

# Apparent causality affects perceived simultaneity

Armin Kohlrausch · Rob van Eijk · James F. Juola ·  
Inge Brandt · Steven van de Par

Published online: 14 August 2013  
© Psychonomic Society, Inc. 2013

**Abstract** The present research addresses the question of how visual predictive information and implied causality affect audio–visual synchrony perception. Previous research has shown a systematic shift in the likelihood of observers to accept audio–leading stimulus pairs as being apparently simultaneous in variants of audio–visual stimulus pairs that differ in (1) the amount of visual predictive information available and (2) the apparent causal relation between the auditory and visual components. An experiment was designed to separate the predictability and causality explanations, and the results indicated that shifts in subjective simultaneity were explained completely by changes in the implied causal relations in the stimuli and that predictability had no added value. Together with earlier findings, these results further indicate that the observed shifts in subjective simultaneity due to causal relations among auditory and visual events do not reflect a mere change in response strategy, but rather result from early multimodal integration processes in event perception.

**Keywords** Audio–visual · Synchrony perception · Asynchrony detection · Apparent causality · Visual predictive information

---

A. Kohlrausch · R. van Eijk · J. F. Juola · I. Brandt  
Human-Technology Interaction Group,  
School of Innovation Sciences, Eindhoven  
University of Technology, Eindhoven,  
The Netherlands

A. Kohlrausch  
Philips Research Europe Eindhoven,  
Eindhoven, The Netherlands

J. F. Juola  
University of Kansas,  
Lawrence, KS, USA

S. van de Par  
Acoustics group, University of Oldenburg,  
Oldenburg, Germany

A. Kohlrausch (✉)  
Eindhoven University of Technology, IPO 1.25,  
P.O. Box 513, 5600 MB Eindhoven, The Netherlands  
e-mail: a.kohlrausch@tue.nl

Simple events, such as an object falling toward the floor or a pair of hands clapping, produce predictable contact sounds. For events happening nearby, visual and acoustic impact information arrive in near synchrony at the ears and eyes of an observer. Humans have evolved to tolerate certain asynchronies in the sensory processing of auditory and visual stimuli. Such tolerance is adaptive, due to the idiosyncrasies that can occur in the relative arrival times of auditory and visual stimuli, most obviously in cases in which the distance from the perceiver varies. In several experiments, we have studied subjective timing judgments in different event contexts (van Eijk, Kohlrausch, Juola, & van de Par, 2007a, 2007b, 2008, 2009, 2010). For example, relative to simple, apparently unrelated clicks and flashes, the visible movement of a ball falling toward a table has reliable predictive and causative implications, allowing for the anticipation of the time of the collision sound from visually observing the event. By varying the types of predictive and postdictive information available in audio–visual stimulus pairs, we have observed systematic shifts in the point of subjective simultaneity (PSS). The PSS is commonly defined as the midpoint of a range of various audio–visual stimulus onset asynchronies (SOAs) at which “synchronous” judgments predominate over “auditory-first” or “visual-first” judgments (Enoki, Washikita, & Yamada, 2006; Fujisaki, Shimojo, Kashino, & Nishida, 2004; Stone et al., 2001; van Eijk et al., 2008; Zampini, Guest, Shore, & Spence, 2005). This apparent synchrony range is thus limited by two boundaries, outside of which judgments of apparent asynchrony prevail. The synchrony ranges and PSSs differ in events with no apparent relation between the audio and visual components from those events that appear to enter into continuous and predictable causative relations. On the basis of the results reported here, we argue that the data are more supportive of the importance of causal directionality, over mere predictability derived from context alone. That is, the attribution of causality is a stronger explanation than visual predictability for the shifts in timing judgments observed in various simulated audio–visual events.

The ability to anticipate the occurrence of the visual component of an audio–visual event reduces an observer’s tolerance of

auditory-leading asynchronies (Enoki et al., 2006; van Eijk et al., 2008).<sup>1</sup> Enoki et al. measured synchrony perception for four different visual stimuli. Three of these included motion patterns in which a ball moved downward or sideways and impacted a bar. These conditions allowed for anticipation of the moment of impact. In the fourth condition, the ball suddenly appeared on the surface of the bar. Participants had to judge single stimulus presentations and to respond with either “synchronous” or “asynchronous” (the “SJ2 procedure,” in the terminology used by van Eijk et al., 2008). Synchrony boundaries at the perceived transition from primarily “synchronous” to primarily “visual-leading” judgments were not affected by visual anticipation, and in the four conditions they ranged from 226 to 238 ms (averaged across 11 participants). The boundaries between “auditory-leading” and “synchronous” judgments were highly similar for the three conditions with visual movement (−90 to −95 ms), but the condition with a sudden appearance had a clearly more negative synchrony boundary of −153 ms. That is, an audio–visual event with the auditory stimulus leading by a few tens of milliseconds is less likely to be judged as synchronous when visual predictive information is available than when no visual predictive information is available.

The same observation was made by van Eijk et al. (2008) in a comparison of synchrony judgments with a bouncing-ball simulation containing visual predictive information, as compared to a flash–click stimulus without visual predictive information. Averaged across the two synchrony judgment procedures—SJ2 (“asynchronous” and “synchronous” judgments) and SJ3 (“auditory-first,” “synchronous” and “visual-first” judgments)—the position of the auditory-leading synchrony boundaries differed significantly by about 30 ms, occurring at −109 ms for the flash–click stimulus and at −78.5 ms for the bouncing-ball stimulus. In contrast, the visual-leading boundary positions were nearly identical for the two stimulus types (152.5 and 152 ms; see Table 2 in van Eijk et al., 2008).

The introduction of visual predictive information shifts not only the auditory-leading synchrony boundary, but also the PSS, toward more-positive SOAs. Indeed, Enoki et al. (2006) showed that visible anticipation of an event resulted in a more positive, visual-leading PSS (68–74 ms) than occurred when no anticipation was possible (36 ms), and van Eijk et al. (2008) found the same (but nonsignificant) trend, showing that a bouncing-ball simulation with visual predictive information tended to yield a more positive PSS (37 ms) than did a flash–click stimulus without visual predictive information (22 ms).

The difference in PSS values for stationary audio–visual stimuli, like flash–click stimuli, and for stimuli based on visual motion leading to impacts is also reflected in the

literature overview in Table 1 of van Eijk et al. (2008). There, PSS values derived from synchrony judgment procedures have a median value of 14 ms across studies (on the basis of ten studies using the SJ2 procedure) for flash–click stimuli, whereas the median value for (simple) motion stimuli amounts to 37 ms, on the basis of eight studies. This between-subjects outcome thus closely agrees with the within-subjects comparison reported in van Eijk et al. (2008).

Two potential explanations exist for the shifts in the auditory-leading synchrony boundaries and PSSs observed by Enoki et al. (2006) and van Eijk et al. (2008). First, the shift could be due to the observer’s ability to predict the occurrence of the auditory event from the ball’s trajectory. An alternative explanation relies on the causal interpretation of the audio–visual stimulus: The bouncing-ball stimulus clearly promotes the impression that the auditory event (an impact sound) is caused by the visual event (a ball hitting a horizontal surface). As a result, the auditory event should not precede the visual event, since that would violate the implied causal relationship present within the audio–visual stimulus. A flash–click stimulus, or other stimuli without visual motion, however, do not promote a clear causal interpretation. Since an audio-leading stimulus no longer violates the implied causal relationship, observers tend to be about as tolerant of an audio-leading temporal interval as of one in which the visual component leads, resulting in a PSS that is closer to 0 ms.

In order to study the effects of visual predictive information and apparent causality on perceived simultaneity in more detail, we performed some studies using a simple, ecologically valid audio–visual stimulus with a common underlying physical event (van Eijk et al., 2007a, 2007b, 2010). The visual stimulus consisted of a simulation of Newton’s cradle toy: five balls suspended from strings, with the leftmost ball initially describing a left-to-right pendulum movement, impacting its neighbor,

**Table 1** Characteristic parameters of the synchrony judgment distributions with a Newton’s cradle stimulus

View of Newton’s Cradle	Audio-First Synchrony Boundary	PSS	Video-First Synchrony Boundary	Threshold, Audio-First Side	Threshold, Video-First Side
Entire	−61	29	120	33	71
Left half	−72	25	122	32	83
Right half	−114	−2	109	71	65

“Synchrony boundaries” are the points at which the proportions of “synchronous” judgments are equal to the proportions of “audio-first” and “video-first” judgments, respectively. The PSS is the point of subjective simultaneity, defined as the midpoint between the two synchrony boundaries, observed in a synchrony judgment task using a simulated Newton’s cradle toy (van Eijk et al., 2010). Also reported are the temporal discrimination thresholds for an audio-leading pair or a video-leading pair against the standard of the 0-ms SOA for the same stimuli (van Eijk, 2008). All values are in milliseconds

<sup>1</sup> In the judgment literature, negative SOAs are typically used for audio-first pairs, and positive SOAs for video-first pairs. The zero point is the point of objective synchrony (POS).

and as a result launching the rightmost ball. The spatial separation between the initial leftmost swing and the resulting rightmost swing allows for independent manipulations of visual predictive and postdictive information, in a physically plausible setting. Visual predictive information was manipulated by either showing or hiding the initial leftmost swing of the pendulum movement, whereas visual postdictive information was manipulated by either showing or hiding the following rightmost swing. As expected, showing the leftmost swing in the stimulus promoted the interpretation that the auditory event (impact sound) was caused by the visual event (the leftmost ball hitting its neighboring ball). That causal interpretation apparently resulted in a higher sensitivity for audio-leading than for audio-lagging delays, producing clearly positive PSSs. Hiding the leftmost swing, however, reversed this apparent causal relation, suggesting the interpretation that the auditory event caused the launching of the rightmost ball. This altered causal interpretation increased the observers' tolerance for audio-leading delays, resulting in a PSS closer to 0 ms (see Table 1).

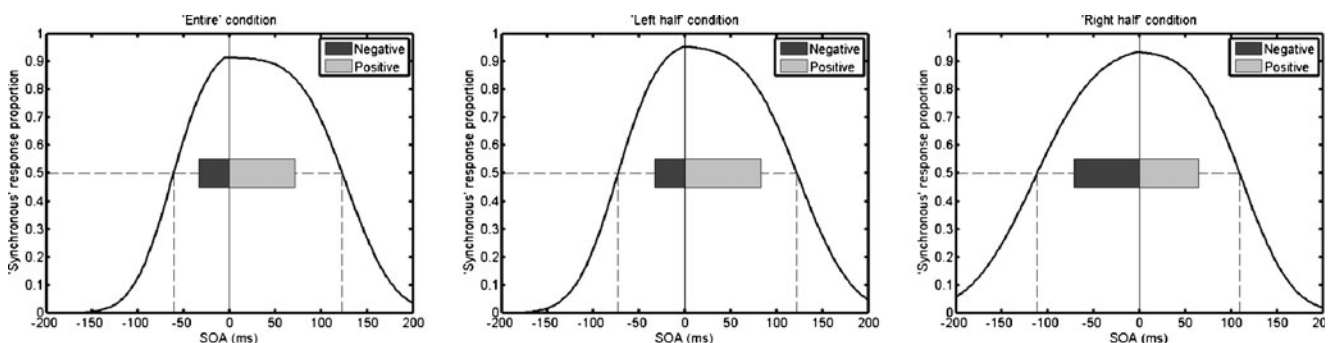
In other experiments, which are described in detail in van Eijk (2008), we evaluated whether the PSS shifts caused by changes in the availability of visual predictive information are truly sensory in nature, or whether they result from a change in response strategy. We had shown before that discrimination sensitivity between audio–visual stimuli with different SOAs is strongly related to the relative locations of the synchrony boundaries for the stimuli (van Eijk et al., 2009). We therefore measured discrimination thresholds using the Newton's cradle toy simulation and a physically synchronous reference stimulus. In each three-interval observation trial, the first interval contained a standard event with an auditory–visual SOA of 0 ms. It was followed by two other audio–visual pairs—one with an SOA also equal to 0 ms, and the other with a negative (audio-leading) or positive (video-leading) SOA—which were tested in different trial blocks. Using a standard staircase procedure (Levitt, 1971) to adapt the SOA in the target interval, we obtained SOA discrimination thresholds for positive and negative SOAs around the SOA of 0 ms (see Table 1). Figure 1 shows, for the three visual conditions, average

discrimination thresholds as horizontal bars that are plotted within the corresponding synchrony response curves. This figure indicates that the stimulus-induced shifts in PSS position are reflected in shifts in threshold asymmetries around 0 ms. This close relationship between PSS location and temporal interval discrimination thresholds indicates that both result at the level of early multimodal integration processes, rather than at later response or decision levels (see also Kim, Peters, & Shams, 2012, for a similar argument).

Manipulation of the visibility of the leftmost ball in the Newton's cradle toy limits the observer's ability to predict the occurrence of the visual event, but simultaneously it affects the apparent causal relationship in the stimulus. That is, if removing the visual predictive information from the stimulus indeed results in a PSS shift toward 0 ms, it cannot be concluded whether that shift is due to the presence versus absence of visual predictive information or to the apparent presence versus absence of a causal relationship in the stimulus.

The motivation for the present study was to develop a new stimulus that would allow us to disambiguate these two explanations. In this stimulus, a ball apparently fell toward and bounced off a bar that was either visible or invisible and that could be located at different vertical positions. As a result, in the invisible-bar conditions, the observer was unable to predict the moment of occurrence of the visual event, whereas in the visible-bar conditions, the moment of occurrence of the apparent impact could be predicted. The causal interpretation of the stimulus remained unaffected by the change in visibility of the bar: A ball bouncing off an invisible bar caused an impact sound identical to that experienced when the bar was visible. This modified bouncing-ball stimulus was used in three different movement patterns that resulted in corresponding changes in the apparent causality between the visual and auditory stimulus components.

The basic stimulus was the same as the one used in van Eijk et al. (2008, 2009). As compared to the original version, a number of variations were introduced, and their general effects were evaluated in pilot experiments. First, the position of the bar was not kept fixed but was chosen randomly, with equal



**Fig. 1** Mean positive (light gray) and negative (dark gray) audio–visual asynchrony detection thresholds for the “entire” (left panel), “left-half” (middle panel), and “right-half” (right panel) visual conditions of the Newton's cradle toy stimulus, plotted within the average synchrony response curves for the same three visual conditions and determined by averaging the parameters of best fit over 11 participants

probabilities, to be at one of five different vertical heights. Given that participants could see the actual vertical position of the bar as soon as the animation started, we expected that this manipulation should have only a minor effect on the synchrony boundaries, and this expectation was supported in pilot experiments involving some of the authors. Second, we varied the motion pattern of the ball in a way that was analogous to the three visual conditions used in research with the Newton's cradle simulation (van Eijk et al., 2010). The "full-motion" pattern contained both visual predictive and postdictive information. The "sudden-stop" motion pattern, showing the ball moving down and impacting the horizontal bar, after which it remained lying on the bar, contained visual predictive information, but no visual postdictive information. The "lift-off" motion pattern showed the ball initially lying on the bar, but then lifting off in a manner analogous to the upward motion in the "right-half" condition of van Eijk et al. (2010).

In the main experiment, two relevant stimulus manipulations were tested. The three motion patterns were used as stimuli in combination with the randomly varying vertical position of the bar. In one condition, the bar had the same luminance as the white disk simulating the bouncing ball ("visible-bar" condition). In the second condition, the bar was invisible because it had the same luminance as the background ("invisible-bar" condition). In both stimulus conditions, a clear impression of a bouncing ball occurred, but in the invisible-bar condition, the visual predictive information was effectively removed from the stimulus.

We expected that the visible-bar conditions would result in PSS shifts—mainly due to shifts in the auditory-leading boundary position—similar to those obtained using the Newton's cradle toy: clearly positive PSS estimates for the full-motion and sudden-stop conditions, but a PSS estimate at less positive SOAs for the lift-off condition. For the invisible-bar conditions, two predictions could be made: If visual predictive information was the major cause for the observed shifts in the auditory-leading synchrony boundary, those differences should disappear in the invisible-bar conditions. If apparent causality was the major cause, however, the boundary differences in the three visual conditions should *not* disappear. Thus, a comparison of the results of the visible-bar conditions with those of the invisible-bar conditions should allow us to disentangle the two possible explanations used to describe our previous synchrony perception results.

## Method

### Design

The experiment had a  $2 \times 3$  design manipulating the visual predictive information of the location of the bounce (two conditions: bar visible or invisible) and the motion pattern of

the bouncing ball (three motion patterns: full motion, sudden stop, or lift off) as independent variables. The combined visibility conditions and motion patterns created six main conditions. In each main condition, the five locations of the bar and the 15 SOAs were combined to produce 75 different stimuli. These stimuli were presented in random order, and the sets of 75 trials were repeated three times. Thus, each SOA was presented a total of 15 times. The relative proportion of reported "synchronous" responses for a given SOA was measured as the dependent variable.

### Stimuli

Visual predictive information was manipulated by showing a white or a black bar against a black background, which, respectively, represented the visible- and invisible-bar conditions. Initially (except for in the lift-off condition), the stimulus showed a ball (white disk) "floating" near the top of the screen for a randomly determined duration of 500–1,000 ms. After that, the ball started falling with an acceleration that was constant between conditions (i.e., the ball's acceleration was completely independent of the position of the vertical bar, and thus offered no information regarding the moment of occurrence of the visual event). The bar was set at one of five different positions. The lowest possible bar position was 530 pixels below the initial "floating" position of the ball. This position was used as a reference for determining the four other bar positions, which were determined such that they corresponded to the ball's position at 5, 10, 15, and 20 frames before the frame showing the impact of the ball with the bar in the "reference position" (lowest bar position). In the lift-off condition, the ball was located at the top of the appropriate bar position for that trial.

Fifteen different audio–visual SOAs in the range from –350 to +350 ms were chosen, with interval differences of 50 ms. In each main condition, the locations of the bar and the audio offsets were combined to produce 75 different stimuli. These stimuli were presented in random order, and the sets of 75 trials were repeated three times.

### Participants

The participants were 13 male and seven female students, 19 to 27 years of age with a mean age of 23 years ( $SD = 2.4$ ). All participants reported normal or corrected-to-normal vision and normal hearing ability and were paid after participating in both sessions (see also the "Procedure" section).

### Apparatus

The stimuli were displayed on a Dell D1025HE CRT computer monitor with a resolution of  $1,024 \times 768$  pixels and a refresh rate of 85 Hz. The sound was produced by a Creative SB Live! sound card, a Fostex PH-50 headphone amplifier, and Sennheiser HD

265 linear headphones. The stimuli were shown on the screen, and the responses were made using a keyboard. The experiment was held in a dimly lit, sound-attenuated room.

## Procedure

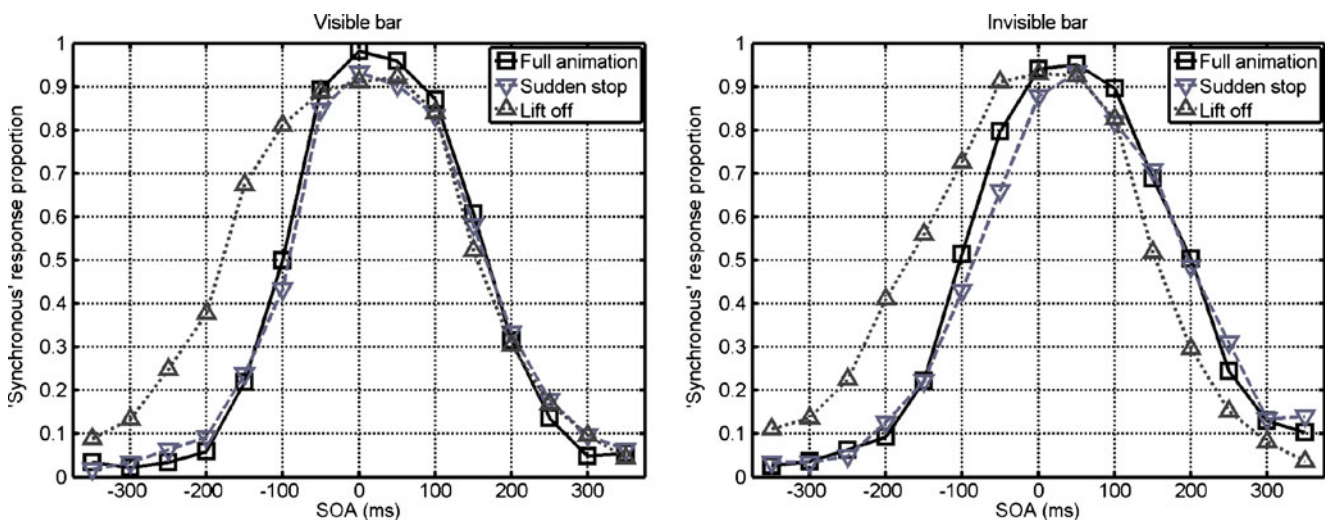
Participants received written instructions prior to the start of the experiment that explained the SJ3 task and the stimulus contents. An experimental session consisted of three main parts, each consisting of a short practice phase and three measurement phases of 75 stimuli each (i.e., one presentation of each of the Bar Position  $\times$  SOA conditions per measurement phase). Participants were allowed to take breaks between phases. A session always started with the full-motion condition, whereas the next two parts contained the sudden-stop and lift-off conditions in counterbalanced order. On each trial, participants were to press one of the numbers 1, 2, or 3 on the keyboard to indicate judgments of “audio first,” “synchronous,” or “video first,” respectively. The participants took part in two sessions in which the visibility of the bar was manipulated in a counterbalanced fashion.

## Results

Two participants were removed from further analysis because their proportions of “synchronous” responses did not fall above the 0.5 level for some conditions. As a result, no synchrony boundaries, and thus no PSS values, could be calculated for these participants. Raw data for the visible-bar conditions, averaged over the 18 remaining participants, are shown in the left panel of Fig. 2, which shows the proportions of “synchronous” responses as a function of the relative delays for all three visual conditions (“audio-first” and “video-first”

response curves are left out for clarity). Full-motion and sudden-stop conditions produced very similar response patterns, whereas the lift-off visual condition resulted in a clear expansion of the synchrony judgment curve to include more negative SOAs. The synchrony response data of each participant were fitted by cumulative Gaussian functions that could have different slopes on the two sides. The PSS values and synchrony boundaries, derived from the individually fitted Gaussian functions and averaged over the 18 remaining participants, are shown in Table 2. As a measure of goodness of fit, we computed the deviance between the model estimates and the data, according to the procedures for computing  $D$  values outlined by Wichmann and Hill (2001), for each condition and for each side of the synchrony curve. The participant-averaged mean  $D$  values ranged from 3.7 to 5.8, without any systematic influence of visibility and visual stimulus condition. These values are all smaller than the number of degrees of freedom (equal to the number of points used to fit the psychometric functions, which was seven or eight) of the corresponding chi-square distribution, indicating that a major part of the variance in the data is explained by the model fits, and that the remainder is not significant.

The result in the left panel of Fig. 2 and the corresponding entries in Table 2 for the visible-bar condition permit a clear prediction for the essential manipulation of this experiment, measuring synchrony perception while making the “floating” bar invisible. If the difference between the lift-off condition and the two other (visible-bar) conditions was caused by the *availability of visual predictive information*, the average synchronous responses for the three movement patterns, obtained in the invisible-bar condition, should be highly similar. Specifically, the auditory-first synchrony boundaries for the full-motion and sudden-stop conditions should overlap with the lift-off auditory-first synchrony boundary. It can be seen from



**Fig. 2** Raw “synchronous” responses for all three motion patterns, averaged over participants ( $N = 18$ ) for the visible-bar (left panel) and invisible-bar (right panel) conditions

**Table 2** Characteristic parameters of the synchrony judgment distributions using a simulation of a ball bouncing off a visible or invisible bar

Animation	Visible Bar			Invisible Bar		
	Audio-First Boundary	PSS	Video-First Boundary	Audio-First Boundary	PSS	Video-First Boundary
Full motion	-106 (8)	31 (5)	169 (10)	-104 (9)	45 (6)	193 (11)
Stop	-104 (7)	32 (6)	168 (13)	-92 (12)	52 (7)	196 (13)
Lift off	-179 (11)	-6 (7)	167 (9)	-169 (16)	-1 (6)	168 (10)

The ball appeared to fall and bounce from the bar (full-motion condition), to fall and stick to the bar (stop condition), or to lift off from the bar (lift-off condition). Data are reported in milliseconds, with standard errors in parentheses

the right panel in Fig. 2 and from the values in Table 2 that this is not the case. In fact, the figure suggests that auditory-first boundaries are hardly affected at all by the manipulation of prior visibility of the impact position.

A 3 (motion pattern)  $\times$  2 (bar visibility)  $\times$  2 (location: audio or video leading) repeated measures analysis of variance (ANOVA) on synchrony boundary estimates resulted in significant effects of bar visibility [ $F(1, 17) = 19.35, p < .001$ ], motion pattern [ $F(1.32, 22.42) = 31.71, \epsilon = 0.66, p < .001$ ], and location [ $F(1, 17) = 358.87, p < .001$ ],<sup>2</sup> as well as a Motion Pattern  $\times$  Location interaction [ $F(1.45, 24.72) = 23.70, \epsilon = 0.73, p < .001$ ]. In line with the visual impression, the auditory-first synchrony boundaries were not affected by bar visibility,<sup>3</sup> but, unexpectedly, video-first synchrony boundaries were,  $F(1, 17) = 13.69, p = .002$ , in a 3 (motion pattern)  $\times$  2 (bar visibility) repeated measures ANOVA on video-first synchrony boundaries. In pairwise  $t$  tests, significant differences were found between visible and invisible video-first synchrony boundaries for full-motion [ $t(17) = 5.19, p < .001$ ] and sudden-stop motion patterns [ $t(17) = 3.29, p = .004$ ], but not for the lift-off motion pattern [ $t(17) = 0.10, p = .92$ ].

In a final analysis, we investigated whether bar height exerted a systematic influence on the auditory-first boundary position. Unfortunately, our fitting procedure, from which we derive the boundary positions and PSS values, requires us to use all available data for a given condition and participant, and does not allow for a separate fit for each height of the bar; the data are much too noisy, with only three responses per bar height per participant. We therefore collapsed all of the data across participants for each bar height condition for the visible and invisible bars separately. These data allowed us to estimate the 50 % point for synchrony judgments on each side of the overall curve, and thereby the auditory-first and video-first synchrony boundaries. An inspection of these data showed no reliable changes in the synchrony boundaries across bar height

<sup>2</sup> All ANOVAs were performed using Greenhouse–Geisser-adjusted degrees of freedom if the sphericity assumption was violated, as tested using Mauchley's test of sphericity. Only when a Greenhouse–Geisser adjustment of the degrees of freedom was applied is the correction factor  $\epsilon$  reported.

<sup>3</sup>  $F(1, 17) = 1.48, p = 0.24$ , in a 3 (motion pattern)  $\times$  2 (bar visibility) repeated measures ANOVA on the auditory-first synchrony boundaries.

conditions, neither for the visible nor for the invisible bar conditions.

## Discussion

The present experiment was concerned with the question of how different types and amounts of contextual information influence the perception of simultaneity for audio–visual stimulus pairs. In some contexts, the stimuli appear to be completely unrelated (as in the laboratory presentation of a flash of light and an audible click). In other cases, the two are highly related, as in when one views a simulation of a ball falling toward and bouncing off a table, or the hammer in a piano hitting a string (Mitterer & Jesse, 2010). In such ecologically valid situations, observers develop a clear expectation of when the sound should occur and the further interpretation that the visible collision actually caused the sound. Here, we presented the results of an experiment designed to determine how variations in event predictability and apparent causality influence the perception of synchrony and further developed the theoretical explanations of why certain stimuli produce temporal shifts of the synchrony boundaries and of the PSS.

In previous research, we used a simulation of Newton's cradle toy to approximate a series of events in which a ball suspended from a string swings in a pendulum movement to strike a series of stationary suspended balls, resulting in the ball on the other side arcing up in response to the strike (van Eijk et al., 2010). The simulation provided a robust perception of action and reaction, with the time of collision and the collision sound both being rather naturally predictable from the visible events. By blanking out the left or right side of the simulated toy, we were able to remove either the predictive and causative action of the falling ball or the resulting visual launching of the rising ball, respectively. By using all three versions of the toy, we found that the judgments of synchrony extended over different ranges of the audio–visual SOA continuum, with the results being largely similar for the entire and left-half versions of the toy, but different from the right-half version. These differences could have resulted either from the differences in predictive information (missing in the right-half case) or by the reversal of causality (as the sound appeared to

be more likely to cause the motion in the right-half case). In all cases, the PSS was closer to physical synchrony for the “right-half” conditions, due almost entirely to a shift of the auditory-first boundary toward more negative SOAs. This result is consistent with the idea that a reduction in visual predictability or visual-event causality resulted in more tolerance for auditory-leading stimulus pairs to be judged as synchronous.

The present experiment was designed to separate these two theoretical explanations for changes in the synchrony boundaries by using variations of a bouncing-ball stimulus. The ball appeared to fall toward and bounce upward from a bar that was either visible (ensuring predictive accuracy for the collision sound) or invisible (eliminating predictability). In both cases, the visible or invisible collision between the ball and some object was the event that apparently caused the sound. In addition, conditions in which the ball stuck to the bar upon contact or initiated its movement from a position sitting on the bar were used to create conditions analogous to the three views of the cradle toy simulation used previously. The results showed the same pattern of effects across the three conditions. In both the Newton’s cradle simulation and the present bouncing-ball conditions, the auditory-leading part of the synchrony judgment curve was *more negative* in the condition without visual predictive information (from the right half of the cradle, or in the lift-off condition in the present study).

In addition, no effect of bar visibility on auditory-first synchrony boundaries was found. That is, the differences in audio-visual synchrony perception between the full-motion and sudden-stop conditions, on the one hand, and the lift-off condition, on the other, cannot be explained by the availability of visual predictive information. However, one specific caveat to this interpretation needs to be discussed. Given that the balls always bounced back from the invisible bar, the uncertainty about the interaction decreased stepwise the longer that the ball had been falling. After having passed the second lowest position of the bar, there was no uncertainty left about the upcoming impact, in contrast to all other impact positions. As we described in the “Results” section, a detailed inspection did not show any reliable changes in the synchrony boundaries across bar height conditions. In particular, for the lowest bar height position in the invisible-bar condition, in which the uncertainty about the impact was much reduced, the large differences in the position of the auditory-first synchrony boundary between the three animation conditions remained the same. Thus, the most parsimonious interpretation of this outcome for the auditory-first boundary position was that participants did not perform differently across the five bar heights, and that visual predictive information played no role as a significant contributor to the observed shifts in the auditory-first boundary estimates.

By extending this interpretation to the results from experiments using Newton’s cradle toy (van Eijk et al., 2010), we concluded that the differences observed between the right-half condition in the cradle experiments, on the one hand, and the

entire and left-half conditions, on the other, can also not be attributed to differences in visual predictive information. This interpretation suggests that the participants’ causal attribution of the scenes, leading to changing cause–effect relations between the visual and auditory stimulus components, has been the major cause of the observed shifts in the auditory-first synchrony boundary.

Unexpectedly, removing visual predictive information from full-motion and sudden-stop conditions did affect the video-first synchrony boundaries for those conditions: The range over which positive SOAs were judged to be “synchronous” was extended for the invisible-bar conditions (in full-motion and sudden-stop conditions only). This is surprising, especially for the sudden-stop condition, since the end of the motion trajectory offers clear sensory information that the visual event has occurred. Possibly, the removal of the bar results in an additional internal delay for registering the stop, or reversal, of the motion, resulting in a larger tolerance for positive SOAs. This hypothesis is supported further by the finding that the video-first synchrony boundary for the lift-off condition was unaffected by bar visibility, and in the invisible-bar condition the boundary was closer to the point of objective synchrony than it was in the full-motion [ $t(17) = 3.11, p = .006$ ] and sudden-stop [ $t(17) = 3.19, p = .005$ ] conditions. Apparently, the motion onset in lift-off conditions is more quickly detected than motion stop in the sudden-stop condition and motion reversal in the full-motion condition. This result is reminiscent of some findings in the flash-lag literature in which a moving object is perceived in a position ahead of that of a stationary flash, even though they are presented physically aligned (e.g., Whitney, 2002). Shen, Zhou, Gao, Liang, and Shui (2007) provided arguments that this is a temporal illusion, in which the position of an object is integrated over location signals occurring within a “time window,” resulting in a discrepancy between the actual and perceived positions of a moving object. Such integrations presumably can produce extensions of the actual position of the moving object over time.

In summary, the apparent causality built into multimodal events can influence our perceptions of synchronous sights and sounds, such that if one event appears to cause the other, we are more likely to judge the time of occurrence of the causative agent to be before that of the resulting action, even in cases in which all temporal parameters of the audio-visual stimulus are held constant.

**Author note** Parts of the results reported here were presented as conference contributions (van Eijk, Kohlrausch, Juola, & van de Par, 2007a, 2007b). Thanks are due Martin Boschman, Constant Hak, Dik Hermes, and Christophe Stoeltinga for technical assistance, as well as three anonymous reviewers for feedback on an earlier version of the manuscript. We also acknowledge the support of a grant from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO; Dutch Organization for Scientific Research) to the third author. The authors thank Charles Spence, who, over a beer at the Eindhoven town market square on the evening preceding the second author’s PhD defense, motivated us to design and

perform the main experiment from this study by suggesting a stimulus design with an unpredictable moment of impact.

## References

- Enoki, K., Washikita, K., & Yamada, M. (2006). *Detection threshold of asynchrony between auditory and visual stimuli for various motion patterns of a ball*. Paper presented at the 9th Western Pacific Acoustic Conference (WESTPAC IX), Seoul, Korea.
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*, 773–778.
- Kim, R., Peters, M. R., & Shams, L. (2012).  $0 + 1 > 1$ : How adding noninformative sound improves performance on a visual task. *Psychological Science*, *23*, 6–12. doi:10.1177/0956797611420662
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477.
- Mitterer, H., & Jesse, A. (2010). Correlation versus causation in multisensory perception. *Psychonomic Bulletin & Review*, *17*, 329–334. doi:10.3758/PBR.17.3.329
- Shen, M., Zhou, J., Gao, T., Liang, J., & Shui, R. (2007). The perceived position of a moving object is not the result of position integration. *Vision Research*, *47*, 3088–3095.
- Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., & Porter, N. R. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society B*, *268*, 31–38.
- van Eijk, R. L. J. (2008). *Audio-visual synchrony perception* (PhD thesis). Eindhoven University of Technology, Eindhoven, The Netherlands.
- van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & van de Par, S. (2007a). *Causal relationships affect perceived audio-visual synchrony: Reversed effect for different judgment methods*. Paper presented at the 8th Annual Meeting of the International Multisensory Research Forum, Sydney, Australia.
- van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & van de Par, S. (2007b). *Causal relationships affect audio-visual asynchrony detection: Opposite trends for different stimuli*. Paper presented at the 8th Annual Meeting of the International Multisensory Research Forum, Sydney, Australia.
- van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & Psychophysics*, *70*, 955–968. doi:10.3758/PP.70.6.955
- van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & van de Par, S. (2009). Temporal interval discrimination thresholds depend on perceived synchrony for audio-visual stimulus pairs. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1254–1263. doi:10.1037/a0014254
- van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & van de Par, S. (2010). Temporal order judgment criteria are affected by synchrony judgment sensitivity. *Attention, Perception, & Psychophysics*, *72*, 2227–2235. doi:10.3758/APP.72.8.2227
- Whitney, D. (2002). The influence of visual motion on perceived position. *Trends in Cognitive Sciences*, *6*, 211–216.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*, 1293–1313. doi:10.3758/BF03194544
- Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics*, *67*, 531–544.