Neural latencies do not explain the auditory and audio-visual flash-lag effect

Roberto Arrighi [a], David Alais [b,*], David Burr [c]

[a] *Dipartimento di Psicologia, Università degli Studi di Firenze, Firenze, Italy*
[b] *Department of Physiology and Institute for Biomedical Research, School of Medical Science, University of Sydney,*
Anderson Stuart Building, Sydney, NSW 2006, Australia
[c] *Istituto di Neuroscienze de CNR, via Moruzzi 1, Pisa, Italy*

Abstract

A brief flash presented physically aligned with a moving stimulus is perceived to lag behind, a well studied phenomenon termed the Flash-Lag Effect (FLE). It has been recently shown that the FLE also occurs in audition, as well as cross-modally between vision and audition. The present study has two goals: to investigate the acoustic and cross-modal FLE using a random motion technique; and to investigate whether neural latencies may account for the FLE in general. The random motion technique revealed a strong cross-modal FLE for visual motion stimuli and auditory probes, but not for the other conditions. Visual and auditory latencies for stimulus appearance and for motion were measured with three techniques: integration, temporal alignment and reaction times. All three techniques showed that a brief static acoustic stimulus is perceived more rapidly than a brief static visual stimulus, while a sound source in motion is perceived more slowly than a comparable visual stimulus. While the results of these three techniques agreed closely with each other, they were exactly opposite that required to account for the FLE by neural latencies. We conclude that neural latencies do not, in general, explain the flash-lag effect. Rather, our data suggest that neural integration times are more important.
© 2005 Published by Elsevier Ltd.

Keywords: Flash-lag effect; Neural latencies; Temporal integration; Auditory motion; Cross-modal

1. Introduction

When a flash is presented adjacent to a moving stimulus the flash is perceived to lag behind (Mackay, 1958; Mateeff & Hohnsbein, 1988; Metzger, 1932; Nijhawan, 1994; for review see Krekelberg & Lappe, 2001). This has been termed the flash-lag effect (FLE). Usually the FLE has been studied with a smooth motion trajectory, either linear (Mackay, 1958; Mateeff & Hohnsbein, 1988; Metzger, 1932; Nijhawan, 1994; van Beers, Wolpert, & Haggard, 2001) or circular (Eagleman & Sejnowski, 2000a). However, Murakami (2001) recently demonstrated that it also occurs for randomly jittering stimuli, when the position of the moving stimulus at one moment cannot be used to predict its position in the next instant of time. On the basis of this study, Murakami concluded that none of the FLE hypotheses was consistent with his findings. His explanation, which is an adaptation of the differential latency hypothesis of Mateeff and Hohnsbein (1988), supposes that briefly flashed stimuli have longer latencies than moving stimuli, so the moving stimulus appears to precede the flash. To account for his data, Murakami proposed that the latency should fluctuate from trial to trial with a

probability density function approximating a Gaussian distribution.

In another recent study it has been shown that the FLE (for smooth continuous motion) occurs not only for vision but also for sound, both for "motion through the scales" (frequency sweeps) and for motion in space (Alais & Burr, 2003), as well as cross-modally. Fig. 1 summarizes these results. Interestingly, both versions of the auditory task produce much larger FLEs than the visual effect, over 150 ms compared with 20 ms for vision in comparable conditions, and an absolute maximum of 80 ms in vision (Krekelberg & Lappe, 2001; Nijhawan, 1994; Whitney, Murakami, & Cavanagh, 2000). The cross-modal version also produced a rather large FLE, roughly half the magnitude of the auditory version. Alais and Burr (2003) questioned whether these large FLEs could be accounted for by neural latencies, since latencies in audition tend to be shorter than those for vision, especially when comparing auditory tones and visual flashes. For example, in the cross-modal version where the "flash" was replaced by a brief auditory tone, a flash-lead effect would have been predicted by the latency hypothesis, but this was never observed.

In the present study, we adapted Murakami's (2001) random-motion technique to investigate the FLE with acoustic stimuli and to examine cross-modal (audio-visual) FLEs. We also apply a variety of techniques to study perceptual latencies and integration times for visual and auditory stimuli, both stationary and moving. The aim of the study is to investigate further the FLE in audition and cross-modally, and to measure the neural latencies that may explain the effect. The results show that not only do the neural latencies fail to account for these results quantitatively, they actually go in the wrong direction.

2. Methods

2.1. Subjects

Two authors, both with normal or corrected vision, served as subjects. The tasks were performed in a dimly lit room.

2.2. Stimuli and procedures

All visual stimuli were generated by a Cambridge VSG 2/3 framestore and presented on a 21 inc. color CRT monitor (Sony GDM-F500 800×600 pixels, refresh rate 100 Hz). Auditory stimuli were digitized at a rate of 65 kHz, and presented through two high quality loud speakers (Yamaha MSP5) flanking the video monitor and lying in the same plane 50 cm from the subject. Speaker separation was 90 cm and stimuli intensity 85 dB at the sound source.

For the first series of experiments the stimuli (both visual and auditory) were displayed in random apparent motion, with a static visual or auditory probe. The moving visual stimulus was a Gaussian luminance blob with a standard deviation of 1° and contrast 50%, caused to jump randomly between three positions located 20° apart on a horizontal axis. The static visual probe was a circular disc of 1° diameter at 100% contrast, flashed briefly just above the central position. The moving auditory signal was created using binaurally correlated white-noise source positioned horizontally by varying the sign and the magnitude of inter-aural temporal delays (temporal resolution 15 μs at 65 KHz digitization), with a spatial resolution of approximately 1°. Like the visual stimulus, the auditory signal also moved randomly between three positions spaced 20° apart. For a static auditory probe, we used a 400 Hz pure tone that segregated well from the moving auditory noise. All auditory signals ramped on and off with a raised cosine over 20 ms.

3. Results

3.1. Experiment 1: flash-lag effect with random motion

This experiment was designed to measure the FLE in audition and cross-modally with a random-motion technique similar to Murakami's objective method. During each trial, a moving stimulus (visual or auditory) jumped randomly between three spatial locations, a central position and 20° left or right of fixation. A trial comprised nine frames. In each frame, stimuli were presented randomly to one of the three possible positions. A brief (20 ms) static stimulus (visual or auditory) was presented on the sixth frame (that we term t_0). Subjects were required to identify the spatial position of the
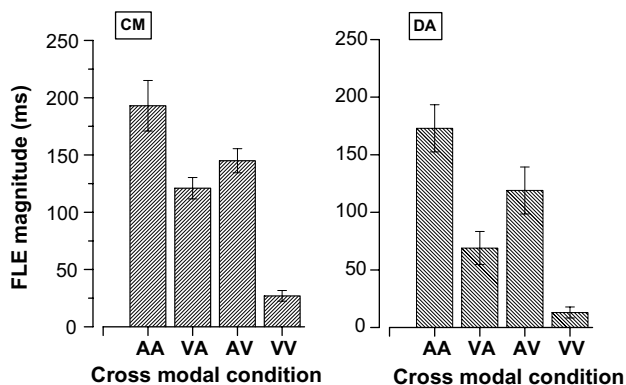

Fig. 1. Data from Alais and Burr (2003) showing the magnitude of the flash-lag effect (in ms) for auditory, visual and cross-modal combinations. The two-letter labels on the abscissa indicate respectively the modality of the brief, static, probe stimulus and the translating motion stimulus. A denotes auditory, V denotes visual.

moving stimulus at that moment. Each session comprised 80 trials, encompassing various frequencies of positional change (ranging from 1 to 20 Hz).

After each trial, the subject's response was cross-correlated with the positions assumed by the moving stimulus during the presentation sequence. Essentially, this involved checking the subject's response (e.g. "left") against the position adopted in the nine frames of that trial. Averaging over many trials results in chance-level performance (33%) for all positions except those where there was subjective synchrony between the random motion stimulus and the probe (Murakami, 2001).

Fig. 2 shows sample data taken from the condition where the motion was visual and the probe was auditory. For three frequencies of random motion, the percentage of correct identification of position is plotted against the nine frames of the random motion sequence. The continuous curves show the best Gaussian fit to the data. At the lowest frequency (2 Hz), the correlation gave chance performance (33%) at all times except for the frame where the probe actually appeared (t_0), where performance was close to perfect. This indicates that observers saw veridically the position of the randomly jumping stimulus at the moment of the probe, the pattern of results expected from a system without hysteresis or memory, whose temporal processing limitations had not been exceeded by the stimulus. The situation for the faster frequency of 4 Hz differs in two aspects: the frame producing best performance was systematically shifted forward (later) in time (as indicated by the peak

of the best-fitting Gaussian), and the width of the Gaussian broadens, indicating reduced spatial precision in locating the instantaneous position of the movement. Both of these effects were even stronger at the fastest rate of 7 Hz.

Fig. 3 shows the results of cross-correlation for eight frequencies of random motion for all conditions. Each point represents the peak of the best-fitting Gaussian of the cross-correlation data illustrated in Fig. 2. If there were no systematic difference in perceived latency of moving stimulus and probe (i.e. no FLE), the data should cluster around frame zero (the frame when the probe was actually presented). If there are perceived delays (FLEs) the cross-correlation peak will not be centered at zero, but move towards later frames. If the latency is constant, then the magnitude of the shift should vary linearly with displacement frequency. We estimated the magnitude of perceptual latency (the FLE) from the slope of the regression of frame-shift against frequency.

For the vision–vision unimodal condition (the visual flash-lag effect from random motion investigated by Murakami), the FLE estimated by this procedure was small: 10 and 11 ms (for RA and DA respectively), about the size of the fitting error and about half that measured under similar conditions for smooth motion (Alais & Burr, 2003: Fig. 1). The audition–audition unimodal condition was also insignificant, as was that for auditory motion and visual probe. However, the FLE for visual-motion auditory-probe cross-modal condition is clearly and significantly different from zero, showing that this condition did produce a cross-modal FLE. Here the estimates are 159 and 79 ms for RA and DA respectively, similar size to that obtained with smooth motion in these conditions (Fig. 1).

3.2. Integration times derived from cross-correlation data

The data from the FLE experiment (Experiment 1) also provide an estimate of temporal integration times for the localization of a randomly moving visual or acoustic stimulus. For this estimate, we took the maximum level of performance (i.e. the peak of the best-fitting Gaussian of Fig. 2) and plotted the height of this peak against the displacement frequency. In most cases, this peak coincided with the position of the random movement at moment t_0 (the only exception being those conditions that produced a flash-lag effect). Peak performance was always greater for slower rates of random motion and declined for higher rates as responses became smeared over time (see Fig. 4). To characterize the drop in peak performance, we fitted the data with Weibull functions. The time constant of the decay yielded threshold values (defined as the displacement frequency that produced 67% correct responses). With visual motion, threshold frequencies
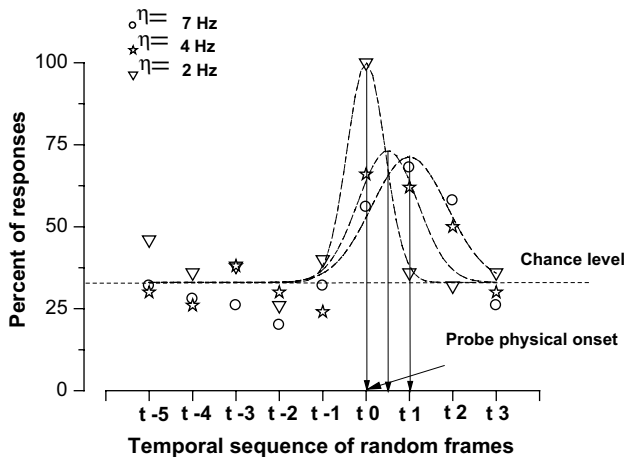

Fig. 2. Data from Experiment 1 showing the best-fitting Gaussian to the cross-correlation data for random motion at rates of 2, 4 and 7 Hz. In the situation shown, the moving stimulus was visual and the probe was acoustic. Most data points sit close to the chance level of 33% except for those near the moment of the probe (t_0). At this moment, performance increases dramatically, although, as the fitted Gaussians show, the peak in correct identification of the random motion's position shifts forward as the displacement rate increases. This forward shift of the Gaussian's peak from t_0 (the actual moment of probe) indicates a flash-lag effect (cross-modal in this case).
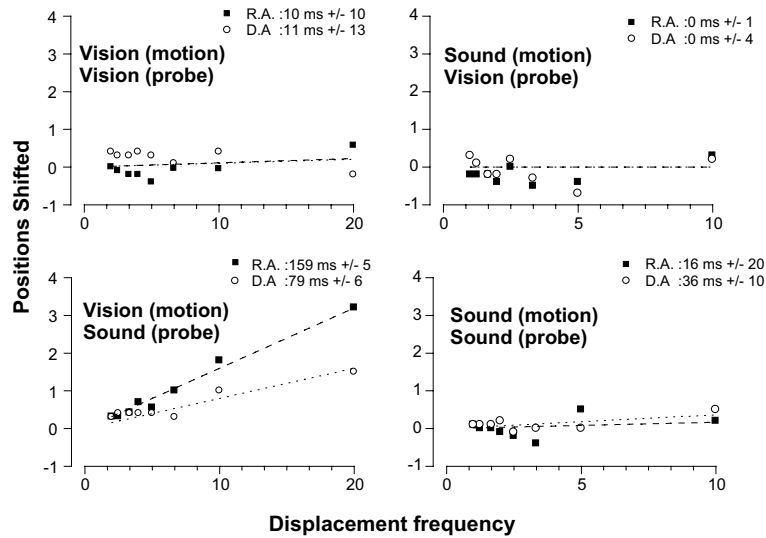
Fig. 3. Data from Experiment 1 showing the FLE of random motion for eight frequencies of displacement. The FLE is calculated by fitting a Gaussian to the cross-correlation data (see Fig. 2) and using the position of the peak to determine the magnitude of the FLE. In this figure, each point corresponds to the peak of the best-fitting Gaussian. The dashed and dotted lines show best fits to these points. The ordinate shows the number of positions the peak of the Gaussian was shifted relative to the position of the probe, where a positive value indicates a lag of the "flash". The slope parameter indicates the estimate of the constant delay corresponding to a FLE. A significant FLE was found only in the vision–sound condition (bottom-left panel), where the moving stimulus was visual and the probe was acoustic. The converse cross-modal condition and both unimodal conditions failed to elicit a significant FLE.
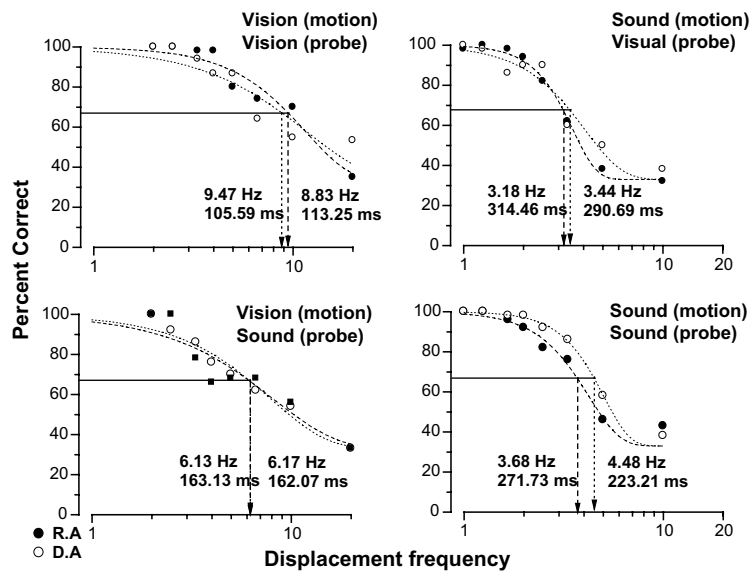

Fig. 4. Data from Experiment 1 replotted to show the level of correct performance at the peak of the best-fitting Gaussian at each of the eight frequencies of random motion. When the randomly moving stimulus is visual, the thresholds are on the order of 8 Hz. However, if the random motion is acoustic, thresholds are approximately half this frequency, around 4 Hz. The inverse of these threshold frequencies is also shown as a number, to provide estimates of temporal integration times for random motion in each modality. See also Fig. 7.

were relatively high: 8–9 Hz for vision–vision, 6 Hz for vision–sound. Auditory motion produced much lower thresholds, 3.5–4.5 Hz for sound–sound and 3–3.5 Hz for sound–vision. Temporal integration periods are given by the inverse of the frequencies, 110–160 ms for visual motion, compared with 250–300 ms for audi-tory motion. The results (also shown in Fig. 7) indicate clearly that the auditory system needs more time to integrate spatial information to produce a location estimate of a randomly moving signal than does the visual system, regardless the modality of the probe stimulus.

3.3. Experiment 2: temporal alignment of visual and auditory stimuli

To explore further the temporal analysis of visual and auditory signals, we investigated temporal alignment more directly with an adjustment technique. In the first condition we measured the temporal offset needed to perceive the onset of a static Gaussian blob and a stationary pure tone as simultaneous. The two stimuli appeared in the same spatial region, (the centre of the computer monitor) for 40 ms, and subjects were required to indicate which stimulus appeared first. The adaptive algorithm Quest (Watson & Pelli, 1983) homed in on the point of subjective alignment, so most data were collected in the most useful range. The data were later fitted with a cumulative Gaussian, whose mean estimated the point of subjective alignment.

In the second condition, the acoustic and visual stimuli oscillated in space, and the point of subjective synchrony was again measured. The stimuli started at opposite sides of the video monitor (20° from the center) and translated with the same speed (80°/s) across the screen and back to complete the cycle. The oscillations were in anti-phase to avoid the risk of "capture" of the auditory stimulus by the visual stimulus which would force a subjective alignment of the stimuli whenever they were in the same spatial neighbourhood. Sub-

jects indicated which stimulus changed its direction first. Quest varied the latencies to home in on the point of subjective alignment, but the final estimate was again obtained as the median of the best fitting cumulative Gaussian. Results of both experimental conditions are shown in Fig. 5. For brief static stimuli (filled symbols in Fig. 5), the onset of the two stimuli was perceived as simultaneous when the auditory stimulus was delayed relative to the visual stimulus by about 50 ms. However, perceptual alignment of motion reversals required the auditory stimulus to begin over 300 ms *before* the visual motion stimulus (open symbols), the exact opposite effect, and much larger.

3.4. Experiment 3: reaction times

Measurement of reaction times (RTs) is a classical technique for investigating neural latencies. Here we measured RTs to the appearance of briefly presented stimuli and to the onset of movement (visual or acoustic). The visual stimulus was a Gaussian blob and the auditory stimulus a brief 400 Hz pure tone, presented (in separate sessions) after a random blank period of 0.5–2 s to avoid predictive responses. In the motion condition the reaction times were in response to the onset of motion of visual or auditory stimulus, after it had appeared statically for a random duration (0.5–2 s). Motion onset was abrupt, at 40°/s Fig. 6 summarizes the results.

RTs were similar for both subjects. In vision the RTs for stimulus onset are close to 250 ms, approximately 30 ms longer than for motion, consistent with suggestions that latencies for moving visual stimuli are shorter than those for flashes, one of the standard explanations for the FLE (Mateeff & Hohnsbein, 1988; Whitney &
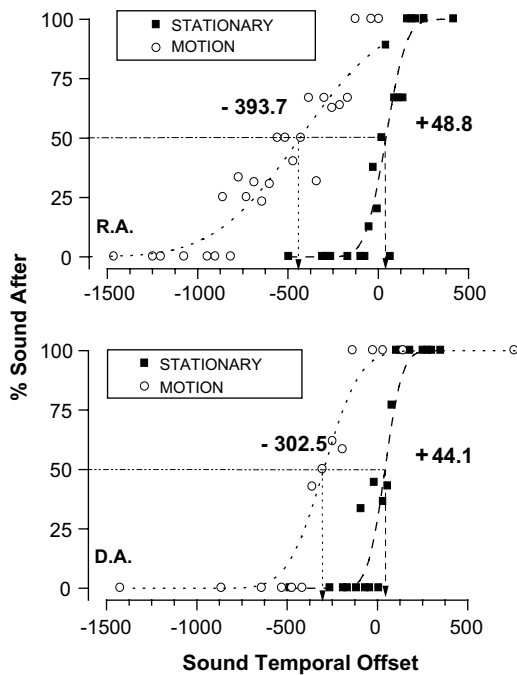

Fig. 5. Data from Experiment 2 showing temporal alignment results. The abscissa shows the temporal offset of the sound relative to the visual stimulus (0 ms indicates physically aligned stimuli) while the ordinate shows the percentage of times that the subject responded that the sound onset (or change in direction) occurred after the visual stimulus. Cumulative Gaussians were fit to the data to yield the point of subjective alignment (median of Gaussian).
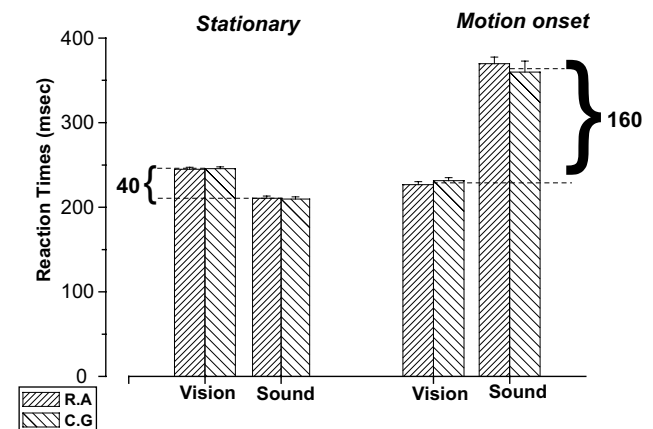

Fig. 6. Data from Experiment 3 showing reaction times for brief static stimuli (left-hand columns) and for motion-onset (right-hand columns) in vision and audition. The time needed to perceive the onset of brief static stimuli is about 40 ms longer for vision than audition. However, the situation is clearly reversed for motion onset, with RTs 160 ms longer in audition than vision.

Murakami, 1998). In audition, however, the situation is reversed, with RTs for motion onset close to 380 ms while for static stimuli they are around 215 ms. Latencies for motion-onset of translating acoustic stimuli are almost twice as long as those for stationary acoustic stimuli. Thus the simple latency argument that has been advanced to explain the visual FLE, consistent with the visual RTs reported here, certainly cannot explain the auditory FLE (Alais & Burr, 2003), and therefore the FLE in general.

3.5. Summary of neural latencies and FLE

The three techniques for measuring temporal latencies produced similar results: brief, static, acoustic stimuli are perceived more quickly than visual flashes, but motion onset for auditory stimuli is perceived more slowly than for visual stimuli.

Fig. 7 attempts to compare directly the latency estimates of the three techniques. One difficulty in plotting these data is that not all techniques give an estimate of absolute latency. The alignment technique estimates only the relative asynchrony of visual and auditory stimuli, whereas the reaction times include non-perceptual latencies, such as decision and motor latencies. However, as we are interested only in relative latencies, difference between audition and vision, and motion-onset and flash, this limitation is relatively unimportant. In order to make the comparison, we chose (somewhat arbitrarily) to anchor all the results to the estimates of visual motion integration (109 ms), the most absolute measure we have (in the sense that it is not contaminated by decision and motor latencies). All latencies for visual motion onset were considered to be 109 ms, and the other values adjusted accordingly to maintain latency differences. The alignment data for flash onset required further anchoring, as this was never related to motion alignment. The anchoring was achieved by making the visual flash latency equal to the adjusted visual reaction time to flash. However, it should be stressed that the anchoring was performed only to aid comparison between all the various conditions, and should not be interpreted too literally.

The normalized data are shown in Fig. 7. While the absolute values of each condition are arbitrary, the differences in the various conditions are real, and consistent between conditions. In the two paradigms where it was measured, latencies to flash onsets were shorter for audition than for vision, by about 40 ms. In all three paradigms, latencies for motion onsets were much longer in audition than in vision, by 100–300 ms. Where measured (only for RTs), latency for visual motion is less than for a visual flash (consistent with the FLE), but auditory latencies are the other way round, predicting a reverse effect (contrary to the actual evidence: Alais & Burr, 2003).

The right-hand cluster of bars shows a set of normalised latencies that would explain the visual, auditory and cross-modal FLEs reported in Fig. 1, if the FLE were to be explained by perceptual latencies. Given the order of the magnitudes of the FLE in various conditions,
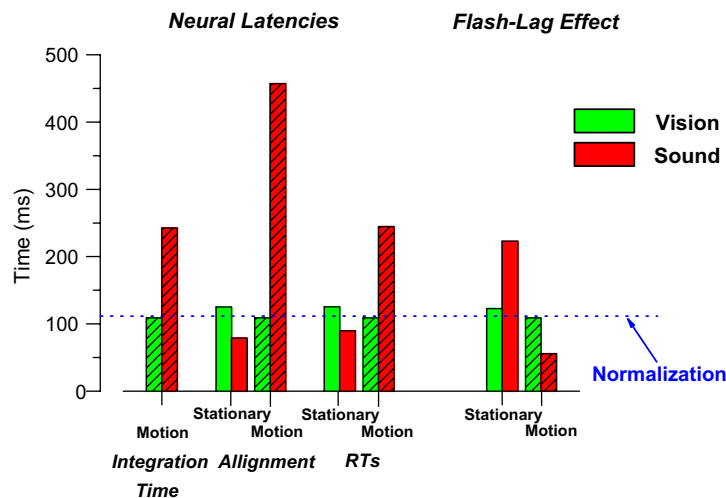

Fig. 7. *On the left:* summary of latency results of this study (averaged over subjects). The absolute levels of all time estimates have been displaced so the visual motion latency corresponds with that estimated by integration (Experiment 1), shown by the dashed line. The alignment data required for flash onset required further anchoring (as this was never related to motion alignment), achieved by making the visual flash latency equal to the adjusted visual reaction time to flash. The data from all three experiments are consistent in showing shorter latencies for acoustic bursts than for visual flashes, but much longer latencies for acoustic motion than visual motion. *On the right:* values of visual and acoustic latency that best predict the data (averaged for subjects) shown in Fig. 1 (visual flash 122 ms, acoustic burst 222 ms, visual motion 109 ms, acoustic motion 55 ms). Again these were anchored at the normalization point used for the other data, as FLE predicts only differences in latencies, not absolute levels. The reader can verify that subtraction of appropriate pairs of values predicts well the data of Fig. 1. However, these predicted latencies are completely different from those obtained by direct measurement (see bars on the left).

auditory motion needs to be fastest, followed by visual flash and auditory motion, and auditory flash the slowest. If the FLE results from latencies, this is the only way for the auditory–auditory FLE to be largest and visual–visual to be smallest, with the others in between (by performing the subtraction of the values given in the figure caption the reader can verify that these values predict quite accurately those reported in Fig. 1). The only problem with this explanation is that the pattern of latencies required is *exactly opposite* to that obtained by direct measurement (left-hand clusters on Fig. 7). Thus it would seem that latencies do not, in general, explain the flash-lag effect.

4. Discussion

This study had two goals: to investigate the acoustic and cross-modal flash-lag effect using the objective technique of Murakami (2001); and to measure the supposed neural latencies by objective techniques—temporal integration, temporal alignment and reaction times—to see if these may explain the FLE.

In the first part of the study, we adapted Murakami's random-motion technique to investigate the FLE. We found a very clear effect in one cross-modal condition, visual motion and a static auditory probe. The slopes of the shift against frequency functions were gave estimates of FLEs of 159 ms for RA and 79 ms for DA. Interestingly, those of DA were very similar to the FLE estimates in this condition using a subjective measure (69 ms, see Fig. 1). This confirms that the technique can be effective in measuring FLEs, and that the cross-modal FLE, previously reported by subjective techniques, is real and large.

We also found a flash-lag effect for the visual–visual condition, but the effect was small—about 10 ms—similar to the error estimate, and therefore not really significant. However, although small, this estimate is in line with Alais and Burr's (2003) data, where subject DA had an effect of 13 ms. So although we do not want to claim a strong effect here, the results are not inconsistent with the previous estimates. We presume the lower precision here, compared with Murikami's original study, was due to using fewer spatial positions, more widely separated. Although this arrangement was clearly not optimal for good visual motion, it was chosen in order to maximize the possibility of auditory motion between clearly resolvable positions.

Although the study was designed to maximize the possibility of auditory motion, with clearly separated positions, each separated by 20° (whereas auditory localization thresholds under these conditions are about 8°: Alais & Burr, 2004), we failed to find a FLE for auditory motion with this technique. In contrast, Alais and Burr (2003) showed a very strong FLE for continuous and smoothly translating sound sources (see Fig. 1). Presumably the failure to obtain an effect is due to the ineffectiveness of randomly jumping sounds to elicit a strong sense of motion, required for the FLE. As auditory motion is not given directly, but calculated from inter-aural timing differences, it is perhaps not surprising that the sense of auditory motion is less robust than that of visual motion.

The second part of this paper tested whether perceptual latencies may account for the FLE under these conditions, using a variety of techniques to examine visual and auditory latencies. Neural latencies refer not only to time taken in neural transmission, but also comprise delays due to sensory transduction, temporal integration and computation. Reaction times (RTs) are one of the oldest techniques for investigating perceptual latencies and their purported neural substrate. With static stimuli our results (Fig. 6) confirm the long-known fact that RTs for brief auditory stimuli are typically 30–50 ms shorter than those for vision (Brebner & Welford, 1980; Galton, 1899; Welford, 1980). Two factors are likely to contribute to this difference: the photo-transduction process in the retina takes about 20 ms at photopic light levels (Lamb & Pugh, 1992), whereas transduction of acoustic waves in the ear occurs almost instantly via a direct mechanical linkage; and the distance from the retina to primary visual cortex is longer than that from the cochlear to primary auditory cortex and therefore causes longer transmission times.

The comparison of RTs for the onset of visual and auditory motion is more interesting (Fig. 6). RTs for onset of visual motion were approximately 220 ms (about 30 ms shorter than those for brief visual probes), compared with 380 ms for the onset of auditory spatial motion. Latencies for visual motion onset are far shorter than those for the auditory motion onset. This very large difference may be related to the fact that space in audition is not a first-order property. The cochlear encodes frequency and intensity. Spatial information must be computed later from a comparison of timing and level differences between the ears to compute azimuth, and of spectral information if elevation is involved (Butler & Humanski, 1992; Carlile, 1996). These extra steps involved in obtaining spatial information in audition may contribute to the slower RTs we report here. Another factor is that the visual system has rapid, mylenated pathways leading to areas specialized for motion processing (Albright, 1984; Britten, Shadlen, Newsome, & Movshon, 1992) whereas such specialization in audition has not been found. Whatever may underlie these latency differences, it is clear that they cannot support a latency account of the FLE.

We also measured temporal synchrony directly (Experiment 2) by perceptual alignment of auditory and visual stimuli, either stationary or moving. The onset asynchronies required to perceptually align the stim-

uli are in close agreement with the data from the RT experiment. Perceptual alignment of the static probes required that the visual stimulus be presented approximately 50 ms earlier than the auditory stimulus, while in stark contrast, perceptual alignment of motion reversals required the auditory stimulus to begin over 300 ms before the visual motion stimulus. Qualitatively, the latencies implied by the alignment technique agree with those of the reaction time technique (Fig. 7).

The data from temporal integration times (Experiment 1) for the localization of a randomly moving visual or acoustic stimulus also agree with this pattern. Average integration times for visual motion are approximately 100 ms (Fig. 4), a figure similar to other estimates of temporal integration of visual movement (Snowden & Braddick, 1991). Integration times for auditory motion, however, were much longer, averaging about 250 ms (Fig. 4). We are not aware of any other studies that have estimated integration times specifically for ITD-defined spatial auditory movement, but the figure seems quite consistent with the figures for auditory movement we derived in the RT and alignment studies. These results demonstrate that the auditory system needs more time to integrate spatial information to derive a location estimate of a motion signal than does the visual system. As noted above, long auditory integration needed for movement may explain why no FLE was obtained in either of the auditory motion conditions in Experiment 1.

While each of the techniques used to measure neural latencies may be prone to criticism, the fact that all three techniques converged on similar estimates gives us confidence that they are measuring something similar. Although the three sets of perceptual latency measurements are consistent with each other, they are not consistent with the latencies required to explain the results of Alais and Burr's (2003) auditory and cross-modal FLE. Fig. 7 shows the visual and acoustic latencies that would be necessary to explain their FLE data: acoustic motion would need to have the shortest latency, followed by visual motion, visual flash and acoustic burst. However, in our present experiments, three different techniques for measuring latencies indicate a different order: acoustic burst, visual motion, visual flash and acoustic motion. These estimates of perceptual latency strongly suggest that the latency account of the FLE cannot be correct in general. Of course these results cannot rule out the possibility that latencies contribute to the visual flash-lag effect, but the fact that they do not constitute a general explain for the FLE makes this less likely.

5. Conclusion

This paper confirms a previous suggestion (Alais & Burr, 2003) that differential latencies were unlikely to explain auditory and cross-modal FLEs. In that paper, it was noted that there remained three theories which could be adapted to account for auditory and cross-modal FLEs, namely postdiction (Eagleman & Sejnowski, 2000b), temporal averaging (Krekelberg & Lappe, 2000) and positional sampling (Brenner & Smeets, 2000). All of these theories depend critically upon temporal integration times. The integration time data we obtained (Fig. 4) were collected partly to assess these theories. We found that integration times in the auditory probe/auditory motion condition were far longer than in the vision probe/vision motion condition, and an intermediate integration time was found for the auditory probe/visual motion condition. These three findings agree with the order of FLE magnitudes reported by Alais and Burr for the same conditions.

The only condition which posed a problem for an analysis in terms of integration times is the visual probe/auditory motion condition. We found this condition yielded the longest integration time of all, and would therefore be expected to produce the largest FLE, while this combination actually produced the second shortest FLE (Alais & Burr, 2003). One possible explanation for this exception is that this condition may have been influenced by the ventriloquism effect—visual capture of auditory by visual stimuli, even when moving (Soto-Faraco, Lyons, Gazzaniga, Spence, & Kingstone, 2002). There may have been a tendency in Alais and Burr's visual probe/auditory motion condition for the location of the translating auditory motion to be attracted to the location of the visual flash, thereby reducing the FLE. By contrast, in our integration time experiment, the wide spacing of the random jump locations (20°) very likely exceeded the spatial limits within which ventriloquism can occur, leaving our estimates of temporal integration unattenuated. This would explain the anomaly between the integration time data in Fig. 4 and the order of FLE effects found by Alais and Burr. It would also support the viability of the three theories based on integration times as plausible general accounts of the FLE within and between modalities.

References

Alais, D., & Burr, D. (2003). The "flash-lag" effect occurs in audition and cross-modally. *Current Biology, 13*, 59–63.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*(3), 257–262.

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology, 52*, 1106–1113.

Brebner, J. T., & Welford, A. T. (1980). Introduction: an historical background sketch. In A. T. Welford (Ed.), *Reaction Times* (pp. 1–23). New york: Academic Press.

Brenner, E., & Smeets, J. B. (2000). Motion extrapolation is not responsible for the flash-lag effect. *Vision Research, 40*(13), 1645–1648.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience, 12*(12), 4745–4765.

Butler, R. L., & Humanski, R. A. (1992). Localization of sounds in the vertical plane with and without high-frequency spectral cues. *Perception Psychophysics, 51*, 182–186.

Carlile, S. (1996). The physical and the psychophysical basis of sound. In S. Carlie (Ed.), *Virtual auditory space: generation and applications*. TX: Lindens Austin, Ch 2.

Eagleman, D. M., & Sejnowski, T. J. (2000a). Motion integration and postdiction in visual awareness. *Science, 287*(5460), 2036–2038.

Eagleman, D. M., & Sejnowski, T. J. (2000b). Flash-lag effect: differential latency, not postdiction. Response. *Science*, 290,1051a.

Galton, F. (1899). On instruments for (1) testing perception of differences of tint and for (2) determining reaction time. *Journal of the Anthropological Institute, 19*, 27–29.

Krekelberg, B., & Lappe, M. (2000). A model of the perceived relative positions of moving objects based upon a slow averaging process. *Vision Research, 40*(2), 201–215.

Krekelberg, B., & Lappe, M. (2001). Neuronal latencies and the position of moving objects. *Trends Neuroscience, 24*(6), 335–339.

Lamb, T. D., & Pugh, E. N. Jr., (1992). A quantitative account of the activation steps involved in phototransduction in amphibian photoreceptors. *Journal of Physiology, 449*, 719–758.

Mackay, D. M. (1958). Perceptual stability of a stroboscopically lit visual field containing self-luminous objects. *Nature, 181*(4607), 507–508.

Mateeff, S., & Hohnsbein, J. (1988). Perceptual latencies are shorter for motion towards the fovea than for motion away. *Vision Research, 28*(6), 711–719.

Metzger, W. (1932). Versuch einer gemeinsamen Theorie der Phänomene Fröhlichs und Hazelhoffs und Kritik ihrer Verfahren zur Messung der Empfindungszeit. *Psychologische Forschung, 16*, 176–200.

Murakami, I. (2001). A flash-lag effect in random motion. *Vision Research, 41*(24), 3101–3119.

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature, 370*(6487), 256–257.

Snowden, R. J., & Braddick, O. J. (1991). The temporal integration and resolution of velocity signals. *Vision Research, 31*(5), 907–914.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. *Brain Research Cognition Brain Research, 14*(1), 139–146.

van Beers, R. J., Wolpert, D. M., & Haggard, P. (2001). Sensorimotor integration compensates for visual localization errors during smooth pursuit eye movements. *Journal of Neurophysiology, 85*(5), 1914–1922.

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception Psychophysics, 33*(2), 113–120.

Welford, A. T. (1980). Choice reaction time: basic concepts. In A. T. Welford (Ed.), *Reaction Times* (pp. 73–128). New York: Academic Press.

Whitney, D., & Murakami, I. (1998). Latency difference, not spatial extrapolation. *Nature Neuroscience, 1*(8), 656–657.

Whitney, D., Murakami, I., & Cavanagh, P. (2000). Illusory spatial offset of a flash relative to a moving stimulus is caused by differential latencies for moving and flashed stimuli. *Vision Research, 40*(2), 137–149.