,59,60</sup> In some of these studies, clear congruence effects were shown, some studies only demonstrated the clear advantageous effect of multimodal compared to unimodal presentation.

Our study produces a new example of auditory-visual interaction that is relevant for the perception of moving objects and underlines the fact that when dealing with patients' complaints about 'blurred' vision, evaluation of static optometric charts may miss the relevant symptom and that in the perception of moving objects, seeing and hearing interact.

As previous research focused mostly on spatial audiovisual motion detection in the horizontal plane, we wanted to complement current findings about motion detection processes with an experiment using abstract auditory stimuli. To be able to compare results from audiovisual motion detection either with a concrete or an abstract auditory stimulus we conducted two similar tasks in the same participants, one experiment with spatial and one with frequency-based auditory motion. In these we investigated the influence of an auditory stimulus on a visual motion task in order to expand current knowledge about cross-modal motion detection. We had a unimodal visual task as control condition as well. We expected lower threshold values for the audiovisual tasks compared to the unimodal task (1) because it was shown that an additional auditory stimulus facilitates visual motion detection.<sup>8,47</sup> Furthermore we hypothesized lower threshold values for congruent trials compared to incongruent trials (2) (referring to matching and nonmatching auditory and visual stimuli), due to significantly better detection for congruent motion than for incongruent motion.<sup>6,7,14,18,49</sup> In our experimental design a third condition was implemented considering the synchrony of auditory and visual stimuli. The spatiotemporal window for cross-modal integration can be up to 1500 ms,<sup>55,56</sup> depending on the complexity of the time window. For our condition, the auditory stimulus was preceding the visual one but continued during visual stimulation. We thought that the preceding auditory stimulus functions as a cue and leads to priming of the participant. Therefore, we expected that a preceding congruent auditory stimulus leads to lower detection thresholds while a preceding incongruent auditory stimulus would lead to a higher detection threshold (3) than in the simultaneous condition. Results

showed that visual motion was influenced by incongruent frequency-based auditory motion and by asynchronous spatial auditory motion leading to a complex and differentiated picture of different forms of audio-visual motion detection.

## Material and methods

### Subjects

Subjects ( $n = 20$ ) were eight male and twelve female healthy volunteers with a mean age of 24.2 ( $\pm 3.5$ ) years. Exclusion criteria were neurological, psychiatric (self-report) or ophthalmological disorders and auditory defects.

All subjects underwent a series of pre-investigations to check for exclusion criteria of visual and auditory defects. Visual acuity was tested for near- and long-distance reading (acuity of 20/20, corresponding to 1.0, was necessary, correction with contact lenses and glasses was allowed). The Lang Test<sup>61</sup> and the Ishihara Color Vision test<sup>62</sup> served as measures for stereo and color vision. An audiometry with 8 frequencies for each ear was applied for assessment of hearing (HTTS software, Berlin, Germany).

### Stimuli

The experiment consisted of visual and auditory stimuli for two different motion directions. Visual stimulation was always concrete. The auditory stimulus was either concrete or abstract. The configurations are described in more detail under 'Experimental Design'.

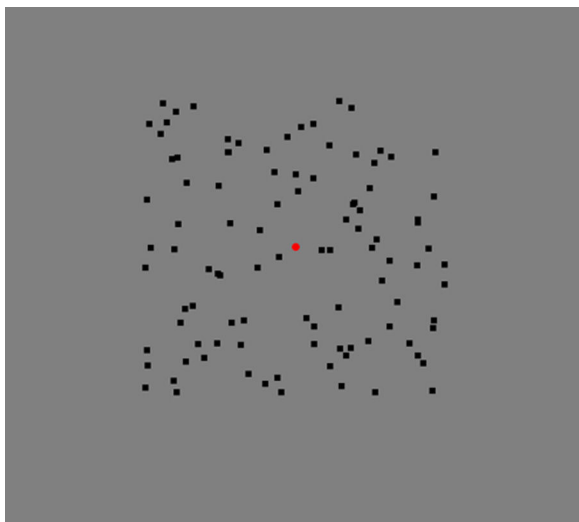
#### Auditory stimuli

In the abstract auditory configuration (vertical) there was one tone with increasing frequency (400–1200 Hz) and one with decreasing frequency (1200–400 Hz). The frequency was linearly increasing/decreasing. Each tone had a duration of 3000 ms. The frequency range was chosen because this range appeared comfortable and very well perceived. A pilot study revealed that tones with higher frequencies in the optimal hearing range were rated uncomfortable by the subjects. Therefore tones similar to those in other studies<sup>22,25</sup> were used. Loudness was the same for all participants.

In the concrete auditory configuration (horizontal) there was only one tone which was changing loudness from one ear to another to induce the perception of a moving tone. This tone had the frequency of 400 Hz.

#### Visual stimuli

A screenshot of the visual stimuli can be seen in Fig. 1. For both configurations the screen was gray with a red fixation dot in the center. The fixation dot in each test had a size of 6 arcmin. Visual stimuli were restricted to a central area of 10 × 10 cm (11.3 × 11.3°). In this area 100 black dots with a size of 10 arcmin each were moving with a velocity of 400 arcmin/s. A percentage of these 100 dots was always moving in either upward or downward direction (vertical abstract configuration) or else in either leftward or rightward direction (horizontal concrete configuration) while the other dots were moving in random directions. Presentation time of the visual stimuli was 3000 ms and dots were visible



**Figure 1** View of the visual stimulus for both configurations. In both configurations the visual stimulus was shown for 3000 ms. The screen is gray and in the center is the red fixation dot. Around the fixation dot there are 100 black dots which are moving in partly random directions.

until they reached the edge of the central presentation area of  $10 \times 10$  cm. The percentage of dots moving together was computed for every trial depending on the response of the participant in the preceding trial. This design corresponds to a threshold value computation by a staircase method: How many dots must be moving together so that the participant can recognize the direction with a performance of 75% correct? Please find a detailed description of this design under 'Data Analysis'.

## Experimental design

### Configurations

Our design comprised two different configurations and three different conditions. The configurations defined the motion of both the visual and the auditory stimuli: in the abstract configuration the visual stimuli were moving up- and downwards and the auditory stimuli had increasing and decreasing frequencies. For the concrete configuration the visual stimuli were moving left- and rightwards and the auditory stimuli were moving from the left to the right headphone or vice versa.

In the following we use the terms 'frequency-based' (abstract auditory stimuli) and 'spatial' (concrete auditory stimuli) to differentiate between the two different configurations.

### Conditions

For each configuration there were three different conditions: a simultaneous audiovisual, an asynchronous audiovisual and a visual control condition. The audiovisual conditions comprised 80 trials and the visual condition comprised 40 trials. In the simultaneous conditions auditory and visual stimuli were presented simultaneously for 3000 ms. In the asynchronous condition the auditory stimuli preceded the visual one by 1500 ms so that there was 1500 ms

**Table 1** Overview of the Experimental Design for audiovisual motion detection. For the configuration three separated experimental tests were conducted by each participant: Visual only; and simultaneous and asynchronous audiovisual tests (both consisting of congruent and incongruent trials).

Configuration	Condition	Trial
Frequency-based	Simultaneous	Congruent
		Incongruent
	Asynchronous	Congruent
		Incongruent
Spatial	Visual only	
	Simultaneous	Congruent
		Incongruent
	Asynchronous	Congruent
		Incongruent
Visual only		

overlap of auditory and visual stimulation. In the visual control condition no auditory stimuli were present. We ended up with six different conditions: up-down simultaneous, up-down asynchronous, up-down visual, left-right simultaneous, left-right asynchronous and left-right visual.

### Trials

For all audiovisual conditions there were two different types of trials: congruent and incongruent ones. In the congruent trials visual and auditory stimuli were moving in the same direction, e.g. dots were moving upwards (or leftwards) and frequency was increasing (or loudness moving left). In the incongruent trials visual and auditory stimuli were moving in different directions, e.g. dots were moving upwards and the frequency was decreasing. This holds true for both configurations. As all audiovisual conditions comprised 80 trials, half of them were congruent and the other half was incongruent. Threshold values were computed for congruent and incongruent trials separately. The visual control condition comprised 40 trials and a threshold value for this condition was computed. We ended up with ten different threshold values, two of them just for visual motion detection and eight for audiovisual motion detection (depicted in Table 1).

## Experimental procedure

All following tests were presented on a computer monitor. Spatial resolution of the monitor was  $1600 \times 1200$  pixels ( $39 \times 31$  degrees) and the frame rate was 75 Hz. Subjects sat in 50 cm viewing distance in front of the monitor with their head in a chin rest. They wore stereo-headphones (Sennheiser HD 201, frequency spectrum of 21–18 000 Hz, sound pressure level of max. 108 dB) when required.

Subjects had two colored buttons in their hands: the red one in the left and the green one in the right hand. In the frequency-based (vertical) configuration the red button should be pressed when a motion was perceived in downward direction and the green button should be pressed for an upward motion direction. In the spatial (horizontal) configuration the right (green) and left (red) hands corresponded to motions to the right and to the left.

All subjects accomplished all configurations in counterbalanced order (four audiovisual and two visual experimental conditions) with short breaks in between. During audiovisual conditions they wore headphones and were informed about hearing an auditory stimulus. For all conditions the subjects had to direct their attention to the visual stimulus on the screen and to respond to the perceived motion direction via button-press. The instruction to the participant was to press the button as fast as possible.

## Data analysis

The experimental design used for this study corresponds to a threshold value computation with a staircase method. During stimulus presentation the actual percentage of dots moving in the same direction was computed by Matlab<sup>66</sup> for every trial. The aim was to measure the threshold of the percentage of dots that must be moving in the same direction so that the subject correctly detects the motion direction with an accuracy of 75% (for each participant for every configuration in the different conditions, like in Table 1). Congruent and incongruent trials for each condition are tested during one experimental trial but yield separate threshold values.

Threshold values for all participants have been compared with SPSS (IBM SPSS Statistics 23). First, an analysis took place of each configuration separately. An Anova for repeated measurements was conducted 1) between the audiovisual and the visual control condition (to detect the general effect of an auditory stimulus), 2) between congruent and incongruent trials (to determine the effect of congruence) and 3) between simultaneous and asynchronous conditions (to determine the influence of a synchrony of stimulus onset). Second, both configurations were compared by paired t-tests to check for differences between frequency-based and spatial auditory stimulation on visual motion detection. Normal distribution as prerequisite for t-test calculation was fulfilled.

The same was done for reaction time data obtained for all measurements.

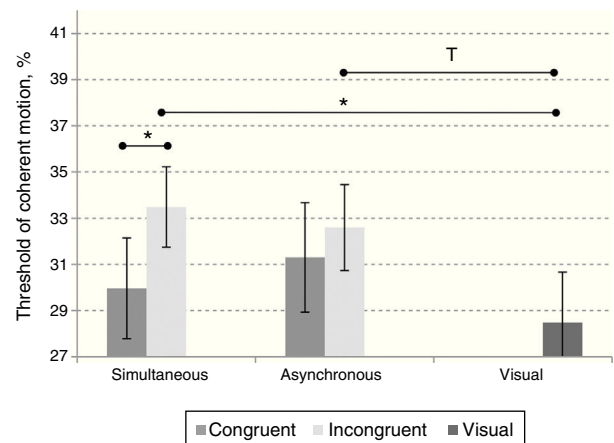
## Results

In this study two experiments (frequency-based and spatial) were conducted with the same participants. Main aim was to expand current knowledge about cross-modal motion detection by investigating the influence of an auditory stimulus on a visual motion task on a more concrete and a more abstract level of processing. We used three different conditions: a unimodal visual condition, a simultaneous audiovisual condition and an asynchronous audiovisual condition. In the following the results for the two different configurations of the motion detection task are presented separately. Normality of the data was verified by the Kolmogorov–Smirnov Test.

### Frequency-based configurations

#### Performance data

Mean values and statistical results for the frequency-based configuration can be seen in Fig. 2. The mean threshold



**Figure 2** Mean threshold values ( $\pm$ standard error) for all frequency-based configurations. Significant differences ( $p < 0.05$ ) are expressed by \* and trends of difference ( $p < 0.1$ ) are expressed by T.

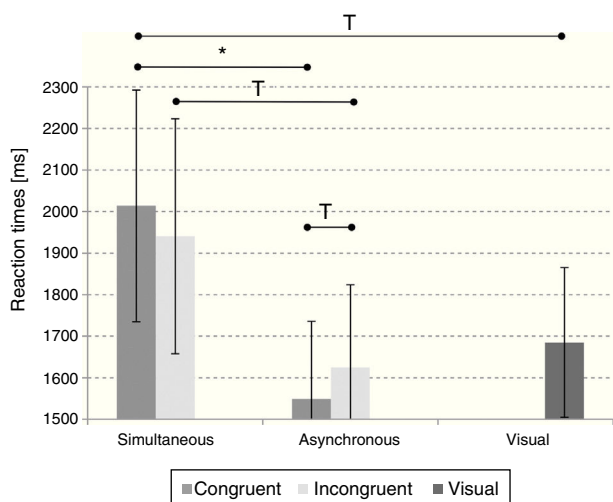
value for the simultaneous conditions in the congruent trials was 30.0 ( $\pm 9.8$ )% and for the incongruent trials it was 33.5 ( $\pm 7.8$ )%. For the asynchronous conditions the mean threshold was 31.3 ( $\pm 10.6$ )% for the congruent trials and 32.6 ( $\pm 8.3$ )% for the incongruent trials. The visual condition yielded a mean threshold value of 28.5 ( $\pm 9.7$ )%.

An ANOVA for repeated measurements for the frequency-based conditions yielded no significant difference,  $F(4,76) = 1.406$ ,  $p = 0.24$ . Post hoc paired t-tests between single conditions yielded significant results: 1) simultaneous congruent versus incongruent conditions ( $p = 0.025$ ) and 2) visual condition versus simultaneous incongruent trials ( $p = 0.016$ ). A trend was observed for visual versus asynchronous incongruent trials ( $p = 0.087$ ). Effect size measures ( $\eta^2 = 0.285$ ) showed a small effect.

#### Reaction time data

Mean values and statistical results for the reaction time data of the frequency-based configuration can be seen in Fig. 3. The mean reaction time for the simultaneous conditions in the congruent trials was 2013.8 ( $\pm 278.8$ ) ms and for the incongruent trials it was 1940.6 ( $\pm 282.9$ ) ms. For the asynchronous conditions the mean reaction time was 1548.6 ( $\pm 187.3$ ) ms for the congruent trials and 1625.2 ( $\pm 198.7$ ) ms for the incongruent trials. In the visual condition the mean reaction time was 1685 ( $\pm 180.4$ ) ms.

An ANOVA for repeated measurements for the reaction times in this configuration yielded no significant difference but a trend,  $F(4,76) = 2.502$ ,  $p = 0.099$ . Post hoc paired t-tests between single conditions yielded significant results between simultaneous and asynchronous congruent trials ( $p = 0.014$ ). Several trends were observed as well: 1) asynchronous congruent versus incongruent conditions ( $p = 0.077$ ), 2) simultaneous versus asynchronous incongruent trials ( $p = 0.058$ ) and 3) visual condition versus simultaneous congruent trials ( $p = 0.093$ ). Effect size measures ( $\eta^2 = 0.116$ ) showed a small effect.



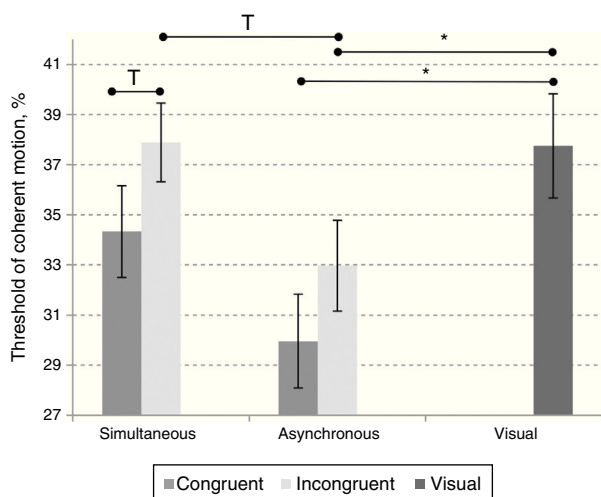
**Figure 3** Mean reaction time data ( $\pm$ standard error) for all frequency-based configurations. Significant differences ( $p < 0.05$ ) are expressed by \* and trends of difference ( $p < 0.1$ ) are expressed by T.

### Spatial configurations

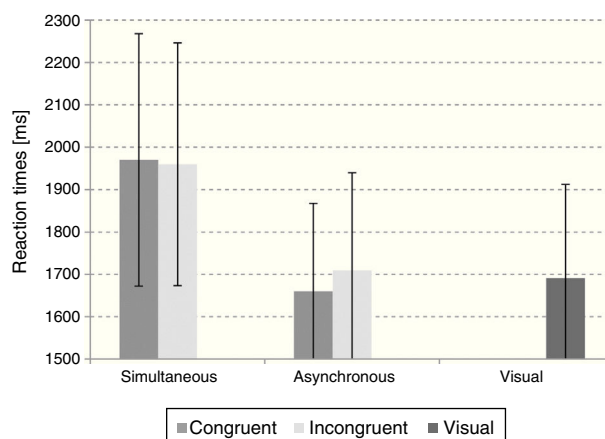
#### Performance data

Mean thresholds were 34.4 ( $\pm 8.18$ )% for the congruent trials and 37.8 ( $\pm 7.04$ )% for the incongruent trials in the simultaneous condition (Fig. 4). For the asynchronous condition the mean threshold values were 30.0 ( $\pm 8.34$ )% for the congruent and 33.0 ( $\pm 8.12$ )% for the incongruent trials. In the visual condition the mean threshold value was 37.8 ( $\pm 9.3$ )%.

An ANOVA for repeated measurements for the frequency-based conditions yielded no significant difference,  $F(4,76) = 2.263$ ,  $p = 0.1$ . Post hoc paired *t*-tests between single conditions yielded significant results: 1) visual versus asynchronous congruent trials ( $p = 0.035$ ) and 2) visual versus asynchronous incongruent trials ( $p = 0.036$ ). Additionally, two trends were observed: 1) simultaneous



**Figure 4** Mean thresholds ( $\pm$ standard error) for all spatial configurations. Significant differences ( $p < 0.05$ ) are expressed by \* and trends of difference ( $p < 0.1$ ) are expressed by T.



**Figure 5** Mean reaction time data ( $\pm$ standard error) for all spatial configurations. No significant differences between conditions were obtained.

congruent and incongruent conditions ( $p = 0.056$ ) and 2) simultaneous incongruent and asynchronous incongruent conditions ( $p = 0.055$ ). Effect size measures ( $\eta^2 = 0.312$ ) showed a medium effect.

#### Reaction time data

Mean reaction times were 1970.2 ( $\pm 297.9$ ) ms for the congruent trials and 1959.9 ( $\pm 286.6$ ) ms for the incongruent trials in the simultaneous condition (Fig. 5). For the asynchronous condition the mean reaction times were 1660.4 ( $\pm 207.3$ ) ms for the congruent and 1709.8 ( $\pm 229.9$ ) ms for the incongruent trials. In the visual condition the mean reaction time was 1691.5 ( $\pm 220.7$ ) ms.

An ANOVA for repeated measurements for the frequency-based conditions yielded no significant difference,  $F(4,76) = 1.199$ ,  $p = 0.311$ . Post hoc paired *t*-tests between single conditions yielded no significant results or trends and effect size measures ( $\eta^2 = 0.059$ ) showed a small effect.

#### Comparison of both configurations

One-sided paired *t*-tests between both configurations showed significant differences for the congruent trials of the simultaneous condition ( $t = -1.800$ ;  $p = 0.044$ ) and for the visual condition ( $t = -4.005$ ;  $p = 0.000$ ). For these comparisons the frequency-based configuration yielded smaller threshold values than the spatial configuration.

Simultaneous and asynchronous conditions from both configurations were compared by a one-sided paired *t*-test which showed a trend ( $t = 1.572$ ;  $p = 0.06$ ). Asynchronous stimulation was slightly better than simultaneous presentation.

Table 2 contains an overview over the results of unimodal visual presentation compared to audiovisual stimulation.

Reaction time data did not significantly differ across configurations.

### Discussion

The main aim of this study was to investigate the influence of a moving auditory stimulus of different abstract levels

**Table 2** Comparison of unimodal visual and audiovisual interactions.

	Simultaneous	Asynchronous
Frequency congruent	∅ <sup>a</sup>	∅
Frequency incongruent	–	(–)
Spatial congruent	∅	+
Spatial incongruent	∅	+

<sup>a</sup> Plus and minus denote significant improvements or impairment, brackets indicate a trend and no effect is shown by ∅.

on visual motion detection. Twenty participants conducted three different subtasks, containing a unimodal visual condition, a simultaneous audiovisual test and an asynchronous audiovisual task, in two configurations (frequency-based and spatial auditory motion).

Main results in the abstract frequency-based configuration were that incongruent trials mostly impaired the detection process compared to congruent and unimodal visual trials (see Fig. 2 and Table 2). A significant incongruent effect was only observed for the simultaneous condition. Preceding auditory stimuli led to an assimilation of congruent and incongruent thresholds, and both tended to be not as good as pure visual stimulation (Fig. 2 and Table 2). In the more direct, spatial auditory motion condition, priming (concurrency effect) improved motion detection compared to unimodal visual presentations (Fig. 4 and Table 2). A small congruence effect was again observed for the simultaneous condition, while the asynchronous condition lacked a significant congruent advantage. Effects were not mediated by reaction times, as these did not differ between conditions in the spatial conditions (Fig. 5) and only showed significant differences in the frequency-based conditions between congruent synchronous and asynchronous trials. Between both configurations no significant differences were detected with regard to reaction times.

### Cross-modality & congruence

The first two hypotheses expected better detection thresholds for audiovisual than for unimodal visual stimuli and that congruent trials lead to better results than incongruent trials. Hypothesis 1 must be rejected for the frequency-based configuration, as the threshold for the unimodal visual condition was lowest. Additionally, incongruence significantly increased the threshold in the simultaneous condition, leading to acceptance of hypothesis 2. In the spatial configuration both congruent and incongruent trials of the asynchronous condition showed significantly lower detection thresholds than the visual task, thus here hypothesis 1 can be partly accepted. Simultaneous and unimodal visual conditions did not differ significantly. Hypothesis 2 cannot be completely accepted as well, because statistics only detected a trend and not a significant difference between congruent and incongruent trials in the simultaneous condition.

Prime and Harris<sup>51</sup> did not find facilitation effects in their experiment either. Prediction of a moving visual stimulus was best in the unimodal condition and impaired in incongruent (auditory spatially displaced) conditions which

resembles results found in our frequency-based configuration but not in the spatial configuration. Further studies dealing with frequency- or pitch-based<sup>6,49</sup> and spatial motion detection<sup>14,47</sup> used apparent motion stimuli, coherent dot paradigms and gratings. They showed a directional visual motion bias in the direction of the auditory stimulus. Other experiments did not find congruence effects<sup>13,60</sup> and argue for a probability summation at the decision level. Another study<sup>24</sup> challenged this view as it demonstrated that an additional auditory stimulus can facilitate visual motion detection, even if it is non-informative. The authors argue that multisensory integration can indeed occur on sensory levels. Sadaghiani et al.<sup>7</sup> performed a combined psychophysical and fMRI study and showed that different brain areas are involved in processing audiovisual motion (horizontal) and speech signals. Audiovisual pitch signals (vertical) were associated with moderate signal increases in both 'motion' and 'speech' areas. The authors argued that audiovisual motion integration emerges at the perceptual level while audiovisual speech integration emerges at the decision level. Maeda et al.<sup>6</sup> also showed that presentation of words did not directly influence visual motion at the perceptual level but at the decision level. Pitch processing is both associated with spatial dimensions<sup>27</sup> but also closely connected to language processing and therefore involves early and higher areas of the brain. In our experiment pitch (frequency-based) and motion (spatial) configurations were mixed. Participants may have adopted a decision-strategy for the frequency-based trials (no facilitation, but impaired by incongruence) and applied this strategy to the spatial trials as well (no facilitation effects; similar to Ref. 60).

### Asynchrony

Hypothesis 3 dealt with the comparison of simultaneous and asynchronous presentation. The priming effect we expected can, if only, be seen in the spatially incongruent conditions, where the preceding auditory stimulus yields a better detection threshold (statistical trend). In both configurations, frequency-based and spatial, the asynchronous congruent and incongruent thresholds did not significantly differ from each other. Simultaneous and asynchronous congruent trials did also not significantly differ from each other. A comparison of all simultaneous trials with all asynchronous ones from both configurations yielded a trend, with slightly better results for asynchronous than for simultaneous presentation. Analysis of reaction time data showed significantly faster reaction times for congruent asynchronous than simultaneous trials and a trend for the incongruent trials in the frequency-based configuration. Therefore, only reaction time data in this configuration slightly resemble the expected results.

Another explanation for slightly better detection in asynchronous trials may be that preceding presentation of an auditory stimulus might have an attentional effect on the perception of the stimuli.<sup>63,64</sup> Thus, the asynchronous conditions might have induced an increased alertness in participants leading to an improved accuracy or faster reaction times just because attention was directed on the task. However, hypothesis 3 must be rejected, as no clear significant effects are found, although the results tend in the

expected direction. In line with other studies,<sup>7,13,21,53,54</sup> we see a reduced – rather small – influence of the auditory stimulus on visual motion detection for congruent and incongruent asynchronous stimuli. Although in our experiment the auditory stimulus was preceding and then concurring with the visual stimulus, integration of both seems not to have taken place.

### Abstract vs. concrete motion

In both configurations we find that congruent and incongruent trials differ in the simultaneous condition (statistical significance in the frequency configuration and a trend in the spatial configuration), but they do not differ in the asynchronous condition. Furthermore, the difference between asynchronous compared to synchronous stimulation across both configurations is nearly significant. Unimodal visual stimulation can be affected in both configurations: either positively by asynchrony in the spatial configuration or negatively by incongruence in the frequency-based configuration (Table 2). This is a very important finding as it shows that concurrency has an effect on spatial audiovisual stimuli and that congruence has an effect on frequency-based audiovisual stimuli.

Our auditory stimuli were comparable to the auditory stimuli used by Jain et al.<sup>49</sup> These authors showed that the motion direction of the auditory modality biased the perceived direction of the visual motion stimulus along three dimensional axes (vertical, horizontal and depth). They also used vertical displacement of auditory stimuli (spatial) and sounds gliding up and down in pitch (abstract) but they used ambiguous gratings (two oppositely moving sinusoidal gratings of the same spatial frequency) as visual stimuli. So they had an unambiguous auditory and an ambiguous visual stimulus while we used only unambiguous stimuli. One might infer that influencing ambiguous visual stimuli by an additional acoustic motion stimulus is more effective than influencing an unambiguous visual stimulus. In the former case the visual motion perception is biased by the auditory stimulus while in the latter case the visual motion detection is impaired/facilitated by the auditory motion direction.

### Conclusion

We conclude that a congruence effect, or more precisely an incongruence effect, was present in the frequency-based configuration. Moreover, temporal asynchrony had an effect on acoustic spatial motion. The preceding (priming) auditory stimulus in the asynchronous condition induced an equalization of congruent and incongruent thresholds in both the spatial and the frequency-based configuration and diminished the congruence effect visible for the simultaneous condition. Cross-modal perception is a highly complex process currently not fully understood and we present results which are not in line with several previous studies casting doubts on the generality of the results of their studies.<sup>6,14,47</sup> One might infer that multimodal facilitation and congruence effects between the visual and acoustic domain differ between frequency-based and spatial acoustic motion. For spatial audiovisual integration asynchrony is an important factor while it is not for frequency- or pitch-based

audiovisual integration. Only for pitch-based acoustic movement the congruence of motion direction of stimuli plays an essential role.

In the context of recovery from visual impairments and improvement of visual perception, multisensory integration and audio-visual interactions play an important role. In hearing-impaired people additional visual input significantly increases speech perception.<sup>33</sup> Additionally, cross-modal distraction is an indicator for hearing loss.<sup>65</sup> Similarly, it may be the case that cross-modal distraction for the visual domain might be an indicator for a visual deficit. Therefore, slight impairments in the visual system could be assessed by an audio-visual motion task in which deficits would become obvious when the influence of the auditory system over the visual system grows. This would mean that incongruent trials would mostly be answered based on auditory information and not on visual information and therefore would lead to false answers and an increased detection threshold. As this topic presents rather complex mechanisms, multisensory tests for improvement of visual functions should be selected and designed carefully and adapted to the need of the patient. Thus, in order to find an appropriate test, the mechanisms of audio-visual motion within different paradigms should be further clarified.

### Ethical approval

The study was approved by the local ethics committee of the University of Bremen and carried out in accordance to the Declaration of Helsinki. All subjects signed a written consent form.

### Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The first author received a PhD grant from the German National Academic Foundation (Studienstiftung des Deutschen Volkes).

### Conflict of interest

The authors declare that they have no conflict of interest.

### Acknowledgments

The authors would like to thank Dennis Trenner for writing the programs for data acquisition and analysis

### References

1. Burr D, Thompson P. Motion psychophysics: 1985–2010. *Vis Res.* 2011;51:1431–1456.
2. Driver J, Spence C. Multisensory perception: beyond modularity and convergence. *Curr Biol.* 2000;10:R731–R735.
3. Hidaka S, Teramoto W, Keetels M, Vroomen J. Effect of pitch-space correspondence on sound-induced visual motion perception. *Exp Brain Res.* 2013;231:117–126.
4. Kitagawa N, Ichihara S. Hearing visual motion in depth. *Nature.* 2002;416:172–174.



5. Koelwijn T, Bronkhorst A, Theeuwes J. Attention and the multiple stages of multisensory integration: a review of audiovisual studies. *Acta Psychol (Amst)*. 2010;134:372–384.
6. Maeda F, Kanai R, Shimojo S. Changing pitch induced visual motion illusion. *Curr Biol*. 2004;14:R990–R991.
7. Sadaghiani S, Maier JX, Noppeney U. Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *J Neurosci*. 2009;29:6490–6499.
8. Sanabria D, Spence C, Soto-Faraco S. Perceptual and decisional contributions to audiovisual interactions in the perception of apparent motion: a signal detection study. *Cognition*. 2007;102:299–310.
9. Shams L, Kim R. Crossmodal influences on visual perception. *Phys Life Rev*. 2010;7:269–284.
10. Shimojo S, Shams L. Sensory modalities are not separate modalities: plasticity and interactions. *Curr Opin Neurobiol*. 2001;11:505–509.
11. Spence C. Multisensory integration – solving the crossmodal binding problem. comment on “crossmodal influences on visual perception” by shams & kim. *Phys Life Rev*. 2010;7:285–286, discussion 295–8.
12. Teramoto W, Manaka Y, Hidaka S, et al. Visual motion perception induced by sounds in vertical plane. *Neurosci Lett*. 2010;479:221–225.
13. Alais D, Burr D. No direction-specific bimodal facilitation for audiovisual motion detection. *Brain Res Cogn Brain Res*. 2004;19:185–194.
14. Baumann O, Greenlee MW. Neural correlates of coherent audiovisual motion perception. *Cerebral Cortex*. 2007;17:1433–1443.
15. Frings C, Spence C. Crossmodal congruency effects based on stimulus identity. *Brain Res*. 2010;1354:113–122.
16. Gleiss S, Kayser C. Oscillatory mechanisms underlying the enhancement of visual motion perception by multisensory congruency. *Neuropsychologia*. 2014;53:84–93.
17. Jaekl P, Perez-Bellido A, Soto-Faraco S. On the ‘visual’ in ‘audio-visual integration’: a hypothesis concerning visual pathways. *Exp Brain Res*. 2014;232:1631–1638.
18. Kim R, Peters MA, Shams L. 0 + 1 > 1: how adding noninformative sound improves performance on a visual task. *Psychol Sci*. 2012;23:6–12.
19. Lewkowicz DJ, Kraebel KS. The value of multisensory redundancy in the development of intersensory perception 67. *Dev Psychol*. 1986;373:373–377.
20. McDonald JJ, Teder-Salejarvi WA, Hillyard SA. Involuntary orienting to sound improves visual perception. *Nature*. 2000;407:906–908.
21. Meyer GF, Wuerger SM, Rohrbein F, Zetzsche C. Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Exp Brain Res*. 2005;166(3–4):538–547.
22. Mossbridge JA, Grabowecy M, Suzuki S. Changes in auditory frequency guide visual-spatial attention. *Cognition*. 2011;121:133–139.
23. Noesselt T, Tyll S, Boehler CN, Budinger E, Heinze HJ, Driver J. Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *J Neurosci*. 2010;30:13609–13623.
24. Spence C, Driver J. Audiovisual links in exogenous covert spatial orienting. *Percept Psychophys*. 1997;59:1–22.
25. Takeshima Y, Gyoba J. Changing pitch of sounds alters perceived visual motion trajectory. *Multisens Res*. 2013;26:317–332.
26. Teder-Salejarvi WA, Munte TF, Sperlich F, Hillyard SA. Intramodal and cross-modal spatial attention to auditory and visual stimuli. an event-related brain potential study. *Brain Res Cogn Brain Res*. 1999;8:327–343.
27. Evans KK, Treisman A. Natural cross-modal mappings between visual and auditory features. *J Vis*. 2011;10, 6.1–12.
28. Spence C. Crossmodal correspondences: a tutorial review. *Atten Percept Psychophys*. 2011;73:971–995.
29. Alais D, Cass J. Multisensory perceptual learning of temporal order: audiovisual learning transfers to vision but not audition. *PLoS One*. 2010;5:e11283.
30. Bendixen A, Grimm S, Deouell LY, Wetzel N, Madebach A, Schroger E. The time-course of auditory and visual distraction effects in a new crossmodal paradigm. *Neuropsychologia*. 2010;48:2130–2139.
31. Soto-Faraco S, Spence C, Kingstone A. Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities. *J Exp Psychol Hum Percept Perform*. 2004;30:330–345.
32. Gilbert JL, Lansing CR, Garnsey SM. Seeing facial motion affects auditory processing in noise. *Atten Percept Psychophys*. 2012;74:1761–1781.
33. Kim J, Sironic A, Davis C. Hearing speech in noise: seeing a loud talker is better. *Perception*. 2011;40:853–862.
34. Maguinness C, Setti A, Burke KE, Kenny RA, Newell FN. The effect of combined sensory and semantic components on audio-visual speech perception in older adults. *Front Aging Neurosci*. 2011;3:19.
35. Rouger J, Lagleyre S, Fraysse B, Deneve S, Deguine O, Barone P. Evidence that cochlear-implemented deaf patients are better multisensory integrators. *Proc Natl Acad Sci U S A*. 2007;104:7295–7300.
36. Michikawa T, Nishiwaki Y, Kikuchi Y, et al. Gender-specific associations of vision and hearing impairments with adverse health outcomes in older Japanese: a population-based cohort study. *BMC Geriatr*. 2009;9:50.
37. Lomber SG, Meredith MA, Kral A. Adaptive crossmodal plasticity in deaf auditory cortex: areal and laminar contributions to supranormal vision in the deaf. *Prog Brain Res*. 2011;191:251–270.
38. Rauschecker JP, Harris LR. Auditory compensation of the effects of visual deprivation in the cat’s superior colliculus. *Exp Brain Res*. 1983;50:69–83.
39. Rettenbach R, Diller G, Sireteanu R. Do deaf people see better? Texture segmentation and visual search compensate in adult but not in juvenile subjects. *J Cogn Neurosci*. 1999;11:560–583.
40. Dundon NM, Bertini C, Ladavas E, Sabel BA, Gall C. Visual rehabilitation: visual scanning, multisensory stimulation and vision restoration trainings. *Front Behav Neurosci*. 2015;9:192.
41. Frassinetti F, Bolognini N, Bottari D, Bonora A, Ladavas E. Audiovisual integration in patients with visual deficit. *J Cogn Neurosci*. 2005;17:1442–1452.
42. Lewald J, Tegenthoff M, Peters S, Hausmann M. Passive auditory stimulation improves vision in hemianopia. *PLoS One*. 2012;7:e31603.
43. Nishida S. Advancement of motion psychophysics: review 2001–2010. *J Vis*. 2011;11:11.
44. Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG. The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci*. 2010;14:400–410.
45. Alink A, Euler F, Kriegeskorte N, Singer W, Kohler A. Auditory motion direction encoding in auditory cortex and high-level visual cortex. *Hum Brain Mapp*. 2012;33:969–978.
46. Hidaka S, Teramoto W, Sugita Y, Manaka Y, Sakamoto S, Suzuki Y. Auditory motion information drives visual motion perception. *PLoS One*. 2011;6:e17499.
47. Meyer GF, Wuerger SM. Cross-modal integration of auditory and visual motion signals. *Neuroreport*. 2001;12:2557–2560.
48. Soto-Faraco S, Spence C, Kingstone A. Assessing automaticity in the audiovisual integration of motion. *Acta Psychol (Amst)*. 2005;118(1–2):71–92.

49. Jain A, Sally SL, Papathomas TV. Audiovisual short-term influences and aftereffects in motion: examination across three sets of directional pairings. *J Vis.* 2008;8, 7:1–13.
50. Caclin A, Bouchet P, Djoulah F, Pirat E, Pernier J, Giard MH. Auditory enhancement of visual perception at threshold depends on visual abilities. *Brain Res.* 2011;1396:35–44.
51. Prime SL, Harris LR. Predicting the position of moving audiovisual stimuli. *Exp Brain Res.* 2010;203:249–260.
52. Radeau M, Bertelson P. Auditory-visual interaction and the timing of inputs. *Psychol Res.* 1987;49:17–22.
53. Soto-Faraco S, Kingstone A, Spence C. Multisensory contributions to the perception of motion. *Neuropsychologia.* 2003;41:1847–1862.
54. Vroomen J, de Gelder B. Sound enhances visual perception: cross-modal effects of auditory organization on vision. *J Exp Psychol Hum Percept Perform.* 2000;26:1583–1590.
55. Colonius H, Diederich A. The optimal time window of visual-auditory integration: a reaction time analysis. *Front Integr Neurosci.* 2010;4:11.
56. Denison RN, Driver J, Ruff CC. Temporal structure and complexity affect audio-visual correspondence detection. *Front Psychol.* 2013;3.
57. Stevenson RA, Wallace MT. Multisensory temporal integration: task and stimulus dependencies. *Exp Brain Res.* 2013;227:249–261.
58. Sugita Y, Suzuki Y. Audiovisual perception: implicit estimation of sound-arrival time. *Nature.* 2003;421:911.
59. Ogawa A, Macaluso E. Audio-visual interactions for motion perception in depth modulate activity in visual area V3A. *Neuroimage.* 2013;71:158–167.
60. Wuerger SM, Hofbauer M, Meyer GF. The integration of auditory and visual motion signals at threshold. *Percept Psychophys.* 2003;65:1188–1196.
61. Lang J. *Ein neuer Stereosehtest.* Vol 182. F. Enke Verlag Stuttgart: Klin. Mbl. Augenheilkunde; 1983:373–375.
62. Ishihara S. *Ishihara's Tests for Colour-Blindness.* Tokyo: Kanehara & Co., LTD; 1986.
63. McDonald JJ, Teder-Sälejärvi WA, Russo FD, Hillyard SA. Neural basis of auditory-induced shifts in visual time-order perception. *Nature Neurosci.* 2005;8:1197–1202.
64. Feng W, Störmer VS, Martinez A, McDonald JJ, Hillyard SA. Sounds activate visual cortex and improve visual discrimination. *J Neurosci.* 2014;34:9817–9824.
65. Puschmann S, Sandmann P, Bendixen A, Thiel CM. Age-related hearing loss increases cross-modal distractibility. *Hear Res.* 2014;316:28–36.
66. MATLAB Statistics Toolbox Release. *The MathWorks Inc.* Massachusetts, United States: Natick; 2013.