Current Biology, Vol. 13, 1–20, January 8, 2003, ©2003 Elsevier Science Ltd. All rights reserved. PII S0960-9822(02)01402-1

The "Flash-Lag" Effect Occurs in Audition and Cross-Modally

David Alais[1,*] and David Burr[1]
Istituto di Neurofisiologia del Consiglio Nazionale
delle Ricerche
Via G. Moruzzi 1
Pisa 56125
Italy

Summary

In 1958 MacKay [1] showed that a rigidly moving object becomes visually fragmented when part of it is continuously visible but the rest is illuminated intermittently. For example, the glowing tip of a lit cigarette moving under stroboscopic illumination appeared to move ahead of the intermittently lit body. Latterly rediscovered as "the flash-lag effect" (FLE) [2], this illusion now is typically demonstrated on a computer monitor showing two spots of light, one translating across the screen and another briefly flashed in vertical alignment with it. Despite being physically aligned, the brief flash is seen to lag behind the moving spot. This effect has recently motivated much fruitful research, prompting a variety of potential explanations, including those based on motion extrapolation [2, 3], differential latency [4, 5], attention [6], postdiction [7], and temporal integration [8] (for review, see [9]). With no consensus on which theory is most plausible, we have broadened the scope of enquiry to include audition and have found that the FLE is not confined to vision. Whether the auditory motion stimulus is a frequency sweep or a translating sound source, briefly presented auditory stimuli lag behind auditory movement. In addition, when we used spatial motion, we found that the FLE can occur cross-modally. Together, these findings challenge several FLE theories and point to a discrepancy between internal brain timing and external stimulus timing.

Results

Brief Tones Lag behind Auditory Spectral Motion

We first studied the FLE in audition by using spectral motion, movement produced by sweeping through the auditory frequency spectrum. Using headphones, observers heard a tone of 1 s duration in one ear. This tone swept over a 2 octave frequency range (from 1 to 4 kHz) in a log-linear frequency glide. In the middle of the frequency sweep, a brief 40 ms tone burst was played to the other ear. Initially, the brief tone was set to 2 kHz, and observers were required to judge whether it was higher or lower than the sweeping frequency at that moment. Based on the observers' responses, an adaptive staircase procedure (Quest [10]) was used for adjusting the burst frequency to home in on subjective equality—the point at which the instantaneous swept frequency and the brief tone are perceived to have the same pitch. As occurs with the visual FLE, subjective alignment of the brief stimulus (tone burst) and the moving stimulus (frequency glide) did not coincide with physical equality. Rather, the brief tone lagged behind the spectral movement and needed to be advanced in the direction of motion in order to be perceived as equal in pitch with the instantaneous sweep frequency (Figure 1A). Data for two subjects (Figure 1B) show that for upward sweeps, the brief tone needed an increment of about 0.35 octaves (approximately 4 semitones) to achieve perceptual alignment, and for downward sweeps, a decrement of 0.27 octaves (approximately 3 semitones) was required. In both cases, then, the "static" burst lagged behind the spectral movement. The experiment was repeated for various sweep speeds, and these conditions revealed that lag magnitude is proportional to the speed of the frequency sweep, with smaller offsets required to match the tone with the sweeping frequency as the sweep speed slows. Importantly, subjects were able to match accurately the dichotic tones when the sweep speed was reduced to zero. The linear fits in Figure 1B show the slope of the speed dependency and indicate a constant temporal differential of about 150–180 ms between the brief and the moving stimuli (Figure 1B). Temporal differentials are often used for quantifying the FLE in vision, although the reported temporal lags are generally much shorter (80 ms or less) [2–9].

Spatial Location Lags behind Spatial Movement, Cross-Modally and Unimodally

In order to relate the auditory "flash-lag" phenomenon more closely to previous visual studies, we also investigated whether the FLE would occur for auditory movement over space (as in visual motion). Using loudspeakers, we moved a low-pass filtered white-noise source smoothly from left to right through an azimuthal angle of approximately $-20°$ to $+20°$. A 20 ms sound burst (1 kHz pure tone) occurred at the temporal mid-point of the motion interval and was initially spatially located at $0°$ azimuth. Observers were required to judge whether the brief tone was located ahead of or behind the motion at the moment it occurred. We varied the actual location of the tone (determined by interaural time difference) from one trial to the next according to a Quest procedure to find the point of subjective alignment with the movement. Analogous to the visual flash-lag effect, the location of the brief tone lagged behind the moving sound source and required a large spatial advance in the direction of motion to be subjectively aligned. This result is shown in Figure 2 for two observers (60 degree/s condition). Again, we manipulated the speed of movement (by varying azimuthal angle) and found that the magnitude of the corrective advance was speed dependent in a linear fashion, with the data suggesting a constant temporal lag of about 200 ms.

*Correspondence: alaisd@physiol.usyd.edu.au
[1]Present address: Department of Physiology, The University Of Sydney, New South Wales 2006, Australia.

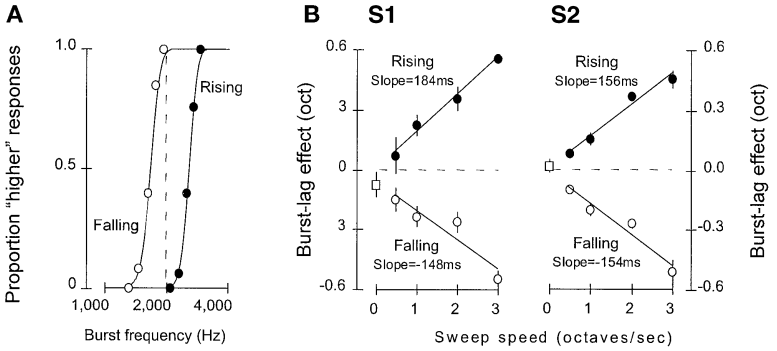

Figure 1. Data Demonstrating that the Visual 'Flash-Lag' Effect Occurs in Audition, with Frequency Glides Used as Motion and Brief Tones as the 'Flash'

(A) In each condition, subjects completed at least three adaptive staircases [10] in order to estimate the point at which the auditory motion and the brief tone were subjectively aligned (in frequency [Figure 1] or in spatial position [Figures 2 and 3]). The data sets were pooled, and psychometric functions were fitted. The example shown represents data from one observer in Experiment 1 showing the probability of perceiving a tone burst higher than the frequency sweep at that instant for sweeps rising and falling at 2 octaves/s. The true point of alignment is 2 kHz (dashed line). Data were fitted with a cumulative Gaussian function whose half-height estimates the point of subjective alignment. In this case, subjective alignment required an increase in the pitch of the tone burst for the rising sweep and a decrease in pitch for the falling sweep. For this listener, the effect size averaged about 0.3 octaves (3–4 semitones). This effect is analogous to the visual flash-lag effect in that the brief tone burst required an advance in the direction of motion to be perceived as aligned.

(B) Data from two subjects showing that the magnitude of this "burst-lag effect" increases with sweep speed. Only the estimated points of subjective alignment taken from the psychometric functions are plotted, together with standard deviation error bars calculated from 500 iterations of a bootstrap procedure [20]. The misalignment increases linearly with sweep speed, suggesting a constant temporal mismatch of about 150–180 ms (slope of linear fits). The square symbol shows that the match to a stationary tone of 2 kHz is near veridical.

Capitalizing on the fact that this version of the auditory FLE and the standard visual FLE both operate over space, we combined both to produce cross-modal conditions. In separate conditions, either spatial auditory motion was paired with a briefly flashed white disc, or a translating white disc was paired with a static tone burst. In both cross-modal combinations, the observer's task was simply to judge whether the spatial position of the briefly presented stimulus was ahead of or behind the spatially translating stimulus. Here again, for both conditions, we found that the location of the brief "flashed" stimulus lagged behind the moving stimulus and needed a spatial advance to be perceptually aligned. As can be seen in Figure 3, the implied temporal lags for both the VA condition (visual flash, auditory motion) and the AV condition (auditory flash, visual motion) are smaller than what is observed for the unimodal auditory version. Both cross-modal conditions, however, produced much larger lags than that obtained with a unimodal visual FLE.

Finally, to test whether the auditory FLE can be described by a motion-extrapolation account [2], we tested half-trajectory conditions: flash-initiated trajectories and flash-terminated trajectories. These have been used previously to show that the visual FLE does not depend on motion information prior to the flash [7]; flash-initiated trajectories (beginning in alignment with the flash at the moment it appears) yield a FLE identical in magnitude to the full trajectory. In contrast, flash-terminated trajectories (ending in alignment with the flash at the moment it appears) yield no FLE at all. We tested the three types of trajectory for all four of the conditions shown in Figure

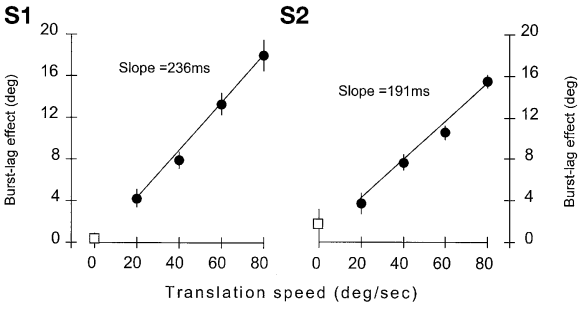

Figure 2. Data Demonstrating that the Auditory 'Flash-Lag' Effect Also Occurs When Spatial Motion Is Used Instead of Spectral Motion

Data from two subjects was used for plotting the speed dependency of the burst-lag effect. Observers consistently heard a brief tone burst as located behind the instantaneous position of a translating low-pass-noise signal. As observed for spectral motion (Figure 1), the magnitude of the burst lag was speed dependent in a linear fashion, with slopes on the order of 200 ms. Again, the linearity of the speed effect suggests a constant temporal mismatch in the perceptual alignment of "flashed" and translating stimuli. The slopes for spatial motion are slightly higher than for spectral motion but are nonetheless comparable and much higher than the values of 40–80 ms typically obtained in vision [2–9].

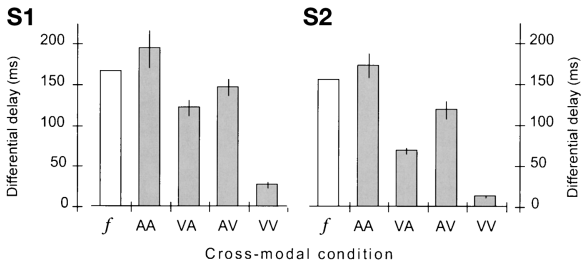

Figure 3. The Flash-Lag Effect Occurs Cross-Modally

Cross-modal versions of the flash-lag stimulus were created by replacing either the translating noise source with a translating white disc or the tone burst with a briefly flashed white disc. All spatial and temporal parameters were matched across the modalities. Data for two subjects are shown (gray columns). Temporal misalignment for the spatial auditory FLE (AA), the visual FLE (VV), and for the two cross-modal conditions is observed. The first letter of the column labels indicates the modality of the flash/burst (**A**uditory or **V**isual), the second that of the translating stimulus. Translation speed for these data was 60 degrees/s. For comparison, averaged data from the spectral auditory FLE conditions are shown (open column).

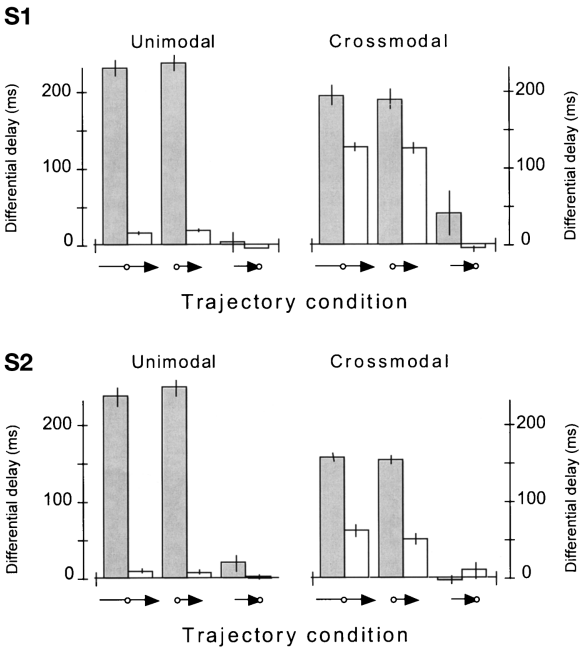

Figure 4. Motion Prior to the 'Flash' Does Not Affect the Flash-Lag Effect

Both unimodal and both cross-modal versions of the flash-lag stimulus were tested with three different motion trajectories, illustrated beneath the abscissae. The first was a full-trajectory condition (left pair of columns), with the flash occurring in the middle of the trajectory. The other two were half-trajectory conditions, either flash-initiated (middle pair of columns) or flash-terminated (right pair of columns). Whether unimodal or cross-modal, an identical pattern of results was obtained: FLEs of equal magnitude for full and for flash-initiated trajectories; no FLE for flash-terminated trajectories. These data rule out a motion-extrapolation account of the FLE in that motion prior to the "flash" does not elicit a flash-lag effect. This result has been previously noted for the visual FLE and is here extended to auditory and cross-modal versions of the FLE.

3. The pattern of results across the three trajectories (Figure 4) matches results obtained with the visual FLE: equal FLEs for full and flash-initiated trajectories and no FLE for flash-terminated trajectories. This pattern held for all four flash-lag stimuli, whether unimodal or cross-modal. As has been found for vision, then, motion prior to the "burst" does not influence the auditory or the cross-modal FLE.

Discussion

Taken together, these findings have important implications for existing theories of the FLE. The cross-modal data (Figure 3) are particularly relevant in that they suggest that the differential latency account [4, 5] cannot be correct. Regarding this hypothesis, the FLE arises because visual latencies for motion are shorter than those for flashes. However, because latencies for audition are shorter than for any visual stimuli (as traditionally measured by reaction times [11] and evoked potentials [12]), this hypothesis would predict a flash-lead effect in the auditory-flash/visual-motion condition since the flash would be perceived first. However, the opposite was observed.

Moreover, from the pattern of differences between conditions (Figure 3), the latency model would dictate the order of latencies for the four stimulus elements as being (from shortest to longest): auditory motion, visual motion, visual flash, and auditory flash. This is because the stimulus with the longest latency is the auditory burst, which lags both visual and auditory motion. The stimulus with the shortest latency, however, would need to be auditory motion because it leads both auditory and visual flashes. If we assign a hypothetical latency of x ms to auditory motion (we cannot know its absolute latency), then latencies for visual motion and visual flashes would have to be (if one works from subject 1 in Figure 3) $x + 56$ and $x + 69$ ms, respectively, and auditory bursts would be $x + 169$ ms. Although these latencies are not impossible, it does seem highly implausible that the auditory system's poor sensitivity to spatial movement should have the fastest latency and that latencies for auditory tones should be slowest. Furthermore, it is odd that latencies for auditory movement should be so much faster than those for visual movement, for which our perceptual system is highly specialized and exquisitely sensitive.

We can also preclude any attentional explanations of the cross-modal data in Figure 3 (AV and VA); for example, we can preclude explanations based on the phenomenon of prior entry or on the modality-shifting effect. Prior entry is the idea that an attended stimulus is perceived to occur earlier than an unattended stimulus (an effect that has been shown to occur cross-modally [17]). The modality-shifting effect [18] refers to the lengthy time period required to shift attention from one modality to another. Because of an asymmetry in the cross-modal conditions, however, neither of these attentional effects could provide a full explanation of the cross-modal data. In the case of visual motion, the FLE increases if the "flash" is extra-modal (c.f., VV and AV), whereas for auditory motion, the FLE decreases if the "flash" is extra-modal (c.f., VA and AA). Modality shifting might explain why the FLE magnitude is larger in the cross-modal AV condition than in the unimodal VV condition, but it cannot explain why FLE magnitude should be *smaller* in the cross-modal VA condition than in the unimodal AA condition. Similarly, the prior-entry hypothesis could explain one result (AV $>$ VV), but not both.

Even if not providing a full account, is it likely that cross-modal attentional effects at least exert an influence on the cross-modal data? There are two reasons to discount this possibility. The first comes from the half-trajectory data. Because both attentional accounts depend on an observer's attention being directed at a pre-existing stimulus in one modality (in this case, the translating motion prior to the flash/burst) before a subsequent extra-modal stimulus (the flash or burst) requires it to be redirected to another modality, the flash-initiated condition provides a crucial test. Here, both stimuli are presented simultaneously, precluding prior attentional allocation to a single modality, and yet the FLE still occurs and does so with a magnitude not statistically different from the full-trajectory condition. A second relevant point is that the preexisting stimulus is not really monopolizing attention in the first place. It would be more accurate to describe the cross-modal condi-

tions as divided-attention tasks because the conditions were tested in a blocked design, meaning subjects knew in advance to divide their attention between vision and audition.

Three theories of the FLE still cannot be dismissed at present on the basis of these data. The temporal averaging [8], postdiction [7], and positional sampling [16] models, with appropriate amendments to extend them into the auditory domain, might be able to offer plausible accounts of the auditory and cross-modal data. The temporal averaging model builds on the principle that neurons integrate input over a brief period before firing, with their output reflecting the average over this period. In this model, the two main parameters determining the flash-lag are the motion-integration period and the duration during which the flash's position signal persists. The flash lag is the time-averaged positional difference between the two. Because integration periods in audition have been reported to be longer (up to several hundred milliseconds [13]) than in vision, they could easily offset the transmission latencies that see auditory signals reach the cortex 20–30 ms before visual signals [14, 15]. Thus, the model could readily predict the large AA effect by positing longer integration periods in audition for brief and for translating stimuli. It could also predict the lesser effect observed in the cross-modal conditions, for example the VA<AA effect. Because both of these conditions share the same motion component, the only assumption required is shorter integration times for flashes than for brief tones (equally, assuming shorter visible persistence than echoic persistence would also explain the result). The same reasoning could explain the AV>VV effect and the small VV effect.

The other remaining models—postdiction [7] and positional sampling [16]—could also be adapted to account for the auditory FLE by incorporating similar assumptions. The postdiction model claims that the FLE arises because the flash serves to reset the process of motion integration. If the flashed and moving stimuli are physically aligned when this process restarts, the first available position for the moving stimulus will inevitably be in advance of the flash, by an amount proportional to the temporal integration period. If the integration period for a brief tone were longer than for a visual flash, the moving stimulus would be further advanced along its trajectory by the time the "reset" signal was generated. This would explain the AA>VA and AV>VV effects. In a similar vein, the positional sampling model proposes that instantaneous position information is not available and that the flash therefore serves to mark the moment when the moving object's position should be sampled. Again, if integration times were greater for auditory than for visual "markers," AA>VA and AV>VV results would be explained. For both models, longer integration times for auditory than for visual motion would explain the rest of the data.

Conclusions

These findings establish that the FLE does indeed occur in audition—the *burst-lag effect*—and can be elicited either by motion through the frequency spectrum or by motion over space (Figures 1 and 2). This has important implications in showing that the FLE is a general phenomenon reflecting sensory processes not specific to vision. As observed with the visual FLE, half-trajectory conditions rule out a motion extrapolation account of the auditory FLE. The data also establish the existence of a cross-modal FLE. A cross-modal FLE poses a serious challenge to the latency model of the FLE and is unlikely to be due to cross-modal attentional effects. Several FLE accounts, namely the postdiction, positional sampling, and temporal integration models, could be readily adapted to account for the auditory FLE and for the cross-modal conditions. Sorting among the remaining models requires data on temporal integration and reaction times to the stimuli used in these experiments. Continuing work in our laboratory is aimed at obtaining these data in order to evaluate these models. A tantalizing possibility is that perceived timing of external stimulus events may not simply depend on physical timing summed with neural delays (due to latencies or to integration). Such a suggestion was made recently after the observation that factors other than neural delays are needed to explain cross-attribute temporal matching in vision [19], and this may well be relevant to the temporal alignment of stimuli in the cross-modal FLE.

Experimental Procedures

All audio signals were digitized at a rate of 65 kHz and had an intensity of 78 dBA. Stimuli in the spectral-motion experiments were dichotically presented with headphones. Sweep duration was 1 s, and tone duration was 40 ms. Both the sweep and the tone were ramped on and off over 20 ms according to a half-cycle raised-cosine profile. Nominally, all frequency sweeps were centered on 2 kHz, and the brief tone occurred halfway through the sweep interval. To prevent subjects from using cues based on frequency or timing when judging whether the tone was higher or lower than the instantaneous sweep frequency, we added a random jitter to each trial to the start/end points of the frequency sweep as well as to the temporal mid-point at which the brief tone occurred. Four sweep speeds were compared in separate blocks, (0.5, 1, 2, and 3 octave/s), with the frequency range expanded or contracted to vary sweep speed. A control condition with no spectral motion (0 octaves/s) was also included.

Spatial movement experiments: auditory stimuli were presented through loudspeakers flanking the video monitor and lying in the same plane 50 cm from the subject in a dimly lit room. We achieved auditory spatial movement by varying the sign and magnitude of interaural temporal delays with a temporal resolution of 15 μs (a digitization rate of 65 kHz), producing a spatial resolution of approximately 1°. The translating sound source was low-pass-filtered white noise (filtered with a fifth-order Butterworth filter attenuating above 1 kHz). The brief burst was a 1 kHz pure tone, which segregated easily from the filtered noise signal. Both signals ramped on and off over 20 ms. Visual stimuli were white Gaussian blobs with a full width at half-height of 1.5°. For visual motion, the blob began and ended its trajectory at the same points as the auditory translation. Motion sequences lasted 0.64 s, and velocity was constant. Four translation speeds were tested, approximately 20, 40, 60, and 80 degrees/s. The start/end points of the trajectories were randomly jittered from trial to trial, and the beep and flash stimuli were randomly jittered in space and time.

Acknowledgments

This work was supported by a Marie Curie Fellowship from the European Commission to D.A.

Received: July 19, 2002
Revised: September 23, 2002
Accepted: October 22, 2002
Published: January 8, 2003

References

1. MacKay, D.M. (1958). Perceptual stability of a stroboscopically lit visual field containing self-luminous objects. Nature *181*, 507–508.
2. Nijhawan, R. (1994). Motion extrapolation in catching. Nature *370*, 256–257.
3. Nijhawan, R. (1997). Visual decomposition of colour through motion extrapolation. Nature *386*, 66–69.
4. Purushothaman, G., Patel, S.S., Bedell, H.E., and Ogmen, H. (1998). Moving ahead through differential visual latency. Nature *396*, 424–426.
5. Whitney, D., Murakami, I., and 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 Res. *40*, 137–149.
6. Baldo, M.V.C., and Klein, S.A. (1995). Extrapolation or attention shift? Nature *378*, 565–566.
7. Eagleman, D.M., and Sejnowski, T.J. (2000). Motion integration and postdiction in visual awareness. Science *287*, 2036–2038.
8. Krekelberg, B., and Lappe, M. (2000). A model of the perceived relative positions of moving objects based upon a slow averaging process. Vision Res. *40*, 201–215.
9. Krekelberg, B., and Lappe, M. (2001). Neuronal latencies and the position of moving objects. Trends Neurosci. *24*, 335–339.
10. Watson, A.B., and Pelli, D.G. (1983). QUEST: a Bayesian adaptive psychometric method. Percept. Psychophys. *33*, 113–120.
11. Welford, A.T. (1980). Reaction Times. (New York: Academic Press).
12. Misulis, K.E. (1994). Spehlmann's Evoked Potential Primer. (London: Butterworth Heinemann).
13. Watson, C.S., and Gengel, R.W. (1969). Signal duration and signal frequency in relation to auditory sesitivity. J. Acoust. Soc. Am. *46*, 989–997.
14. Nowak, L.G., Munk, M.H.J., Girard, P., and Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. Vis. Neurosci. *12*, 371–384.
15. Heil, P. (1997). Auditory cortical onset responses revisited. I. First-spike timing. J. Neurophysiol. *77*, 2616–2641.
16. Brenner, E., and Smeets, J.B.J. (2000). Motion extrapolation is not responsible for the flash-lag effect. Vision Res. *40*, 1645–1648.
17. Spence, C., Shore, D.I., and Klein, R.M. (2001). Multisensory prior entry. J. Exp. Psychol. Gen. *130*, 799–832.
18. Spence, C., Nicholls, M.E.R., and Driver, J. (2000). The cost of expecting events in the wrong sensory modality. Percept. Psychophys. *63*, 330–336.
19. Nishida, S., and Johnston, A. (2002). Marker correspondence, not processing latency, determines temporal binding of visual attributes. Curr. Biol. *12*, 359–368.
20. Efron, B., and Tibshirani, R.J. (1993). An Introduction to the Bootstrap. (New York: Chapman & Hall).