Spatial Congruity in Audiovisual Synchrony Judgments

Rainer Guski
Dept. of Psychology, Ruhr-University Bochum, Germany
Rainer.Guski@uni-bochum.de

Abstract
The systematic analysis of the perception of audiovisual synchrony has shown that auditory delays are tolerated to a certain extent in synchrony judgments, but the variation of spatial separation between light and sound has brought conflicting results: While Lewald & Guski (2003) did not find a significant influence of spatial separation in synchrony judgments, Spence et al (2003) did find one in temporal order judgments. This may be due to methodological differences between synchrony and temporal order judgments, but it may also be due to spatial congruity, which was randomized by Lewald & Guski, and was confounded with spatial separation by Spence et al. (2003). The present contribution asks if audiovisual synchrony judgments are influenced (1) by the sound reproduction system, and (2) by the spatial congruity of light and sound. The results of an experiment are presented that asks participants for quick Yes/No-Responses with respect to audiovisual synchrony and varies systematically three factors: (a) the audiovisual synchrony (-250 to +250 ms), (b) the sound reproduction system (earphones vs. loudspeakers), and c) the spatial congruity (left and right) of light and sound stimuli. The results do not show a systematic influence of spatial compatibility of the audiovisual stimuli on raw synchrony judgments.

Introduction
Audiovisual judgments of synchrony or temporal order depend on a number of influences, e.g. the type of stimuli used (with speech producing greater just noticeable differences (JNDS) than abstract stimuli), the type of task (with synchrony judgments (SJ) producing greater JNDS than temporal order judgments (TOJ)). The spatial separation between light and sound is also claimed to have an influence (Bertelson & Aschersleben 2003; Spence et al. 2003; Spence & Read 2003; Zampini et al. 2003a,b), but Lewald & Guski (2003) did not find a significant influence of spatial separation on synchrony judgments. The difference between results from different laboratories may be due to methodological differences between synchrony and temporal order judgments, but it may also be due to spatial congruity, which was randomized by Lewald & Guski (2003), and was confounded with spatial separation by Spence et al. (2003). The present contribution asks whether audiovisual synchrony judgments and decision times for these judgments are influenced by the spatial congruity of light and sound.

Brief overview previous experiments
Bertelson & Aschersleben (2003) presented sound bursts and light flashes with varying stimulus onset asynchronies (SOAs) delivered either in the same or in different locations. Participants judged the order of occurrence (TOJ) without feedback. Stimulus presentation was organized using randomly mixed psychophysical staircases, by which the SOA was reduced progressively until a point of uncertainty was reached. This point was reached at longer SOAs with the sounds in the same frontal location as the flashes than in different places, showing that spatial disparity increases the slope of audiovisual TOJs. The authors conclude that "timing and spatial layout of the inputs play to some extent interchangeable roles in the pairing operation at the base of crossmodal interaction" (p.147).

Lewald & Guski (2003) presented flashing light spots from LEDs and tone bursts from loudspeakers with various spatio-temporal disparities. The position of the LED and the active loudspeaker varied randomly (+/-20 deg in 4 deg steps). The SOA between light and sound varied randomly too (+/-250 ms in steps of 50 ms). Subjects either scaled their impression of the likelihood of a common cause (Exp 1) or spatial alignment (Exp 2) or synchrony of sound and light (SJ, Exp 3). In all three experiments the subjects' judgments depended significantly on temporal disparity whereas influences of spatial disparity were significant only in Exp 1 and 2. That is, the spatial disparity had no systematic influence on SJs.

Spence et al. (2003) presented pairs of 8 ms light spots from two LEDs and 8 ms white noise from loudspeakers (82 dBa with 75 dBa background noise) at varying SOAs (+/-200 ms in 10 unequal steps without physical synchrony). The LEDs and loudspeakers were 26 cm apart (distance from the participants = 62 cm), the positions of light and sound varied randomly, i.e., in half of the trials light and sound had the same position, and in the other half, they occupied different positions. The 10 participants were to give TOJ without feedback. After rejecting the data of two participants because of a high "attentional blink rate", data were fitted to a psychophysical model (using the Swanson & Birch 1992 procedure), and PSS (point of subjective simultaneity) and JNDS were calculated. It turned out that PSS was significantly different from Zero only for the different-position data, and not for the same-position data. JNDS were significantly smaller (41.5 ms) when sound and light occupied different positions than when they occupied the same positions (53.3 ms). The authors conclude that "that people can use redundant spatial cues to facilitate their performance on multisensory TOJ tasks" (p.318).

Taken together, the three results reported so far seem partially conflicting: Two audiovisual TOJ tasks show a small but significant influence of the spatial position of stimuli, and one SJ task does not show a significant spatial influence. The difference may be due to several factors, the most obvious being the judgment task, another may be the response mode (dichotomous vs. scaling), and a third may be the spatial compatibility of the audiovisual stimuli: Lewald & Guski (2003) varied the positions of light and sound independently, while the
others explicitly used same/different locations as an independent variable.

A new experiment
We first concentrated on the location compatibility as a potential influence on synchrony judgments. Auditory and visual stimuli with varying SOAs were presented on either the same side or on different sides, and participants were to judge the audiovisual synchrony as fast as possible. It was hypothesized that synchrony judgments are both easier to perform (i.e., require less decision time), and are less tolerant for physical asynchrony in the case of spatial incongruity than in the case of spatial congruity.

Apparatus and procedure: Flashes of big white vertical stripes were presented for 30 ms on a 15” TFT screen and 800 Hz bursts were presented either via loudspeakers or headphones. The location of the sound was either the same or different side as that of the visual stimulus. The visual stimulus was a white 100 X 300 pixel stripe, appearing either left or right on the screen (+/- 18 degrees from the fixation point in the middle of the screen) on a black background. The sound was presented either by means of loudspeakers to the right and left of the screen (+/- 45 degrees from the fixation point), or via headphones (Sennheiser K500, +/- 90 degrees from the fixation point). The level sound of the headphone stimuli was 75 dB(A), measured with an artificial ear in “fast” position (B&K 2226), the level of the loudspeaker stimuli was 70 dB(A), measured at the head of the participants. The physical difference between headphone and loudspeaker levels controls for the poor sound attenuation of the laboratory room and was judged to match headphone and loudspeaker loudness. The stimuli were controlled by means of a 2 MHz PC and a custom-written program using the “Presentation” programming language. The experimental system was measured to reach a precision of +/- 2.4 ms, even with simulatenous video and sound output. The method of constant stimuli was used, and SOAs varied between +/- 250 ms in equal steps of 50 ms, including Zero SOA. The order of SOAs and the audiovisual locations were randomized within the trials of each participant. The source of the audio stimulus (loudspeakers and headphones) was blocked, and the order of the source randomized within each participant. Each of the 22 SOA-location combinations was presented 5 times. The participants were to judge audiovisual synchrony by means of two response keys (yes/no).

22 participants (psychology students) served as subjects for course credits. They were naive with respect to the experimental questions, and the concept of audiovisual location compatibility was never mentioned. They were first given a systematic overview of the range of SOAs and then practiced 20 trials with feedback, and another 20 trials without feedback. The participants were told that the auditory and the visual stimuli could appear on different sides, but that this were of no importance for the task. They should respond as fast as possible, and their decision times were measured. Results: After discarding 5 participants because they constantly were unable to discriminate between synchronous and asynchronous audiovisual events over a range of 250 ms SOA, the individual and collective synchrony judgments and decision times of 17 participants show systematic relations to SOA in each of the presentation conditions. But the raw data do not show a systematic effect of location compatibility – neither in the loudspeaker condition, nor in the earphone condition. Figure 1 gives the mean probability of synchrony judgment in the loudspeaker condition as a function of SOA and location compatibility. Both compatible and incompatible audiovisual locations show maximal synchrony judgments at Zero SOA, and a monotonous decrease of judged synchrony with increasing SOA. In the incompatible condition, the peak of synchrony judgments at Zero SOA appears somewhat sharper than the comparable peak in the compatible condition, but there is neither a systematic difference of the means of the individual synchrony judgments at mid-range SOAs (-100 .. +100 ms) nor of the weighted sums of mid-range SOAs between compatible and incompatible audiovisual locations. This holds both for loudspeaker, and earphone conditions. Future statistical analyses will explore the relations between synchrony judgments and spatial audiovisual compatibility by means of fitting the data to different psychophysical models (e.g., that proposed by Swanson & Birch 1992).

Figure 1: Mean synchrony judgments in the loudspeaker condition. Explanation: ”comp“ means: audio and visual signals from the same side; ”incomp“: audio and visual signals from contrary sides.

Reaction times (or decision times) show a systematic bimodal relation to SOAs (with minima at very big, and very small SOAs), and there seems to be a tendency for compatible audiovisual locations to show a somewhat broader minimum of decision times near threshold, as compared to the compatible audiovisual locations (see Fig. 2). But the statistical analysis does not show a
significant effect of the spatial audiovisual compatibility, neither with loudspeakers nor with headphones.

Figure 2: Mean decision times for spatial compatible and incompatible audiovisual locations as a function of SOA in the loudspeaker condition.

Discussion

So far, the statistical analyses of raw data do not show a significant influence of the audiovisual spatial compatibility. This is in contrast with both the Bertelson & Aschersleben (2003) study and all of the studies published by the Spence-group in 2003. But it is in accordance with the studies by Lewald & Guski (2003). The present study and the studies reported by other laboratories differ with respect to (a) the type of judgment (synchrony vs. temporal order), (b) the spatial alignment of auditory and visual stimuli, and (c) the use of psychophysical models to handle raw data.

(a) Type of judgment (synchrony vs. temporal order): Bertelson & Aschersleben (2003) and all of the Spence studies use TOJ, while Lewald & Guski (2003, Exp. 3) and the present experiment used SJ. Generally, it is believed that TOJs pose different questions to perceivers than SJs (Exner 1875; Hirsh & Sherrick 1961), and it seems that perceivers are somewhat more tolerant to audiovisual asynchrony in synchrony judgments than they are in judging temporal order (Kohlrausch & VandePar 1999, 2000; VandePar & Kohlrausch 2000). It may well be that perceivers are also more tolerant with respect to discrepancies in audiovisual spatial locations when judging synchrony as compared to temporal order. Fraisse (1957, p.110) points out that "the perception of simultaneity is easiest to obtain when stimuli occur at the same place".

(b) Spatial alignment of auditory and visual stimuli: Bertelson & Aschersleben (2003) and all of the Spence studies use identical locations for audio and visual signals, while Lewald & Guski (2003) vary the audio location independently of the visual location, and the present experiment uses greater distances between the visual fixation point and the audio locations than between the fixation point and the visual locations. The use of identical audio and visual locations may have helped the perceivers in TOJs to discriminate between same and different locations, and this in turn might have helped to discriminate the temporal order between audio and visual stimuli. On the other hand, using spatial alignment confounds "location" with "side" and cannot be used to study the effect of spatial compatibility, which was the purpose of our present study. Future studies will have to explore possible differences between "location identity" and "spatial compatibility" in audiovisual synchrony and temporal order judgments.

(c) Use of psychophysical models to handle raw data: Spence et al. (2003) explicitly point out that they extended the common procedure of fitting data to a psychophysical model by means of the Weibull function (see Quick 1974) by assuming "extraneous noise" (by inattentiveness of the participants) to operate: "If the subject responds to irrelevant aspects of the experimental situation or mistakenly makes the wrong response even though the stimulus was detected, R(x) (the probability of a correct response for 2AFC with stimulus level x) may never reach 100 % correct, even if the range of stimulus levels is correct and the number of trials is large. To model the effects of extraneous noise, Equation 1 ((the usual way of calculating the probability of a correct response)) was modified: R(x) = gamma * P(x) + 0.5 [1-P(x)], in which gamma is the upper asymptote for R(x)" (Swanson & Birch 1992, p.410). This model assumes that the poor performance of some participants in "easy" trials can be treated as statistical errors and be "corrected" by using individual psychophysical models, depending on individual maximum discrimination. The statistical analyses presented in this paper refrain from fitting data to models in order to see whether there is a strong effect of audiovisual location compatibility. The analyses did not show any effect, and we believe that if there is any effect at all, it is not very strong.

(d) Use of a video screen vs. LED: All former studies quoted in this paper used LEDs for visual signals, and the present study used a big white rectangle on a video screen instead. LEDs can be easily aligned with loudspeakers, they have almost no delay, but they are of restricted size and intensity, compared with video signals on computer screens. Size and intensity of stimuli do have a relation with spatial position: small and dim video signals seem to be farther away than big and bright signals (Sedgwick 1986); and loud audio signals seem to be closer than soft audio signals (Warren 1977). Although it is very difficult to predict the comparative distance of visual and audio signals without distance judgments of perceivers, we believe that single LEDs produce greater perceived distance than bright light video signals at the same physical distance.
References


