

Recalibration of audiovisual simultaneity

Waka Fujisaki¹, Shinsuke Shimojo^{1,2}, Makio Kashino¹ & Shin'ya Nishida¹

To perceive the auditory and visual aspects of a physical event as occurring simultaneously, the brain must adjust for differences between the two modalities in both physical transmission time and sensory processing time. One possible strategy to overcome this difficulty is to adaptively recalibrate the simultaneity point from daily experience of audiovisual events. Here we report that after exposure to a fixed audiovisual time lag for several minutes, human participants showed shifts in their subjective simultaneity responses toward that particular lag. This 'lag adaptation' also altered the temporal tuning of an auditory-induced visual illusion, suggesting that adaptation occurred via changes in sensory processing, rather than as a result of a cognitive shift while making task responses. Our findings suggest that the brain attempts to adjust subjective simultaneity across different modalities by detecting and reducing time lags between inputs that likely arise from the same physical events.

People have the clear sense of a single audiovisual source while watching movies, but that sense is disrupted if the audio and video tracks are misaligned. Finding the relationship of simultaneity across separate sensory channels is a challenge for the brain, especially in the case of audiovisual integration^{1,2}. The difficulty arises because the temporal congruency at the source of an audiovisual event is contaminated by differential delays in the transmission—both physical and neural—of signals. Because of the large difference in the speeds of sound and light, physical arrival time of audiovisual signals changes with distance from the event. Even if a light stimulates the retina and a sound stimulates the eardrum at the same time, brain activation occurs roughly 30–50 ms earlier for the auditory signal^{3,4}. Human participants seem to compensate for these neural^{5,6} and physical⁷ lags when they judge audiovisual simultaneity. How, then, does the brain compensate for the lags and appropriately bind audiovisual signals stemming from a single event? Here we report a new type of cross-modal aftereffect that demonstrates the brain's ability to recalibrate audiovisual simultaneity. Our findings suggest that the brain may attain compensation, at least partially, by reducing the constant audiovisual lag accompanying correlated signals, without explicitly adjusting for event distance and neural delay.

RESULTS

Simultaneity judgments

Each session began with an adaptation phase during which participants were repeatedly presented with a tone pip and a ring flash, separated by a set time lag (Fig. 1). The initial adaptation period lasted 3 min, and each test trial was preceded by a 10-s re-adaptation. In a test trial, participants judged the simultaneity of tone-flash stimuli presented with various audiovisual time lags (ternary choice of 'simultaneous,' 'related but not simultaneous' and 'not related'). From the frequency distribution of simultaneous responses plotted as a function of the time lag, we estimated the point of subjective simultaneity (Fig. 2a).

The results show that subjective simultaneity, which was –10 ms (on average; negative sign indicates tone before flash) after adaptation to the zero-lag, shifted to –32 ms after adaptation to the –235 ms lag, and to +27 ms after adaptation to the +235 ms lag. Thus, the adaptation effect, estimated by the difference between the ± 235 ms conditions, was 59 ms. The finding that lag adaptation shifted subjective simultaneity in the direction of the adapted audiovisual lag is consistent with the hypothesis that the brain constantly recalibrates subjective audiovisual simultaneity in accordance with real-world audiovisual inputs.

Additional findings

In our pilot studies, the ± 235 ms settings created nearly the largest adaptation effects (Fig. 2b). According to the simultaneity judgment data, however, participants rarely perceived stimuli with lags of these magnitudes as simultaneous. Instead, these lags typically fell in the range where stimuli were judged to be 'related'. This range was ~100 ms wider than the range of simultaneity response on both sides (compare Fig. 2c and 2d). Perceptual grouping of audiovisual events may be required for lag adaptation to be effective.

We also noted that 'simultaneous' judgments were obtained for a small range of audiovisual lags that were about ± 100 ms, on average, around the point of subjective simultaneity. The lag adaptation not only shifted, but also widened the range in which simultaneity was perceived. In the group-averaged data (Fig. 2c,d), the combined effect of lag adaptation can be described as selectively extending the simultaneity range (and the 'related' range) in the direction of the adapted lag. In individual data, however, the position shift and the range extension did not always co-occur.

Although we collected data during a session that lasted about 30 min, the development of adaptation did not take that long. As estimated from the responses of the first 13 trials of each session (78 trials), the difference between the ± 235 ms conditions was 59 ms. This

¹NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan. ²Division of Biology, California Institute of Technology, Pasadena, California 91125, USA. Correspondence should be addressed to S.N. (nishida@bri.ntt.co.jp).

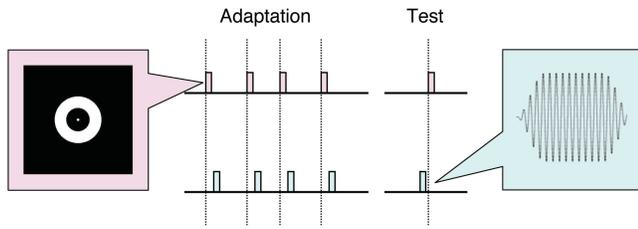


Figure 1 The time course of the stimulus sequence used to test the effects of audiovisual lag adaptation on simultaneity judgments. The left-hand box shows the configuration of the visual stimulus, and the right-hand box shows the waveform of the auditory stimulus.

implies that the adaptation effect was present even immediately after the initial 3-min adaptation.

We obtained similar adaptation effects when binary temporal order judgments, collected at the same time, were used for subjective simultaneity estimation (54-ms difference between the +235 ms and -235 ms conditions), although the estimation was less stable and some participants did not show the adaptation effect.

There was a small but significant ($P = 0.037$) difference in the point of subjective simultaneity between the no-adaptation condition and the zero-lag adaptation condition (Fig. 2a). One possible explanation is that the no-adaptation condition might have been more affected by pre-adaptation to the natural environment, in which audio signals tend to be delayed relative to visual signals.

Effect of attention and sound-presentation method

It is well known that attention affects temporal perception: the events of the attended modality seem to occur earlier than the events of the unattended modality (prior entry effect^{8,9}). A large individual variation in subjective simultaneity (indicated by large error bars) could be, at least partially, ascribed to the participant-specific tendency to attend to one modality more than to the other.

Given the strong effect of attention, one may suspect that the lag adaptation shifts the subjective simultaneity by biasing a participant's attention primarily to the modality of the second stimulus. If so, the lag adaptation effect should diminish when the participant has to attend to

one modality during the test period; but in a subsidiary experiment, we found that this was not the case. To direct participants' attention primarily to audition, we asked them to make a fast button press at the onset of the auditory test stimulus before making an audiovisual simultaneity judgment. This attentional control successfully shifted subjective simultaneity in the direction predicted by the prior entry effect, but it did not affect the lag adaptation effect itself (Fig. 2e).

In another subsidiary experiment, we presented the sounds using a hidden speaker placed immediately below the visual stimuli so that the auditory and visual stimuli seemed to come from nearly the same physical location. The lag aftereffect obtained in this situation was nearly the same as that obtained with headphone presentation (Fig. 2e).

Stream/bounce illusion

The estimation of subjective simultaneity based on participants' explicit judgment of simultaneity did not exclude the possibility that the aftereffect resulted from a change in cognitive decision criterion rather than a change of perception. The second experiment therefore tested the effects of audiovisual lag adaptation functionally by using a perceptual phenomenon that depended on audiovisual simultaneity¹⁰⁻¹². Two balls moving across each other on a screen can be perceived either as streaming through or bouncing off each other, although the former is typically dominant. Presentation of a 'collision' sound timed near the crossover of the balls facilitates the perception of balls' bouncing off each other. The perception is spontaneous and the judgment is effortless; it does not involve any explicit simultaneity judgment. Thus, this phenomenon was particularly suitable for measuring the brain's implicit processing of audiovisual simultaneity.

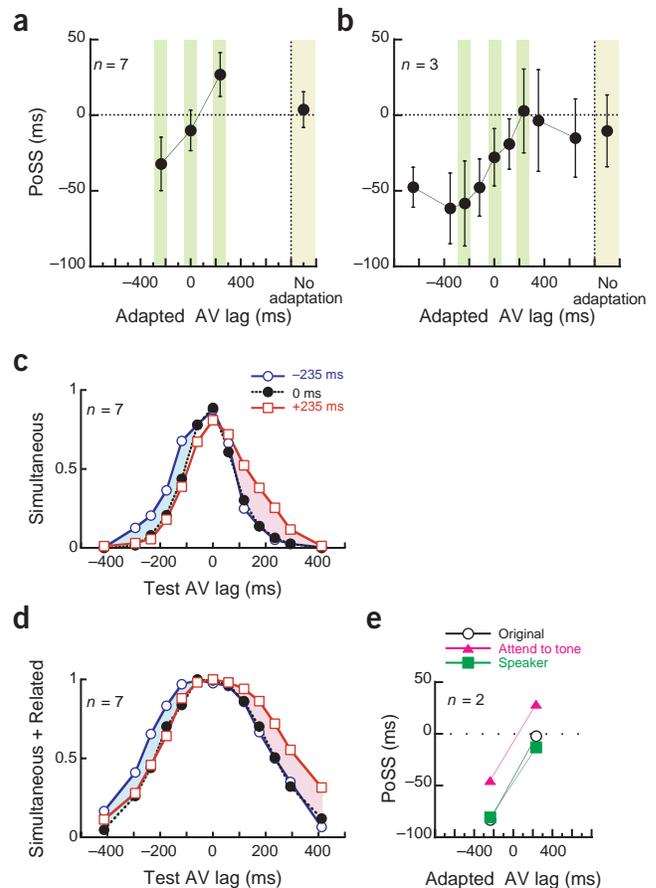


Figure 2 The effects of audiovisual (AV) lag adaptation on simultaneity judgments. (a) The point of subjective simultaneity (PoSS, the center of the fitted Gaussian function averaged over participants; error bars indicate standard error (s.e.m.) across participants) for the three adaptation conditions. A one-way analysis of variance (ANOVA) indicated that the main effect of the adapted AV lag was highly significant ($F_{2,12} = 22.5$, $P < 0.001$). The P -values obtained by *post hoc* analysis (Tukey HSD test) and the number of participants who showed significant differences in the bootstrapping analysis ($P \leq 0.05$, one-tailed) are as follows: $P = 0.068$ and 4/7 for -235 vs. 0 ms; $P = 0.004$ and 7/7 for 0 vs. +235 ms; $P < 0.001$ and 7/7 for -235 vs. +235 ms. (b) The effects of a wider range of adaptation lags. (c) The probability of 'simultaneous' response as a function of the test audiovisual lag. The response probability for each lag was computed for each participant, and then averaged across participants. The main effect of adaptation is an increase in the probability of simultaneity on the side of adapted lag (shaded areas). (d) The probability that the participants made either 'simultaneous' or 'related' response. (e) The results of subsidiary experiments. 'Original': original condition. 'Attend to tone': the participant's attention was directed to the tone pip in the test stimuli. 'Speaker': the sounds were made by a hidden speaker. These manipulations had no significant effects. In all cases, the lag adaptation effect was statistically significant for both participants.



Figure 3 Effects of audiovisual (AV) lag adaptation on the stream/bounce illusion. The pattern of results was quite similar to those obtained with simultaneity judgments (Fig. 2). (a) Left, configuration of the adaptation display. Center and right, space-time plots of the adaptation and test stimuli (central area), shown with a tone pip presented at zero delay. (b) The centroid of bounce responses for the three adaptation conditions. A one-way ANOVA indicated that the main effect of the adapted AV lag was significant ($F_{2,16} = 8.6$, $P = 0.003$). The P -values obtained by Tukey HSD test and the number of participants who showed significant differences in the bootstrapping analysis are as follows: $P > 0.1$ and 2/9 for -235 vs. 0 ms; $P = 0.012$ and 5/9 for 0 vs. $+235$ ms; $P = 0.004$ and 5/9 for -235 vs. $+235$ ms. (c) The effect of a wide adaptation lag change. (d) The probability of bounce response as a function of the test audiovisual lag.

In the experiment, the adaptation phase was an unambiguous bounce display in which a tone pip was timed with a white and a black ball bouncing off each other. Test trials consisted of an ambiguous motion display of two black or white balls, and participants reported whether they saw the balls ‘bounce’ or ‘stream’ through each other (Fig. 3a). The index of simultaneity was the centroid of the ‘bounce’ response distribution. The estimated simultaneity might be biased by temporal tuning specific to this illusion¹², which potentially includes the delay required for the sound to affect the visual percept. As long as the bias is constant across different adaptation conditions, however, it should not affect the estimation of the adaptation effect.

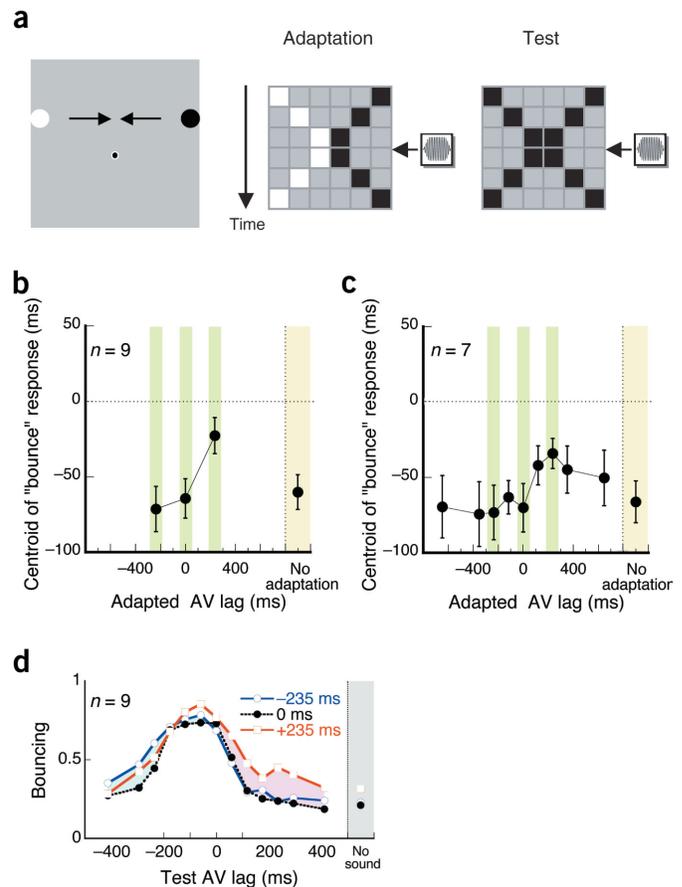
The results showed lag adaptation effects similar to those found in the first experiment. The centroid of the bouncing response, which was -64 ms after adaptation to the zero-lag condition, shifted to -71 ms and -23 ms after adaptation to the -235 ms lag and $+235$ ms lag, respectively (Fig. 3b). The difference between the ± 235 ms conditions was 48 ms. For a wider range of adaptation lags, the adaptation effect again showed a similar tuning function (Fig. 3c). The lag adaptation also broadened the distribution of the bounce response (Fig. 3d). The results of the second experiment therefore suggest that the lag adaptation effect has a perceptual origin.

Cross adaptation

Additional experiments showed that the lag adaptation was effective even when the stimulus was fundamentally changed between adaptation and test (Fig. 4). We found lag adaptation when using the tone-flash adaptation stimulus of the first experiment and then when testing with the bounce-stream task of the second experiment. When the stimulus in the adaptation phase was changed to tone pips timed with a ball bouncing off of the inner walls of a square (‘Wall’ display), tests either with the bounce-stream task or with the tone-flash simultaneity task showed lag adaptation. For some of the above conditions, we also tested the presentation of the adaptation and test tones to different ears, and the lag aftereffect remained. Although the magnitude of the aftereffect was slightly smaller for some cross-adaptation conditions, the difference across conditions did not reach statistical significance. These transfer results indicate that the lag adaptation effects on the subjective simultaneity judgment and the stream/bounce illusion have a common perceptual origin, at least partially. A more systematic investigation of stimulus specificity may be required to draw a definite conclusion, but our present results suggest that lag adaptation occurs neither at a peripheral sensory stage that is sensitive to low-level stimulus properties, nor at a higher cognitive stage that takes into account the content correspondence of adapted event pairing.

DISCUSSION

We found novel psychophysical adaptation effects in which exposure to a fixed audiovisual time lag for several minutes shifts subjective



simultaneity toward the adapted lag. Recalibration of audiovisual simultaneity, demonstrated by the present adaptation effects, is a useful mechanism for the human brain to compensate for the processing delay of visual information, relative to auditory information^{3,4}. Such a compensation method could be affected by events during development, damage to neural mechanisms and by the physical characteristics of one’s current environment (for example, the average distance and signal intensity of events).

Although their effects were slightly smaller than those of positive audiovisual lags, negative audiovisual lags, which rarely occur in a natural environment, still induced the aftereffect. It seems that the brain does not strictly limit the range of recalibration by following the physical rule of sound-light asymmetry.

Given that the aftereffect is a result of recalibration, one may wonder why the post-adaptation shift was only about 10% of the adapted lag. One reason may be a hardware limitation of the adaptation mechanism¹³, but an alternative and more compelling idea is that the adaptation mechanism takes into account the long history of ‘veridical’ sensory inputs that it has received throughout the lifetime of the participant, outside of these short adaptation experiments. Although our data indicate that a few minutes of adaptation suffices to shift audiovisual simultaneity, it is possible that much longer adaptation on the order of hours, days or even months (especially during particular periods in development) markedly increases the adaptation effect through a more long-term mechanism. Note also that the post-adaptation shift is generally much smaller than the adapted magnitude in sensory aftereffects^{14–16}, which are said to have similar functional roles of recalibrating the internal norm to the current environment¹⁷.

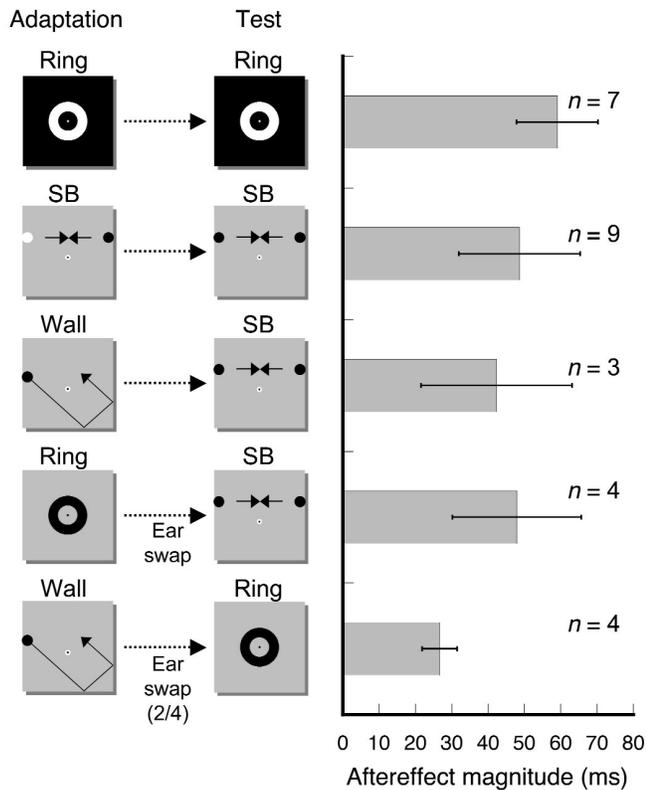


Figure 4 The lag aftereffect for various combinations of adaptation and test stimuli. Each bar indicates the difference in the point of subjective simultaneity between the -235 ms lag condition and the $+235$ -ms lag condition. Error bar represents s.e.m. across participants. 'Ring-Ring' is from **Figure 2**, and 'SB-SB' is from **Figure 3** (SB, stream/bounce). For the other conditions, the visual display fundamentally changed between adaptation and test. 'Wall-SB' is a variant of 'SB-SB', in which the adaptation display consisted of a ball bouncing in a square. 'Ring-SB' and 'Wall-Ring' were the conditions in which we tested the transfer of the lag adaptation effect between flash events and bounce events. In the 'Ring' display of the 'Ring-SB' condition, we flashed a black pattern on the gray background to maintain background luminance between adaptation and test. Additionally, for all the participants of 'Ring-SB', and for two of the four participants of 'Wall-Ring', the tone pip was presented monotonically, and the presentation ear was swapped between adaptation and test (for 'Wall-Ring', there was no effect of ear swap: 27 ms for ear-swapped participants and 26 ms for diotically hearing participants). The auditory stimulus was the same (1,800-Hz pip) for all the conditions. A statistically significant ($P < 0.05$, except $P = 0.089$ for 'Wall-SB', one-tailed t -test) lag aftereffect was obtained regardless of the similarity between adaptation and test stimuli.

The effects of lag adaptation cannot be accounted for by distance-dependent calibration of audiovisual simultaneity⁷ (but also see refs. 6,18) because the adaptive changes occurred independently of depth perception or its changes. Our finding is also distinct from a recently reported audiovisual temporal aftereffect¹⁹, as it affects apparent flicker/flutter rate, not simultaneity.

Yet, several related phenomena may be involved. It has been reported within visual²⁰ and auditory²¹ modalities that adaptation to a constant temporal order sequence can bias the temporal order judgment toward the opposite direction. Although these effects, unlike ours, can be ascribed to adaptation of low-level stimulus-change detectors, they are functionally similar to the audiovisual lag aftereffect. With regard to sensorimotor coordination, when a constant delay is inserted between motor response and visual feedback, participants can gradually adapt to the delay, and then a negative aftereffect is apparent when the delay is removed^{22,23}. In the space domain, recalibration of the audiovisual spatial map is well known as the ventriloquism aftereffect^{24,25}, and prism adaptation studies have demonstrated dramatic cross-modal remapping effects^{16,26,27}. We believe that adaptation is a general mechanism that the biological system uses in order to adjust spatiotemporal congruency across separate channels, which provides critical cues for feature binding.

Our findings demonstrate a functional similarity of feature binding mechanisms operating within and across modalities and thus, together with the recent findings of strong cross-modal interactions^{28,29}, argue against such a stringent modular view that sensory modalities are "informationally encapsulated from each other"³⁰.

Concerning the underlying neural mechanisms, it is debatable whether the lag adaptation alters neural transmission time of one modality relative to the other, given the short time scale of the experiments. As suggested in other contexts, subjective simultaneity is a result of the brain's interpretation of external events, not a simple

reflection of the physical simultaneity between neural signals^{31,32}. Although subjective audiovisual simultaneity may be represented in complex neural activities in higher cortical areas, it also could be simply represented as the activity pattern of neurons sensitive to various temporal lags of audiovisual signals (which are conceptually similar to those sensitive to interaural temporal lags^{33,34} or visual spatiotemporal lags^{35,36}). In the latter case, response magnitudes and/or temporal tunings of the neurons would be modifiable by adaptation^{13,37}. Those neurons are likely to exist somewhere in multimodal areas, including the superior colliculus (SC), insula and prefrontal cortex, which show activities correlated with the percept of audiovisual simultaneity³⁸ and sound-induced bouncing³⁹. A possible candidate is multisensory SC neurons that show audiovisual interactions even for temporal disparities of several hundred milliseconds, and a variation in the shape of their temporal tuning⁴⁰. Although these neurons (and the multisensory-evoked eye movements possibly related to their activity⁴¹) are sensitive to the spatial alignment of audiovisual inputs, our results indicate that the adaptation mechanism is not very selective to the sound-presentation method (headphones versus a speaker) or the swap of the presentation ear. The effect of spatial localization on the lag aftereffect certainly deserves a closer examination in future studies.

METHODS

Subjects and setup. Participants were three of the authors and six paid volunteers (four, six and four in the first, second and third experiments, respectively) who were unaware of the purpose of the experiments. All had normal or corrected-to-normal vision and hearing. Informed consent was obtained after the nature and possible consequences of the studies were explained. We ran the experiments on Apple Macintosh PowerBook G3s and G4s, using Matlab (Mathworks) with Psychophysics Toolbox extensions^{42,43}. In a quiet dark room, the participant sat at the distance of 64.5 cm from the monitor (Sony GDM-F500, 85Hz), wearing headphones (Sennheiser HDA 200).

First experiment: simultaneity judgments. The visual stimulus was a white ring (outer diameter, 5.0° ; inner, 2.5° ; 83.1 cd/m²) that flashed for one monitor frame at the center of a black square area (11.6° , 0.8 cd/m²) surrounded by a white background. A fixation marker was presented at the square's center. The auditory stimulus was a diotically presented (to both ears) tone pip (1,800 Hz, 70 dB SPL) lasting for 10 ms with a 2.5-ms raised-cosine ramp at the onset and offset.

A session started with a 3-min initial adaptation, followed by test trials, each preceded by a 10-s top-up adaptation. During adaptation, the audiovisual pair was repeatedly presented with a constant lag (between the onsets of flash and

pip, accuracy < 1 ms). Before each test, the circle fixation marker changed to a cross. After a 2-s pause, a ring-pip pair was presented with a lag randomly chosen from 13 values between -412 and +412 ms. The participants judged simultaneity as 'simultaneous,' 'related but not simultaneous' or 'neither simultaneous nor related,' and then judged temporal order as 'auditory stimulus first' or 'visual stimulus first' (a temporal order judgment was recorded even if they chose 'simultaneous' in the first judgment). Each session, lasting about 30 min, consisted of 78 test trials. The trials were divided into six blocks, with each block containing one repetition of the 13 different test lag conditions in random order.

The adaptation lag changed across sessions. For the four participants who ran only three adaptation conditions, the interval between the adjacent adaptation pairs was randomly varied within the range of 776 ± 259 ms. For the three participants who ran the full nine lag conditions, the adaptation interval was $1,518 \pm 506$ ms (but it was fixed at 1,518 ms for the ± 647 ms lag conditions). These conditions were chosen to maximize stimulus density while avoiding unintended audiovisual grouping (by placing the stimuli of an intended pair temporally closer to each other than to adjacent stimuli of a different pair). During adaptation, to direct participants' attention to both the audio and visual stimuli, we had them detect odd stimuli (a smaller ring that was two-thirds the size of the original or a 1,500-Hz pip) that appeared with a probability of ~5%. Each participant ran four sessions (24 trials for each test lag) for each adaptation condition.

In the data analysis, for each participant and adaptation condition, the rate of simultaneity responses was plotted as a function of test lag. Using the maximum likelihood method⁴⁴, the response distribution was fitted by a truncated Gaussian function,

$$y = \min\{1, a \times \exp[-(x - m)^2/2\sigma^2]\}$$

where a is the amplitude, m is the mean (the estimate of subjective simultaneity), σ the standard deviation of the Gaussian function and the min function gives an upper bound of 1, even when $a > 1$. The correlation coefficient of the fitting was ≥ 0.92 .

In each test trial of the first subsidiary experiment, participants first made a button press at the onset of a tone pip as fast as possible, then made an unsped ternary judgment on audiovisual simultaneity. The temporal order judgment was not requested. In the second subsidiary experiment, the sounds were presented by a speaker located immediately below the ring stimulus. We wrapped the speaker in a black cloth to make it invisible in the dark experimental room.

Second experiment: the stream/bounce illusion. The display consisted of two balls (0.4° in diameter), presented within a gray square (41.6 cd/m^2 , 9.5°) surrounded by a white background. Each ball moved from one side of the square to the other along a horizontal path. The balls crossed each other at the center of the path, which was located 2.07° above the fixation marker and 1.4° above the square center. The balls' movement ($11.1^\circ/\text{s}$) was produced by a 0.4° position shift every three monitor frames ($35 \text{ ms} = 1 \text{ image frame}$). In the movie sequence of 24 image frames, the two balls moved toward each other, attached at the 12th frame, switched positions at the 13th frame, and then moved away from each other. We defined '0 lag' as the tone (1,800-Hz pip) synchronized with the 13th frame. (If we had defined the 12th frame as the origin, the centroid of bounce response would have shifted upward by 35 ms in Fig. 3b,c, and the difference in the point of simultaneity between the first and second experiments would have been greatly diminished). In the adaptation stimulus, the two balls had different colors (black and white) to facilitate bouncing perception. (The color difference did not perfectly exclude streaming perception for some participants, but the adaptation effect was evident even for them.) An audiovisual display with a constant lag was presented once every $1,647 \pm 400$ ms. In the test stimulus, the two balls had the same contrast polarity chosen randomly from trial to trial. Both during the adaptation and test phases, the participants made a binary judgment on motion perception (stream or bounce). Each session consisted of 84 test trials (six repetitions of 13 lags, plus no sound). The procedures were otherwise the same as in the first experiment. In the data analysis, we computed the centroid (weighted mean) of the distribution of bounce responses. We also tried Gaussian fitting, but the fit was poor ($r < 0.8$) for some data.

Third experiment: cross adaptation. In the 'Wall' display, a black ball (0.4°) bounced against inner walls of a gray square (9.57°), on average 1.7 times per second. Participants had to detect odd stimuli (a brief flash of the ball at bouncing or a 1,500-Hz pip).

ACKNOWLEDGMENTS

We thank D. Arnold, M. Changizi, T. Hirahara, A. Johnston and D. Wu. This work was partially supported by the Human Frontier Science Program (RGP0070/2003-C).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 22 January; accepted 27 April 2004

Published online at <http://www.nature.com/natureneuroscience/>

- Pöppel, E. *Grenzen des Bewußtseins: Über Wirklichkeit und Weiterfahrung* (Deutsche Verlags-Anstalt GmbH, Stuttgart, 1985).
- Spence, C. & Squire, S. Multisensory integration: maintaining the perception of synchrony. *Curr. Biol.* **13**, R519–R521 (2003).
- King, A.J. & Palmer, A.R. Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Exp. Brain. Res.* **60**, 492–500 (1985).
- Regan, D. *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine* (Elsevier, New York, 1989).
- Tappe, T., Niepel, M. & Neumann, O. A dissociation between reaction time to sinusoidal gratings and temporal-order judgment. *Perception* **23**, 335–347 (1994).
- Stone, J.V. *et al.* When is now? Perception of simultaneity. *Proc. R. Soc. Lond. B Biol. Sci.* **268**, 31–38 (2001).
- Sugita, Y. & Suzuki, Y. Audiovisual perception: implicit estimation of sound-arrival time. *Nature* **421**, 911 (2003).
- James, W. *Principles of Psychology* (Holt, New York, 1890).
- Spence, C., Shore, D.I. & Klein, R.M. Multisensory prior entry. *J. Exp. Psychol. Gen.* **130**, 799–832 (2001).
- Sekuler, R., Sekuler, A.B. & Lau, R. Sound alters visual motion perception. *Nature* **385**, 308 (1997).
- Watanabe, K. & Shimojo, S. When sound affects vision: effects of auditory grouping on visual motion perception. *Psychol. Sci.* **12**, 109–116 (2001).
- Shimojo, S. & Shams, L. Sensory modalities are not separate modalities: plasticity and interactions. *Curr. Opin. Neurobiol.* **11**, 505–509 (2001).
- Dragoi, V., Rivadulla, C. & Sur, M. Foci of orientation plasticity in visual cortex. *Nature* **411**, 80–86 (2001).
- Mather, G., Verstraten, F.A.J. & Anstis, S.M. *The Motion Aftereffect: A Modern Perspective* (MIT Press, Cambridge, Massachusetts, 1998).
- Kashino, M. & Nishida, S. Adaptation in the processing of interaural time differences revealed by the auditory localization aftereffect. *J. Acoust. Soc. Am.* **103**, 3597–3604 (1998).
- Dolezal, H. *Living in a World Transformed: Perceptual and Performatory Adaptation to Visual Distortion* (Academic, New York, 1982).
- Barlow, H.B. & Földiák, P. Adaptation and decorrelation in the cortex. in *The Computing Neuron* (eds. Durbin, R., Miall, C. & Mitchison, G.) 54–72 (Addison-Wesley, Boston, 1989).
- Lewald, J. & Guskir, R. Auditory-visual temporal integration as a function of distance: no compensation for sound-transmission time in human perception. *Neurosci. Lett.* **357**, 119–122 (2004).
- Recanzone, G.H. Auditory influences on visual temporal rate perception. *J. Neurophysiol.* **89**, 1078–1093 (2003).
- Bennett, R.G. & Westheimer, G. A shift in the perceived simultaneity of adjacent visual stimuli following adaptation to stroboscopic motion along the same axis. *Vision Res.* **25**, 565–569 (1985).
- Okada, M. & Kashino, M. The role of spectral change detectors in temporal order judgment of tones. *Neuroreport* **14**, 261–264 (2003).
- Cunningham, D.W., Billock, V.A. & Tsou, B.H. Sensorimotor adaptation to violations of temporal contiguity. *Psychol. Sci.* **12**, 532–535 (2001).
- Cunningham, D.W., Chatziastros, A., von der Heyde, M. & Bulthoff, H.H. Driving in the future: temporal visuomotor adaptation and generalization. *J. Vis.* **1**, 88–98 (2001).
- Canon, L.K. Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *J. Exp. Psychol.* **84**, 141–147 (1970).
- Recanzone, G.H. Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proc. Natl. Acad. Sci. USA* **95**, 869–875 (1998).
- Knudsen, E.I. & Knudsen, P.F. Vision guides the adjustment of auditory localization in young barn owls. *Science* **230**, 545–548 (1985).
- Zwiers, M.P., Van Opstal, A.J. & Paige, G.D. Plasticity in human sound localization induced by compressed spatial vision. *Nat. Neurosci.* **6**, 175–181 (2003).
- Shams, L., Kamitani, Y. & Shimojo, S. Illusions. What you see is what you hear. *Nature* **408**, 788 (2000).
- Kitagawa, N. & Ichihara, S. Hearing visual motion in depth. *Nature* **416**, 172–174 (2002).

30. Fodor, J.A. *The Modularity of Mind* (MIT Press, Cambridge, Massachusetts, 1983).
31. Dennett, D.C. & Kinsbourne, M. Time and the observer: the where and when of consciousness in the brain. *Behav. Brain Sci.* **15**, 183–247 (1992).
32. Nishida, S. & Johnston, A. Marker correspondence, not processing latency, determines temporal binding of visual attributes. *Curr. Biol.* **12**, 359–368 (2002).
33. Jeffress, L.A. A place theory of sound localization. *J. Comp. Physiol. Psychol.* **41**, 35–39 (1948).
34. McAlpine, D. & Grothe, B. Sound localization and delay lines – do mammals fit the model? *Trends Neurosci.* **26**, 347–350 (2003).
35. Reichardt, W. Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. in *Sensory Communication* (ed. Rosenblith, W.A.) 303–317 (MIT Press, Cambridge, Massachusetts, 1961).
36. Lu, Z.L. & Sperling, G. Three-systems theory of human visual motion perception: review and update. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **18**, 2331–2370 (2001).
37. Barlow, H.B. & Hill, R.M. Evidence for a physiological explanation of the waterfall phenomenon and figural after-effects. *Nature* **200**, 1345–1347 (1963).
38. Bushara, K.O., Grafman, J. & Hallett, M. Neural correlates of auditory-visual stimulus onset asynchrony detection. *J. Neurosci.* **21**, 300–304 (2001).
39. Bushara, K.O. *et al.* Neural correlates of cross-modal binding. *Nat. Neurosci.* **6**, 190–195 (2003).
40. Meredith, M.A., Nemitz, J.W. & Stein, B.E. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J. Neurosci.* **7**, 3215–3229 (1987).
41. Corneil, B.D., Van Wanrooij, M., Munoz, D.P. & Van Opstal, A.J. Auditory-visual interactions subserving goal-directed saccades in a complex scene. *J. Neurophysiol.* **88**, 438–454 (2002).
42. Brainard, D.H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
43. Pelli, D.G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
44. Watson, A.B. Probability summation over time. *Vision Res.* **19**, 515–522 (1979).