

Research report

Visual illusion induced by sound

Ladan Shams^{a,*}, Yukiyasu Kamitani^b, Shinsuke Shimojo^{a,c}

^aCalifornia Institute of Technology, Division of Biology, MC 139-74, Pasadena, CA 91125, USA

^bHarvard Medical School, Beth Israel Medical Center, Department of Neurology, 330 Brookline Ave. KS-454, Boston, MA 02215, USA

^cNTT Communication Science Laboratories, Human and Information Science Laboratory, Atsugi, Kanagawa 243-0198, Japan

Accepted 15 August 2001

Abstract

We present the first cross-modal modification of visual perception which involves a phenomenological change in the quality—as opposed to a small, gradual, or quantitative change—of the percept of a non-ambiguous visual stimulus. We report a visual illusion which is induced by sound: when a single flash of light is accompanied by multiple auditory beeps, the single flash is perceived as multiple flashes. We present two experiments as well as several observations which establish that this alteration of the visual percept is due to cross-modal perceptual interactions as opposed to cognitive, attentional, or other origins. The results of the second experiment also reveal that the temporal window of these audio–visual interactions is approximately 100 ms. © 2002 Elsevier Science B.V. All rights reserved.

Theme: Sensory systems

Topic: Visual psychophysics and behavior

Keywords: Crossmodal interaction; Auditory–visual interaction; Visual illusion; Illusory flashing; Multisensory integration; Audio–visual perception

1. Introduction

Our perception of the world clearly benefits from the information delivered by multiple modalities. A usual strategy in examining the relative weight of individual sensory modalities to the overall perception is to make the information conveyed by two modalities in conflict with each other. Results of these studies identify vision as the most important or dominant modality, and often suggest that the signals of the competing modality are ignored. Two well-known examples of this paradigm are the ventriloquism effect [4] and visual capture [3]. The former involves a conflict between spatial location of auditory and visual signals. The perceived location of the overall event is determined predominantly by the location of the visual stimulus. Similarly, visual capture involves a spatial localization task when the visual information is in conflict with that of another modality—namely, proprioceptive

information. Again the perceived location is determined predominantly by visual information.

There are conflict paradigms in which vision does not dominate, but nevertheless, modifies the percept in the other modality. A well-known example is McGurk's effect [5] where visual information significantly alters the auditory phoneme perception. Another study has shown that the modification of auditory perception by a conflicting visual stimulus is not unique to speech signals and occurs also with musical note perception [8].

While the best-known examples of cross-modal interactions involve modification of other modalities by vision, there exists a number of studies in the literature that report cross-modal interactions in the opposite direction. The majority of these findings involve modification of perceived temporal characteristics of the visual stimulus such as duration [14], frequency [2,12,15], and timing [1,9] by sound. Temporal characteristics are not the only attribute of visual stimuli subject to modification, however. Stein et al. reported that the perceived intensity of the visual stimulus is enhanced in the presence of sound [13]. Note that all the aforementioned reports of modification of vision by other modalities involve small quantitative changes as opposed to radical and phenomenological

*Corresponding author. Tel.: +1-626-395-2362; fax: +1-626-844-4514.

URL and E-mail addresses: <http://neuro.caltech.edu/~lshams> (L. Shams), ladan@caltech.edu (L. Shams).

changes in the quality of the percepts. One study has shown that sound can alter the visually perceived direction of motion [10]. Here sound causes a phenomenological change in the percept, however, the motion direction of the visual stimulus is inherently ambiguous and can be interpreted in two different ways. The effect of sound is, therefore, to bias the interpretation in favor of one of the two alternatives. It remains to be seen whether the visual perception can be altered by other modalities qualitatively even when there is no ambiguity in the visual stimulus. Building upon a recently discovered visual illusion [11], here we report data that firmly establish that the visual perception is seriously malleable by signals of other modalities, and motivate new hypotheses about cross-modal interactions.

2. Experiment 1

The purpose of this experiment is to investigate a recently discovered phenomenon [11]: when a single flash of light is accompanied with multiple beeps, it is perceived as multiple flashes. In the following experiment we examined whether this phenomenon is a perceptual illusion or whether it is due to artifacts (Fig. 1).

2.1. Materials and methods

2.1.1. Participants

Eight naïve volunteers participated in the experiment (six females, two males). Their ages ranged from 24 to 41 years. Participants gave their informed consent before inclusion in the study.

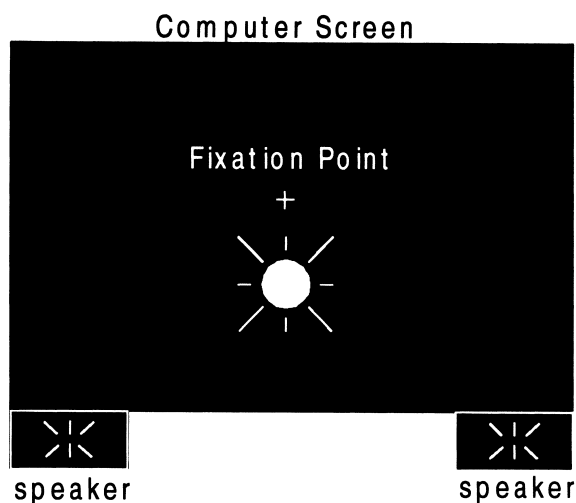


Fig. 1. Stimulus configuration for Experiments 1 and 2. A white uniform disk is displayed against a black background at some eccentricity below the fixation point which is at the center of the screen. Approximately at the same time some beeps are played from two speakers directly beneath and to the sides of the screen.

2.1.2. Stimuli

In each trial a uniform white disk (with a luminance of 108 cd/m^2) subtending 2° of visual field at 5° eccentricity was flashed on a black computer screen (with a luminance of 0.02 cd/m^2) one to four times. In single flash trials, the flash was accompanied with 0–4 beeps and, in multiple flash trials, flashes were accompanied with 0 or 1 beep. The beeps had a 95 dB SPL and 3.5 kHz frequency. The pitch was chosen arbitrarily, as the pilot data indicated that the sound pitch does not make any difference in the results. We will henceforth refer to trials with one flash accompanied by 2–4 beeps as illusion trials. The first beep always preceded the first flash by 23 ms. Each beep had a duration of 7 ms and consecutive beeps were spaced 57 ms apart (Fig. 2). We made the successive flashes tightly spaced in order to match the perceptual impression of the illusory multiple flashes. The aforementioned time durations and intervals were chosen fairly arbitrarily otherwise.

2.1.3. Procedure

Participants sat at a viewing distance of 57 cm from the computer screen and speakers which presented the stimuli. Throughout the trials there was a constant fixation point at the center of the screen. The observer's task was to judge the number of flashes s/he saw on the screen. The experiment consisted of five trials of each condition, amounting to a total of 60 trials, ordered randomly. Notice that the 15 illusion trials were dispersed randomly within 45 trials which did not involve the illusion. We used such a setting to ensure that the observers employed the same strategy (for judging the number of flashes) in illusion trials as they did in the other trials.

2.2. Results and discussion

The main result of experiment 1 is shown in Fig. 3. The figure shows the data for trials in which a single flash was presented. The number of perceived flashes is plotted against the number of beeps in each trial (averaged across observers; the error bars represent the standard error of the

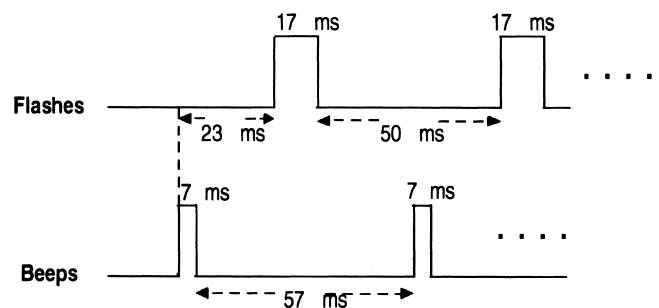


Fig. 2. Temporal profile of the stimuli in Experiment 1. This diagram shows the relationship between the timing of the beep(s) and flash(es) as well as the time duration and spacing of the signals. In each trial there were one or more (up to four) flashes accompanied with zero or more (up to four) beeps.

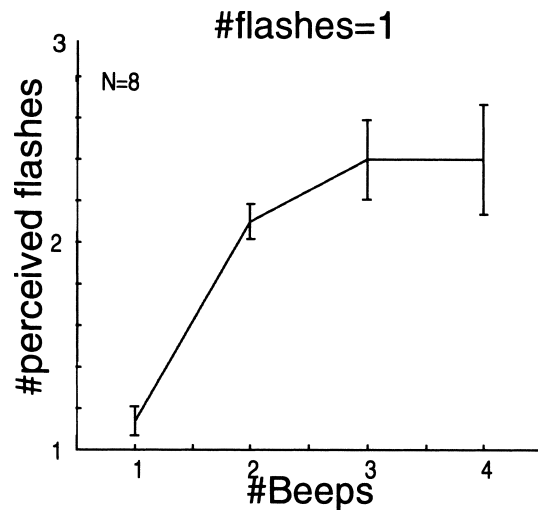


Fig. 3. Illusory flashing. The average number of perceived flashes across eight observers is plotted against the number of beeps, for trials in which the visual stimulus consisted of one single flash. Observers report seeing two or more flashes, when the single flash is accompanied by two or more beeps.

mean). The observers report seeing one flash, i.e. the veridical value, when the number of accompanying beeps is one. However, they report seeing two or more flashes when the flash is accompanied with two or more beeps. The perceived number of flashes in trials with a single flash and two, three, or four beeps is significantly greater than that of trials with single flash and one (or no) beep ($P < 0.001$). We refer to this phenomenon as sound-induced illusory flashing. The results of illusion trials suggest that multiple beeps change the percept of a single flash into multiple flashes.

To examine whether these results are due to difficulty of the task, we turn to the control condition in which the sound is absent, and the number of physical flashes varies. The data for this condition is displayed in Fig. 4a. In this figure, the number of perceived flashes is plotted against the actual number of flashes. The observers performed the task of judging the number of flashes very well in the absence of sound. These results indicate that the task of judging the number of flashes was not overly difficult for the observers and that the visual stimuli were not ambiguous.

Looking back at Fig. 3, one can see that the number of perceived flashes increases with the number of beeps. This observation may lead to a suspicion that the reported number of flashes has been in response to auditory as opposed to visual perception. To investigate this possibility, we examine the ‘catch trials.’ These are trials, other than the illusion trials, in which there is a discrepancy between the number of flashes and beeps: the number of beeps is one and the number of flashes varies ranging from two to four. The results of these trials are shown in Fig. 4b. In this figure, the number of perceived flashes is plotted against the number of physical flashes. Had the observers’

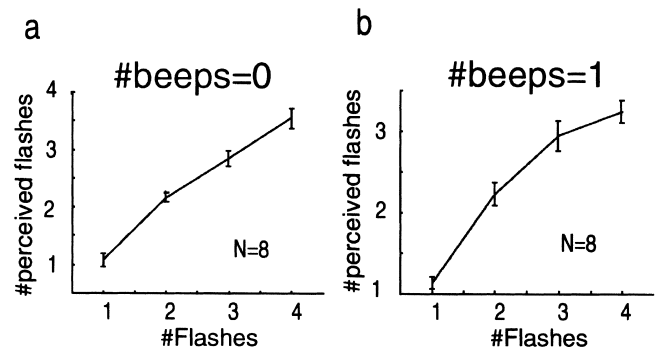


Fig. 4. Control data for Experiment 1. The number of perceived flashes is plotted against the number of actual flashes displayed in corresponding trial, for trials in which auditory stimulus was absent or consisted of one beep, shown in (a) and (b), respectively. The results displayed in (a) demonstrate that the observers could judge the correct number of flashes in the absence of sound. The portion of data in (b) corresponding to 2–4 flashes (on the horizontal axis) illustrate the results of the catch trials. The responses of the observers in these trials are in contrast to the number of beeps they heard in those trials, indicating that the responses were not cognitively influenced by the auditory derived information.

responses been determined by the number of beeps, we would expect to obtain, as the number of perceived flashes, a flat line intersecting the vertical axis at one, in agreement with the number of beeps. As can be clearly seen, this is not the case, and the observers’ responses are consistent with the number of actual flashes and in conflict with the number of beeps (for trials with more than one beep). The perceived number of flashes in trials with two, three, or four flashes and one beep is significantly greater than that of trials with one flash and one (or no) beep ($P < 0.001$). These results indicate that the observers’ responses were indeed based on their visual perception, and were not determined by cognitive biases derived from the auditory perception.

The results discussed thus far indicate that the sound-induced flashing is indeed a visual perceptual illusion and is not due to artifacts such as the difficulty of the task or cognitive biases. The next natural question to ask is how comparable an illusory flash is to a real physical flash. To explore this question, we compare the reported perceived number of flashes across different conditions. Fig. 5 combines the three plots shown in Figs. 3 and 4. For all three plots, the vertical axis represents the perceived number of flashes, while the horizontal axis denotes the number of beeps and the number of flashes for the gray plot and the control plots (solid and broken), respectively. As can be seen in the overlap of the three plots at the second data point, the responses of the observers were the same whether they were exposed to one flash accompanied with two beeps (the gray plot), or two flashes accompanied with one or no beeps (solid and broken plots, respectively). It is also notable that the participants (including non-naïve observers who were aware of the physical stimuli presented—results are not shown here) reported after the

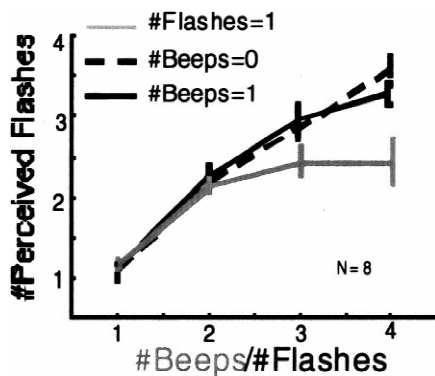


Fig. 5. Comparison of different conditions in Experiment 1. The horizontal axis represents the number of beeps for the gray plot (corresponding to the single flash condition) and the number of flashes for the broken and solid plots (corresponding to the control conditions with a constant number of beeps). The overlap of all three plots at the second data point corresponding to one flash and two beeps for the gray plot and two flashes and one or no beeps for the solid and broken plots, respectively, suggests that the former condition is perceptually equivalent with the latter two so far as visual perception is concerned.

experiment that they could not distinguish the illusory double flash trials from the physical double flash trials. The results of data comparisons taken together with these reports suggest that a single flash accompanied with two beeps is perceptually equivalent to two flashes accompanied with one or no beeps.

3. Experiment 2

Experiment 1 established that the auditory stimuli altered the visual perception. To investigate how distant in time the auditory beeps can be from the flashes and still interfere with visual perception we performed the following experiment. This experiment uses the illusory flashing paradigm of the previous experiments to behaviorally measure the temporal window within which sound can alter the vision.

3.1. Materials and methods

3.1.1. Participants

Eight naïve volunteers participated in the experiment (five females, three males). Their ages ranged from 19 to 27 years. None had participated in Experiment 1. Participants gave their informed consent before inclusion in the study.

3.1.2. Stimuli and procedure

We used the same stimulus configuration as in Experiment 1. But the number of flashes and beeps was now the same across trials. In each trial one flash was ‘accompanied’ by two beeps. One beep was always physically simultaneous with the flash, while the timing of the other

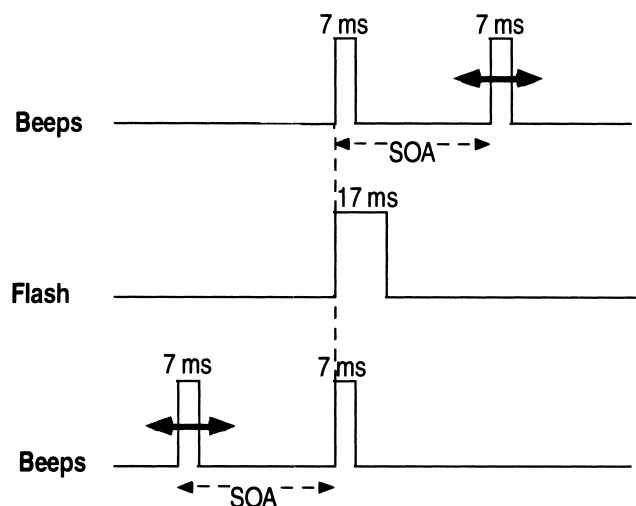


Fig. 6. Temporal profile of stimuli in Experiment 2. In each trial one flash is accompanied with two beeps. One beep is always simultaneous with the flash, but the other can occur either after or before the flash, as depicted in the top and bottom profiles, respectively. The timing of the non-simultaneous beep varies, ranging from 25 to 250 ms (at six equal intervals) before or after the flash.

beep varied from trial to trial with stimulus onset asynchronies (SOAs): 25, 70, 115, 160, 205, 250 ms either before or after the flash (see Fig. 6). The observer’s task was to judge the number of flashes he/she sees on the screen in a 2-AFC paradigm (one or more flashes). Each participant was presented five sets of each combination amounting to a total of 60 trials. The order of the 60 trials was random.

3.2. Results

The results of Experiment 2 are displayed in Fig. 7. The vertical axis represents the percentage of trials in which observers saw more than one flash. This measure can be thought of as amount or strength of the illusion. The horizontal axis denotes the timing of the variable-time beep from the flash. Zero denotes the timing of the flash and positive and negative numbers reflect the timing of the variable beep when occurring after or before the flash, respectively. The illusion starts degrading from ± 70 ms onwards, however, it is still strong (at about 33% and 23%) at ± 115 ms. This ~ 100 ms temporal window of interaction is interesting as it is consistent with integration window of polysensory neurons in the mammalian brain [6].

4. General discussion

The results of the two experiments described above (as well as other observations) dismiss possible alternative explanations for the observed illusory flash effect. The illusory flash phenomenon does not seem to be due to

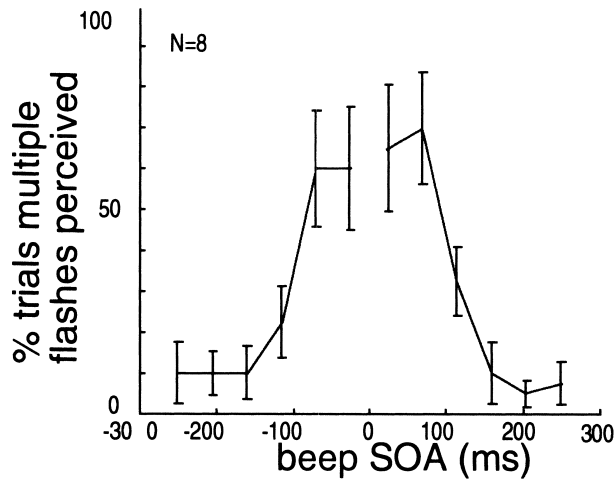


Fig. 7. Results of Experiment 2. The horizontal axis represents the timing of the variable-time beep from the flash. Zero denotes the time of the flash and positive and negative numbers denote the time of the variable beep when it occurs after or before the flash, respectively. The vertical axis is a measure of the strength of the illusion. The illusion remains strong within 115 ms of the flash.

general attentional enhancement caused by auditory stimulation, as there is no illusory flash elicited by a single beep (Experiment 1). It is not due to eye movements, as the effect is stronger with shorter flash durations (data not shown here), persists with very large disk size, and degrades with decrease in disk contrast. It is not caused by cognitive biases as shown by catch trials in Experiment 1. Other results and observations also dismiss a cognitive top down origin: the illusion vanishes when the second beep falls outside the window of interaction (Experiment 2) and gets stronger with increased eccentricity of the disk in the visual field.

Thus, the only explanation for the findings is that the auditory stimuli (beeps) altered the percept of the visual stimulus (flash) through bimodal perceptual interaction. This alteration is most conspicuous in the case of a single flash accompanied with multiple beeps perceived as multiple flashes. The reverse modulation, that is, the fusing of two physical flashes into one, when accompanied by a single beep, is negligible, however. This asymmetry in modulation is interesting, because it cannot be explained by the ‘modality appropriateness’ hypothesis, a well-established theory which holds that the direction of crossmodal interactions depends on the ‘appropriateness’ of the involved modalities for the given task; whichever modality is more attuned for carrying out a given task will dominate in that context. The modality appropriateness hypothesis cannot explain the asymmetry in our data, as neither the task (judging the number of flashes on the screen) nor the modalities involved (vision and audition) were changed across the conditions. The results suggest instead that the direction of the crossmodal interactions depends at least partly on the characteristics of the stimuli.

Interestingly, we found the same type of asymmetry in the data reported in another paper [8] although the modalities of the stimuli were the opposite (vision altering audition). We noticed in the published data that the influence of the discontinuous stimulation (cello plucking video) on the percept of the continuous stimulus (bow sound) was much stronger than the effect of continuous (bowing video) on discontinuous stimulus (plucking sound). These results taken together suggest that the dependency of the crossmodal interactions on the stimulus nature may be characterized as follows: the discontinuous stimulus in one modality alters the percept of the continuous stimulus in the other modality and not as strongly vice versa (Fig. 8).

Finally, we would like to address the relationship between our findings and a phenomenon referred to in the literature as ‘auditory driving’ [2,7,12,15]. The phenomenon can be described as follows: the frequency of a fluttering sound influences the perceived frequency of a flickering light. It should be pointed out that auditory driving does not necessarily imply that the auditory flutter breaks a single flash into two or more flashes resulting in a perceived higher flicker frequency. An alternative and simpler explanation for this phenomenon is that the perceived duration of each flash or the gap between two successive flashes is altered by accompanying flutter. Indeed such alteration of duration and gap of flashes by sound has been shown in other studies [14]. A recent study argues that the temporal modification of the visual percept is the mechanism underlying auditory driving by demonstrating that the auditory flutter ‘drives’ the perceived timing of the flashes [1]. Moreover, auditory driving works symmetrically, i.e. flutter is as effective in making the flicker rate perceived lower as it is in making it perceived higher. This effect cannot be accounted for by breaking a

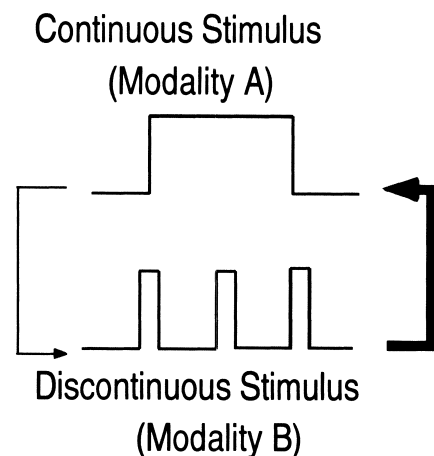


Fig. 8. Dependency of the direction of crossmodal interactions on the characteristics of stimuli. The discontinuous stimulus in one modality is highly effective in changing the percept of the continuous stimulus in another modality, but not vice versa.

flash into two but rather would require perceptual fusion of two flashes into one, an effect which we found to be quite weak. We therefore suspect that auditory driving is primarily based on modification of perceived temporal attributes of the visual stimulus [1], and thus, not entirely related to sound-induced illusory flash phenomenon.

5. Conclusion

A single flash accompanied by multiple beeps is perceived as multiple flashes. This phenomenon clearly demonstrates that sound can alter the visual percept qualitatively even when there is no ambiguity in the visual stimulus. The settings within which this radical alteration of vision by sound occurs is not at all convoluted or complex. The stimuli and the task used in our experiments were both very simple. More importantly, the illusion was found to be surprisingly robust to variation of many parameters. Moderate manipulation of the relative and absolute timings of the auditory and visual stimuli, the shape, color, brightness and size of the visual flash, the frequency and intensity of the auditory beeps, the spatial disparity between beeps and flash, and so forth do not disrupt the illusion. Such degree of robustness as well as simplicity of the eliciting stimuli suggests that sound-induced illusory flashing reflects the working of a mainstream circuitry in the brain as opposed to accidental or marginal neural activity. They also suggest that the cross-modal interactions may be the rule rather than the exception in our perception of the world. If true, our understanding of these interactions is an integral part of understanding perception. We are currently investigating the neural levels at which audio–visual interactions occur, and thereby hope to get new insights into the mechanisms involved in the multimodal integration of perceptual information.

Acknowledgements

This work was supported by NIH grant HD08506.

References

- [1] R. Fendrich, P.M. Corballis, Auditory capture of the timing of visual events. *Investigative Ophthalmology and Visual Science*, Vol. 40, Fort Lauderdale, FL, 1999, p. S47.
- [2] J.W. Gebhard, G.H. Mowbray, On discriminating the rate of visual flicker and auditory flutter, *Am. J. Psychol.* 72 (1959) 521–528.
- [3] J.C. Hay, H.L. Pick, K. Ikeda, Visual capture produced by prism spectacles, *Psychonomid. Sci.* 2 (1965) 215–216.
- [4] I.P. Howard, W.B. Templeton, *Human Spatial Orientation*, Wiley, London, 1966.
- [5] H. McGurk, J.W. MacDonald, Hearing lips and seeing voices, *Nature* 264 (1976) 746–748.
- [6] M.A. Meredith, J.W. Nemitz, B.E. Stein, Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors, *J. Neurosci.* 10 (1987) 3215–3229.
- [7] A.K. Myers, B. Cotton, H.A. Hilp, Matching the rate of concurrent tone bursts and light flashes as a function of flash surround luminance, *Percept. Psychophys.* 30 (1981) 33–38.
- [8] H.M. Saldaña, L.D. Rosenblum, Visual influences on auditory pluck and bow judgments, *Percept. Psychophys.* 54 (1993) 406–416.
- [9] C.R. Scheier, R. Nijwahan, S. Shimojo, Sound alters visual temporal resolution, *Investigative Ophthalmology and Visual Science*, Vol. 40, Fort Lauderdale, FL, 1999, p. S4169.
- [10] R. Sekuler, A.B. Sekuler, R. Lau, Sound alters visual motion perception, *Nature* 385 (1997) 308.
- [11] L. Shams, Y. Kamitani, S. Shimojo, What you see is what you hear, *Nature* 408 (2000) 788.
- [12] T. Shipley, Auditory flutter-driving of visual flicker, *Science* 145 (1964) 1328–1330.
- [13] B.E. Stein, N. London, L.K. Wilkinson, D.D. Price, Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis, *J. Cognit. Neurosci.* 8 (1996) 497–506.
- [14] J.T. Walker, K.J. Scott, Auditory–visual conflicts in the perceived duration of lights, tones and gaps, *J. Exp. Psychol.: Hum. Percept. Perform.* 7 (1981) 1327–1339.
- [15] R.B. Welch, L.D. Duttonhurl, D.H. Warren, Contributions of audition and vision to temporal rate perception, *Percept. Psychophys.* 39 (1986) 294–300.