Available online at www.sciencedirect.com
SCIENCE DIRECT®

Vision Research 45 (2005) 1275–1284

Vision Research

www.elsevier.com/locate/visres

Timing sight and sound

Derek H. Arnold [a,b,*], Alan Johnston [a,b], Shinya Nishida [c]

[a] *Department of Psychology, University College London, Gower Street, London WC1E 6BT, United Kingdom*
[b] *Institute of Cognitive Neuroscience, University College London, Gower Street, London WC1E 6BT, United Kingdom*
[c] *NTT Communication Science Laboratories, NTT Corporation, Japan*

Received 9 April 2004; received in revised form 30 September 2004

Abstract

It has been proposed that there is a perceptual compensation for the difference between the speeds of light and sound. We examined this possibility using a range of auditory–visual tasks, in which performance depends on the relative timing of auditory and visual information, and manipulated viewing distance to test for perceptual compensation. We explored auditory–visual integration, cross modal causal attributions, and auditory–visual temporal order judgments. We observed timing shifts with viewing distance following loudspeaker, but not headphone, presentations. We were unable to find reliable evidence of perceptual compensation. Our findings suggest that auditory and visual signals of an event that reach an observer at the same point in time tend to become perceptually bound, even when the sources of those signals could not have occurred together.

Keywords: Cross modal; Timing differences; Perceptual awareness

1. Introduction

Sound travels more slowly than light. In a thunderstorm we can often see lightning seconds before we hear thunder. While this timing difference is obvious for distant sources of sight and sound, it has been suggested that, for proximate sources within ~15 m, physically coincident sources of sight and sound appear simultaneous because of a perceptual compensation for the slower speed of sound (Kopinska & Harris, 2004; Sugita & Suzuki, 2003). This would be an impressive computational feat. The calculation requires accurate information concerning absolute viewing and hearing distance in addition to the speed of sound within the environmental setting. It is not clear how the brain might gain access to this information or implement the calculation that is required.

If there is a perceptual compensation for the slower speed of sound, relative to light, the point of subjective simultaneity between temporally concurrent sources of sight and sound should not vary as a function of viewing distance. Such compensation has been found by Sugita and Suzuki (2003) and by Kopinska and Harris (2004) but not by Stone et al. (2001) or by Lewald and Guski (2004).

To date, researchers who have examined the possibility of perceptual compensation for differences between the speeds of light and sound have used explicit timing judgments. Subjects were asked, *did the visual event occur before or after the auditory event* (Lewald & Guski, 2004; Sugita & Suzuki, 2003), *was the sound or light stimulus presented first* by (Kopinska & Harris, 2004) or *did the light and sound come on at the same time* (Stone et al., 2001)? These experimental tasks provide information concerning an observer's ability to determine the relative

---

[*] Corresponding author. Address: Department of Psychology, University College London, Gower Street, London WC1E 6BT, United Kingdom. Tel.: +44 79 8686 4498.
E-mail address: derek.arnold@ucl.ac.uk (D.H. Arnold).

0042-6989/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2004.11.014

timing of sights and sounds, but do not necessarily provide any insight into the perceptual integration of signals from the two modalities. Perceptual integration of information may involve different mechanisms. Arguably, the later may be more important in our daily lives. For instance we rarely contemplate the relative ordering of visual and auditory events. However, we often need to identify ambiguous speech, a task that can be facilitated by the integration of vision and audition (McGurk & MacDonald, 1978).

In contrast to previous reports, we decided to examine the possibility of a perceptual compensation for the slower speed of sound by focussing upon the perceptual integration of information. In the stream/bounce illusion two dots move toward, become superimposed, and then move away from one another. The dots can be seen to either pass through, or to bounce off, one another (Bertenthal, Banton, & Bradbury, 1993). Usually, if a brief tone is presented at, or shortly before, the point at which the two dots become superimposed, observers are biased to see the dots as bouncing—the auditory–visual stream/bounce illusion (Sekuler, Sekuler, & Lau, 1997; Watanabe & Shimojo, 2001). The tuning of this phenomenon may therefore be used as an implicit measure of perceptual simultaneity (Fujisaki, Shimojo, Kashino, & Nishida, 1994).

If there is a perceptual compensation for the different speeds of light and sound, given a common source of auditory and visual information, the optimal time for tone onset to produce a *bounce* percept should not vary with viewing distance. This theoretical compensation should be observed when a loudspeaker, placed in close proximity to the visual display monitor, is used for sound presentations. However, if the sound were presented to the observer by headphones, according to data obtained by Sugita and Suzuki (2003), we could expect the optimal time for the tone to induce a bounce percept to vary by an amount that is consistent with the required perceptual compensation. In these circumstances, an auditory event that is physically synchronous with a visual event should appear to be too early due to the lack of delay in the sound reaching the ear for far viewing distances.

2. Experiment 1: viewing distance and auditory–visual integration

2.1. Methods

Four observers participated in this experiment, the first author and three others who were naïve as to the purpose of the study. Visual stimuli were displayed on a 19 in. Sony Trinitron Multiscan 400 PS monitor, with a refresh rate of 100 Hz, driven by a VSG 2/5 (Cambridge Research Systems).

On each trial, observers watched two dots oscillating back and forth at a constant retinal velocity of 1.67°/s and a period of 600 ms. As the stimulus was viewed from different distances, we ensured that the retinal properties of the stimulus remained constant by adjusting the physical speed and size of the moving dots on the screen—as viewing distance increased the physical size and speed of the dots increased. Each trial commenced at a random point within this cycle. The two dots could appear black (1.3 cd/m^2) or white (60.3 cd/m^2), determined at random on a trial-by-trial basis, and moved against a grey background (30.3 cd/m^2). The physical size of the individual dots subtended 0.35° of visual angle at each of the viewing distances sampled.

On each trial, a 400 Hz tonal pip was presented at a particular point during the cycle and observers were required to indicate if the two dots were passing through, or bouncing off, one another by pressing one of two response levers. The tone onset was sudden (no ramping) and was actively driven for ~15 ms. When presented by loudspeaker (dan Technology K 8285150), the tone dissipated over ~35 ms (see Fig. 1) whereas it dissipated over ~5 ms when presented diotically by headphones (Sony MDR-006). During a run of trials, the tone was presented 10 times at each of 12 phases (see Fig. 2) in a pseudo-random order. Each trial run therefore provided 120 responses, from which an estimate of the temporal tuning of the auditory–visual stream/bounce illusion was derived.

In different trial runs, we presented the tone by headphones (peak level 80 dB SPL, measured using a Bruel & Kjaer Precision Sound Level Meter Type 2203) or a loudspeaker (peak level, 100 dB SPL from 57 cm) placed

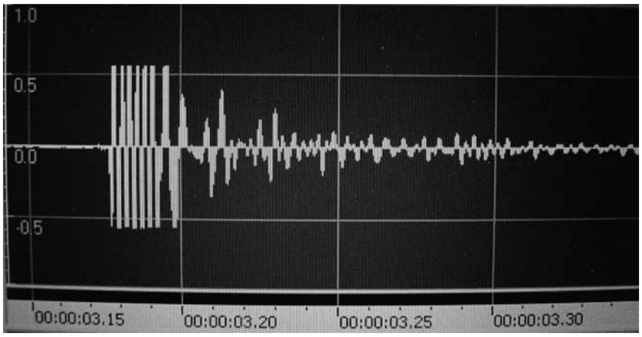

Fig. 1. Plot showing the recorded waveform of a 400 Hz tonal pip presented over the loudspeaker. Although the sound source was actively driven for just 15 ms, the tone takes a further ~35 ms to dissipate because of reverberations within the corridor (see Fig. 3) and persistent reverberations of the speaker. As a result, sounds presented over loudspeaker provided a number of distance cues (including loudness changes, spectrum changes and reverberations within the corridor) that were not available following headphone sound presentations or in a previous study where compensation was observed (Sugita and Suzuki, 2003).

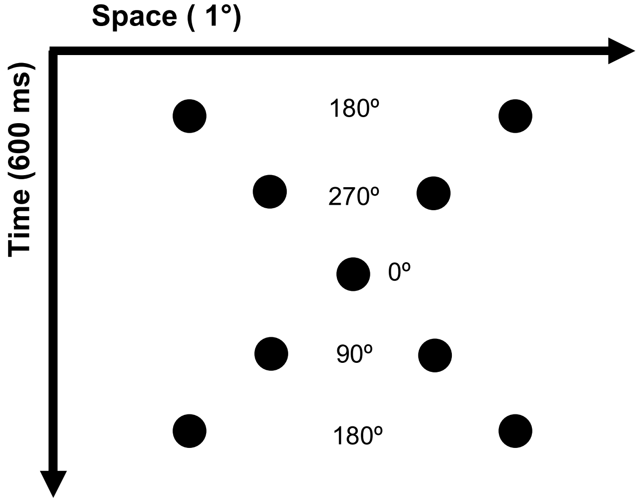

Fig. 2. Space–time plot depicting the stimulus presentation. Two dots move towards one another, become superimposed, and then move away from one another. Each trial commenced at a random point within this sequence, which repeated until the observer made a response. In different trials, the tone was presented at different points within the cycle. At a phase relationship of 0°, the tone was presented when the two dots were superimposed. At phase relationships of 90° and 270°, the tone sounded 150 ms before and therefore 450 ms after (270°), or 150 ms after and therefore 450 ms before (90°) the dots became superimposed. At a phase relationship of 180°, the tone sounded 300 ms before and after the dots were superimposed. These phase relationships reflect the relative timings at which the signals were generated, not the times at which they reached the observer.

15 cm below the monitor screen on which the visual stimulus was displayed. During each mode of presentation, two experienced observers completed four separate trial runs from each of four different viewing distances, 114 cm, 513 cm, 1026 cm and 1486 cm. To establish the robustness of the findings, we tested two additional observers, who were also naïve as to the purpose of the study. Each of these observers completed a trial run for each mode of sound presentation from 114 cm and 1482 cm. The experiment was conducted in a corridor under artificial illumination (see Fig. 3).

2.2. Results

Responses during each run of trials provided a distribution of the percentage of *bounce* responses as a function of phase. We fitted a centroid to the distribution determined in each trial run. As these distributions are symmetrical the fitted centroids, which provide a measure of central tendency, also provide an estimate of the time within the cycle at which the tone is most likely to induce the *bounce* percept. Fig. 4 shows four polar plots each fitted to data obtained in four trial runs that were completed by observer D.A. Two of these plots (A and C) are fitted to data obtained during trial runs in which headphones were used for sound presentation. The other two plots (B and D) are fitted to data obtained during trial runs in which a loudspeaker was used for sound presentation. Plots A and B were fitted to data obtained during trial runs completed from a viewing distance of 114 cm whereas plots C and D were fitted to data obtained during trial runs completed from a viewing distance of 1482 cm.

Fig. 3. Photo showing an observer completing a trial run from a viewing distance of 1482 cm. Note that the experiment was conducted in a lighted corridor which provided abundant visual depth cues to the distance of the testing apparatus.

For observer D.A., there is little difference between the optimal timings for tone presentation determined from viewing distances of 114 cm and 1482 cm with headphone sound presentations. The optimal times determined in these circumstances were 15.98 ms and 18.39 ms before the superimposition of the two dots respectively. However, there is a substantial difference between the optimal times determined at the same viewing distances with loudspeaker sound presentations. In these circumstances, the optimal times were 21.95 ms and 60.88 ms *before* the superimposition of the two dots respectively. The difference between these times (38.93 ms) closely approximates the greater time (~41 ms) required for the tone to reach the ear from the further viewing distance.

In Fig. 5 we report the time of the tone, relative to the point of superposition of the dots, that maximised the stream/bounce illusion minus the baseline time measured at the nearest viewing distance for two experienced observers. Data points with a positive value signify that, at the relevant viewing distance, *later* tones induced the bouncing percept relative to the optimal timing at a viewing distance of 114 cm. Negative values signify the reverse. As shown in Fig. 4, the optimal timing for the tone to induce a *bounce* percept varied with viewing distance when presented by loudspeaker (D.A. $F_{3,12} = 13.33$, $p < 0.001$; J.N. $F_{3,12} = 4.05$, $p = 0.033$). As viewing distance increased, progressively earlier tones were required. However, this did not happen with headphone presentation (D.A. $F_{3,12} = 0.08$, $p = 0.97$;

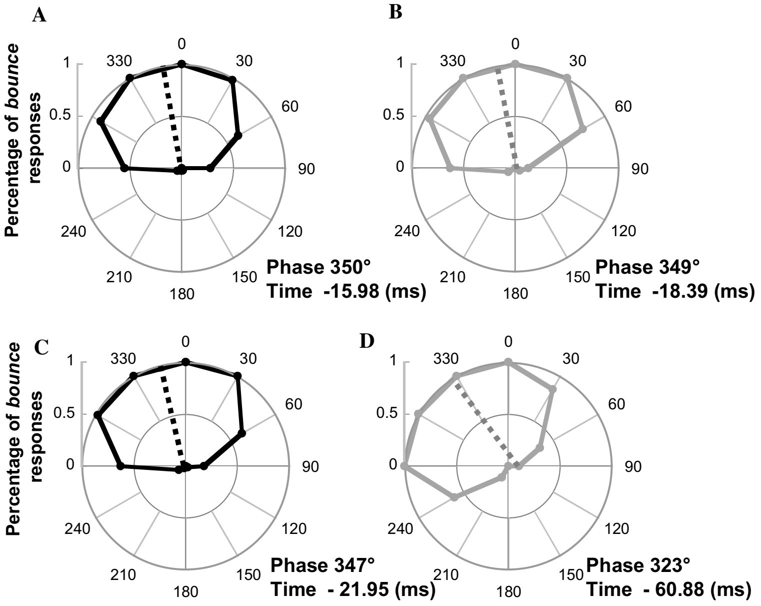

Fig. 4. Polar plots depicting the percentage of times that observer D.A. reported the two dots as *bouncing*, at each of the phase relationships tested, during four different trial runs. Plots A and B show data gathered from a viewing distance of 114 cm, C and D from 1482 cm. Data shown in plots A and C were determined during trial runs wherein the tone was presented over headphones, B and D from a speaker. Dotted lines show the average value of the centroids fitted to the four trial runs that contribute to the distribution. This value provides an estimate of the relative timing for the tone that maximally influenced visual perception. These estimates are ~16 ms (A) 18 ms (B) 22 ms (C) and 61 ms (D) before the two dots became superimposed. The centroids were fitted to data obtained in individual trial runs according the equation $\phi_c = \tan^{-1}\left(\frac{\sum_{i=1}^{N} M_i \sin\phi_i}{\sum_{i=1}^{N} M_i \cos\phi_i}\right)$ where $\phi = 0°$; $\phi_2 = 30°\ldots$; $\phi = 330°$; M_i = percent bounce responses; ϕ_c = phase of centroid.

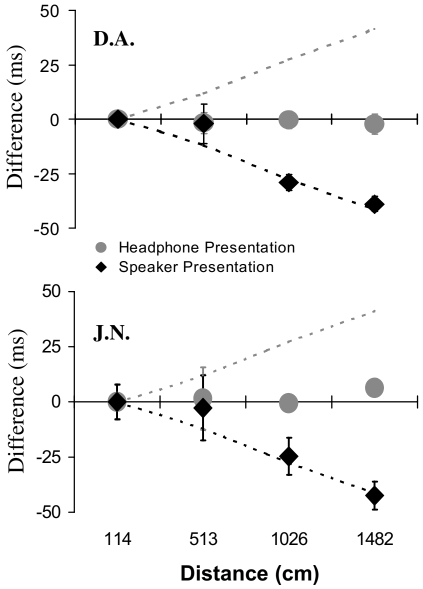

Fig. 5. The time of the tone relative to the point of superposition of the dots that maximised the stream/bounce illusion minus the baseline time measured at the nearest viewing distance. Each data point shows the average for four trial runs. Error bars show the standard error between the four estimates determined at each of the viewing distances. Dark dotted line shows the predicted times *without* compensation for the loudspeaker presentation. The grey dotted line shows the predicted times *with* compensation for the headphone presentation.

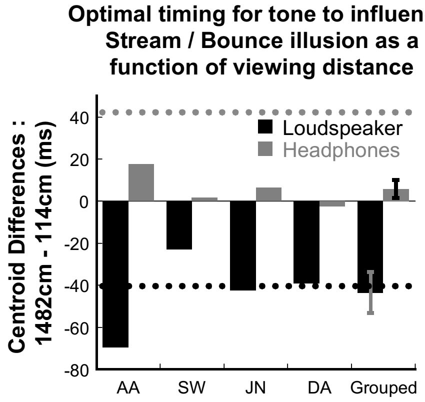

Fig. 6. The time of the tone, relative to the superposition of the pair of dots, that maximised the auditory–visual stream/bounce illusion at a viewing distance of 1482 cm minus the baseline time measured at 114 cm. Data is shown individually for four observers, the first author and three observers who were naïve as to the purpose of the experiment. The grouped average performance of these observers is also shown. Error bars depict the standard error between the observers four difference scores for each experimental condition. With loudspeaker presentations, the tone needed to be presented ~43.44 ms sooner relative to a viewing distance of 114 cm ($t_3 = -4.49$, $p = 0.021$). With headphone presentations, there was no significant difference between these times ($t_3 = 1.34$, $p = 0.273$). The dark dotted line shows the predicted time *without* compensation following loudspeaker presentations. The grey dotted line shows the predicted time *with* compensation following headphone presentations.

J.N. $F_{3,12} = 0.14$, $p = 0.933$). As shown in Fig. 6, this pattern of results was also confirmed by the performance of two additional observers in the same task.

2.3. Discussion

With loudspeaker presentation of sound, increasing viewing distance induced temporal shifts in the tuning function of the auditory–visual stream/bounce illusion that closely approximated the difference between the speeds of light and sound. This observation is not consistent with a perceptual compensation for the different speeds of light and sound.

One interesting quality of the data obtained in Experiment 1 is that it suggests that perceptual integration can occur over a broad range of temporal offsets between sight and sound (±150 ms, see Fig. 4). This observation is consistent with previous studies that have examined the perceptual integration of vision and audition (Sekuler et al., 1997; Watanabe & Shimojo, 2001). The broad range over which sights and sounds can become integrated suggests that, in this context, it may be unnecessary to correct for the small timing differences that can arise because of the difference between the speeds of light and sound within a viewing range of ∼15 m.

When sound is presented by loudspeaker, the sound pressure level is reduced as a squared function of the viewing distance. In Experiment 1, the reduction in sound pressure level between the nearest and furthest viewing distances was ∼28 dB SPL. Therefore, it could be argued that there is a perceptual compensation for the slower speed of sound that is mitigated by a reduction in sound intensity. This is unlikely as there was no variation in sound intensity following headphone presentations and no perceptual compensation was observed for this mode of presentation. Moreover, any perceptual compensation for the impact of viewing distance would have to compensate for both reductions in the intensity and the relative delay of sound to vision. Individually, either form of compensation would be insufficient. Of course it may not be possible to compensate for a reduction that renders a sound inaudible, but this did not occur in our experiment. From all viewing distances, the sound remained clearly audible and exerted a robust influence upon the perceptual experience of the normally ambiguous stream/bounce illusion (see Fig. 4).

Another possible reason why we failed to observe compensation is because our auditory stimulus did not provide accurate distance cues. However, successful compensation for physical auditory/visual timing differences has previously been observed using an auditory stimulus that did not provide any distance cues whatsoever (Sugita & Suzuki, 2003) and with an auditory stimulus that was very similar to that used in Experiment 1 (Kopinska & Harris, 2004). Therefore, our failure to observe compensation is not a necessary consequence of poor or lacking auditory distance cues—clearly successful compensation can be determined solely on the basis of visual distance cues which were readily available during our experiment (see Fig. 3). However, it is possible that the provision of auditory cues might increase the probability of compensation.

The task used in the first experiment can be regarded as an implicit measure of visual/auditory simultaneity as it does not require that the observer make any explicit judgment concerning the two sources of information. In contrast, the tasks used by Sugita and Suzuki (2003) and by Kopinska and Harris (2004) were explicit measures. The discrepancy between the results of the first experiment and those obtained by Sugita and Suzuki (2003) and by Kopinska and Harris (2004) may be explained by the discrepancy between the measures of auditory/visual simultaneity—whether they were implicit or explicit. Perhaps a perceptual compensation for the different speeds of light and sound only occurs when observers are required to make explicit judgments about two sources of information that are perceived to be causally linked, i.e., when the sound is perceived to be caused by the visual event. To explore this possibility we conducting a second experiment where observers performed an explicit two-alternative forced choice task in which observers were required to attribute an auditory event to one of two possible visual events.

3. Experiment 2: viewing distance and auditory–visual causal attributions

3.1. Methods

Six observers participated in this experiment, the first author and five others who were naïve as to the purpose of the study.

As in the first experiment, the individual dots subtended 0.35° of visual angle from each viewing distance. The moving dots oscillated back and forth at a constant velocity of 1.67°/s with a periodicity of 600 ms. In this experiment, subjects viewed two pairs of dots that became superimposed and maximally separated with a phase difference of 180° (see Fig. 7). As a consequence, when one pair of dots was superimposed, the other pair would become superimposed after 300 ms and had been superimposed 300 ms before.

During each trial, one pair of dots was black and the other white. Whether the upper or lower pair was black was randomly determined on a trial-by-trial basis. On each trial, observers were required to indicate if the tone sounded like it was being caused by the collision of the upper, or lower, pair of dots. As in the first experiment, we manipulated the time of the tone with respect to the visual cycle. The details concerning this manipulation

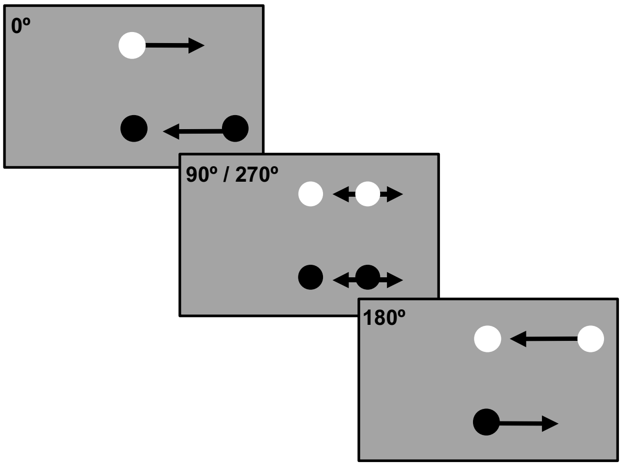

Fig. 7. Depiction of the stimulus configuration used in the second experiment. There were two pairs of vertically displaced dots. To assist the perceptual segregation of the two pairs, one pair was black and the other white. One of each pair of dots was static and horizontally positioned in the middle of the screen. The other dots moved towards, become superimposed with, and then moved away from the static dots. The direction of the displacement was randomised from trial to trial. There was a phase difference of 180° between the movements of the different dots. At a phase relationship of 0°, the tone was presented when the upper pair of dots was superimposed and the lower pair was maximally separated. At phase relationships of 90° and 270°, the tone sounded 150 ms after and therefore 450 ms before (90°), or 150 ms after and therefore 450 ms before (270°) the upper pair of dots became superimposed. The opposite pattern was true of the lower pair. At a phase relationship of 180°, the tone sounded when the upper pair was maximally displaced and the lower pair of dots was superimposed. Each trial commenced at a random point within this sequence, which repeated until the observer made a response.

were similar to those of the first experiment (see Fig. 7). Performance during a run of trials provided an estimate of the relative timing at which auditory and visual events were perceptually paired.

As in the first experiment, in different trial runs, the tone was presented over headphones (80 dB SPL) or a loudspeaker (100 dB SPL from 57 cm) placed below the visual display monitor. For each mode of sound presentation, the two observers who had participated in the first experiment completed four separate trial runs from each of the four different viewing distances, 114 cm, 513 cm, 1026 cm and 1486 cm. To establish the robustness of these findings we also tested four additional observers, who were naïve as to the purpose of the study. Each of these observers completed a trial run for each mode of sound presentation from 114 cm and 1482 cm. The experiment was conducted in a corridor under artificial illumination (see Fig. 3).

3.2. Results

Responses during each run of trials provided a distribution, as a function of phase, of the percentage of times that the tone was paired with the perceived collision of the upper pair of dots. We fitted a centroid to the distribution determined during each trial run to provide an estimate of the time at which the tone was most likely to be paired with the perceived collision of the upper pair of dots.

As shown in Fig. 8, when the tone was presented by loudspeaker, the time at which the tone was paired with the perceived collision of the upper pair of dots varied as a function of viewing distance (D.A. $F_{3,12} = 12.05$, $p = 0.001$; J.N. $F_{3,12} = 6.43$, $p = 0.008$). However, this did not occur when the tone was presented over headphones (D.A. $F_{3,12} = 0.04$, $p = 0.988$; J.N. $F_{3,12} = 0.31$, $p = 0.815$). This pattern of results is entirely consistent with those observed in the first experiment. As shown in Fig. 9, this pattern of results was also confirmed by the performance of four additional observers in the same task.

3.3. Discussion

The results of the first and second experiments suggest that the visual and auditory signals of an event that reach an observer at the same time tend to become perceptually bound, even if they could not have been generated at the same time.

The first and second experiments have focussed upon the integration of visual and auditory information and have failed to find evidence of perceptual compensation for the difference between the speeds of light and sound. However, seemingly contradictory data has been obtained in experiments that focussed upon relative timing

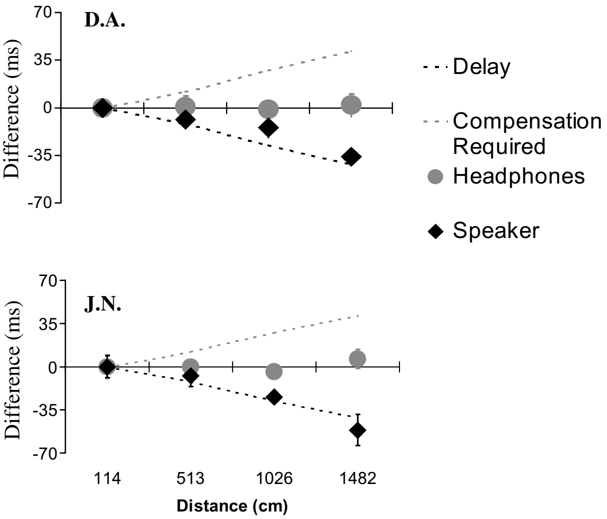

Fig. 8. Difference in temporal tuning of perceptual pairing judgments, as a function of viewing distance, for the same observers who participated in the first experiment. The details of the figure are the same as those for Fig. 3.

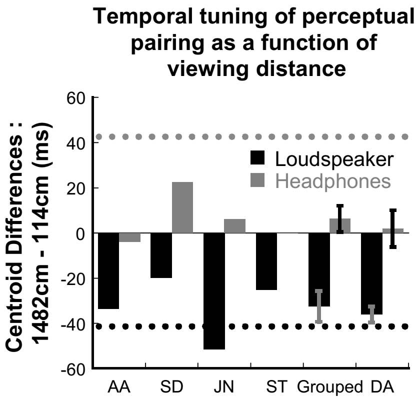

Fig. 9. The time of the tone, relative to the superposition of the upper pair of dots, that maximised perceptual pairing judgments at a viewing distance of 1482 cm minus the baseline time measured at 114 cm. Data is shown individually for four observers, who were naïve as to the purpose of the experiment. The grouped average performance of the observers is also shown. Observer D.A.'s data, contrasting the same conditions from Fig. 7, are shown for the purposes of comparison. Following loudspeaker presentations, the tone needed to be presented ~32.5 ms sooner relative to a viewing distance of 114 cm ($t_3 = -4.7$, $p = 0.018$). With headphone presentations, there was no significant difference between these times ($t_3 = 1.06$, $p = 0.367$). The dark dotted line shows the predicted time *without* compensation following loudspeaker presentations. The grey dotted line shows the predicted time *with* compensation following headphone presentations.

judgments (Kopinska & Harris, 2004; Sugita & Suzuki, 2003).

Both Sugita and Suzuki (2003) and Kopinska and Harris (2004) exposed observers to transient auditory and visual events and required them to judge their temporal order. Sugita and Suzuki (2003) used headphones for sound presentation, so sound was not physically delayed and observers were required to *imagine* that the sound was originating from the same location as the visual event. However, Kopinska and Harris (2004) used a loudspeaker for sound presentations. In both studies, as viewing distance increased, there was a shift in the relative timing at which sights and sounds were judged as being coincident. These shifts were consistent with the compensation required for the physical timing differences between sight and sound (Sugita & Suzuki, 2003; Kopinska & Harris, 2004).

While Sugita and Suzuki (2003) and Kopinska and Harris (2004) have found evidence for compensation for timing differences, both Stone et al. (2001) and Lewald and Guski (2004) have failed to find such compensation. All of these studies focussed upon relative timing. Why then are the results of the studies contradictory? We explored this controversy by adopting a similar methodology to those that have been used in previous studies—forced choice temporal order judgments between visual events and tones presented either by headphones, or loudspeaker.

4. Experiment 3: viewing distance and auditory–visual temporal order judgments

4.1. Methods

Five observers participated in this experiment, the first author and four others who were naïve as to the purpose of the study.

The stimulus was composed of a single pair of dots. These could be black or white, determined at random on a trial-by-trial basis. One of the two dots was static and positioned in the centre of the screen. At the start of each trial, the other dot was located to the left or right (determined at random) and then moved toward, became superimposed upon, and then moved away from the central static dot. This sequence lasted 600 ms. As in the first and second experiments, the individual dots subtended 0.35° of visual angle and the moving dot translated at a constant velocity of 1.67°/s.

The time at which the tone was presented during a trial was manipulated according to the method of constant stimuli. During a run of trials, the tone was presented at one of 11 temporal offsets, ranging ∓200 ms from the point at which the moving and static dot were superimposed. During a run of trials, each temporal offset was sampled on 10 occasions. Following each stimulus presentation, the observer was required to indicate if they felt that the timing of the tone was too early or too late to be consistent with the collision of the dots. A psychometric function was fitted to the data obtained during each trial run (see Fig. 10) and the 50% point of

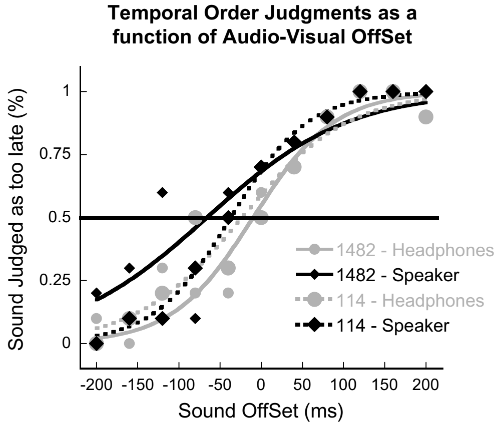

Fig. 10. Psychometric functions fitted to data obtained in four trial runs completed by observer A.K. from viewing distances of 1482 cm and 114 cm. This observer was naïve as to the purpose of the experiment. Functions fitted to data obtained from a viewing distance of 1482 cm are plotted as broken lines whereas functions fitted to data obtained from 114 cm are plotted as solid lines. The bold-black horizontal line depicts points where the observer was equally likely to report that the tone was too early or too late to be consistent with the superimposition of the two dots. Tone offset refers to amount of time by which the beginning of the tone either preceded or lagged the superimposition of the two dots.

the function was taken as an estimate of the temporal relationship at which the tone was subjectively coincident with the perceived collision.

The first author completed 16 trial runs, four for each mode of sound presentation from viewing distances of 114 cm and 1482 cm. Naïve observers completed four trial runs, one for each mode of sound presentation at each viewing distance.

4.2. Results

Fig. 11 shows the difference between the times of perceived simultaneity of tones and dot collisions from viewing distances of 1482 cm and 114 cm. Data for 114 cm were subtracted from the equivalent measure for 1482 cm, so a negative value indicates that earlier tones were perceived as being synchronous with visual events from a distance of 1482 cm as opposed to 114 cm. Error bars for D.A. show the standard error between four difference-scores. Error bars for the grouped naïve data show the standard error between the individual difference scores calculated for each observer, which are also shown.

Following headphone tone presentations, of the naïve observers, only A.K. and S.T. displayed a trend that was consistent with perceptual compensation, whereas S.D. and A.A. displayed the opposite effect. As a consequence, for these observers, the effects of viewing distance following headphone tone presentation were not significant ($t_3 = -1.08$; $p = 0.359$). However, for the grouped naïve observers, there was a significant effect of viewing distance following loudspeaker tone presentation. If tones were to be judged as coincident with the visual *collision*, they had to be presented significantly earlier (~37.69 ms, $t_3 = -3.78$; $p = 0.0324$) when the stimulus was viewed from 1482 cm as compared to 114 cm. Again there is no substantive evidence for compensation.

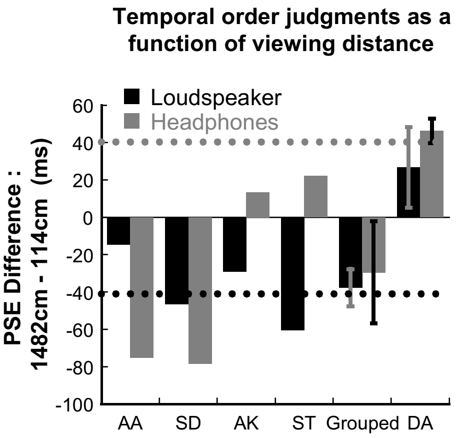

Fig. 11. The time at which a tone was judged to be perceptually coincident with the apparent collision of a pair of dots at a viewing distance of 1482 cm minus the baseline time measured at 114 cm. Data is shown individually for the first author and four additional observers who were naïve as to the purpose of the experiment. Data from the four additional observers is also shown grouped. Error bars show the standard error between four difference scores. The dark dotted line shows the predicted time *without* compensation following loudspeaker presentations. The grey dotted line shows the predicted time *with* compensation following headphone presentations.

In a final attempt to resolve the discrepancy the first author adopted an explicit strategy of trying to *imagine* that the sound and visual stimuli had originated from a common source. In this case, there was a significant difference between the timings of tone presentation that were subjectively consistent with the perceptual collision of the dots ($t_3 = 4.22$; $p = 0.002$). There were no significant effects of viewing distance following loudspeaker tone presentation ($t_3 = 1.41$; $p = 0.267$).

4.3. Discussion

If there were a perceptual compensation for physical timing differences, we might expect the point of subjective coincidence between sights and sounds to vary with viewing distance when headphones are used for sound presentation (Sugita & Suzuki, 2003), but not if a loudspeaker is used (Kopinska & Harris, 2004). The performances of the naïve observers in this experiment were inconsistent with this. This seems to contradict the findings obtained by Sugita and Suzuki (2003) and Kopinska and Harris (2004). What are we to make of these contradictory findings?

When an observer is forced to make a decision concerning a single stimulus they must adopt a response criterion. For instance, in the task that we used in the third experiment, observers were required to decide under what circumstances they would accept a tone as being too early or late. With headphone sound presentation, there was no physical delay of sound relative to vision and observers were required to *imagine* that the sound and visual stimuli had originated from a common source. From 1482 cm, the observer could decide that sound should lag the perceived collision and adopt an appropriate response bias. To show that this was possible, the first author intentionally adopted this strategy during this experiment. While this may effectively compensate for the physical timing difference between sight and sound, the origin of the compensation is probably cognitive, not perceptual.

Of course, it is also possible that observers could adopt a biased pattern of response when a loudspeaker is used for sound presentation. For instance they could decide that, because sound is physically delayed, that it is appropriate to accept delayed sounds as being concurrent with visual events. Again, to show that this was possible, the first author intentionally adopted this strategy during the experiment. However, observers may be less inclined to adopt a biased pattern of response in this context because the task does not require that the observer use their *imagination*. As the visual and auditory stimuli originate from proximate sources, observers can simply judge the relative order (Lewald & Guski, 2004) or simultaneity (Stone et al., 2001) of the two events.

It is interesting to note that, for the naïve observers, the patterns of results obtained following headphone and loudspeaker sound presentations differed. Following loudspeaker sound presentations, there was a significant influence of viewing distance upon judgments of temporal order—when viewing distance increased all observers judged earlier tones as being coincident with the visual stimulus. This is not consistent with a perceptual compensation for the different speeds of light and sound (Kopinska & Harris, 2004), but is in agreement with some previous findings (Lewald & Guski, 2004; Stone et al., 2001). However, following headphone sound presentations the pattern of results was highly variable. As viewing distance increased two observers judged earlier tones, whereas other observers judged later tones, as being coincident with the visual stimulus. We do not believe that this variability is characteristic of a perceptual process and we suggest that the variability is actually indicative of variable cognitive strategies being adopted by observers who, in this experimental condition, were instructed to imagine that a proximate sound source was originating from a distal location.

Sugita and Suzuki (2003) suggested that they had eliminated the possibility of biased patterns of response by using a two-alternative forced choice task. In this sort of task, the observer is presented with two stimuli and must decide which of the two possesses a certain quality, or signal. If there is an equal probability of the two stimuli containing the signal, the experimenter can assess the observers' ability to detect the signal in a manner that has a greatly reduced probability of being influenced by biased patterns of response. In Sugita and Suzuki (2003) experiment, observers were presented with a single stimulus and responded to the question: *did the visual event occur before or after the auditory?* This is a binary forced choice task, not a two-alternative forced choice task. Kopinska and Harris (2004) used the same sort of experimental task. This type of task is susceptible to the influence of biased patterns of response (Campion, Latto, & Smith, 1983).

Given the conflicting results obtained in the third experiment, it is tempting to give greater credence to the perceptual task that is probably less likely to encourage a cognitive bias: the temporal order judgment concerning visual events and sounds emanating from a loudspeaker. The performances of the naïve observers in this task were entirely consistent with previous results obtained by Stone et al. (2001) and by Lewald and Guski (2004). The performance of the first author was inconsistent, but this observer intentionally adopted a biased pattern of response.

The implication of the results obtained in Experiment 3 is that the origin of compensation for the difference between the speeds of light and sound is cognitive and not perceptual. According to this view, observers can make use of their knowledge of the physical properties of the world when they make judgments concerning timing and they do not necessarily rely solely upon the relative times at which different signals are perceived. Of course this strategy could only provide an accurate compensation for the difference between the speeds of light and sound if the observer were able to accurately estimate the distance of an auditory–visual stimulus. At this point, it is unclear if this estimation is necessarily based upon visual or auditory information, or a combination of both. In future, this could be clarified by manipulating the relative accuracy of the available depth cues.

Another implication of the results of Experiment 3 is that it is not sufficient to obtain evidence of compensation if we are to clarify the cause of that compensation. Any evidence of compensation could have at least two possible causes—a cognitive strategy that taps the observers' knowledge of the physical properties of the world or a process that actively re-orders your sensory experience so that appropriate sounds and sights are experienced as occurring together in time and space. Even if compensation is observed, these possibilities need to be differentiated.

4.4. Conclusions

It has been proposed that there is a perceptual compensation for the slower speed of sound relative to light (Kopinska & Harris, 2004; Sugita & Suzuki, 2003). We examined this possibility using a range of auditory–visual tasks and manipulated viewing distance to test for perceptual compensation. In all tasks, visual displays were magnified with viewing distance to keep the retinal stimulus constant and sounds were either presented by headphones or by a loudspeaker near the monitor.

Following loudspeaker sound presentations, for all tasks, we observed timing shifts with viewing distance that closely approximated the difference between the speeds of light and sound. There were no shifts following headphone sound presentations. Perceptual compensation was only observed when one of the authors tried to *imagine* that the sound was emanating from the display and adopted an intentional response bias. We were unable to find any reliable evidence of perceptual compensation. This suggests that the visual and auditory signals of an event that reach an observer at the same point in time tend to become perceptually bound, even when the sources of those signals could not have occurred together.

References

Bertenthal, B. I., Banton, T., & Bradbury, A. (1993). Directional bias in the perception of translating patterns. *Perception, 22*, 193–207.

Campion, J., Latto, R., & Smith, Y. M. (1983). Is blindsight an effect of scattered-light, spared cortex and near-threshold vision. *Behavioural and Brain Sciences, 6*, 423–447.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (1994). Recalibration of audiovisual simultaneity. *Nature Neuroscience, 7*, 773–778.

Kopinska, A., & Harris, L. R. (2004). Simultaneity constancy. *Perception, 33*, 1049–1060.

Lewald, J., & Guski, R. (2004). Auditory–visual temporal integration as a function of viewing distance: no compensation for sound-transmission time in human perception. *Neuroscience Letters, 357*, 119–122.

McGurk, H., & MacDonald, J. (1978). Hearing lips and seeing voices. *Nature, 264*, 746–748.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385*, 308.

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., et al. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society of London Series B: Biological Sciences, 268*, 31–38.

Sugita, Y., & Suzuki, Y. (2003). Audiovisual perception: Implicit estimation of sound-arrival time. *Nature, 421*, 911.

Watanabe, K., & Shimojo, S. (2001). Postcoincidence trajectory duration effects motion event perception. *Perception and Psychophysics, 63*, 16–28.