Exp Brain Res (2009) 198:233–244
DOI 10.1007/s00221-009-1917-z

RESEARCH ARTICLE

Stimulus duration influences perceived simultaneity in audiovisual temporal-order judgment

Lars T. Boenke · Matthias Deliano · Frank W. Ohl

Received: 30 September 2008 / Accepted: 17 June 2009 / Published online: 10 July 2009

Abstract The temporal integration of stimuli in different sensory modalities plays a crucial role in multisensory processing. Previous studies using temporal-order judgments to determine the point of subjective simultaneity (PSS) with multisensory stimulation yielded conflicting results on modality-specific delays. While it is known that the relative stimulus intensities of stimuli from different sensory modalities affect their perceived temporal order, we have hypothesized that some of these discrepancies might be explained by a previously overlooked confounding factor, namely the duration of the stimulus. We therefore studied the influence of both factors on the PSS in a spatial-audiovisual temporal-order task. In addition to confirming previous results on the role of stimulus intensity, we report that varying the temporal duration of an audiovisual stimulus pair also affects the perceived temporal order of the auditory and visual stimulus components. Although individual PSS values varied from negative to positive values across participants, we found a systematic shift of PSS values in all participants toward a common attractor value with increasing stimulus duration. This resulted in a stabilization of PSS values with increasing stimulus duration, indicative of a mechanism that compensates individual imbalances between sensory modalities, which might arise from attentional biases toward one modality at short stimulus durations.

L. T. Boenke · M. Deliano · F. W. Ohl (✉)
Leibniz Institute for Neurobiology (IfN),
Brenneckestr. 6, 39118 Magdeburg, Germany
e-mail: frank.ohl@ifn-magdeburg.de

F. W. Ohl
Otto von Guericke University, Magdeburg, Germany

Keywords Audiovisual · Crossmodal · Temporal-order judgment · Temporal processing · Time perception

Introduction

For the meaningful interpretation of objects and events in a complex environment, neural systems can take advantage of integrating cues from different sensory modalities. Psychophysical studies in humans have demonstrated various forms of crossmodal integration and interaction (for review see Calvert et al. 2004; Spence and Driver 2004). To perceive a compound, multisensory event, temporal differences in the occurrence of multimodal cues, which are thought to arise mainly from modality-dependent physical transmission and sensory processing times, have to be compensated for (for the integration of audiovisual stimuli: Alais and Carlile 2005; Lewald and Guski 2004; Spence and Squire 2003; Sugita and Suzuki 2003). Since the second half of the nineteenth century with Exner (1875), who was among the first to carry out systematic psychophysical experiments on the perceived temporal relationship of bimodal signals, there has grown a large body of psychophysical literature dealing with the temporal aspects of multisensory integration. Thereby, many studies determining the point of subjective simultaneity (PSS), i.e., the relative timing of a stimulus in one sensory modality to a stimulus presented in another modality, for being perceived as simultaneous, have shown that the actual value of the PSS depends on various factors such as stimulus intensity (Neumann et al. 1992; Smith 1993), task conditions (Shore et al. 2001; van Eijk et al. 2008; Zampini et al. 2005a, b) and cognitive factors (Johnston and Nishida 2001; Titchener 1908; Zampini et al. 2003a; for reviews see: Jaśkowski 1996; Neumann and Niepel 2004; Spence et al. 2001).

Moreover, exposure to asynchronous events in two different sensory modalities can lead to a recalibration of the perception of temporal order (Fujisaki et al. 2004; Vatakis et al. 2007, 2008; Vroomen et al. 2004). Compared to other modality pairings, this recalibration effect is highly pronounced in the processing of audiovisual stimuli, as indicated by prominent plastic changes in temporal integration of asynchronous auditory and visual cues (cf. Harrar and Harris 2008; but see Hanson et al. 2008).

As the actual value of the PSS depends on various factors, it is not surprising that the crossmodal PSS values determined in different studies vary to a large degree, yielding conflicting results on the relative timing of stimuli in different modalities required for simultaneous perception. Still, some studies claim that, at least in TOJ studies, there exist general patterns of temporal integration, which are specific for two interacting sensory modalities. For example, in their review, Neumann and Niepal (2004) stated that the typical pattern observed in TOJ tasks is that acoustic stimulation has to precede visual stimulation in order to be perceived as simultaneous (cf. Neumann et al. 1992; Rutschmann and Link 1964; Smith 1933). One might conclude from this that physiological and psychological processes leading to the detection or perception of a stimulus act more rapidly for a visual stimulus. However, Jaśkowski (review: 1996) and more recently Zampini et al. (2003a, 2005b) have demonstrated the opposite effect, i.e., visual stimulation has to precede acoustic stimulation to be perceived as simultaneous, suggesting that in their TOJ tasks, auditory perception is faster than visual perception.

To determine the relative processing speed for different sensory modalities, most of these studies have chosen fixed settings of stimulus parameters and task conditions, which differ between studies. This might provide an explanation for the aforementioned conflicting results, as the results indicate that the PSS is not fixed in a modality-specific way, but instead depends on stimulus and task conditions. One factor influencing the PSS in TOJ tasks seems to be the relative stimulus intensity. In the studies of Jaśkowski et al. (Jaśkowski et al. 1990; Jaśkowski 1996) and Zampini et al. 2003a, 2005b), reporting that the visual stimulus has to be presented before the auditory stimulus for perceived simultaneity, the intensities of the acoustic stimuli were generally higher [$>\sim$70 dB(A)] and/or the visual stimulus intensities were lower than in the studies of Smith (1933) and Rutschmann and Link (1964) arguing for the opposite. Neumann et al. (1992) further determined audiovisual PSS values for combinations of three different sound intensities and three different light intensities. Their results indicate that the stimulus of higher intensity was perceived earlier than the stimulus of lower intensity, and that the magnitude of this effect is a positive function of the intensity difference between the stimuli. Despite this intensity effect, Neumann and Niepel maintained that for TOJ tasks, the general pattern of observation indicates that the auditory stimulus has to be presented before the visual stimulus for perceived simultaneity (Neumann and Niepel 2004, p. 255), although with high sound intensities, auditory perception in their experiment finally became faster than visual perception, comparable to the results in the experiments of Jaśkowski et al. (Jaśkowski et al. 1990; for review: Jaśkowski 1996) and Zampini et al. (2003a, 2005a, b).

However, relative stimulus intensity does not seem to fully account for the observed discrepancies between the studies of Neumann et al. (1992) and Zampini et al. (2003a, b, 2005b). In the former study, audiovisual (AV) stimuli were presented in pairs with all combinations drawn from the set of auditory intensities of 45, 48 and 68 dB(A), and of visual intensities of 0.03, 0.1 and 4.93 cd/m^2. The latter study used intensities of 82 dB(A) for auditory, and 64.3 cd/m^2 for visual stimulation. This suggests that, besides relative stimulus intensity, other factors might influence the crossmodal PSS. Notably, in TOJ studies concluding that the visual stimulus has to precede the auditory one for perceived simultaneity (i.e., positive PSS values; e.g., Jaśkowski et al. 1990, Jaśkowski 1996; van Eijk et al. 2008; Zampini et al. 2003a, b, 2005a, b), typically much shorter stimuli ($\leq$12 ms in the cited studies) had been used than in the studies showing a visual lead (i.e., negative PSS values) for 40 ms stimulus duration (Neumann et al. 1992). While other factors such as interindividual differences (Stone et al. 2001), retinal position of the visual stimulus and psychophysical task (van Eijk et al. 2008) might have various effects on the PSS (see "Discussion"), a longer stimulus duration might be hypothesized to contribute to perceptual speed in advantage of the visual system.

Hence, to further explain the observed variability of PSS values obtained in different TOJ tasks across laboratories, we conducted a psychophysical experiment to investigate the combined effects of stimulus duration and intensity on the temporal integration of AV stimuli using an AV-TOJ task. For this purpose, we used AV stimuli with duration (9, 40 and 500 ms) and light intensity (0.14 and 0.64 cd/m²), which varied independently. Based on the literature cited above, we expected that an increase in stimulus duration would reduce the necessary presentation delay of the auditory stimulus relative to the visual stimulus for perceived simultaneity measured by a TOJ paradigm.

Methods

Participants

A total of 29 naïve healthy participants from the participant pool of the Leibniz Institute for Neurobiology (18 females) aged between 19 and 32 years (mean 23.3 ± 0.5 SED) took part in the experiment. Of these, 28 participants were right-handed, and one left-handed. All participants reported normal or corrected-to-normal vision, and gave their written consent. The Ethical Committee of the Otto-von-Guericke University in Magdeburg had given its permission for the study. The study conformed to the 1964 Declaration of Helsinki.

Apparatus

The experiment was conducted in a dark and sound attenuated room with an ambient sound level of 29 dB(A). To avoid unintended reflection of the light signal, the setup was covered with a black velvet cloth. A green light-emitting diode (LED), which was placed in front of the participants at eye level and at a distance of about 165 cm was used for fixation. Two boxes were placed symmetrically, left and right of the center of the LED (slightly below the horizontal line). These two boxes contained a sound source, a speaker with white coverage, and a light source, a white LED. The white LED was placed above the speaker. In front of the speaker was an aperture (diameter 4 cm), which allowed the light and sound source to be perceived, stemming virtually from the same location. The centers of the apertures of the boxes were approximately at a distance of 38 cm from the center of the fixation LED, corresponding to a visual angle of ~12.5°. The auditory stimuli consisted of white noise bursts with 1.5 ms onset and offset ramps. The visual stimuli were presented by white LEDs with a time course of rectangular shape. Responses were given on a specifically designed hand-held response box. The stimuli were presented and recorded by a program written in psychtoolbox (PT-2) within Matlab R14 environment and stored on an IBM 486-compatible microcomputer. The timing of the stimuli was controlled by a National Instruments card (PCI-6071E, Austin, TX, USA) with checked high temporal precision and presentation accuracy better than 1 ms.

Stimuli and design

We conducted a TOJ task using the method of constant stimuli. In literature, there are two different basic varieties of crossmodal TOJ tasks. Participants can be asked to report the modality they perceived first, when light and sound are presented at a single or two separate spatial locations (cf. Zampini et al. 2003a). Alternatively, participants can be asked to report the location of the first stimulus, on- or offset, when light and sound are delivered from two spatially different sources. Here, we decided on the latter and asked for the side of stimulus onset to avoid bias to one modality (see Spence et al. 2001; Zampini et al. 2003a). In each trial, one auditory and one visual stimulus were presented (AV stimulus pair). One AV stimulus pair always consisted of an auditory and a visual stimulus with the same duration (see below). One of the two stimuli was presented on the left side and the other on the right side. The applied auditory and visual stimuli had the following stimulus onset asynchronies (SOAs): ±20, ±60, ±100, ±140 and ±240 ms (negative values indicate that the auditory stimulus was leading). There were two orthogonal factors: "duration" (9, 40, and 500 ms) and "light intensity" (high intensity versus low intensity; see Fig. 1). The first two durations were chosen because they were used in former studies (see above), and the long duration was chosen as a control beyond the critical durations for both modalities where strong temporal summation effects were expected (for vision cf. Nilsson 2006; for audition cf. Moore 2003; for an alternative view on temporal summation effects see Heil and Neubauer 2003; Heil et al. 2008; Neubauer and Heil 2008). Stimulus intensities were determined in a pilot study to obtain a good level of comfort for the participants with respect to the 500 ms light flash of high intensity, and the 500 ms noise burst. No attempt for further matching between auditory and visual stimuli was carried out (Zampini et al. 2003a). All auditory stimuli were kept constant in intensity at 50 dB(A), which was measured at the

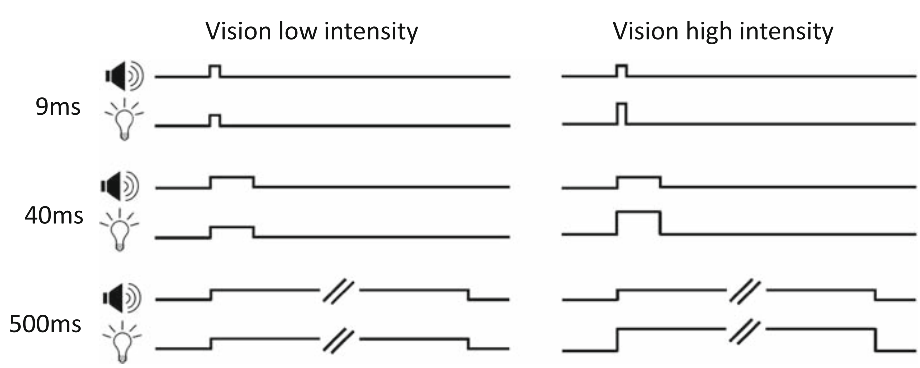

Fig. 1 Schematic illustration of the experimental design. Two conditions of relative stimulus intensities (columns 0.14, 0.64 cd/m^2) were combined with three conditions of stimulus duration (rows 9, 40, 500 ms duration). The auditory (*speaker symbol*) and visual (*light bulb symbol*) stimulus elements were separated by a stimulus onset asynchrony of 20–240 ms (not shown)

participants' head position with a sound level meter. For the light flashes, an intensity of 0.64 cd/m^2 was chosen for the visual high-intensity condition, and 0.14 cd/m^2 for the visual low-intensity condition. Notably, the terms "high" and "low" here refer to a relative intensity difference between conditions, and do not qualify as absolute light intensities. For technical reasons, luminance and sound pressure level were obtained for stimuli with a duration of 5 s. Therefore, to avoid any immeasurable technical intensity bias for the shortest duration between the left and the right source of auditory and visual stimuli, the boxes containing loudspeakers and LEDs were swapped in half of the participants. All configurations of trials were repeated 20 times and were pseudo randomized, in such a way that no more than three identical configurations occurred in succession. The experiment was performed on 1 day in four experimental blocks yielding a total of 2,400 trials [5 SOAs (20, 60, 100, 140, 240 ms) × 2 stimulus first (A, V) × 2 side (left, right) × 3 durations (9, 40, 500 ms) × 2 intensity conditions (A–V (low intensity), A–V (high intensity)) × 20 repetitions].

Procedure

A green fixation LED (0.5 cd/m^2) was illuminated throughout the experiment and participants were instructed to maintain their view on the LED during stimulus presentation. In each trial, participants were asked to report on which side (left or right) they perceived the first out of two stimulus onsets by pressing the button on the corresponding side of a handheld response device with their left or right thumb. Participants were asked to respond as accurately as possible. Further, they were instructed to make their best guess in cases of uncertainty about the correct sequence. The first trial started 4–5 s after a verbal instruction. The following trials were self-paced and could be interrupted for a little break at any time the participant needed. After a response given by pressing one of the defined response buttons, an intertrial interval was initiated with a duration of between 1.5 and 2 s chosen randomly from a uniform distribution.

Each experiment was preceded by a training block lasting approximately 10 min, which also served as adaptation for the participants. At the very beginning of this training block, the participants were asked to indicate their response verbally, and feedback was given for the first five presentations of ±240 ms SOAs to ensure that the participants understood the task. Training blocks were not further analyzed.

Results

In our experiment, a spatial TOJ task was preferred over other paradigms to minimize the possibility of a modality bias (Zampini et al. 2003a). By randomizing and balancing the side on which auditory or visual stimuli were presented first, we prevented a location bias (Scharlau 2004). Since the intensity conditions were also randomized and balanced across sides, we could exclude further a possible bias toward low intensities (Jaśkowski and Verleger 2000). Furthermore, one type of response bias, which is often referred to in the context of prior-entry research (Frey 1990; Hikosaka et al. 1993; Jaśkowski 1993; Rorden et al. 1997; Stelmach and Herdman 1991) is related to a possible preactivation of the response side by a cue (Shore et al. 2001, p. 205). We will refer to it as exogeneous response bias. This exogeneous response bias is not expected to play a role in our study, because there was no consistent cue predicting particular stimulus conditions due to our randomized presentation. Since the duration and intensity conditions were balanced across sides, endogenous response bias could be quantified by the proportion of "right" responses on the right side. In 3 (of the 29) participants this proportion exceeded the expected value (50%) by a predefined criterion of 5%. These participants were excluded from further analysis. The remaining 26 participants were then tested for side-specific differences in temporal perception. For each participant, we calculated a rank correlation between response proportions on the left and the right side across SOAs, separately for each duration and intensity condition. Only participants with significant ($p < 0.05$) rank correlation in each condition were included in the analysis. By this, two further participants had to be excluded from the analysis. Thus, in the remaining 24 participants, we were able to exclude any endogenous response bias and were able to pool the data irrespective of the side of first stimulation. This increased the number of observations from 20 to 40. The demonstration that the great majority of participants does not show side-specific differences in temporal perception is in line with other studies (Keetels and Vroomen 2005; Zampini et al. 2005a).

To estimate the PSS values for each of the participants and for each of the six experimental conditions (2 light intensities × 3 durations), psychometric functions for the response "vision first" were calculated using a Bayesian inference procedure developed by Kuss et al. (2005). In this procedure, lapses are taken into account. In 2 out of the remaining 24 participants, psychometric functions were flat and the estimated PSS exceeded the SOA range tested (>240 ms). This indicates that these two participants had difficulties in handling the task. Thus, these participants were excluded from further analysis (see Zampini et al. 2003b for similar exclusion criteria). For all experimental conditions, mean and standard errors of PSS determined from these psychometric functions across all remaining 22 participants are displayed in Fig. 2. Values of mean and the standard error of the PSS values for each experimental

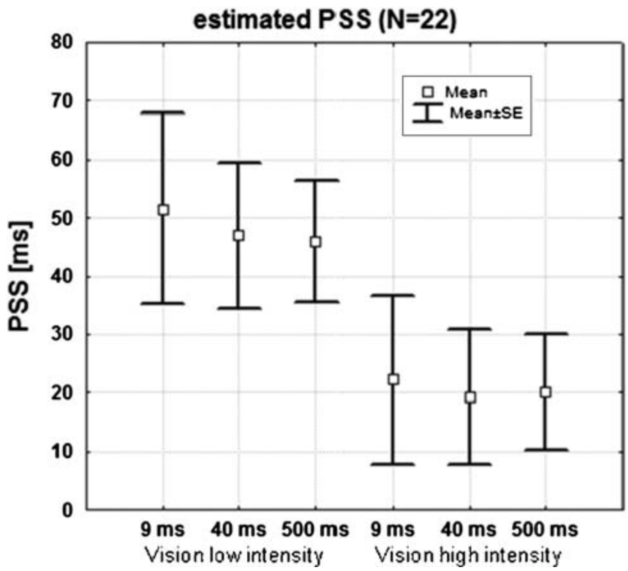

Fig. 2 Grand mean values for the points of subjective simultaneity (PSS) and standard error of mean across experimental conditions indicated by the abscissa labels. The *left* three conditions are of vision low intensity and the three conditions on the *right* of vision high intensity. Positive PSS means that the physical visual stimulus onset has to precede the physical auditory onset for a simultaneous perception of the stimuli. An effect of intensity variation on the PSS is evident. An effect of duration variation on the PSS is not evident in the analysis of mean values, but see Figs. 4 and 5

Table 1 Grand mean values of the point of subjective simultaneity (PSS) and standard errors of means (SE) for the different experimental conditions

Experimental condition	PSS ± SE (ms)
Vision high intensity (VH) 9 ms	22 ± 15
Vision high intensity (VH) 40 ms	19 ± 12
Vision high intensity (VH) 500 ms	20 ± 10
Vision low intensity (VL) 9 ms	52 ± 16
Vision low intensity (VL) 40 ms	47 ± 13
Vision low intensity (VL) 500 ms	46 ± 11

Across conditions, relative stimulus intensity of the visual stimulus (V) could be either "high" (H) or "low" (L) (see "Methods") and durations of the audiovisual compound stimulus was 9, 40 or 500 ms

condition are given in Table 1. For the determined PSS values, a two-way repeated measures ANOVA with the main within-participant factors "duration" (9, 40, and 500 ms) and "light intensity" (vision high intensity versus vision low intensity) was calculated. A significant main effect was found for the factor "light intensity" [$F(1,21) = 86.8$, $p = 0.01$]. The PSS averaged across all durations differed by 27 ms between vision low intensity and vision high intensity. In the vision low intensity condition, the mean PSS (48 ± 7.6 ms, mean ± standard error) was significantly shifted to more positive values as compared to the vision high-intensity condition (21 ± 6.9 ms). In other words, the necessary lead of the visual stimulus relative to the auditory stimulus for perceived simultaneity was larger in the vision low-intensity condition than in the vision high-intensity condition. Unexpectedly (see "Introduction"), no significant effect was found for the main factor "duration" [$F(2,42) < 1$, n.s.]. The mean PSS for the durations across intensities was: 37 ± 11.1 ms for 9 ms duration, 33 ± 8.7 ms for 40 ms duration, and 33 ± 7.4 ms for 500 ms duration. ANOVA analysis of just noticeable differences (JNDs) derived from the estimated psychometric functions yielded no significant effects for the factors "light intensity" [$F(1,21) = 1$] and "duration" [$F(2,42) < 1$]. In Fig. 3, mean proportions of "vision first" responses across participants are displayed as a function of SOA for each duration and intensity. As can be seen, increasing the visual stimulus intensity resulted in a shift of the mean functions toward lower SOA values, without a change of slope. This illustrates the shift of the PSS without a change of JND.

Stone et al. (2001) reported a high interindividual variability of AV PSS values using a simultaneity judgment (SJ) task. To assess the variability across participants in our TOJ paradigm, we plotted individual PSS values for each participant and for all durations and intensities (Fig. 4). Interestingly, although there was no significant effect of stimulus duration on the mean PSS, inspection of PSS values of single participants revealed systematic changes of PSS with duration. As shown in Fig. 4, PSS values varied between negative and positive values across participants. As a general pattern, the magnitude of both positive and negative PSS values tended to decrease with increasing duration of the AV stimulus pair. In other words, magnitude and sign of the individual PSS values determined how strong and in what direction, respectively, these PSS values changed with stimulus duration. This pattern of PSS change with stimulus duration was consistently found with high and low visual stimulus intensities (Fig. 4, left and right panel). It should be noted that this particular dependence of individual PSS values on stimulus duration cannot be retrieved by the ANOVA based on mean values across participants. For further analysis, we quantified the direction and the amount of PSS change with increased stimulus duration. For this purpose, we calculated, in the visual low-intensity and the visual high-intensity condition for each participant, the range of the duration-related PSS shift by calculating the difference of the PSS measured with 500 ms stimulus duration and the PSS measured with 9 ms stimulus duration. This range of PSS shift was correlated with an estimate of the overall location of the individual PSS curve, given by the PSS value measured with 40 ms stimulus duration. Notably, this was done for quantitative approximation of the PSS change without the need for invoking any particular model of the PSS dependence on stimulus duration.

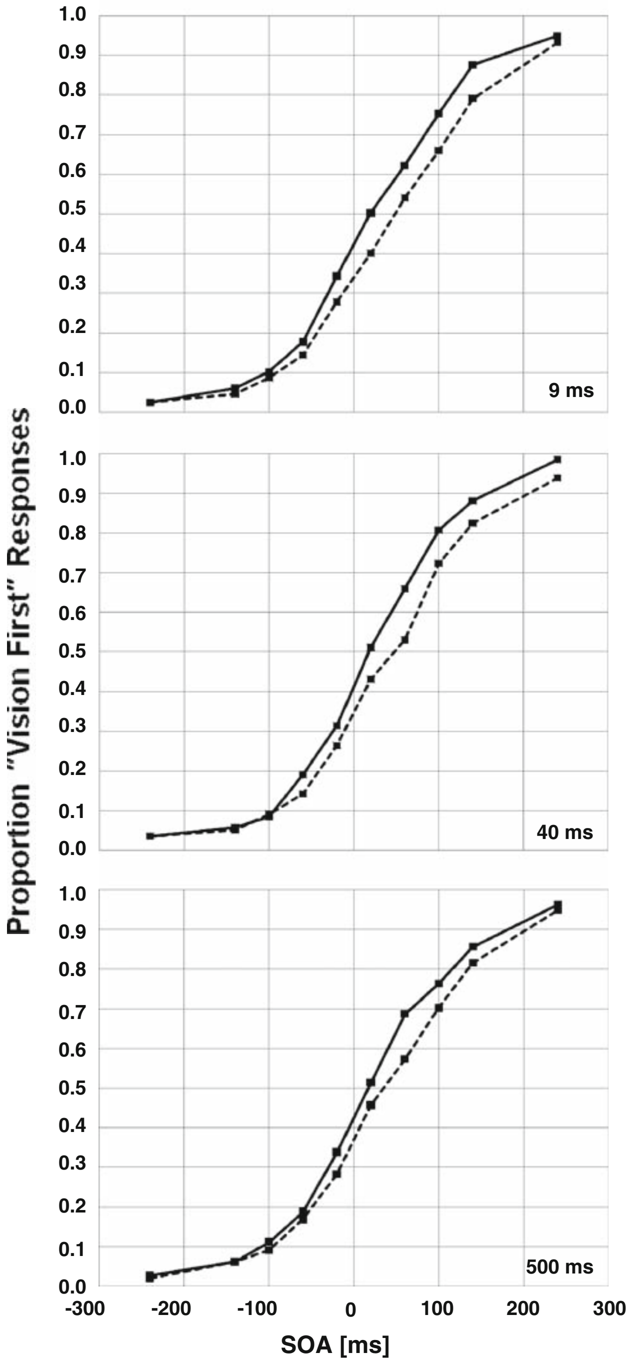

Fig. 3 Proportion of "vision first" responses as a function of SOA (positive values indicating that the visual stimulus component was presented before the auditory component) for the two intensity conditions (vision low intensity *dashed line*, vision high intensity *continuous line*). Data obtained from the three duration conditions (9, 40, 500 ms) are plotted separately. In the vision low-intensity condition the PSS is shifted to more positive values compared to the vision high-intensity condition

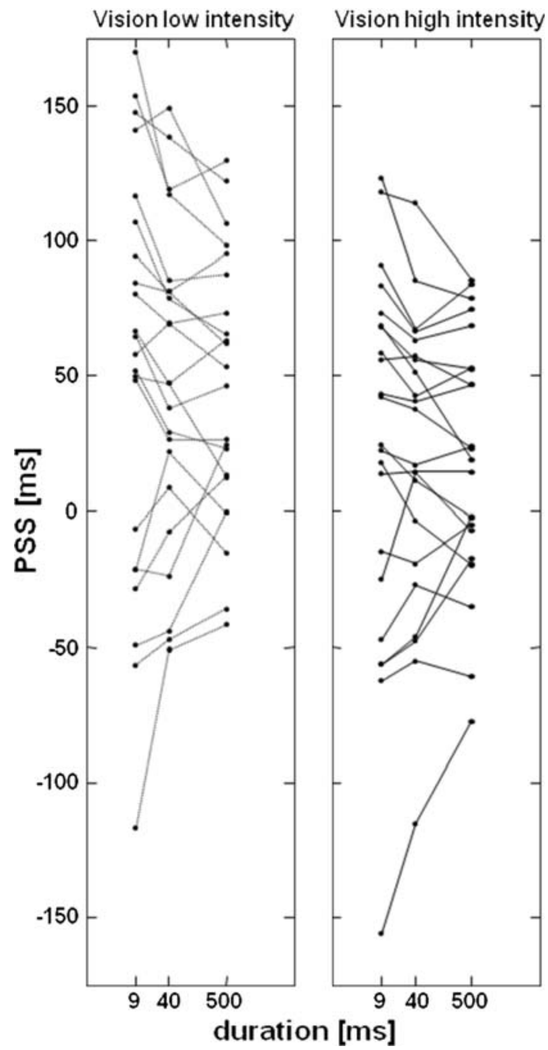

Fig. 4 All PSS values for all participants and all durations are shown separately for the vision low-intensity condition (*left panel*) and the vision high-intensity condition (*right panel*). Positive PSS values indicate that the physical onset of the visual stimulus has to precede the physical onset of the auditory stimulus onset for simultaneous perception, whereas negative PSS values indicate that the physical onset of the auditory stimulus has to precede the onset of visual stimulus. Comparison of the two intensity conditions (*left and right panel*) indicates the overall shift of PSS values into negative direction with increased intensity across subjects. Within each intensity condition, single-participant PSS values change either into positive or negative direction, depending on the sign and magnitude of the individual PSS values. The population of duration curves suggests a general pattern of convergence with increasing stimulus duration. See text for further explanation

In Fig. 5, the PSS shift with stimulus duration for all participants is plotted against the individual PSS values measured with 40 ms stimulus duration. This makes evident the aforementioned relationship of sign and magnitude of individual PSS values on the one hand, and the direction and amount of PSS shift with stimulus duration, on the other hand. The dependence can be characterized by a highly significant linear correlation in both the vision low-intensity condition ($r = 0.76$, $p < 10^{-4}$) and the vision high-intensity condition ($r = 0.77$, $p < 10^{-4}$). With more negative individual PSS values (for 40 ms stimulus duration),

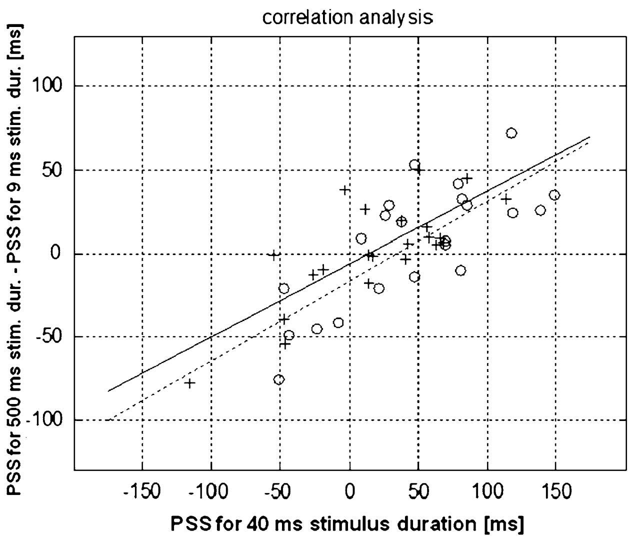

Fig. 5 Correlation between the PSS shift (difference between PSS values measured for 500 ms stimulus duration and for 9 ms stimulus duration) (ordinate) and PSS value measured for 40 ms stimulus duration (abscissa). The correlation analysis was done separately for the two intensity conditions (vision low intensity: o, *dashed line*; vision high intensity: +, *continuous line*). See text for further explanation

the duration-related PSS shifts were more negative, whereas with more positive individual PSS values (for 40 ms stimulus duration), the shifts were more positive. That is, with increasing stimulus duration, positive PSS shifted toward more negative values, and negative PSS shifted toward more positive values. This implies an attractor effect for the variation of PSS with stimulus duration. The attractor is found where the PSS shift vanishes (range = 0) at regressed values of PSS of 35.1 ms (vision low-intensity condition) and 14.9 ms (vision high-intensity condition). To test for spurious correlations (similar to the well-known regression to the mean artifact), induced merely by the systematic change of the population variance of measured PSS values with stimulus duration, we applied a permutation test. For this purpose, we permutated 10,000 times the PSS values obtained from the three duration conditions separately across participants. For each permutation, we calculated the correlation between PSS shift and PSS value for the 40 ms condition. The distribution of the *r* values in the set of permutations is plotted in Fig. 6 for the visual low-intensity condition (left) and the visual high-intensity condition (right) separately, together with the *r* value corresponding to a significant correlation on the 0.05-level (red dotted line). The empirically found *r* values, shown as a solid blue line, indicate a highly significant correlation ($p < 10^{-4}$). Accordingly, we rejected the null hypothesis that the observed correlation between PSS shift and PSS value for the 40 ms condition relied on a spurious population effect irrespective of interindividual differences in PSS values across participants (similar to the regression to the mean artifact). To further test the reliability of our PSS measurement in single participants, we calculated the rank correlation between the independently measured PSS values of the vision low-intensity and the vision high-intensity conditions across participants, separately for all three durations. The correlation coefficients were $r = 0.98$, 0.97 and 0.91 for the durations 9, 40 and 500 ms, respectively. This also suggests that the duration effect was due to specific PSS shifts within participants, and not to some kind of unspecific measurement error between participants.

Discussion

Using an audiovisual temporal-order judgment task we have addressed the question of how stimulus duration and relative stimulus intensities influence perceived temporal order. Our results show that both factors play a role in the temporal perception of AV stimulus pairs. First, we found a strong effect for relative intensity, confirming previous results (Jaśkowski 1996; Neumann and Niepel 2004). In addition, we could demonstrate that duration of AV stimuli

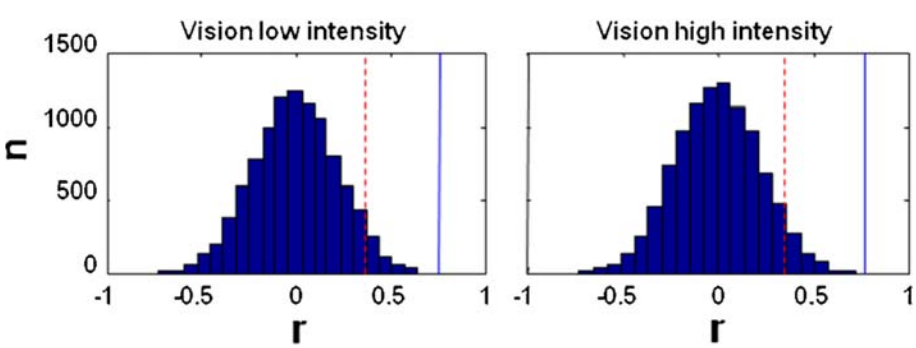

Fig. 6 Distribution of frequency of occurrence of *r* values for the linear correlation of PSS shifts with duration, using a permutation test (see text for explanation) in the two intensity conditions (vision low intensity: on the *left*; vision high intensity: on the *right*). The *dotted line* indicates a significant correlation on the 0.05 level; the *continuous line* depicts the *r* value observed in the experimental data

also has an influence on the perception of their temporal order, and that the size and the direction of the effect depended on interindividual differences of PSS levels between participants.

While the effect of stimulus duration on the TOJ could not be retrieved by a variance analysis of the mean PSS values, each single participant showed a clear influence of stimulus duration on his/her individual PSS value. This could be quantified using regression analysis and a permutation test. Specifically, increasing stimulus duration shifted individual PSS values toward an attractor value that existed for the entire population. Moreover, this attractor value was dependent on the intensity of the visual signal (35.1 ms for vision low intensity and 14.9 ms for vision high intensity). In other words, in participants with large positive or negative PSS value, there was a strong shift of the PSS value into the negative or positive direction, respectively, with increasing stimulus duration, which reduces interindividual differences. In participants with smaller PSS values (closer to the attractor value), the PSS shifts were also smaller. Measures were taken to verify that this correlation was not due to a "regression to the mean" artifact (see "Results"). This effect is indicative of a mechanism by which perceived asynchronies between visual and auditory stimuli are stabilized by increasing the stimulus duration. Hypothetically, such a mechanism could serve important functions for the integration of near-coincident auditory and visual stimuli.

Discordant results on audiovisual PSS values in previous studies

The mean PSS values across participants ranged between 19 and 52 ms (Table 1, Fig. 3) and thus were clearly positive in all conditions. At first glance, this result supports previous findings stating a relatively faster processing of auditory signals compared to visual signals in TOJ tasks (e.g., Jaśkowski 1996; Zampini et al. 2003a). It is in contrast to studies that conclude that in AV-TOJ tasks, vision generally leads audition (Neumann et al. 1992; Neumann and Niepel 2004). The observed overall PSS differences between the studies of Neumann et al. (Neumann et al. 1992; Neumann and Niepel 2004) and our own study are notable, given the highly similar stimulus intensities and durations applied in some conditions of both studies: medium intensities in the studies of Neumann et al. (1992) were in the same range as our low-intensity condition (0.1 cd/m^2 and 48 dB(A) versus 0.14 cd/m^2 and 50 dB(A)), also using a stimulus duration of 40 ms. However, whereas Neumann et al. (1992) found a negative PSS of ~ -15 ms (estimated from their Fig. 4), we found a PSS of 47 ms under comparable conditions. This suggests that, besides intensity and duration, additional factors play a role for multisensory temporal integration. A possible factor, which could cause the observed discrepancy, is stimulus eccentricity. Neumann et al. presented the visual stimuli centrally, whereas we presented the stimuli peripherally. This was done to avoid a modality bias of the participants, as discussed extensively in Spence et al. (2001). Although Posner et al. (1980) have demonstrated that attention leads to a benefit in processing speed irrespective of the location of visual events (foveal or peripheral), several studies have shown that visual information processing of stimuli in the periphery is faster compared to processing of stimuli presented foveally (Carrasco et al. 2003), and slower responses to peripheral than to foveal visual stimulation have been found in reaction time tasks (Kopinska and Harris 2004; Rains 1963; Schiefer et al. 2001). Furthermore, peripheral stimulation seems to be more sensitive to intensity differences as measured by PSS in a visual TOJ task (Arden and Weale 1954). Some researchers have already suggested a possible influence of this factor on the crossmodal PSS: Zampini et al. (2003a) found a relatively large positive PSS in a spatial AV-TOJ task[1] with peripheral stimulation and assumed that the positive PSS would shift toward smaller values if audio and visual stimuli had been presented foveally. Further support for this notion comes from the so-called unity effect, i.e., the finding that audiovisual events are more likely to be bound together when they are in spatiotemporal proximity (e.g., Bertelson and de Gelder 2004; Koppen and Spence 2007a, b; Lewald et al. 2001; Slutsky and Recanzone 2001; Stein and Meredith 1993; Vatakis and Spence 2007). With foveally presented visual stimuli and acoustic stimuli presented via headphones as in the studies of Neumann et al., percepts might be spatially more contiguous than in our study, where visual stimuli were always presented peripherally and clearly spatially distinct from the sound. However, only changing eccentricity of the visual stimulus (0° versus 20° of visual angle) in an AV-TOJ task had no effect on perceived simultaneity (Kopinska and Harris 2004). In another study, the spatial disparity between audio and visual events was varied, which led to smaller, just noticeable difference (JND) at higher disparity, but yielded no differences in the PSS (Keetels and Vroomen 2005). Since the former study varied only the eccentricity of the visual stimuli, and the latter study investigated disparity and not directly eccentricity, the question of how much eccentricity affects the relative processing speed in an AV-TOJ task cannot be resolved, yet.

Another difference in the experimental design, which might contribute to the clearly different PSS is the size of the visual stimuli, which was about ten times smaller in our study compared to Neumann et al.'s ($\sim$ 1.3° versus 13°).

[1] Zampini et al. (2003a) found the PSS in the range of 60–80 ms versus other studies using central stimulation that found PSS in the range of 20–40 ms.

Neumann et al., using an orthogonal design and eight conditions with factors tone intensity, light intensity and size of visual stimulus (2 × tone intensities, 2 × light intensities, and 2 × size of visual stimuli), also found a positive PSS for all dim visual stimuli and a negative PSS for three out of four bright stimuli (see Jaśkowski 1996). Only with bright and small visual stimuli presented together with loud tones (brightness: 0.03 cd/m^2 and 4.93 cd/m^2, size: 0.8° and 13°, sound pressure level: 45 dB(A) and 68 dB(A), Niepel, personal communication), mean PSS was also positive, as in the present study. Finally, there are additional differences between our experimental design and that of Neumann et al., which might also influence the mean PSS. For example the use of a 2,300-Hz pure tone in contrast to white noise bursts might have had an effect, because recently it has been shown that the spectral characteristics of auditory stimuli can have an effect on multisensory integration (Tajadura-Jiménez et al. 2009). White noise bursts and white light flashes in our study were chosen to avoid any effects by color or frequencies or effects introduced by their relationships (for color speed differences see e.g., von Helmholtz 1867; Parise and Spence 2009). Additionally, the relationship of target luminance and background luminance could influence visual perception speed (see e.g., Roufs 1974). However, Neumann et al. (Niepel, personal communication) carried out their experiments in an almost completely dark environment, as we did. Still, possible differences in the adaptation state between the participants of Neumann et al.'s study and our participants might contribute to the observed differences in the relative perception speed (see Jaśkowski 1996). The broad variance of PSS values (ranging from positive to negative values) across participants in our data also suggest that interindividual differences might contribute to the observed diversity of PSS estimation results in different studies, especially for short stimulus durations.

A previously overlooked effect of stimulus duration on the audiovisual TOJ

Some insight into the dependence of PSS values from physical stimulus characteristics could be obtained from a single-participant analysis of the data. First of all, single-participant analysis revealed that, while mean PSS values were all positive for the range of stimulus parameters tested, the PSS values of individual participants broadly scattered from negative values <-50 ms to positive values >100 ms. Despite this variance across participants, which is in agreement with the results from Stone et al. (2001) using a SJ task, there was a systematic variation of PSS values with stimulus duration in our TOJ study. This systematic variation was characterized by a shift of the PSS value toward an attractor value that was consistent across the population with increasing stimulus duration. The fact that this dependence cannot be retrieved using variance analysis of the data (Fig. 2) might have contributed to its oversight in previous studies.

Possible mechanisms underlying the duration effect

While the underlying mechanism for the observed PSS shift with duration of the AV stimulus pair is yet to be investigated, it is conceivable that the effect of a shifted onset perception with stimulus duration, known from the auditory modality (for review see Schimmel and Kohlrausch 2008), might contribute to it. Such a mechanism could explain those cases in which we have observed decreases of PSS values with increasing duration. However, as we have observed both decreases and increases of the PSS value with increasing stimulus duration, our data cannot be solely explained by such a mechanism. At present, it is difficult to assess the precise role of such a mechanism, as it has not yet been determined whether a similar effect holds for the visual modality and how these two effects would then interact in an audiovisual compound stimulus. This could be addressed by further studies employing independent duration variations of the two sensory modalities.

The duration dependence of individual PSS values indicates that participants did not base their TOJ solely on the stimulus onsets, as they were instructed to do. Presently, it is still an open question, on which perceptual cue participants base their TOJs, and which factors influence the selection of this cue (stimulus- or strategy-dependent factors). For example, Jaśkowski (1991) has shown that participants, when asked to match the onset of two visual stimuli with unequal duration, failed in doing so by shifting the onset of the shorter stimulus to the offset of the longer stimulus, at least for stimulus durations up to 300 ms. Moreover, participants might use the offset of a stimulus as a cue for TOJ. However, in the visual modality, Jaśkowski (1993) could not find differences on temporal-order perception using the method of adjustment between experimental conditions with and without stimulus offsets. Nevertheless, stimulus offsets might affect temporal-order perception in multisensory stimuli (Efron 1970). Furthermore, perceived stimulus duration could have an influence on temporal-order perception (cf. Behar and Bevan 1961; Loeb et al. 1966). However, as these effects generally would only explain a shift of PSS in one direction, they are, in isolation, insufficient to account for the present results.

It is further important to consider the temporal overlap between auditory and visual stimuli in the different experimental conditions. For the 9 ms duration condition, there was no temporal overlap. In the 40 ms condition, overlap occurred at SOAs of 20 ms. For the 500 ms condition, stimuli were always temporally overlapping. As shown by

Meredith et al. (1987), temporal overlap has a strong influence on the response of superior colliculus neurons, which receive converging multimodal inputs. Thus, responses of these neurons were enhanced when there was a temporal overlap of peak discharge periods, evoked by different modalities, whereas suppression of responses was found for non-overlapping discharge periods. In response to stimuli with longer durations, neuronal discharge periods are more likely to overlap, which could lead to an enhanced interaction between the modalities. Thus, temporal overlap of multisensory stimuli might promote their integration. As a consequence, similar to the unity effect (see above), the magnitude of the PSS might decrease with increasing stimulus duration. To what extent the observed effect of stimulus duration relies on the duration per se, or on the temporal overlap between stimuli, needs to be explored in further experiments. However, such a mechanism could only partially explain our data, i.e., those cases in which the magnitude of the PSS decreased with increasing stimulus duration.

Finally, interindividual differences in which cues are selected for TOJ might further contribute to the PSS variance observed in TOJ tasks (for discussion see Jaśkowski 1996; Johnston and Nishida 2001; Shore et al. 2001). Here, an important factor determining TOJ might be the ability of one sensory modality to "grab attention" better than another modality (Spence et al. 2001). In fact, in our study, a post hoc interview of the participants revealed that participants, whose PSS at short durations was more negative relative to the attractor value, had the impression that conditions with a leading visual stimulus were more frequent than conditions with a leading auditory stimulus; participants whose PSS was shifted to more positive values had the opposite impression. If this variance across participants was caused by individual modality biases at short stimulus durations, the described duration effect could be seen as a compensating effect that exploits stimulus duration to neutralize individually present attentional imbalances for the two sensory modalities.

Conclusion

We studied the role of stimulus intensity and stimulus duration in a spatial audiovisual TOJ. We have replicated and confirmed the previously reported intensity effect, manifest in a shift of the PSS of a visual and auditory stimulus with variation of their relative intensities. In addition, we have described a further, previously overlooked, confounding effect in TOJ tasks (duration effect). This effect is manifest in a shift of a participant's individual PSS value with increasing stimulus duration toward an attractor value that is valid for the entire population. This dependence is indicative of a mechanism that stabilizes perceived stimulus asynchronies with increasing stimulus duration, which might facilitate crossmodal integration processes. It also seems to suggest that some participants tend to be biased in favor of one modality at shorter stimulus durations. As we currently cannot exclude or quantify a possible contribution of crossmodal integration/interaction by the temporal overlap of the stimuli, further research is needed. However, with these results, we provide evidence that the choice of stimuli of different durations might have contributed to the high variability of PSS values found across studies of audiovisual TOJs. Moreover, the high variability of PSS values across participants strongly suggests the analysis of individual data in addition to that of the grand mean in further TOJ studies. Our results extend the findings of Stone et al. (2001) that the interindividual difference in absolute values seems to be larger for shorter stimuli. This is true at least for TOJ paradigms. In further research it would be of interest to investigate if the duration effect has a similar influence on PSS values in SJ tasks.

Acknowledgments We would like to thank the reviewers and Charles Spence for their valuable comments. Furthermore, we would like to thank Cees van Leeuwen and Andrey R. Nikolaev for critical discussion of a previous version of the manuscript. Finally, we would like to thank Anna Fiedler and Felix Ball for assistance during data collection. This study was supported by a grant from the European Community ("DIRAC", FP6-IST-027787).

References

Alais D, Carlile S (2005) Synchronizing to real events: subjective audiovisual alignment scales with perceived auditory depth and speed of sound. Proc Natl Acad Sci USA 102:2244–2247

Arden GB, Weale RA (1954) Variations of the latent period in vision. Proc R Soc Lond Series B Biol Sci 142:258–268

Behar I, Bevan W (1961) The perceived duration of auditory and visual intervals: cross-modal comparison and interaction. Am J Psychol 74:17–26

Bertelson P, de Gelder B (2004) The psychology of multimodal perception. In: Spence C, Driver J (eds) Crossmodal space and crossmodal attention. Oxford University Press, Oxford, pp 140–177

Calvert GA, Stein BE, Spence C (eds) (2004) The handbook of multisensory processing. MIT Press, Cambridge

Carrasco M, McElree B, Denisova K et al (2003) Speed of visual processing increases with eccentricity. Nat Neurosci 6:699–700

Efron R (1970) Effect of stimulus duration on perceptual onset and offset latencies. Percept Psychophys 8:231–234

Exner S (1875) Experimentelle Untersuchung der einfachsten psychischen Processe. III. Abhandlung. Der persönlichen Gleichung zweiter Theil [Experimental examination of the most simple psychological processes. III. Treatise Second part of the personal equation]. Pflüger's Archiv für die gesammte Physiologie des Menschen und der Thiere 11:403–432

Frey RD (1990) Selective attention, event perception and the criterion of acceptability principle: Evidence supporting and rejecting the doctrine of prior entry. Hum Mov Sci 9:481–530

Fujisaki W, Shimojo S, Kashino M et al (2004) Recalibration of audiovisual simultaneity. Nat Neurosci 7:773–778

Hanson JV, Heron J, Whitaker D (2008) Recalibration of perceived time across sensory modalities. Exp Brain Res 185:347–352

Harrar V, Harris LR (2008) The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. Exp Brain Res 70:807–817

Heil P, Neubauer H (2003) A unifying basis of auditory thresholds based on temporal summation. Proc Natl Acad Sci USA 100:6151–6156 Erratum in: Proc Natl Acad Sci USA 101:3323

Heil P, Neubauer H, Brown M, Irvine DR (2008) Towards a unifying basis of auditory thresholds: distributions of the first-spike latencies of auditory-nerve fibers. Hear Res 238:25–38

Hikosaka O, Miyauchi S, Shimojo S (1993) Focal visual attention produces illusory temporal order and motion sensation. Vis Res 33:1219–1240

Jaśkowski P (1991) Perceived onset simultaneity of stimuli with unequal durations. Perception 20:715–726

Jaśkowski P (1993) Selective attention and temporal-order judgment. Percept Psychophys 22:681–689

Jaśkowski P (1996) Simple reaction time and perception of temporal order: dissociations and hypotheses. Percept Mot Skills 82:707–730

Jaśkowski P, Verleger R (2000) Attentional bias toward low-intensity stimuli: an explanation for the intensity dissociation between reaction time and temporal-order judgment? Conscious Cognit 9:435–456

Jaśkowski P, Jaroszyk F, Hojan-Jezierska D (1990) Temporal-order judgments and reaction time for stimuli of different modalities. Psychol Res 52:35–38

Johnston A, Nishida S (2001) Time perception: Brain time or event time? Curr Biol 11:427–430

Keetels M, Vroomen J (2005) The role of spatial disparity and hemifields in audio-visual temporal-order judgments. Exp Brain Res 167:635–640

Kopinska A, Harris LR (2004) Simultaneity constancy. Perception 33:1049–1060

Koppen C, Spence C (2007a) Audiovisual asynchrony modulates the Colavita visual dominance effect. Brain Res 1186:224–232

Koppen C, Spence C (2007b) Spatial coincidence modulates the Colavita visual dominance effect. Neurosci Lett 417:407–411

Kuss M, Jäkel F, Wichmann FA (2005) Bayesian inference for psychometric functions. J Vis 5:478–492

Lewald J, Guski R (2004) Auditory–visual temporal integration as a function of distance: no compensation for sound-transmission time in human perception. Neurosci Lett 357:119–122

Lewald J, Ehrenstein WH, Guski R (2001) Spatio-temporal constraints for auditory–visual integration. Behav Brain Res 121:69–79

Loeb M, Behar I, Warm JS (1966) Cross-modal correlations of the perceived durations of auditory and visual stimuli. Psychonomic Sci 6:87

Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J Neurosci 7:3215–3229

Moore CJ (2003) An introduction to the psychology of hearing, 5th edn. Emerald Group Publishing, UK

Neubauer H, Heil P (2008) A physiological model for the stimulus dependence of first-spike latency of auditory-nerve fibers. Brain Res 1220:208–223

Neumann O, Niepel M (2004) Timing of "perception" and perception of "time". In: Kärnbach C, Schröger E, Müller H (eds) Psychophysics beyond sensation: laws and invariants of human cognition. Erlbaum, Mahwah, pp 245–269

Neumann O, Koch R, Niepel M et al (1992) Reaktionszeit und zeitliches Reihenfolgeurteil: Übereinstimmung oder Dissoziation? [Reaction time and temporal-order judgment: Correspondence or Dissociation]. Z Exp Angew Psychol 39:621–645

Nilsson T (2006) Transient effects in vision. In: Karwowski W (ed) International encyclopedia of ergonomics and human factors, 2nd edn. CRC Press, USA, pp 520–533

Parise CV, Spence C (2009) 'When birds of a feather flock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes. PLoS ONE 4(5):e5664

Posner MI, Snyder CR, Davidson BJ (1980) Attention and the detection of signals. J Exp Psychol 109:160–174

Rains JD (1963) Signal luminance and position effects in human reaction time. Vis Res 3:239–251

Rorden C, Mattingley JB, Karnath H-O, Driver J (1997) Visual extinction and prior entry: impaired perception of temporal-order with intact motion perception after unilateral parietal damage. Neuropsychologia 35:421–433

Roufs JAJ (1974) Dynamic properties of vision-V. Perception lag and reaction time in relation to flicker and flash thresholds. Vis Res 14:853–869

Rutschmann J, Link R (1964) Perception of temporal-order of stimuli differing in sense mode and simple reaction time. Percept Mot Skills 18:345–352

Scharlau I (2004) Evidence against response bias in temporal-order tasks with attention manipulation by masked primes. Psychol Res 68:224–236

Schiefer U, Strasburger H, Becker et al (2001) Reaction time in automated kinetic perimetry: effects of stimulus luminance, eccentricity, and movement direction. Vis Res 41:2157–2164

Schimmel O, Kohlrausch A (2008) On the influence of interaural differences on temporal perception of noise bursts of different durations. J Acoust Soc Am 123:986–997

Shore DI, Spence C, Klein RM (2001) Visual prior entry. Psychol Sci 12:205–212

Slutsky DA, Recanzone GH (2001) Temporal and spatial dependency of the ventriloquism effect. Neuroreport 12:7–10

Smith WF (1933) The relative quickness of visual and auditory perception. J Exp Psychol 16:239–257

Spence C, Driver J (eds) (2004) Crossmodal space and crossmodal attention. Oxford University Press, Oxford

Spence C, Squire S (2003) Multisensory integration: maintaining the perception of synchrony. Curr Biol 13:519–521

Spence C, Shore DI, Klein RM (2001) Multisensory prior entry. J Exp Psychol Gen 130:799–832

Stein BE, Meredith MA (1993) The merging of the senses. MIT Press, Cambridge

Stelmach LB, Herdman CM (1991) Directed attention and perception of temporal-order. J Exp Psychol: HPP 17:539–550

Stone JV, Hunkin NM, Porill J et al (2001) When is now? Perception of simultaneity. Proc Biol Sci 268:31–38

Sugita Y, Suzuki Y (2003) Audiovisual perception: implicit estimation of sound-arrival time. Nature 421:911

Tajadura-Jiménez A, Kitagawa N, Väljamäe A et al (2009) Auditory–somatosensory multisensory interactions are spatially modulated by stimulated body surface and acoustic spectra. Neuropsychologia 47:195–203

Titchener EB (1908) Lectures on the elementary psychology of feeling and attention. Macmillan, New York

Van Eijk RLJ, Kohlrausch A, Juola JF et al (2008) Audiovisual synchrony and temporal-order judgments: effects of experimental method and stimulus type. Percept Psychophys 70:955–968

Vatakis A, Spence C (2007) Crossmodal binding: evaluating the "unity assumption" using audiovisual speech stimuli. Percept Psychophys 69:744–756

Vatakis A, Navarra J, Soto-Faraco S et al (2007) Temporal recalibration during asynchronous audiovisual speech perception. Exp Brain Res 181:173–181

Vatakis A, Navarra J, Soto-Faraco S et al (2008) Audiovisual temporal adaptation of speech: temporal-order versus simultaneity judgments. Exp Brain Res 185:521–529

von Helmholtz H (1867) Handbuch der Physiologischen Optik. (Algem Enz d Physik Bnd 9). Leopold Voss, Berlin

Vroomen J, Keetels M, de Gelder B et al (2004) Recalibration of temporal-order perception by exposure to audio-visual asynchrony. Cogn Brain Res 22:32–35

Zampini M, Shore DI, Spence C (2003a) Audiovisual temporal-order judgments. Exp Brain Res 152:198–210

Zampini M, Shore DI, Spence C (2003b) Multisensory temporal-order judgments: the role of hemispheric redundancy. Int J Psychophysiol 50:165–180

Zampini M, Guest S, Shore DI et al (2005a) Audio-visual simultaneity judgments. Percept Psychophys 67:531–544

Zampini M, Shore DI, Spence C (2005b) Audiovisual prior entry. Neurosci Lett 381:217–222