

Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type

ROB L. J. VAN EIJK

Eindhoven University of Technology, Eindhoven, The Netherlands

ARMIN KOHLRAUSCH

*Eindhoven University of Technology, Eindhoven, The Netherlands
and Philips Research Laboratories, Eindhoven, The Netherlands*

JAMES F. JUOLA

*Eindhoven University of Technology, Eindhoven, The Netherlands
and University of Kansas, Lawrence, Kansas*

AND

STEVEN VAN DE PAR

Philips Research Laboratories, Eindhoven, The Netherlands

When an audio–visual event is perceived in the natural environment, a physical delay will always occur between the arrival of the leading visual component and that of the trailing auditory component. This natural timing relationship suggests that the point of subjective simultaneity (PSS) should occur at an auditory delay greater than or equal to 0 msec. A review of the literature suggests that PSS estimates derived from a temporal order judgment (TOJ) task differ from those derived from a synchrony judgment (SJ) task, with (unnatural) auditory-leading PSS values reported mainly for the TOJ task. We report data from two stimulus types that differed in terms of complexity—namely, (1) a flash and a click and (2) a bouncing ball and an impact sound. The same participants judged the temporal order and synchrony of both stimulus types, using three experimental methods: (1) a TOJ task with two response categories (“audio first” or “video first”), (2) an SJ task with two response categories (“synchronous” or “asynchronous”; SJ2), and (3) an SJ task with three response categories (“audio first,” “synchronous,” or “video first”; SJ3). Both stimulus types produced correlated PSS estimates with the SJ tasks, but the estimates from the TOJ procedure were uncorrelated with those obtained from the SJ tasks. These results suggest that the SJ task should be preferred over the TOJ task when the primary interest is in perceived audio–visual synchrony.

Most everyday events give rise to both visual and auditory sensations, such as when we listen to someone speaking in front of us or observe a book falling to the floor. Because of the relatively low speed of sound, the auditory component of a perceived event will reach an observer later than the visual component, and this difference increases with physical distance. Thus, when auditory and visual components of a single audio–visual event reach an observer’s sensory receptors, there will always be a physical delay between the leading visual component and the following auditory component. At least part of this difference in arrival times can be made up (or even reversed, for proximal stimuli) by the faster processing of auditory information once it arrives at the sensory receptors, since brain activation occurs about 30–50 msec earlier for sounds than for visual stimuli that arrive simultaneously

with them (see, e.g., Arrighi, Alais, & Burr, 2006; King & Palmer, 1985; Pöppel, 1988).

As a result of the different propagation speeds of light and sound, observers should be more tolerant of lagging audio than of lagging video when integrating the sensed components of an event. This tolerance can be measured in terms of the milliseconds of auditory delay over which integration into perception of a unitary common event is satisfactorily achieved. By convention (see, e.g., Arrighi et al., 2006; Aschersleben & Müsseler, 1999; Enoki, Washikita, & Yamada, 2006; Vatakis & Spence, 2006a, 2006b; Zampini, Shore, & Spence, 2003a, 2003b), such auditory delays are measured in positive values (with 0 msec indicating physical synchrony of the auditory and visual stimulus components), whereas negative values are used for the far-less-frequent occurrence of an auditory compo-

R. L. J. van Eijk, r.l.j.v.eijk@tue.nl

ment of some event arriving before its visual counterpart. We can estimate the point of subjective simultaneity (PSS) from the midpoint of the range of delays within which synchrony is perceived. Furthermore, we expect that the PSS should occur near the point of physical synchrony or at some positive audio delay—that is, with the auditory component lagging behind the visual component.

Research in the area of perceived audio–visual synchrony has made use of a wide range of stimulus types (see, e.g., Arrighi et al., 2006; Enoki et al., 2006; Keetels & Vroomen, 2005; Vatakis & Spence, 2006a) and experimental methods (see, e.g., Dixon & Spitz, 1980; Exner, 1875; Vatakis, Navarra, Soto-Faraco, & Spence, 2008; Vroomen, Keetels, de Gelder, & Bertelson, 2004). Stimuli have varied from simple (e.g., a flash of light accompanied by an audible click; see, e.g., Aschersleben & Müseler, 1999; Hamlin, 1895; Jaśkowski, Jaroszyk, & Hojan-Jeziarska, 1990) to complex (e.g., a video of a person speaking or playing a musical instrument; see, e.g., Dixon & Spitz, 1980; Hollier & Rimell, 1998; Vatakis & Spence, 2006a). Asking observers to judge whether or not the auditory and visual components of stimuli with variable onset asynchronies are synchronous (a synchrony judgment [SJ] task) yields a relatively direct measure of perceived synchrony (see, e.g., Fujisaki, Shimojo, Kashino, & Nishida, 2004; Stone et al., 2001; Zampini, Guest, Shore, & Spence, 2005).

Other methods have been used to estimate the PSS, and perhaps most important among these is the temporal order judgment (TOJ) task (see, e.g., Aschersleben & Müseler, 1999; Spence, Baddeley, Zampini, James, & Shore, 2003; Sternberg & Knoll, 1973; Zampini et al., 2003a, 2003b). In the TOJ method, observers are asked to indicate which of two sensory modalities was stimulated first. As relative delays are systematically varied, it is possible to estimate the PSS as the point at which the proportion of “audio first” judgments equals the proportion of “video first” judgments. Comparing PSS estimates derived from different experimental methods shows that the TOJ procedure often yields different estimates of the PSS than does the SJ task. This is true whether the SJ task involves two judgment categories (simply “synchronous” or “asynchronous”; we will call this the SJ2 task) or three (“audio first,” “synchronous,” or “video first”; the SJ3 task).

A nonexhaustive overview of PSS values reported by or estimated from various studies is shown in Table 1. Given the context of the present study, the overview in Table 1 is restricted to publications about perceived timing relations between the auditory and visual modalities (i.e., studies of other modalities and unimodal studies are excluded) that investigate the perception of audio–visual synchrony under “normal conditions” (i.e., studies are also excluded that attempt to manipulate synchrony perception by, for example, manipulating attention or exposing participants to asynchronous audio–visual adaptation stimuli).

The most striking results shown in Table 1 are the negative PSS values, which represent a situation in which the auditory stimulus had to lead the visual stimulus for the pair to be interpreted as synchronous. The PSS values reported in Table 1 are generally measured at the position

of the observer (with the exception of the study by Bald, Berrien, Price, & Sprague, 1942, which reported delays measured at the stimulus source). Negative “external” delays thus are highly unnatural, because they represent an audio–visual timing relationship that does not occur in everyday events.

From Table 1, one can see that both negative and positive PSS values are reported for individual participants for both TOJ and SJ tasks, but negative *overall* PSS values are reported almost exclusively when using the TOJ paradigm. Thus, from the review of the literature presented in Table 1, it may be concluded that the two tasks (TOJ and SJ) might measure different things. That is, the SJ task could emphasize the judgment of “synchrony” versus “successiveness,” whereas the TOJ task could emphasize the judgment of “order,” which requires the perception of successiveness for correct perception (Allan, 1975; Hirsh & Sherrick, 1961). Indeed, Shore, Gray, Spry, and Spence (2005, p. 1260) reported that their “present findings corroborate the claim (Allan, 1975; Hirsh & Sherrick, 1961) that judgments of temporal order and judgments of simultaneity (vs. successiveness) are fundamentally different.” In the context of their unimodal experiments on tactile temporal processing, Shore et al. (2005, p. 1252) stated that “it has been argued that TOJs require more information about the stimuli before a correct response can be made” and that “this increased processing requirement might reveal more subtle effects than the simpler simultaneity judgments used in previous studies.” Furthermore, Zampini et al. (2003a, p. 208) noted that TOJ and SJ tasks “may reflect very different processes/mechanisms (i.e., one related to multisensory binding, and the other related to temporal discrimination instead . . .).” Such differences could call into question whether estimates of parameters such as the PSS are independent of the experimental method.

Explanations for the differences in PSS values shown in Table 1 can be based on differences in experimental methods. Shore, Spry, and Spence (2002) and Zampini, Guest, et al. (2005) indeed suggested that differences between PSS estimates derived from TOJ and SJ tasks should be experimentally investigated. Such research, however, has been carried out mainly in between-subjects designs (Smith, 1933; Vroomen et al., 2004) or in the context of a temporal recalibration study (Fujisaki et al., 2004; Vatakis et al., 2008; Vroomen et al., 2004). Fujisaki et al. measured recalibration of audio–visual simultaneity using both TOJ and SJ tasks in a within-subjects design and “obtained similar adaptation effects” (p. 774) for both tasks, although the effect was less stable for the TOJ task. Smeele (1994) found a significant difference between PSS values obtained from TOJ and SJ data for 9 out of the 10 speech stimulus types she used, by comparing the results for the 6 participants common to both experiments. Interestingly, she found a very high correlation between TOJ and SJ PSS values and a constant shift between TOJ and SJ PSS values of 94 msec (with the TOJ PSS values more negative). A between-subjects design comparing an SJ2 with a TOJ task was employed by Smith, and he reported consistent results between the two tasks, in that both produced negative overall PSS values, although

he did report that individual differences were somewhat greater in the SJ2 task. Vatakis et al. found a significant PSS shift when participants were exposed to an audio-visual speech video in which the auditory speech lagged behind the visual stream. The PSS shift, however, was only observed in the SJ task, not in the TOJ task. Vroomen et al. reported a between-subjects study in which TOJ and SJ2 tasks were compared using the same stimuli; they found similar shifts in the PSS using the two methods after adaptation to a series of stimulus pairs with specific offsets of their audio-visual components. In summary, the literature comparing the effects of experimental method on audio-visual synchrony perception is not only limited, but also has produced inconsistent results.

Research has shown that stimulus complexity may influence perceived simultaneity (Arrighi et al., 2006; Enoki et al., 2006; Vatakis & Spence, 2006a). Furthermore, Aschersleben (1999) suggested that differences in the ecological validity of the stimuli could explain some of the conflicting results reported in the TOJ and SJ literature. She considered the bouncing ball stimulus used by Lewkowicz (1996) to be more ecologically valid than the flash-click stimuli used in other studies. A ball bouncing off a surface promotes a clear impression of a moving stimulus, leading to an event with a causal interpretation (that the bouncing sound is caused by the ball hitting the surface), and thus a more predictable temporal relation: The auditory component is not expected to precede the visual impact of the ball. Furthermore, the ecologically invalid occurrence of sound preceding video should be more easily detected because, unlike in a flash-click stimulus, the occurrence of the visual event (the moment of impact) can be predicted from the context of the ball's trajectory, which is continuously available over some observation period.

The study reported here is the first, to our knowledge, that compares the effects of both stimulus complexity and three (TOJ, SJ2, and SJ3) temporal judgment methods in a completely within-subjects design without exposing the participants to asynchronous audio-visual adaptation stimuli. We used a simple stimulus, without motion or visual anticipatory information (i.e., a light flash and a click), as well as a more complex stimulus with natural motion and visual anticipatory information (i.e., a bouncing ball and a subjectively appropriate impact sound). Measurements for both stimulus types were made using TOJ, SJ2, and SJ3 tasks in separate, counterbalanced sessions with the same set of participants. The aim was to collect evidence within a single study to reveal causes for the apparent discrepancies in the literature.

METHOD

Two stimulus types were used in the present experiment: a flash-click and a bouncing ball stimulus. The measurements for the two stimulus types were initially carried out in two separate parts, but they are described here as a single experiment.

The flash-click stimulus has a long history in SJ and TOJ research (e.g., Exner, 1875; Hirsh & Sherrick, 1961; Jaśkowski et al., 1990). Given the absence of motion cues, anticipatory information, and apparent causality (but see Whipple, 1899), the flash-click stimulus may very well be the most simple stimulus in auditory-visual timing

perception research. As such, it is an excellent reference stimulus for the comparison of stimulus types of different complexity.

To investigate the effects of stimulus type on the PSS and on judgment sensitivity, a second stimulus was used: a simulation of a bouncing ball. This stimulus differs from the flash-click stimulus in two important, and somewhat related, aspects: (1) motion and (2) predictive information. Whereas the flash-click stimulus contains no motion, but merely a suddenly appearing and disappearing sphere, the bouncing ball stimulus contains a sphere that is continuously present in the visual scene and that produces a visual event by moving toward, and apparently contacting, a horizontal bar (and then bouncing back up again).

Several years passed between our bouncing ball measurements for the 4 participants described in van de Par, Kohlrausch, and Juola (2002) and for the 8 new participants. As a result, some equipment changes resulted in minor differences in the experimental method. Described here is the method used for the most recent measurements. Given that the changes in method were relatively minor (cf. van de Par et al., 2002) and that new measurements for 1 of the 4 original participants yielded results almost identical to those in the previous study, we think that the changes had little or no influence on the results and conclusions presented here.

Participants

Flash-click stimulus. Three female and 8 male participants, including the 4 authors, participated voluntarily. All participants had previously completed measurements with the bouncing ball stimulus. The participants varied in age from 26 to 58 years, with a mean of 35. All reported normal or corrected-to-normal vision and normal hearing.

Bouncing ball stimulus. Four female and 8 male participants, including the 4 authors, participated voluntarily. Eleven of these participants also took part in the flash-click portion of the experiment. The participants varied in age from 23 to 53 years, with a mean of 32. All reported normal or corrected-to-normal vision and normal hearing.

Design

Three parameters were systematically manipulated: (1) stimulus type (bouncing ball and flash-click), (2) relative delay (15 values, from -350 to +350 in steps of 50 msec, with negative values indicating audio first, positive values indicating video first, and 0 msec indicating physical synchrony), and (3) response rule (TOJ, SJ2, and SJ3). The experiment was divided into two parts, one for each of the stimulus types, with each part consisting of three sessions. The bouncing ball measurements were made before those for the flash-click. The response rules were each used in a separate session, with some time between sessions (typically 1 day). The orders in which the response rules were assigned were counterbalanced.

Stimulus

Flash-click. The visual part of the flash-click stimulus consisted of a white disk (97 cd/m² as measured using an LMT L1003 luminance meter) shown for one frame (12 msec) at a central position on the screen. The disk had a diameter of 49 pixels and subtended an area of about 1.4° at an unconstrained viewing distance of about 60 cm. The total duration of the visual stimulus was 2 sec, during which four corners of a surrounding square were visible, in order to indicate the central location of the flash. The square was presented to give the participants spatial and temporal information about the upcoming flash and click that was equivalent to the information they had when viewing the bouncing ball stimuli. The temporal occurrence of the flash was randomized, with the restriction that it occurred within the time window of 500–1,500 msec after the onset of the surrounding square. The acoustic part of the stimulus consisted of a 12-msec white-noise burst with a sound pressure level of 67 dB.

Bouncing ball. The visual part of the bouncing ball stimulus consisted of a white disk (identical to the disk in the flash-click stimulus) against a black background. An animation showed the

Table 1
Point of Subjective Simultaneity Values (PSS; in Milliseconds) for Audio-Visual Stimuli Reported by or Estimated From Studies Using Different Methods and Stimulus Types

Study	Note on Stimulus	TOJ			S12			S13		
		PSS	Range	N	PSS	Range	N	PSS	Range	N
Flash-Click (Stationary) Stimulus										
Aschersleben & Müssele (1999)		-13	n/r	16						
Bald et al. (1942) ^a	Hidden sound source	-9	n/r	32						
	Visible sound source	-1	n/r	30						
Bloch (1887; in Hamlin, 1895) ^b		-4	n/r	n/a						
Dinnerstein & Zlotogura (1968)		+71	±61	23		+24 ... +50	11			
Enoki et al. (2006)	Sudden appearance									
Exner (1875)										
Fujisaki et al. (2004)	No adaptation									
	Adaptation to 0 msec									
Hamlin (1895) ^c		-19	-34 ... -2	2						
Hirsh & Fraisse (1964) ^d		+29	n/r	8						
Hirsh & Sherrick (1961) ^e		+5	n/r	5						
Jaskowski et al. (1990) ^f		+48	+36 ... +69	3						
Keetels & Vroomen (2005)		+8	n/r	15						
Rutschmann & Link (1964)		-43	-46 ... -40	2						
Smith (1933) ^g		-8	n/r	40						
Spence et al. (2003) ^h		+20	±24	8						
Stone et al. (2001) ⁱ		-7	-30 ... +23	5						
Teatini, Farné, Verzella, & Berruocos (1976) ^j	Visual stimulus left	+5	-27 ... +27	5						
	Visual stimulus right	-10	-20 ... +1	6						
Tracy (described in Hamlin, 1895)		+1	n/r	13						
Vatakis et al. (2008)	Adaptation to 0 msec	-6	±24	10						
Vroomen et al. (2004)	Adaptation to 0 msec	-13	-73 ... +6	5						
Whipple (1899) ^k	Single presentation	-4	-16 ... +6	6						
	Repeated presentation									
Zampini, Guest, et al. (2005) ^l	Same stimulus location									
	Different stimulus location									
Zampini et al. (2003a) ^m	Same stimulus location	+60	±17	9						
	Different stimulus location	+75	±19	9						

disk apparently moving down over a distance of 460 pixels toward a horizontal bar, impacting the bar, and then bouncing back up again with the same (constant) acceleration. The total duration of the visual stimulus was 2 sec, whereas the duration of the movement was varied randomly between 1 and 2 sec. As a result, the disk appeared to float for a random time before and after the actual animation took place. The impact could occur only when a flash in the flash-click stimulus was also possible (i.e., the occurrence of the visual event in both stimuli was randomized within the same time window of 500–1,500 msec after stimulus onset). The maximum speed of the disk (achieved just before impact) varied between 1.6 arcmin/msec, when the movement lasted for the entire 2-sec duration of the stimulus, and 3.3 arcmin/msec, when the movement lasted for only 1 sec. The auditory component of the stimulus consisted of a 500-Hz sinusoid with a rectangular onset and an exponentially decaying envelope $e^{-t/T}$, with $T = 20$ msec. At the moment of onset, the auditory stimulus had a sound pressure level of 80 dB.

Apparatus

The visual stimulus was shown on a Dell D1025HE CRT monitor at a resolution of $1,024 \times 768$ pixels and at an 85-Hz refresh rate. The auditory stimulus was played through a Creative SB Live! sound card, a Fostex PH-50 headphone amplifier, and Sennheiser HD 265 linear headphones. Participants were seated in front of the monitor at an approximate distance of 60 cm and responded using a keyboard. The setting was a dimly lit, sound-attenuating room.

The timing control of the auditory and visual signals was determined by using a photocell that measured the light emitted by a flashing circle on the screen. The position of the flashing circle corresponded to the region in which the visual event was presented during the experiments. The electrical output of the photocell was shown as a trace on an oscilloscope, together with a pulsed tone played in synchrony with the flashing circle, using the same equipment that presented the stimuli. In these calibration conditions, the synchrony between the flash and the tone was shown to be accurate to within ± 2 msec.

Some research has shown that using loudspeakers to produce a spatial separation between the physical sources of auditory and visual information can result in increased sensitivity to audio-visual asynchrony (Keetels & Vroomen, 2005; Spence et al., 2003; Zampini et al., 2003a, 2003b). However, Zampini et al. (2003b) did not

find such a performance difference when they presented the auditory stimulus over headphones (i.e., in a different stimulus position), rather than over a loudspeaker that was effectively in the same position as the visual stimulus. Therefore, we can expect that the presentation of the auditory stimulus over headphones in the present study (as opposed to using speakers) did not affect the resulting judgments.

Procedure

Prior to the start of the experiment, participants received written instructions. They were informed about the content of the audio-visual stimulus and, depending on the experimental method, were instructed to judge the temporal order (TOJ task) or (a)synchrony of the auditory and visual components of the stimulus (SJ2 and SJ3 tasks), using numeric keys on the keyboard for their responses. These responses were unsped, and the response choice was briefly indicated on the screen after each keypress, but no feedback was provided on its correctness. The experiment started with a practice block, which consisted of 30 randomly presented trials appropriate to the subsequent experimental session (i.e., the same response rule was used); across these trials, each of the 15 audio-visual delays was presented twice. After practice, four measurement blocks were presented. Within each block of 225 trials, all 15 delays were presented 15 times each, resulting in a total of 60 judgments per delay \times response rule combination for each participant.

RESULTS AND DISCUSSION

The main data of interest are the proportions of judgment responses as a function of the audio-visual delay within stimulus pairs. We will treat the TOJ data first, followed by the SJ data, and will conclude with a comparison of the data for the two types of response rule.

TOJ Data

Raw TOJ data averaged over all participants are shown in Figure 1, which shows the proportions of “video first” responses as a function of relative delay for the flash-click (left panel) and bouncing ball (right panel) stim-

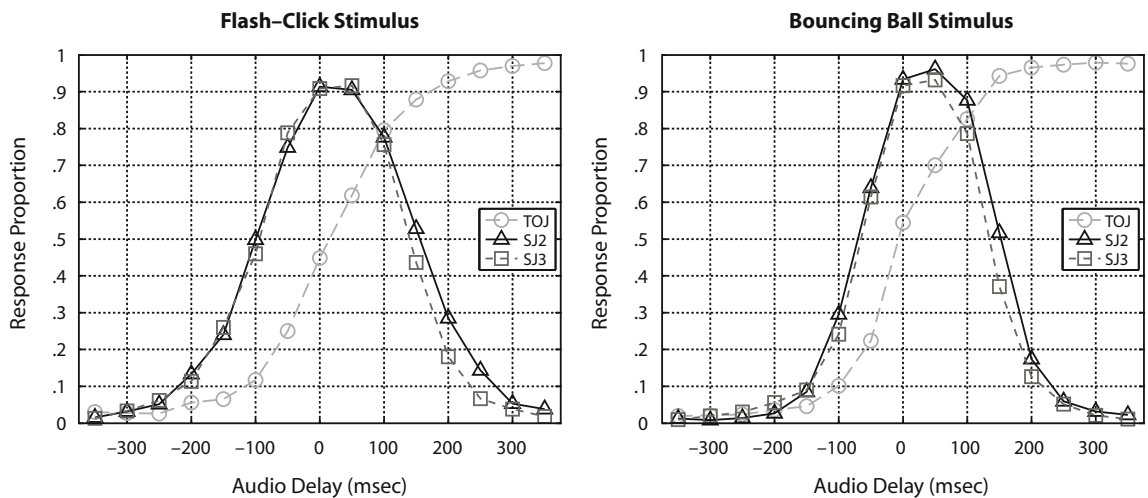


Figure 1. Raw “video first” response proportions from the temporal order judgment (TOJ) task (light gray circles and dashed lines) and raw “synchronous” response proportions from two-choice (SJ2; black triangles and solid lines) and three-choice (SJ3; dark gray squares and dotted lines) synchrony judgment tasks as a function of audio delay (in milliseconds), for the flash-click (left panel) and bouncing ball (right panel) stimuli. Response proportions were averaged over participants ($N = 11$ for the flash-click TOJ data, $N = 10$ for the flash-click SJ2 and SJ3 data, and $N = 12$ for all of the bouncing ball data).

uli. These mean results show that the 50% point falls at about +15 msec for the flash-click stimulus and at about -7 msec for the bouncing ball stimulus. Furthermore, the slopes of the video-first curves appear to be shallower at positive than at negative audio delays. These average data are included mainly as an illustration, since all analyses were performed on the data of individual participants.

The TOJ data were fitted using a psychometric function of the form $\gamma + (1 - \gamma - \lambda)F(\alpha, \beta)$, with F being the cumulative Gaussian distribution with mean α and standard deviation β . Fitting was done using the *psignifit* toolbox, which implements the maximum-likelihood method described by Wichmann and Hill (2001). Stimulus-independent “lapses” (e.g., pressing the wrong key or blinking during stimulus presentation) are fitted by the γ (lower horizontal asymptote) and λ (higher horizontal asymptote) parameters. The γ and λ parameters later are removed from the theoretical or underlying psychometric function, which is assumed to represent the actual perception of the participant (rather than the observed performance; see also Spence et al., 2003, who used a somewhat different approach than the one described here).

The TOJ 50% point is used as a PSS estimate, since at this point participants presumably report the temporal order of the two modalities by a random guess, suggesting that the stimuli are perceived to be synchronous. The steepness of the slope at the TOJ 50% point is used as a sensitivity measure. The PSS and slope values are summarized in Table 2 for the 10 participants who produced useful data for both stimulus types (see below for details).

The effect of stimulus type was tested using pairwise t tests on the 10 participants who performed all judgment tasks for both stimulus types. No difference in PSSs was revealed between the two stimulus types [$t(9) < 1$]. Stimulus type also had no effect on TOJ slopes [$t(9) = 1.15, p = .28$].

SJ Data

Raw SJ data averaged over all participants are shown in Figure 1, which plots the proportion of “synchronous” responses as a function of relative delay for the flash-click (left panel) and bouncing ball stimuli (right panel);

the audio-first, video-first, and asynchronous response curves are left out for clarity). Since 1 participant was excluded from the analysis (see below), data for this participant are not included in the figure. It can be seen in Figure 1 that the response patterns for both SJ tasks are very similar, although participants responded “synchronous” more often in the SJ2 task than in the SJ3 task when the auditory component followed the visual component (i.e., for positive delays).

In order to allow for different slopes in the two halves of the curve, two cumulative Gaussians were used to fit the overall synchrony response curve (similar to the procedure of Smeele, 1994). The raw synchronous data were split into two parts by determining the maximum synchronous response proportion and including this data point (or these points) in both halves of the synchrony curve. The underlying psychometric (synchronous) curve was determined by removing the γ parameter from the fitted curve and by keeping the λ parameter, which indicates the maximum “synchronous” response proportion. For SJ2, the asynchronous response curve is by definition complementary to the synchronous curve, and as a result that asynchronous curve does not necessarily reach a proportion of 0. For SJ3, the audio-first and video-first curves were fitted separately. In order to keep the SJ3 curve-fitting procedure as similar as possible to the procedure for SJ2, the audio-first and video-first curves were also not constrained to reach a proportion of 0.

For the SJ procedures, we determined the *synchrony range* by calculating the intersection points of the synchrony curve with its neighboring asynchrony judgment curve(s). The synchrony range has an intuitive importance to synchrony perception, because within this range an audio-visual delay is more often judged to be synchronous than not. Therefore, we will call the intersection points *synchrony boundaries*. The synchrony boundaries may be considered asynchrony detection thresholds, and the width of the synchrony range can thus be seen as a measure of sensitivity to audio-visual temporal intervals. The two synchrony boundaries are distinguished by referring to the response curves with which they intersect—that is, with audio first and video first. These labels correspond

Table 2
Mean Point of Subjective Simultaneity (PSS) Values, Mean Boundaries and Widths of the Synchrony Range, and Mean (Absolute) Slopes Determined at Audio-First and Video-First Synchrony Boundaries for Participants Common to the Flash-Click and Bouncing Ball Measurements (N = 10)

	Flash-Click Stimulus						Bouncing Ball Stimulus					
	SJ2		SJ3		TOJ		SJ2		SJ3		TOJ	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
PSS (msec)	25	11	18	10	21	11	41	7	33	7	7	11
Boundaries and Widths of Synchrony Range (msec)												
Audio first	-111	16	-107	17			-79	11	-78	14		
Video first	161	12	144	8			160	5	144	8		
Width	271	19	251	17			239	10	222	18		
Slope ($\times 10^{-2}$ δ proportion/msec)												
Audio first	0.83	0.10	0.84	0.12			1.43	0.27	1.25	0.17		
Video first	0.75	0.04	0.76	0.06	0.52	0.06	0.98	0.10	0.96	0.10	0.64	0.08

Note—SJ2 and SJ3, synchrony judgment task with two or three response categories; TOJ, temporal order judgment.

to the relative delays at which the synchrony boundaries were located (i.e., an audio-first synchrony boundary was always located at a negative relative delay, which means that the auditory component preceded the visual component, and the video-first boundary was always located at a positive relative delay).

As with the TOJ data, slopes are used here as measures of sensitivity to changes in relative delay at the synchrony boundaries. The resulting average synchrony boundaries for the two SJ response rules are shown in Table 2, together with the average slopes at those synchrony boundaries. Since flash-click “synchronous” response proportions for 1 of the participants were always below either the audio-first or the video-first curve, no synchrony range could be determined for this participant using our definition. Since the PSS is defined as the center of the synchrony range, also no PSS could be calculated, and therefore this participant’s data were excluded from further analyses.

In summary, Table 2 contains two types of sensitivity measures. The synchrony boundaries define the range of delays that are predominantly judged to be synchronous, and thus these boundaries can be seen as asynchrony detection thresholds. The slopes measured at the synchrony boundaries define how sensitive people are at the transition from perceived synchrony to perceived asynchrony.

In order to test whether the two SJ tasks resulted in different “synchronous” response patterns, the locations of the synchrony boundaries and the slopes at these boundaries were analyzed via two repeated measures ANOVAs, with the factors stimulus type (flash-click or bouncing ball), SJ task (SJ2 or SJ3), and location (audio- or video-first synchrony boundary) as independent variables. A significant main effect of SJ rule was found [$F(1,9) = 8.37, p = .018$], indicating that video-first synchrony boundaries were shifted farther in the direction of larger, more positive audio delays for SJ2 (+160 msec) than for SJ3 (+144 msec) [$t(9) = 3.64, p = .005$, when we compared SJ2 and SJ3 synchrony boundaries averaged over stimulus types in a paired t test]. Furthermore, a marginally significant effect of stimulus type on the location of the synchrony boundaries was found [$F(1,9) = 4.27, p = .069$]: The audio-first bouncing ball synchrony boundaries (−78 msec) were shifted more toward the point of objective simultaneity than were the flash-click synchrony boundaries (−109 msec) [$t(9) = 3.09, p = .013$, when we compared flash-click and bouncing ball synchrony boundaries averaged over SJ rules in a paired t test], whereas no differences were found between video-first boundaries [$t(9) < 1$]. The finding that audio-first bouncing ball synchrony boundaries were shifted more toward the point of objective simultaneity than were flash-click synchrony boundaries is in line with the findings reported by Enoki et al. (2006), who found no difference between video-first boundaries when comparing stimuli with and without visual anticipatory information, but who did report that audio-first synchrony boundaries were closer to physical synchrony for stimuli with visual anticipatory information than for stimuli without such information.

Overall, the audio-first synchrony boundary was closer to physical synchrony (−94 msec) than was the video-

first synchrony boundary (+152 msec), as tested in a 2 (stimulus type) \times 2 (SJ task) \times 2 (location) repeated measures ANOVA [$F(1,9) = 13.70, p = .005$]. That is, the synchrony range was asymmetrically distributed around physical synchrony.

The effects of stimulus type and SJ task on the resulting PSS values were tested in a 2 (stimulus type) \times 2 (SJ task) repeated measures ANOVA. Identical to the main effects on the locations of the synchrony boundaries, a marginally significant effect of stimulus type on PSS [$F(1,9) = 4.27, p = .069$] and a significant effect of SJ task [$F(1,9) = 8.37, p = .018$] were found: Bouncing ball PSS values (+37 msec) tended to be larger, or shifted farther away from physical synchrony, than flash-click PSS values (+22 msec), and SJ2 PSS values (+33 msec) were larger than SJ3 PSS values (+25 msec).

With respect to the different measures of sensitivity to audio-visual temporal intervals, the widths of the synchrony ranges were not significantly different when we compared the bouncing ball (230 msec) and flash-click (261 msec) stimulus types [$F(1,9) = 4.17, p = .072$]. The slope data, however, did show a significant effect of stimulus type [$F(1,9) = 13.98, p = .005$] and a marginally significant effect of location [the synchrony boundary at which the slope was measured; $F(1,9) = 4.24, p = .069$], but no effect of SJ task [$F(1,9) < 1$]. Slopes were steeper for the bouncing ball stimulus (1.16×10^{-2} vs. 0.80×10^{-2} δ proportion/msec), and overall, audio-first slopes (1.09×10^{-2} δ proportion/msec) tended to be steeper than video-first slopes (0.86×10^{-2} δ proportion/msec). In separate analyses for the two stimulus types, the difference between the audio-first and video-first slopes was significant for the bouncing ball stimulus [$F(1,11) = 6.68, p = .025$] but not for the flash-click stimulus [$F(1,9) < 1$].

Comparison of TOJ and SJ Response Rules

Since one of the primary goals of this study was to investigate the effect of experimental method on the resulting PSS values, we first compared PSS values estimated from the TOJ, SJ2, and SJ3 data. The effects of experimental method on the sensitivity measures will be discussed later.

The PSS values were estimated from TOJ data by taking the 50% point of the fitted video-first curve. The SJ PSS values were determined by finding the midpoint of the two synchrony boundaries (see Table 2). The correspondence between the resulting PSS values for the three response rules is plotted in Figure 2, which shows SJ3 PSS values (squares) and TOJ PSS values (crosses) plotted against SJ2 PSS values for each individual participant both for the flash-click (left panel) and the bouncing ball (right panel) stimuli. As can be seen from Figure 2, all but three of the SJ PSS values are positive (i.e., represent relative delays at which the visual event precedes the auditory event), whereas about half of the TOJ PSS values are negative. The ranges of the PSS values (from −61 to +72 msec for SJ tasks and from −48 to +72 msec for the TOJ task) are similar to those of (individual) PSS values reported in the literature for flash-click and simple stimuli (see the examples in Table 1), with the exception of the highly negative flash-click SJ PSS values reported for 1 partici-

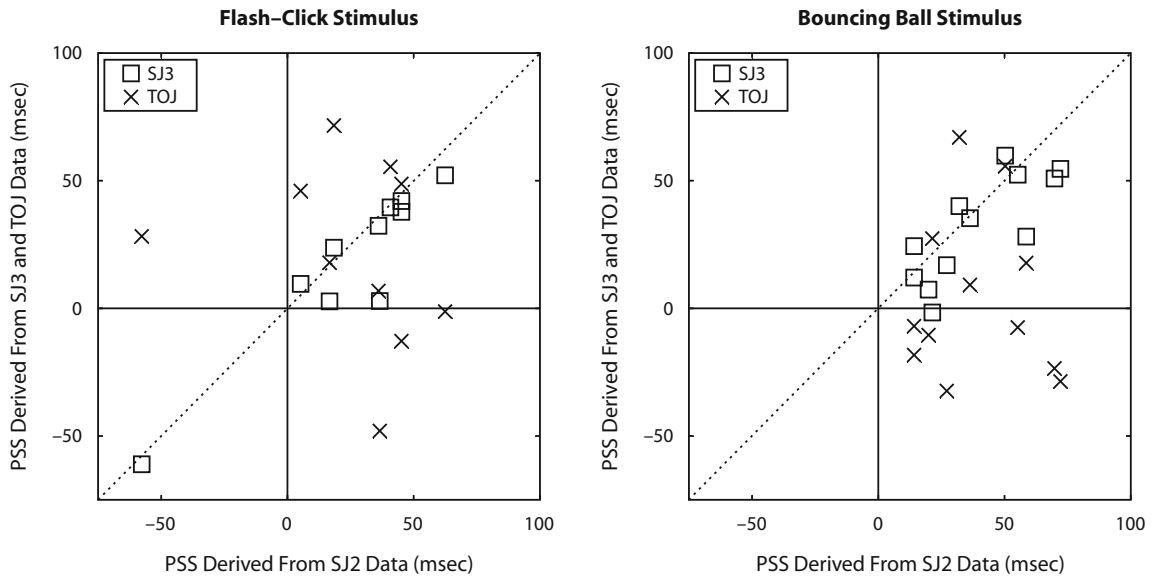


Figure 2. Three-choice synchrony judgment (SJ3; squares) and temporal order judgment (TOJ; crosses) point of subjective simultaneity (PSS) values, in milliseconds, plotted against two-choice synchrony judgment (SJ2) PSS values, for the flash-click (left panel) and bouncing ball (right panel) stimuli. The dotted lines indicate perfect correspondence between PSS values.

pant, which not only stand out in the present study but also appear to be a rarity in the literature (but see Vatakis et al., 2008).

A 2 (stimulus type) \times 3 (response rule) repeated measures ANOVA of the PSS values, with Greenhouse–Geisser-adjusted degrees of freedom,¹ did not yield significant effects of stimulus type [$F(1,9) < 1$] or response rule [$F(1.08, 9.68) = 1.97, p = .19; \epsilon = 0.54$]. Separate analyses for both stimulus types confirmed this null effect for the flash-click stimulus [$F(1.04, 9.36) < 1; \epsilon = 0.52$]. For the bouncing ball stimulus, however, a main effect of response rule was found [$F(1.17, 12.86) = 8.00, p = .012; \epsilon = 0.58$], indicating a systematic effect of experimental method on the resulting mean PSS. A multiple comparison using the Holm–Šidák procedure (see Arrighi et al., 2006; Ludbrook, 1998) showed that the PSS values for SJ2 (+39 msec, on average) and SJ3 (+32 msec) differed marginally [$t(11) = 1.96, p = .076$], but that TOJ PSS values (+4 msec) were significantly different from those for both SJ2 [$t(11) = 3.04, p = .011$] and SJ3 [$t(11) = 2.65, p = .022$].

For the flash-click stimulus, average SJ and TOJ PSS values were quite similar (+25 msec for SJ2, +18 msec for SJ3, and +21 msec for TOJ). Nevertheless, the TOJ task yielded PSS values that did not correlate with those obtained using an SJ2 ($r = -.25, n.s.$) or SJ3 ($r = .04, n.s.$) task, whereas SJ2 and SJ3 PSS values correlated highly ($r = .94, p < .001$; however, $r = .78, p = .013$, without 1 participant who clearly provided an influential observation in terms of the correlation, as can be seen from the left panel of Figure 2). For the bouncing ball stimulus, as can be seen in the right panel of Figure 2, the within-subjects consistency in PSS estimates obtained from the two SJ

tasks (plotted as squares) was quite high, which was reflected in a significant correlation ($r = .79, p = .002$). If the PSS data obtained from TOJ and SJ2 tasks (plotted as crosses) are compared, however, no apparent relationship emerges. Indeed, no significant correlations were found between TOJ PSS values and the PSS estimates obtained from the SJ2 ($r = -.08, n.s.$) and SJ3 ($r = .13, n.s.$) tasks. Consistency in SJ PSSs was also found by Stone et al. (2001), who found a high correlation ($r = .95$) between individual PSS values that were obtained using an SJ2 task for a flash-click stimulus presented (in two separate sessions for each participant) at distances of 0.5 and 3.5 m.

Although no correlation was found between the TOJ and SJ PSS values, it should be noted that the TOJ 50% points for both stimulus types were always located within each individual's synchrony range. The resulting PSS values clearly demonstrate an effect of experimental method on the resulting (individual) PSS values, because the TOJ task yielded PSS values that did not correlate with those obtained using an SJ task, whereas those from SJ2 and SJ3 tasks correlated highly. To test whether the correlation between SJ PSS values is a consequence of determining the PSS by averaging the synchrony boundaries, SJ2 and SJ3 PSS values were also determined from the mean of individually fitted truncated Gaussians (see Fujisaki et al., 2004) using maximum-likelihood estimation (Myung, 2003). The correlation analysis yielded similar results, and thus suggests that the consistency found between SJ PSS values indeed can be attributed to the experimental method itself. From the lack of correlation between SJ and TOJ PSS values, we conclude that PSS values may very well reflect different underlying perceptions being measured by the TOJ and SJ tasks.

The effect of experimental method on estimates of sensitivity to audio–visual timing differences was tested by correlating the widths of the synchrony ranges from SJ data with the slopes from TOJ data (see Table 2). A relatively wide synchrony range would indicate a high tolerance for audio–visual asynchronies, and thus relatively low sensitivity. Similarly, a relatively shallow slope, indicated by a small slope value, would also indicate low sensitivity. If the two sensitivity measures derived from SJ and TOJ data are related, a negative correlation would be expected. For the bouncing ball stimulus, a significant correlation indeed was found between sensitivity measures derived from the SJ2 and TOJ data ($r = -.62, p = .042$), but not between measures derived from SJ3 and TOJ data ($r = -.35, p = .29$). For our flash–click data, no significant correlations were found between TOJ slopes and the widths of the synchrony ranges obtained from SJ2 ($r = -.02, p = .95$) and SJ3 ($r = .17, p = .64$) data.

GENERAL DISCUSSION

The goal of the study described here was to investigate the influence of experimental method and stimulus type on estimates of audio–visual timing perception. The main goal was to determine whether previously reported results of ecologically invalid (i.e., negative or auditory-leading) PSS values can be attributed to the experimental method used to measure them. Although the PSS values reported here, averaged over all participants, were always positive, individual PSS values sometimes were negative. The vast majority (11 out of 14) of the negative individual PSS values were obtained from TOJ data, with negative SJ PSS values all stemming from 1 participant. Interestingly, whereas this participant produced highly negative PSS values for the SJ2 (–58 msec) and SJ3 (–61 msec) tasks with flash–click stimuli, for the TOJ task he showed a clearly positive PSS (+28 msec). Overall, the PSS values estimated from the two SJ tasks are consistent with each other within experiments (i.e., they correlated significantly across individual participants). The PSS values obtained from TOJ data, on the other hand, did not correlate with the SJ PSS values within each experiment. If PSS values were independent of the experimental method, we would expect similar PSS estimates for each of the three methods used here, but this is clearly not the case. From the SJ PSS values, it can be concluded that the concept of PSS is a relatively stable one, which, given the lack of correlation between SJ and TOJ PSS values, is reflected less well in the TOJ task.

An alternative conclusion from our PSS values here might be that SJ tasks are less sensitive to changes in audio–visual synchrony perception, and hence result in more consistent PSS values, whereas the observed inconsistency for the TOJ task might reflect the higher sensitivity of this method. Indeed, Shore et al. (2005, p. 1252) stated that TOJs “might reveal more subtle effects than the simpler simultaneity judgments.” As such, the TOJ task may be preferred over an SJ task, depending on the specific objectives of a study. Vatakis and Spence (2007), for example, used a TOJ task when evaluating the “unity

assumption” in audio–visual speech stimuli, because “any bias to assume that the matched auditory and visual stimuli should be bound together could not have biased participants toward either a ‘vision first’ or ‘sound first’ response as would have been the case had we used a simultaneity judgment task” (p. 750). Although we do not deny the potential usefulness of the TOJ task, the present study suggests that the results of a single TOJ experiment should not be taken at face value when one is primarily interested in the perception of synchrony itself.

From a theoretical perspective, one could argue that differences in PSS estimates between SJ and TOJ tasks are due to differences in the underlying perceptions that are being measured. Although there is no general agreement on whether successiveness discrimination is a *sufficient* condition for the perception of temporal order, it is generally accepted that the perception of successiveness is a *necessary* requirement for such order to be correctly perceived (Allan, 1975; Hirsh & Sherrick, 1961; Mitrani, Shekerdjiiski, & Yakimoff, 1986). Given that successiveness is a precondition for the perception of temporal order, it follows that there is a range of relative delays, which we call the *synchrony range*, within which TOJ performance is expected to be at about chance level, and that relative delays within this range “will be coded as simultaneous” (Allan & Kristofferson, 1974, p. 40). The present findings show that the PSS estimate from the TOJ task can occur anywhere within the synchrony range, and thus can be expected to be susceptible to individual response strategies, rather than to reflect the “true” point of subjective simultaneity. Indeed, it has been noted by Sternberg and Knoll (1973, p. 632) that “under this definition [i.e., the TOJ 50% point] stimuli at the PSS need not necessarily give rise to a perception of simultaneity.” The lack of correlation between the observed TOJ and SJ PSS values can be explained by differences in response strategy when judging a relative delay that cannot reliably be perceived as being successive. Indeed, Schneider and Bavelier (2003) reported differences between PSS values from unimodal (visual) TOJ and SJ tasks and attributed those differences to response biases or criterion shifts present in the TOJ, but not the SJ, task.

The observation that the TOJ 50% points occur virtually anywhere in the synchrony range, combined with the consistency in SJ response patterns observed across the two SJ tasks, offers support for the notion that perception of successiveness is at least a necessary requirement for the correct perception of temporal order (Allan, 1975; Hirsh & Sherrick, 1961; Mitrani et al., 1986). In this respect, it is interesting to note that the video–first synchrony boundary is shifted farther in the direction of larger, more positive audio delays for the SJ2 than for the SJ3 task. If perception of successiveness is a necessary, but not a sufficient, condition for the correct perception of temporal order, the opposite may have been expected. Whereas the SJ2 task allows participants to respond “asynchronous” for asynchronous delays just outside the synchrony range, the SJ3 task forces them to either guess the temporal order or choose “synchronous.” Such guessing behavior would have resulted in synchrony boundaries that are shifted farther away from

physical synchrony in the SJ3 than in the SJ2 task. Interestingly, the reverse was found, which may be interpreted as support for the notion that perception of successiveness is not only necessary, but also sufficient for the correct perception of temporal order, as assumed by Allan.

A possible explanation for the small observed difference between SJ2 and SJ3 synchrony boundaries can be offered by the response frequency equalization tendency (Erlebacher & Sekuler, 1971; Parducci & Haugen, 1967; Sekuler & Erlebacher, 1971; Tune, 1964). Such a tendency is more likely to come into play in the SJ2 than in the SJ3 task, since a participant is more likely to notice the predominant use of one response category over the other(s) when only two categories are available. Indeed, the audio delays used in the present experiment were judged to be asynchronous (an average of 605 responses over both stimulus types and SJ tasks) more often than they were judged to be synchronous (average of 295 responses). Although we cannot rule out that the proportion of "synchronous" responses (approximately .33) was due to response frequency equalization in the SJ3 task, the fact that no significant difference² in the frequency of "synchronous" responses was found between SJ2 and SJ3 tasks (305 and 285 "synchronous" responses, respectively, averaged over both stimulus types) suggests that a response frequency equalization tendency had little if any effect on the data in the present study. This conclusion is further supported by the finding that in the SJ3 task the audio-first response category was used more frequently³ than the video-first category—that is, the response frequencies were not equal (336 and 279 responses, respectively, averaged over both stimulus types).

Directly related to the topic of response frequency is the range of audio delays used in the present study. A wider range of delays is likely to result in a relative increase in the number of subjectively asynchronous delays, and as such might influence the effect that a potential response frequency equalization tendency has on the resulting responses. Because the effect of response frequency equalization is limited to the region of uncertainty, however, one would expect the effect of narrowing or widening the range of audio delays to be limited, if present at all. Indeed, Zampini, Guest, et al. (2005) failed to demonstrate an effect of using a wider range of audio delays on the width of the synchrony range, and so concluded that "the participants' performance was based on perceptual rather than decisional factors" (p. 538). As such, the specific range of audio delays used in the present study should have had, at most, little influence on the results.

Similar to the lack of correlation found between SJ and TOJ PSS values, sensitivity measures showed no consistent correlation for the SJ and TOJ tasks. Furthermore, our observation that slopes for the bouncing ball stimulus were steeper at the audio-first synchrony boundary than at the video-first synchrony boundary challenges the implicit assumption in TOJ studies that there exists only one value of sensitivity to audio-visual timing delays. (When fitting TOJ data with a cumulative Gaussian with two slope parameters, no significant slope differences were found between those on either side of the 50% point.) It

seems that sensitivity to audio-visual timing delays depends on the position within the synchrony range at which it is determined, as well as on the type of stimulus used in the judgment task. Since sensitivity estimates are derived from judgments of single audio-visual timing delays, they provide an indirect measure of the ability of observers to discriminate between two audio-visual timing delays. Therefore, future research should examine whether sensitivity to audio-visual timing is different at different relative positions in the synchrony range. Such research could also benefit from a more direct measure of the sensitivity to audio-visual timing differences, such as by using discrimination threshold measures.

We found that the SJ3 task may result in very small proportions of "synchronous" responses, as reported for 1 participant's flash-click data (see also Jaśkowski, 1993; Stelmach & Herdman, 1991). A maximum proportion of "synchronous" responses below .5 is difficult to interpret in terms of perceived simultaneity, and participants with such a response pattern may need to be removed from the data set. Removal of participants for their inability to perform the task is not restricted to the SJ task, however. Spence, Shore, and Klein (2001), Vatakis and Spence (2006b), and Zampini et al. (2003a, 2003b), for example, reported excluding participants' TOJ data from further analysis, because the participants' PSS values were beyond the range of the tested audio-visual delays. As such, both SJ and TOJ tasks can result in data that are unsuitable in the context of the specific aim of an experiment.

Inherent in the use of a within-subjects design is the experience that is gained by participants during the experiment, which consequently might influence the results (see, e.g., Shore et al., 2002; Spence et al., 2001; Zampini, Brown, et al., 2005). Since here our different experimental methods were counterbalanced between subjects, the possible influence of increasing experience during the course of the experiment should not have affected the results of the comparisons between the different experimental methods. Since the two experiments were performed sequentially, however, it is possible that the experience gained during the bouncing ball measurements influenced the flash-click measurements in a systematic way. Since the authors were generally more experienced than the other participants in psychophysical tasks, and more specifically in the temporal order and synchrony judgment tasks used here, the effect of experience was more closely investigated by reanalyzing the data with author/nonauthor as an additional between-subjects variable. The results indicate that the authors were more sensitive overall (i.e., produced steeper slopes) than naive participants, demonstrating an effect of experience on performance in audio-visual temporal performance.

No influence of stimulus type on PSS estimates was found in an overall comparison of both SJ and TOJ PSS values. Therefore, Aschersleben's (1999) suggestion that the conflicting PSS estimates reported in the TOJ and SJ literature (see Table 1) can be explained by differences in the ecological validity of the stimuli is not supported (although it should be noted that the ecological validity of the bouncing ball stimulus is somewhat limited in com-

parison with, for example, the video clips of audio–visual speech and musical performances used by Vatakis & Spence, 2006a). The absence of an overall main effect of stimulus type is not surprising, however, given the lack of consistency in PSS estimates obtained from the TOJ data in the present experiments. When comparing PSS values estimated from SJ data, however, we found a trend toward larger (more positive) PSS values for the bouncing ball stimulus than for the flash–click stimulus. This finding is in line with Aschersleben's (1999, p. 424) suggestion that, for ecologically valid stimuli, "our cognitive system takes into account the different conduction times for visual and auditory stimuli." When we restricted the PSS comparison to TOJ PSS values, an opposite trend emerged: The average PSS obtained from bouncing ball data was shifted toward physical synchrony (i.e., is less positive; the trend was nonsignificant, however) as compared with the average PSS from flash–click data. Overall, stimulus type seems to have had only a marginally significant effect on the resulting PSS for the stimuli used in the present experiments, and this effect was limited to data obtained using an SJ task.

An important difference between the two stimuli used here is that the bouncing ball stimulus allowed observers to anticipate the moment of occurrence of the visual impact, whereas the flash–click stimulus offered no such information. Another difference between the two stimuli is the higher complexity of the bouncing ball stimulus introduced by its (apparently natural) motion. Since those two stimulus characteristics are closely related to each other in the stimuli used in the present article, they need to be disentangled in future research in order to investigate the individual effects of predictive value and motion cues on estimates of human performance in the perception of audio–visual timing. Stekelenburg and Vroomen (2007) reported that visual information may facilitate auditory processing, but only when vision predicts when the sound occurs. Since both auditory leads and lags were present in our study as part of the experimental manipulation, visual predictive information could not reliably be used to predict the occurrence of the auditory event. As such, further investigation into multisensory integrative mechanisms is required in order to be able to assess the influence of visual anticipatory information on the processing of the asynchronous auditory component of an audio–visual stimulus.

As a result of the presence of visual anticipatory information in the bouncing ball stimulus, it might be easier for observers to detect a leading audio component (i.e., when the auditory component is presented before the visual impact has occurred) than to detect a lagging audio component, perhaps because of the violation of ecological validity, or the naturally occurring cause-and-effect relationship. Although the predictive value in the bouncing ball stimulus, following from the continuous presence of the visual component (the bouncing ball), is closely related to the higher relative ecological validity of that stimulus, it is not necessarily the ecological validity per se that results in the larger, more positive PSS values found for the bouncing ball stimulus. If the increase in ecological validity, rather than the predictive

value of the stimulus, were the main cause for the larger PSS values we found using an SJ task, such values should also be found for more complex ecologically valid stimuli, such as a video of a person speaking. Smeele (1994) used video recordings of the face of a speaker pronouncing 10 nonsense consonant–vowel–consonant syllables in an SJ2 task. Averaged over the 10 syllables, Smeele found a slightly negative overall "synchronization point" (–12 msec), with half of the individual syllables resulting in negative PSS values, and the other half resulting in positive PSS values. The finding of an ecologically invalid (negative) PSS for an ecologically valid stimulus in an SJ task indicates that the influence of the stimulus is more dependent on the specific stimulus contents than on the ecological validity of the stimulus per se.

In summary, given the inconsistency between PSS estimates obtained from the TOJ and SJ tasks, it is confusing and, in the case of a TOJ, perhaps even incorrect to refer to the different estimates with the same term *point of subjective simultaneity*. Depending on the specific objectives of a study, either a TOJ or an SJ may be preferred over the other task. When the goal is to investigate the point of subjective simultaneity for audio–visual stimuli, however, the SJ task should be preferred over the TOJ task. Given that the SJ2 task may be more susceptible to response bias than the SJ3 task, the latter task should be considered superior, at least within the range of stimuli and methods reported here.

AUTHOR NOTE

Parts of these results have been presented as conference contributions (Juola, van Eijk, Kohlrausch, & van de Par, 2007; van de Par et al., 2002; van Eijk, Kohlrausch, van de Par, & Juola, 2006). Thanks are due Martin Boschman, Constant Hak, and Christophe Stoeltinga for technical assistance, and Jean Vroomen and three anonymous reviewers for valuable comments on previous versions of the manuscript. We also acknowledge support from a grant by the NWO (Dutch Organization for Scientific Research) to the third author. Correspondence relating to this article may be sent to R. L. J. van Eijk, Den Dolech 2, IPO 1.25, P.O. Box 513, 5600 MB Eindhoven, The Netherlands (e-mail: r.l.j.v.eijk@tue.nl).

REFERENCES

- ALLAN, L. G. (1975). The relationship between judgments of successiveness and judgments of order. *Perception & Psychophysics*, **18**, 29–36.
- ALLAN, L. G., & KRISTOFFERSON, A. B. (1974). Successiveness discrimination: Two models. *Perception & Psychophysics*, **15**, 37–46.
- ARRIGHI, R., ALAIS, D., & BURR, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision*, **6**, 260–268.
- ASCHERSLEBEN, G. (1999). On the asymmetry of the temporal contiguity window: Commentary on Lewkowicz. In G. Aschersleben, T. Bachmann, & J. Müssele (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 421–424). Amsterdam: Elsevier.
- ASCHERSLEBEN, G., & MÜSSELE, J. (1999). Dissociations in the timing of stationary and moving stimuli. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 1709–1720.
- BALD, L., BERRIEN, F. K., PRICE, J. B., & SPRAGUE, R. O. (1942). Errors in perceiving the temporal order of auditory and visual stimuli. *Journal of Applied Psychology*, **26**, 382–388.
- DINNERSTEIN, A. J., & ZLOTOGURA, P. (1968). Intermodal perception of temporal order and motor skills: Effects of age. *Perceptual & Motor Skills*, **26**, 987–1000.

- DIXON, N. F., & SPITZ, L. (1980). The detection of auditory visual desynchrony. *Perception*, **9**, 719-721.
- ENOKI, K., WASHIKITA, K., & YAMADA, M. (2006). Detection threshold of asynchrony between auditory and visual stimuli for various motion patterns of a ball. In *Proceedings of the 9th Western Pacific Acoustics Conference (WESPAC IX)* [CD-ROM]. Seoul: Acoustical Society of Korea.
- ERLEBACHER, A., & SEKULER, R. (1971). Response frequency equalization: A bias model for psychophysics. *Perception & Psychophysics*, **9**, 315-320.
- EXNER, S. (1875). Experimentelle Untersuchung der einfachsten psychischen Prozesse: III. Der persönlichen Gleichung zweiter Theil. *Archiv für die gesammte Physiologie des Menschen & der Thiere*, **11**, 403-432.
- FUJISAKI, W., SHIMOJO, S., KASHINO, M., & NISHIDA, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, **7**, 773-778.
- HAMLIN, A. J. (1895). On the least observable interval between stimuli addressed to disparate senses and to different organs of the same sense. *American Journal of Psychology*, **6**, 564-575.
- HIRSH, I. J., & FRAISSE, P. (1964). Simultanéité et succession de stimuli hétérogènes. *L'Année Psychologique*, **64**, 1-19.
- HIRSH, I. J., & SHERRICK, C. E., JR. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, **62**, 423-432.
- HOLLIER, M. P., & RIMELL, A. N. (1998, September). *An experimental investigation into multi-modal synchronization sensitivity for perceptual model development*. Paper presented at the 105th Convention of the Audio Engineering Society, San Francisco, CA.
- JAŚKOWSKI, P. (1993). Selective attention and temporal-order judgment. *Perception*, **22**, 681-689.
- JAŚKOWSKI, P., JAROSZYK, F., & HOJAN-JEZIERSKA, D. (1990). Temporal-order judgments and reaction time for stimuli of different modalities. *Psychological Research*, **52**, 35-38.
- JUOLA, J. F., VAN EIJK, R. L. J., KOHLRAUSCH, A., & VAN DE PAR, S. (2007, July). *Synchrony detection and temporal order judgments for bimodal stimuli*. Paper presented at the joint meeting of the Experimental Psychology (U.K.) & Psychonomic (U.S.A.) Societies, Edinburgh, Scotland.
- KEETELS, M., & VROOMEN, J. (2005). The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Experimental Brain Research*, **167**, 635-640.
- KING, A. J., & PALMER, A. R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research*, **60**, 492-500.
- LEWKOWICZ, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 1094-1106.
- LUDBROOK, J. (1998). Multiple comparison procedures updated. *Clinical & Experimental Pharmacology & Physiology*, **25**, 1032-1037.
- MCGRATH, M., & SUMMERFIELD, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America*, **77**, 678-685.
- MITRANI, L., SHEKERDIJISKI, S., & YAKIMOFF, N. (1986). Mechanisms and asymmetries in visual perception of simultaneity and temporal order. *Biological Cybernetics*, **54**, 159-165.
- MYUNG, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, **47**, 90-100.
- NEUMANN, O., & NIEPEL, M. (2004). Timing of "perception" and perception of "time." In C. Kaernbach, E. Schröger, & H. Müller (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition* (pp. 245-269). Mahwah, NJ: Erlbaum.
- PARDUCCI, A., & HAUGEN, R. (1967). The frequency principle for comparative judgments. *Perception & Psychophysics*, **2**, 81-82.
- PÖPPEL, E. (1988). *Mindworks: Time and conscious experience* [Trans. of *Grenzen des Bewußtseins* by T. Artin]. Boston: Harcourt Brace Jovanovich.
- RIHS, S. (1995). The influence of audio on perceived picture quality and subjective audio-video delay tolerance. In R. Hamberg & H. de Ridder (Eds.), *Advanced methods for the evaluation of television picture quality: Proceedings of the MOSAIC Workshop* (pp. 133-137). Eindhoven, The Netherlands: Institute for Perception Research.
- RUTSCHMANN, J., & LINK, R. (1964). Perception of temporal order of stimuli differing in sense mode and simple reaction time. *Perceptual & Motor Skills*, **18**, 345-352.
- SCHNEIDER, K. A., & BAVELIER, D. (2003). Components of visual prior entry. *Cognitive Psychology*, **47**, 333-366.
- SEKULER, R., & ERLEBACHER, A. (1971). The invalidity of "invalid results from the method of constant stimuli": A common artifact in the methods of psychophysics. *Perception & Psychophysics*, **9**, 309-311.
- SHORE, D. I., GRAY, K., SPRY, E., & SPENCE, C. (2005). Spatial modulation of tactile temporal-order judgments. *Perception*, **34**, 1251-1262.
- SHORE, D. I., SPRY, E., & SPENCE, C. (2002). Confusing the mind by crossing the hands. *Cognitive Brain Research*, **14**, 153-163.
- SMEELE, P. M. T. (1994). *Perceiving speech: Integrating auditory and visual speech*. Unpublished PhD thesis, Technische Universiteit Delft, The Netherlands. Retrieved April 11, 2007, from repository.tudelft.nl/file/257315/201176.
- SMITH, W. F. (1933). The relative quickness of visual and auditory perception. *Journal of Experimental Psychology*, **16**, 239-257.
- SPENCE, C., BADDELEY, R., ZAMPINI, M., JAMES, R., & SHORE, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one. *Perception & Psychophysics*, **65**, 318-328.
- SPENCE, C., SHORE, D. I., & KLEIN, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology: General*, **130**, 799-832.
- STEKELENBURG, J. J., & VROOMEN, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, **19**, 1964-1973.
- STELMACH, L. B., & HERDMAN, C. M. (1991). Directed attention and perception of temporal order. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 539-550.
- STERNBERG, S., & KNOLL, R. L. (1973). The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 629-685). New York: Academic Press.
- STONE, J. V., HUNKIN, N. M., PORRILL, J., WOOD, R., KEELER, V., BEANLAND, M., ET AL. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society B*, **268**, 31-38.
- TEATINI, G., FARNÈ, M., VERZELLA, F., & BERRUECOS, P., JR. (1976). Perception of temporal order: Visual and auditory stimuli. *Giornale Italiano di Psicologia*, **3**, 157-164.
- TUNE, G. S. (1964). Response preferences: A review of some relevant literature. *Psychological Bulletin*, **61**, 286-302.
- VAN DE PAR, S., KOHLRAUSCH, A., & JUOLA, J. F. (2002). Some methodological aspects for measuring asynchrony detection in audio-visual stimuli. In A. Calvo-Manzano, A. Perez-Lopez, & J. S. Santiago (Eds.), *Proceedings of the Forum Acusticum 2002* [CD-ROM]. Porto, Portugal: Universidade do Porto, Faculdade de Engenharia.
- VAN EIJK, R. L. J., KOHLRAUSCH, A., VAN DE PAR, S., & JUOLA, J. F. (2006, June). *The influence of psychophysical procedure and stimulus type on estimates of human performance in detecting audio-visual asynchrony*. Paper presented at the 7th International Multisensory Research Forum (IMRF), University of Dublin, Trinity College, Dublin, Ireland.
- VAN WASSENHOVE, V., GRANT, K. W., & POEPEL, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, **45**, 598-607.
- VATAKIS, A., NAVARRA, J., SOTO-FARACO, S., & SPENCE, C. (2008). Audiovisual temporal adaptation of speech: Temporal order versus simultaneity judgments. *Experimental Brain Research*, **185**, 521-529.
- VATAKIS, A., & SPENCE, C. (2006a). Audiovisual synchrony perception for music, speech, and object actions. *Brain Research*, **1111**, 134-142.
- VATAKIS, A., & SPENCE, C. (2006b). Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neuroscience Letters*, **393**, 40-44.
- VATAKIS, A., & SPENCE, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics*, **69**, 744-756.
- VROOMEN, J., KEETELS, M., DE GELDER, B., & BERTELSON, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, **22**, 32-35.
- WHIPPLE, G. M. (1899). On nearly simultaneous clicks and flashes. *American Journal of Psychology*, **10**, 280-286.
- WICHMANN, F. A., & HILL, N. J. (2001). The psychometric function:

- I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, **63**, 1293-1313.
- ZAMPINI, M., BROWN, T., SHORE, D. I., MARAVITA, A., RÖDER, B., & SPENCE, C. (2005). Audiotactile temporal order judgments. *Acta Psychologica*, **118**, 277-291.
- ZAMPINI, M., GUEST, S., SHORE, D. I., & SPENCE, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics*, **67**, 531-544.
- ZAMPINI, M., SHORE, D. I., & SPENCE, C. (2003a). Audiovisual temporal order judgments. *Experimental Brain Research*, **152**, 198-210.
- ZAMPINI, M., SHORE, D. I., & SPENCE, C. (2003b). Multisensory temporal order judgments: The role of hemispheric redundancy. *International Journal of Psychophysiology*, **50**, 165-180.

NOTES

1. In the following results, a correction factor ϵ is reported only when a Greenhouse-Geisser adjustment of the degrees of freedom was applied.
2. As tested in a 2 (stimulus type) \times 2 (SJ task) repeated measures ANOVA [$F(1,9) = 2.61, p = .14$].
3. As tested in a 2 (stimulus type) \times 2 (response category: audio or video first) repeated measures ANOVA [$F(1,9) = 5.87, p = .038$].

(Manuscript received May 7, 2007;
revision accepted for publication February 26, 2008.)