Atten Percept Psychophys (2013) 75:1583–1599
DOI 10.3758/s13414-013-0511-4

Quantifying temporal ventriloquism in audiovisual synchrony perception

Irene A. Kuling · Armin Kohlrausch · James F. Juola

Published online: 19 July 2013

Abstract The integration of visual and auditory inputs in the human brain works properly only if the components are perceived in close temporal proximity. In the present study, we quantified cross-modal interactions in the human brain for audiovisual stimuli with temporal asynchronies, using a paradigm from rhythm perception. In this method, participants had to align the temporal position of a target in a rhythmic sequence of four markers. In the first experiment, target and markers consisted of a visual flash or an auditory noise burst, and all four combinations of target and marker modalities were tested. In the same-modality conditions, no temporal biases and a high precision of the adjusted temporal position of the target were observed. In the different-modality conditions, we found a systematic temporal bias of 25–30 ms. In the second part of the first and in a second experiment, we tested conditions in which *audiovisual* markers with different stimulus onset asynchronies (SOAs) between the two components and a visual target were used to quantify temporal ventriloquism. The adjusted target positions varied by up to about 50 ms and depended in a systematic way on the SOA and its proximity to the point of subjective synchrony. These data allowed testing different quantitative models. The most satisfying model, based on work by Maij, Brenner, and Smeets (Journal of Neurophysiology 102, 490–495, 2009), linked temporal ventriloquism and the percept of synchrony and was capable of adequately describing the results from the present study, as well as those of some earlier experiments.

Keywords Audiovisual synchrony · Temporal ventriloquism · Temporal alignment

I. A. Kuling · A. Kohlrausch (✉) · J. F. Juola
Human-Technology Interaction Group, School of Innovation Sciences, Eindhoven University of Technology, IPO 1.25, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: a.kohlrausch@tue.nl

I. A. Kuling
MOVE Research Institute Amsterdam, Faculty of Human Movement Sciences, VU University, Amsterdam, The Netherlands

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

J. F. Juola
Department of Psychology, University of Kansas, Lawrence, KS 66045, USA

Perceiving temporal discrepancies between the various sensory components of specific events rarely happens in daily life. However, everyone knows that in technology-driven events—that is, speech on television or communication by webcams—a temporal conflict can occur when the sound is not exactly aligned with the accompanying visual information. Furthermore, the asynchrony is noticed only when one of the sensory inputs is delayed by an amount that is larger than the temporal delay that the human brain can integrate. This window of temporal integration depends on many factors—that is, context and stimulus identity (Van Eijk, Kohlrausch, Juola, & Van de Par, 2008, 2010). The fact that different propagation speeds for light and sound for distal stimuli are combined with different neural conduction times for proximal stimuli means that the brain must have some tolerance for audiovisual (AV) temporal discrepancies (Vroomen & Keetels, 2010).

Multiple studies focusing on temporal conflicts between visual and auditory information have concluded that in the temporal domain, audition is the more dominant sensory modality (Aschersleben & Bertelson, 2003; Bertelson & Aschersleben, 2003; Morein-Zamir, Soto-Faraco, & Kingstone, 2003). Shams, Kamitani, and Shimojo (2000, 2002) studied the illusory flash effect, in which multiple auditory beeps can change the perception of a single flash into two or more flashes. For *periodically modulated* auditory and visual stimuli, it has been long known that, in the case of temporal-rate disparities, the perceived rate of visual

modulation is strongly influenced by the rate of auditory modulation, while the reverse effect is nearly absent (for an early demonstration, see Gebhard & Mowbray, 1959). More recent research on this topic has shown that the amount of cross-modal interaction in perceived rate can be controlled by adjusting the within-modality rate discrimination sensitivity—for instance, by changing the depth of modulation (Roach, Heron, & McGraw, 2006). Also, the perceived time of occurrence of a visual signal can be influenced by asynchronous sounds (Bertelson & Aschersleben, 2003; Morein-Zamir et al., 2003). In these two papers, this dominance of the auditory modality in perceiving temporal positions was called *temporal ventriloquism*.

Morein-Zamir et al. (2003) designed an experiment in which participants had to decide which of two lights was turned on first in a visual temporal order judgment (TOJ) task. Data were analyzed in terms of percentage of "onset of light 1 first" responses as a function of the delay between the two onsets, and just noticeable differences (JNDs) were derived from the slope of the resulting psychometric function. By accompanying the onsets of the lights with irrelevant sounds, the authors influenced the JND between the onsets. In fact, by presenting the two sounds such that one appeared before the first light onset and the other after the second onset, the psychometric function became steeper, indicating an increase in sensitivity. The interpretation of this observation was that the *perceived* temporal positions of the onsets of the lights were attracted by the presence of the sounds. This means that the onsets were pulled away from each other when the sounds were presented before the first light onset and after the second onset. Similarly, the perceived onsets were pulled toward each other when the sounds were presented in between the onsets. Morein-Zamir et al. reported different shift sizes for different time delays between light onsets and sounds and found significant decreases in JND for delays of 75–225 ms. To test whether this effect was due to the influence of the first or the second sound, another experiment was performed in which one of the sounds was presented simultaneously with one of the light onsets and the other sound preceded the onset of the first or trailed the onset of the second light by 100, 200, 450, or 600 ms. The authors observed a significant decrease in JND only for the conditions in which the first sound occurred simultaneously with the onset of the first light and the second sound trailed the onset of the second light by 100 or 200 ms.

Perceived timing of a stimulus is not easy to measure, due to the absence of an obvious absolute time reference, in contrast to the effects of spatial ventriloquism, where the three-dimensional space around the participant forms a reference frame. Therefore, the demonstration of temporal ventriloquism is often based on well-known visual effects, like four-dot masking (Vroomen & Keetels, 2009), the flash-lag effect (Vroomen & de Gelder, 2004), and the visual Ternus effect (Shi, Chen, & Müller, 2010). All these stimulus conditions share the following perceptual properties: In a visual-only presentation, two different visual percepts are possible depending on the value of a temporal parameter in the stimulus. Typically, if the time interval is short, one percept is seen, and when the time interval is longer, the other percept is seen. When the probability of each percept is determined as a function of the temporal parameter, a psychometric function is obtained, allowing derivation of the parameter value leading to an equal probability (50% probability of the psychometric function) of both percepts. When sounds are added in spatial and temporal proximity to the visual stimuli, the time interval at which this 50% probability is obtained can be influenced, showing that the perceived temporal distance between the visual stimuli is affected by the sounds. However, there are different possible causes for such an observation. In order to evaluate whether this effect of sound was merely attentional or reflected a true influence on the perceived timing, in various of the cited studies, the influence of only one (preceding) sound was compared with that of two sounds (preceding and trailing). If the effects were completely due to attention, both conditions should show the same result. Although, in most studies, a single preceding sound had an influence on the visual timing, the effects were stronger in the condition with a preceding and a trailing sound (e.g., Morein-Zamir et al., 2003; Shi et al., 2010; Vroomen & Keetels, 2009). In these studies, the authors concluded that attention plays a role in their experiments, but in addition, a temporal ventriloquism effect seemed to influence the observers' percepts.

All the experiments mentioned above demonstrate the same basic effect: Temporal relations between two visual percepts can change when they are accompanied by sounds. Sounds can influence the steepness of the psychometric function for visual temporal discriminations, which has led some authors to conclude that the sounds pull the visual stimuli along the temporal domain, consistent with the temporal ventriloquism hypothesis. The fact that the perceived temporal occurrence of visual information is influenced by accompanying auditory information has been found in many different settings, but a quantitative description of this effect as a function of the temporal distance between auditory and visual stimulus components has not yet been attempted (Vroomen & Keetels, 2010).

In recent years, two studies have been published that addressed temporal ventriloquism in AV stimuli in a different experimental setting that enabled quantification of this effect (Burr, Banks, & Morrone, 2009; Hartcher-O'Brien & Alais, 2011). Whereas previous studies had measured the shift in perceived moment of the visual component across a wide range of cross-modal stimulus onset asynchronies (SOAs) up to 200 ms, the two recent studies measured the perceived

moment of occurrence of AV stimuli with relatively small SOAs. Although this distinction has not been made explicit in the recent studies, we consider it to be of relevance for the discussion of experimental effects and underlying mechanisms.

In the first experiment published by Burr et al. (2009), using a short high-frequency auditory stimulus, participants performed a temporal bisection task between two AV marker stimuli that had a temporal separation of 800 ms. A third AV signal (the target) was presented within a range of specific delays near the middle between the two markers, and participants had to indicate whether the central target appeared closer to the first or the second marker. In the target stimulus, the audio and visual components were always in physical synchrony. In the two marker stimuli, an SOA between the flash and the audio signal of ±120 ms in steps of 20 ms was introduced, while keeping the mean temporal position constant. The dependent variable was the temporal position of the central component at the point of subjective equality (PSE), defined as that temporal position that bisected the interval between the two markers.

The first observation was that in the baseline condition, when all three stimuli were presented in cross-modal synchrony, the PSE appeared about 60 ms earlier than the physical midpoint between the markers. This means that the second interval had to be 120 ms longer than the first one to lead to the percept of two intervals with equal durations. The authors argued that "this was a common feature in all the data . . . and consistent with much other data in the literature, showing that the first interval in a sequence is perceived as longer than the others (Rose and Summers 1995; Tse et al., 2004)" (Burr et al., 2009, p. 52). However, Tse, Intriligator, Rivest, and Cavanagh (2004) concluded from their data that, in the comparison of two successive intervals of equal duration, the second interval is judged as longer, when the durations are greater than 150 ms. Thus, for the interval durations of 400 ms each, as used by Burr et al., the two studies showed opposite results.

When Burr et al. (2009) introduced discrepancies between the auditory signal and the visual flash in the markers, the position of the target was displaced relative to the baseline value. The resulting position of the PSE was dominated by the position of the auditory component of the markers and was influenced only to a minor extent by the position of the visual component. The relation between SOA and adjusted position could be well fitted by a linear function, indicating that the relative weights of the auditory and the visual components on the perceived timing of the AV markers were constant over the SOA range from +120 to −120 ms. The authors summarized their observations by stating that in AV stimuli, the temporal positions of the auditory components tend to dominate the perceived timing of the AV stimuli, but the domination was not total.

Hartcher-O'Brien and Alais (2011) performed an experiment similar to that of Burr et al. (2009) using a two-alternative forced choice procedure. Two observation intervals were defined by auditory stimuli of 1,250-ms duration. In each interval, a short auditory, visual, or AV stimulus was added. In one interval, this added stimulus occurred exactly in the temporal center of the frame; in the other interval, it had a temporal offset. Participants had to indicate in which of the two observation intervals the added short component occurred later. From these raw data, psychometric functions were constructed and fitted with cumulative Gaussian functions, allowing derivation of the mean (estimating the shift with respect to the reference stimulus position) and standard deviation (estimation of sensitivity). When using AV stimuli, the reference interval contained synchronized AV stimuli at the center position of the interval, while the test interval contained AV stimuli with an SOA of ±80 ms, centered on the middle of the interval (thus, each stimulus component was at most 40 ms displaced from the center of the interval). For maximal leading or trailing positions of the audio component, the perceived moment of the AV stimulus was shifted by an absolute amount of 15–25 ms away from the center position, indicating a much greater weight for the auditory than for the visual component in determining the perceived moment of the AV stimulus. Overall, these data agree with the observation from Burr et al. about the weighted contribution of the two unimodal components on the perceived times of AV stimuli.

Due to the inconsistent conclusions about the position of the PSE in interval bisection procedures (see above), and as a first attempt to quantify the temporal shift in perceiving onsets of visual stimuli, we introduce a different method in the present study to quantify the temporal integration of auditory and visual information. The method has been developed in speech perception and production research to establish the perceptual center hypothesis (Marcus, 1976) and is based on adjusting perceived temporal positions of target stimuli in a sequence of isochronous marker sounds. A variant of this method was used by Schimmel and Kohlrausch (2006, 2008) to measure temporal positioning biases for noise burst stimuli of different durations. In this variant, a target pulse has to be adjusted to a regular temporal pattern of a total of four marker pulses, two leading and two trailing the target. Both of these methods are known to allow very accurate and, in contrast to the method used by Burr et al. (2009), bias-free adjustments of the target positions, if identical signals are used as markers and target(s) (Marcus, 1976; Schimmel & Kohlrausch, 2006, 2008).

In the experiments described here, the stimuli consisted of an isochronous sequence of five audio, visual, or AV pulses. On the first, second, fourth, and fifth positions of the sequence, a marker pulse was placed. The target pulse was placed around the third position (randomly between the

second and fourth positions). The participant's task was to temporally align this adjustable target to the missing third position (see top row of Fig. 1). The temporal displacement between the (physically) isochronous position of the target pulse and the adjusted (i.e., perceptually isochronous) position gives an absolute measure for the shift in perceptual occurrence, relative to those of the marker stimuli.

We used this rhythmic judgment method for different combinations of modalities of marker and target pulses. Two within-modality conditions (audio–audio, video–video) served as baselines to verify the absence of a temporal bias, and they allowed us to establish the precision with which temporal positions can be established. The between-modality conditions should indicate whether our results are consistent with known differences in peripheral nervous system transduction times between the auditory and the visual modalities, in much the same way as they presumably contribute to differences in reaction times (e.g., Riggs, 1971). Given that the rhythmic judgment method proved to be sensitive enough to produce significant and symmetric displacement effects in the between-modality conditions, we used the same method in further studies in which we combined a visual target among AV markers. The SOAs of the AV markers were systematically varied, but the participants were asked to adjust the target to the visual components of the marker stimuli while trying to ignore the auditory components. Auditory dominance in temporal AV integration should be revealed by a systematic shift in the alignment of the visual target. These expectations were tested in the second part of Experiment 1 and were extended in Experiment 2.

Experiment 1

The first part of Experiment 1 was designed to establish a baseline for the perception of rhythmic sequences of flashes and noise bursts. By using unimodal markers and targets in the same or different modalities, the precision of the procedure and potential temporal offsets (reflecting at least the relative processing times necessary to perceive auditory vs. visual stimuli) should be found. We expected that the precision of the aligned target position would increase with the number of auditory stimuli (marker or target), because it is reportedly more difficult to recognize visual stimulus rhythms than auditory ones (Glenberg & Jona, 1991; Grahn, 2012; Guttman, Gilroy, & Blake, 2005; Repp & Penel, 2002). Since visual rhythmic patterns activate the auditory cortex (Guttman et al., 2005), the investigation of rhythmic perception of mixed auditory and visual stimuli appears worthwhile. For the initial four unimodal conditions, we expected the following order from low to high precision: visual markers–visual target (V–V), visual markers–auditory target (V–A), auditory markers–visual target (A–V), and auditory markers–auditory target (A–A). No temporal offsets are expected for the same-modality conditions, V–V and A–A, because in these conditions, the processing times for markers

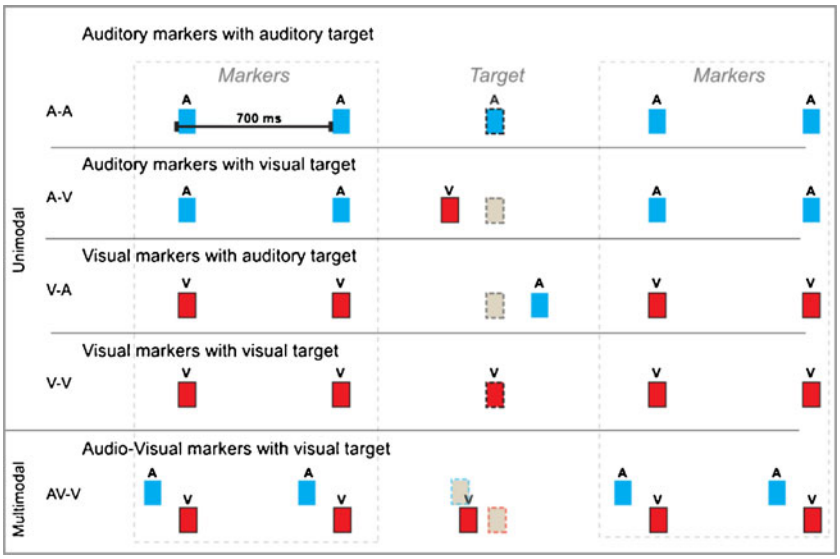

Fig. 1 Design of Experiment 1. The five rows schematically indicate the temporal course of stimulus presentation. In each condition shown in the first four rows, four identical markers from the same modality (either A or V) are presented at positions 1, 2, 4, and 5. At position 3, the target stimulus from the same (rows 1 and 4) or the other (rows 2 and 3) modality, as compared with the markers, is presented. The light gray bars with dashed borders indicate the positions of the physical midpoint between the marker components. The plotted position of the target stimulus is the hypothesized end position. In the bottom row, audiovisual markers with a specific onset asynchrony are combined with a visual target, and physical midpoints for both marker components are indicated. The target had to be aligned to the rhythm of the visual marker components, and the indicated end position of the target reflects a temporal shift of the visual marker components caused by the auditory marker components

and targets should be equal. For the A–V and V–A conditions, an offset with the same magnitude, but in opposite directions, is expected, reflecting a contribution from differences in processing times between auditory and visual stimuli. Because peripheral transduction of auditory stimuli requires less time than that of visual stimuli (Arrighi, Alais, & Burr, 2006; King, 2005), the following expectations can be formulated: The auditory target between visual markers is expected to be placed later in time than at the bisection of the temporal positions of the second and fourth markers; in the same vein, the visual target between auditory markers is expected to be placed earlier in time than at the bisection of the temporal positions of the second and fourth markers. These two shifts are expected to be equal, in opposite directions, because the underlying differences in processing times are the same in the V–A and the A–V conditions (cf. Fig. 1).

The second part of Experiment 1 used multimodal markers and visual targets. The visual target had to be aligned with the visual components of the markers. By structurally changing the SOAs of the visual and auditory parts of the markers, we expected to determine the parameters of auditory influences in temporal ventriloquism.

On the basis of the assumption that temporal ventriloquism is due to auditory capture and auditory capture occurs only when the components of a multimodal stimulus are perceived as being in synchrony, predictions can be made for the second part of Experiment 1. When the marker SOAs are within the AV synchrony window, we expect that the perceived timing of the visual marker component will be shifted toward the auditory component and that this shift will be reflected in the adjusted time of the visual target. This shift should be stronger for large SOAs and should be zero for a marker SOA corresponding to the Point of Subjective Simultaneity (PSS). For marker SOAs outside the synchrony window, however, the capture of the visual marker component should be reduced or absent, and the visual target should be aligned with the visual component of the markers without significant bias.

Method

The perceptual rhythm of the stimulus was defined by five stimuli with an interonset interval of 700 ms. The first, second, fourth, and fifth stimuli were the rhythmic markers. The target stimulus was initially set at a random position between the second and fourth markers, and the participant's task was to align the target such that the whole marker–target sequence had an isochronous rhythm (Fig. 1). The stimuli used for both markers and targets were simple flashes, noise bursts, or their combinations. These basic stimuli have been used previously in cross-modal synchrony experiments, since there are no context cues that could help participants to predict or anticipate the stimuli (Sternberg & Knoll, 1973; Van Eijk et al., 2008).

Participants

Twelve participants took part (4 females). Three of the participants were experienced in this research area (the authors) and voluntarily joined the experiment. The other participants were naïve about the experiment and received a payment of 30 Euros. All participants reported (corrected-to-) normal vision and normal hearing. The participants varied in age from 20 to 69 years, with a mean of 32 years. The experiments conformed to the requirements of the World Medical Association as laid down in the Declaration from Helsinki 1964.

Stimulus

The visual part of the rhythmic stimulus consisted of a white disk (97 cd/m^2 as measured using an LMT L1003 luminance meter) shown for one frame (11.8 ms) at a central position on the screen. The disk had a diameter of 49 pixels and subtended a diameter of about 1.4° at an unconstrained viewing distance of about 60 cm. During the experiment, four corners of a surrounding square were visible, in order to indicate the central location of the flashes. The temporal occurrence of the markers was set with an interonset interval of 700 ms. In the conditions with AV markers, the temporal position of the flash was kept fixed to the 700-ms rhythm, and the temporal positions of the noise bursts were changed to create specific marker SOAs. The acoustic part of the stimulus consisted of an 11.8-ms white-noise burst with a sound pressure level of 67 dB. The temporal calibration between auditory and visual stimulus onsets followed the approach in van Eijk et al. (2008) and was accurate to within ±2 ms.

Apparatus

The visual stimuli were shown on a Dell D1025HE CRT monitor at a resolution of 1,024 × 768 pixels and at an 85-Hz refresh rate. The auditory stimuli were 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-attenuated room (as in Van Eijk et al., 2008).

Design

Every participant was assigned two different tasks distributed over three sessions. In the first and third sessions, the participants executed a temporal adjustment task with AV markers and a visual target. Within each trial block in a session, the relative SOA between the flashes and noise

bursts of the markers was varied. There were five values (−150, −50, 0, +50, and +300 ms), with negative values indicating a leading audio component and positive values indicating a trailing audio component. Physical synchrony was defined as an SOA of 0 ms. These delays were chosen because the middle three are expected to be within the synchrony window, while the first and last conditions are expected to be outside the synchrony window, as measured in previous experiments using similar stimuli (Van Eijk et al., 2008). In the second session, the participants performed an adjustment task with unimodal markers and targets. There were four different combinations of markers and targets (A–A, A–V, V–A, and V–V). These conditions were blocked and ordered in a balanced Latin-square design. We chose this order of the two tasks across the three sessions in order to minimize potential order effects.

Procedure

The participants received written instructions about the response options and the use of the keyboard. On the instruction sheet, schematic representations of sequences of markers with an embedded target were given with an example of a temporally misaligned target and an example of a well-aligned target, similar to Fig. 1. The task was described as realizing a temporal sequence in which the target stimulus was well aligned to the rhythm of the marker stimuli. Furthermore, they were informed how they could shift the target position in small and large steps in either temporal direction and how they could terminate an adjustment process (for details, see below). In the first and third sessions, the target (flash) had to be aligned with the visual component of the AV markers (flashes). The auditory components were distractors, and participants were told to use only the visual components in adjusting the target position. In the second session, the target (flash or noise) had to be aligned with the markers (flashes or noise), without any distractors present. At the start of a trial, the target was placed at a random position between the second and fourth marker positions. After perceiving the marker–target sequence with this start position, the participants had to adjust the temporal position of the target toward optimal alignment by pressing keys on the keyboard. They had five response possibilities: "1," which resulted in a large step earlier; "2," a small step earlier; "3," repeat the current adjustment; "4," a small step later; and "5," a large step later. Large and small steps represented 59.0 ms (five frames) and 11.8 ms (one frame), respectively. After each response, the stimulus sequence was played again, and the participant could judge the rhythmic sequence. Participants were allowed to make as many adjustments as necessary to be satisfied with the rhythmic sequence. When they were satisfied with the result, "f" was to be pressed on the keyboard to finish the trial and start the next one.

A trial began when the participant pressed the start button (which was the same as the finish button of the previous trial). In the unimodal conditions, the first marker was presented 700 ms after the start button was pressed. In the AV–V conditions, the first visual component was presented 700 ms after the start button was pressed, but because in these conditions the sound could precede or trail the first visual component, the actual trials started between 550 ms (most advanced sound) and 700 ms (trailing sound) after the start button was pressed. In all sessions, participants could take a small break after each trial, as they needed it.

In the first session, participants started with the AV markers and visual targets. There were five audio delays, and each delay was presented 12 times, which resulted in 60 trials. The trials were presented in random order. This session lasted about 40–55 min. In the second session, the unimodal conditions were measured. The four conditions (2 × 2 combinations of marker and target modalities) were presented in blocks of 5 identical trials. Each block was presented 4 times in a semirandom order (four different blocks were presented in random order, then again four different blocks in random order, etc.), yielding a total of 80 trials. The second session lasted around 50–60 min. The third session was comparable to the first session, but now each condition was presented 15 times, resulting in 75 trials (45–60 min). This slight increase in trial numbers was chosen to maximize the amount of data given the available participant time, because we realized in session 1 that participants performed the task somewhat faster than we had concluded from pilot experiments. The three sessions were run on 3 different days within a 2-week period. After the third session, the participants were asked some general questions about their age and whether they played a musical instrument, and they were asked to describe their strategies in the AV sessions.

Results

First, the raw data of the unimodal conditions from the second session were analyzed. For each participant, the mean adjusted values and standard deviations for the different conditions were calculated (see Table 1). The results of participant 5 were excluded from further analyses for two reasons: The standard deviations of this participant were about twice as large as the mean of the standard deviations of all other participants, and they were large even in the simplest unimodal conditions (A–A and V–V); and in the final debriefing, this participant expressed great difficulties in performing the task. The exclusion reduced the number of participants to 11, with an age range from 20 to 64 years and a mean of 28 years.

The mean adjusted values showed that there was hardly any shift away from physical isochrony for the conditions with markers and target in the same modality. Also, the

Table 1 Results of the unimodal conditions of Experiment 1: Mean adjusted values and standard deviations for each condition of each participant

Condition	A–A		A–V		V–A		V–V	
Participant	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
P1	−4.7	36.4	−79.1	102.2	21.0	62.7	24.2	79.4
P2	11.8	70.7	−39.5	105.0	3.5	106.4	−7.1	61.2
P3	2.4	25.0	−36.6	54.4	15.3	68.8	−3.0	32.9
P4	5.3	23.7	1.2	56.4	29.5	33.0	−11.2	26.4
P5	−40.1	119.6	7.1	159.7	24.9	134.7	33.0	136.0
P6	−11.8	39.0	−73.2	94.8	101.5	57.1	18.3	88.1
P7	−9.4	22.2	−54.9	85.1	38.9	57.6	−4.7	44.4
P8	10.0	46.8	−14.8	63.0	11.8	92.2	20.1	62.8
P9	−4.1	16.4	2.4	30.5	44.6	41.6	−5.3	25.8
P10	−10.0	30.5	−39.5	85.3	−14.4	71.0	−18.3	42.2
P11	11.8	19.1	5.3	105.0	78.7	68.6	−4.7	29.0
P12	−0.6	11.8	−15.9	56.7	−40.1	56.2	−33.6	32.8
Mean	0.1		−31.3		26.4		−2.3	
SE	2.6		8.6		11.5		5.0	

Note. The first letter of the condition indicates the modality of the markers, the second the modality of the target. The values are in milliseconds. Participant 5 was excluded from further analysis (see the text for details), and, therefore, these data were not included in calculating the overall means and standard errors.

precision for these conditions was quite high (as indicated by the small standard errors). The mean adjusted values of the A–V and V–A conditions were shifted in opposite directions by about the same amount. When the markers were auditory and the target was visual, the target was placed 31 ms earlier than the physical midpoint. In the reverse condition, with visual markers and an audio target, the target was placed 26 ms later than the physical midpoint.

The mean adjusted values for all participants were analyzed in a repeated measures, 2 (marker modality) × 2 (target modality) ANOVA. The analysis found a significant main effect for both marker, $F(1, 10) = 6.7$, $p < .05$, partial $\eta^2 = .40$, and target, $F(1, 10) = 19.9$, $p < .01$, partial $\eta^2 = .67$, modality. The interaction effect was not significant, $F(1, 10) < 1$. The results show that the adjusted values are more negative (earlier on the time axis) for auditory markers than for visual markers (−15.91 ± 5.27 ms vs. 11.28 ± 7.78 ms, $p < .05$) and also that the adjusted values are more positive (later on the time axis) for auditory targets than for visual targets (11.77 ± 6.07 ms vs. −16.40 ± 4.08 ms, $p < .01$). Follow-up *t*-tests showed that neither adjusted value differed from 0 for the conditions with identical marker and target modalities [A–A, $t(10) = -0.148$, and V–V, $t(10) = -0.268$], but in the different-modality conditions, the adjusted values differed significantly from 0 [condition A-V, $t(10) = -3.53$, $p < .01$, effect size .74; condition V–A, $t(10) = 1.96$, $p < .05$, effect size .53, one-tailed].

We interpret these shifts in adjusted position as an indication of the differences in processing speeds for auditory and visual stimuli. It takes less time to process a sound than a flash, and therefore the conditions with the combined modalities have nearly equal shifts of about 25–30 ms, but in opposite directions (Fig. 2).

To analyze the data of the second part of the experiment—that is, the data from the condition using AV markers and visual targets—the mean adjusted values and standard deviations were calculated for each participant for both sessions. Initial tests (repeated measures ANOVA, 2 [session] × 5 [SOA]) showed no difference in the adjusted values between the two sessions, $F(1, 10) < 1$, and, therefore, the data were

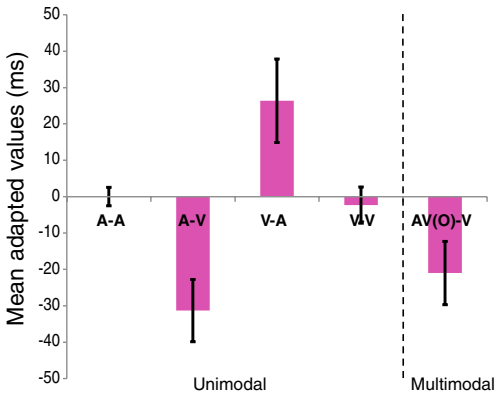

Fig. 2 The bars show the mean adjusted values of unimodal conditions and of the audiovisual condition with an SOA of 0 ms. The first letter indicates the marker modality, the second the target modality. The mean values are the positions relative to the physical midpoint between the second and fourth markers. The values are in milliseconds. Error bars represent *SE*s

combined. Again, the data of participant 5 were excluded from further analysis for the reasons given above. Between subjects, we found a negative correlation between the number of adjustment steps (mean = 7.5) and the standard deviation of the resulting adjustments (R^2 = .54) (Table 2).

Of initial interest is the adjusted value of the AV–V condition with no audio delay in the marker stimulus (SOA = 0 ms). If the participants had been able to ignore the sound and align the flash solely to the visual components of the markers (flashes), the adjusted value should be at zero, as in the V–V condition in the unimodal experiment. On the other hand, if participants were not able to align the flash with the flashes but aligned it with the auditory components, we would have seen the same shift as in the A–V condition in the unimodal experiment, presumably due to the differences in internal processing speed. Initial tests show that the adjusted values for an SOA of 0 ms are in between the results of the V–V and A–V conditions in the unimodal experiment but not significantly different from either (Fig. 2). This result suggests that the adjusted value is influenced by the temporal position of both components of the markers. Such a combined influence could be realized by forming a weighted mean across the temporal positions of the auditory and visual marker components, to which the target is then adjusted. For the observed shift, the weight of the auditory component would then be about twice as large as the weight for the visual component. Alternatively, one could see this shift as a reflection of temporal ventriloquism. Due to the presence of the auditory marker component, the perceived moment of the visual component was shifted by about 20 ms, and the target was adjusted to this shifted position of the visual marker components.

The next step in the analysis explored the effects of changes in the AV SOAs. For the overall effect of audio delay, the mean adjusted values of the participants were analyzed in a one-way (audio delay) repeated measures ANOVA. The analysis found a significant main effect for audio delay, $F(4, 40) = 15.9$, $p < .001$, partial $\eta^2 = .63$. This result shows the relative influence of the various audio delays on the perceived isochronous rhythm of the flashes. In a first analysis, we performed a linear regression including all SOAs. In that situation, the audio delay significantly predicted the mean adjusted value of the flash, $b = .106$, $t(10) = 4.22$, $p < .05$. The audio delay also explained a significant proportion of variance in the adjusted values, $R^2 = .85$, $F(1, 10) = 17.8$, $p < .05$ (see Fig. 3, data marked by a star).

This simple analysis follows the approach taken by Burr et al. (2009), who had data for 13 equally spaced SOAs within the range ±120 ms. However, such a linear fit ignores the expected influence of the temporal synchrony window on the capture effect. If we look at the data from this point of view, we can see that the SOAs of the markers of −50, 0, and 50 ms are, without any doubt, within the synchrony window, within which people are not able to identify which of these components (auditory or visual) had an earlier onset (e.g.,

Table 2 Results of the audiovisual condition of Experiment 1. Mean adjusted position, relative to the isochronous position of the visual marker component, and standard deviations for each condition for each participant

Audio delay	−150 ms		−50 ms		0 ms		+50 ms		+300 ms	
Participant	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
P1	−31.5	63.4	−26.2	73.6	−20.1	83.6	−27.5	85.8	15.7	96.6
P2	−53.3	86.9	−62.5	76.4	−53.8	96.2	−22.7	84.3	4.4	79.2
P3	−40.2	42.3	−17.9	38.1	−5.9	37.2	2.6	38.5	−8.2	46.9
P4	−62.5	36.0	−31.9	30.5	−17.9	25.2	3.5	28.5	16.2	29.1
P5	−24.9	139.7	12.7	142.3	8.3	147.7	−16.6	164.1	42.8	125.2
P6	−70.4	79.7	−23.6	81.6	−73.9	80.4	−28.8	124.7	−30.2	100.6
P7	−49.4	50.4	−42.8	40.4	−42.8	44.7	−12.2	41.3	11.4	60.7
P8	−57.3	38.4	−17.9	66.1	6.1	56.3	8.7	53.7	34.5	87.3
P9	−15.3	27.0	−9.2	24.3	3.5	31.2	−6.6	31.6	−19.7	32.1
P10	−17.9	58.1	−41.1	54.0	−5.7	47.5	7.4	53.1	8.3	59.1
P11	−6.6	42.8	0.9	55.5	21.0	50.0	39.3	43.2	27.1	37.6
P12	−76.5	32.6	−58.6	29.0	−41.5	38.8	−17.9	37.7	−13.5	27.5
Mean	−43.7		−30.1		−21.0		−4.9		4.2	
SE	7.0		6.0		8.7		6.1		6.0	

Note. All conditions had audiovisual markers and a visual target. "Audio delay" indicates the temporal delay of the audio component of the markers, relative to the visual component (negative values mean that the audio component occurred first). The values are in milliseconds. Participant 5 was excluded on the basis of the large standard deviations, and therefore, these data were not included in calculating the overall means and standard errors.

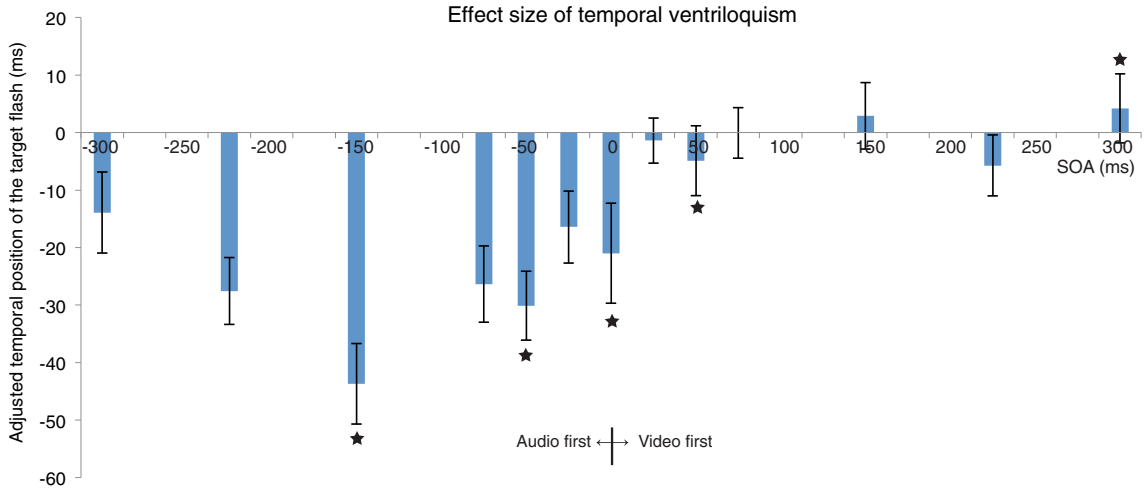

Fig. 3 Mean adjusted values of (the audiovisual conditions of) Experiment 1 (indicated with stars) and Experiment 2. Conditions represent audiovisual markers with an audio delay relative to the visual component as indicated on the *x*-axis. The mean values are the positions relative to the physical midpoint between the visual components of the second and fourth markers. The values are in milliseconds. Error bars represent *SE*s

Van Eijk et al., 2008). The SOA of −150 ms was expected to be outside the synchrony window, but this might have been a problematic choice based on a compromise between experimental time and expected effects, because synchrony boundaries for these types of stimuli could be around −120 ms (e.g., Van Eijk et al., 2008). Therefore, the SOA of −150 ms might not have been large enough to reduce the influence of the sound on the adjusted value of the flash completely. However, the influence (relative to the SOA) is already reduced, as compared with the three SOA values in the center of the synchrony window. At the SOA of 300 ms, the adjusted value was not different from zero and could be interpreted as not having been influenced by the auditory components of the markers at all. This means that the data found in Experiment 1 are compatible (although not exclusively) with our initial hypothesis.

Discussion

The first part of Experiment 1 was designed to quantify the relative time interval between physical and perceived stimulus onsets for auditory and visual stimuli. We found, as was expected, that there was no offset in adjusted position for the condition with markers and target in the same modality (A–A and V–V). Furthermore, the method showed a similar bias for both combinations of auditory and visual markers and targets. The auditory target between visual markers was placed later in time than at the bisection of the temporal positions of the second and fourth markers, and the visual target between auditory markers was placed earlier in time than at the bisection of the temporal positions of the second and fourth markers. The size of this relative delay was found to be about 25–30 ms, which is slightly larger than the delay between the auditory and visual stimuli that led to the strongest synchrony percept (10–15 ms) in previous experiments with the same stimuli (Kuling, Van Eijk, Juola, & Kohlrausch, 2012; Van Eijk et al., 2008). In our previous experiments, the methodology involved synchrony perception, in which bimodal stimuli were judged to be synchronous over a range of SOAs around the point of physical synchrony. It is possible that temporal ventriloquism operates in the synchrony judgment paradigm to pull the apparent time of the visual onset toward an earlier auditory onset, thus leading to an estimate smaller than conduction time differences when measured with single-modality targets. It is also possible that part of the observed difference in results between tasks is due to the fact that there was a task shift in our experiment, which was not present in synchrony judgment tasks. In synchrony judgments, the participant is focused on determining which modality is stimulated first. Although the synchrony judgment task uses two modalities, no modality switch had to be made during the trial. In our experiment, participants had to switch actively between the marker modality and the target modality and back. Therefore, any switch costs (e.g., Yeung & Monsell, 2003) might have been added to the relative transductive delays.

The standard deviations of the adjusted positions also lead to some interesting conclusions about relative timing sensitivity. The precision of the A–A condition was in the same range as that found in previous auditory experiments (Schimmel & Kohlrausch, 2006, 2008). The standard deviations increased when more visual components were added (A–V and V–A), but for the V–V condition, the standard deviation did not increase any further but, rather, fell between those for the A–A and the two combined-modality conditions. This means that it is possible to adjust the

temporal position of a visual target to create an accurate isochronous visual rhythm.

The data of the second part of the experiment are more difficult to interpret. In this part of the experiment, the adjustment method with isochronous rhythmic stimuli was used to quantify the influence of auditory components on the perceived moment of occurrence of visual components and to determine the range of SOAs over which auditory capture occurs. Analysis of the data shows that mainly two explanations are possible. First, we could say that the auditory influence on visual stimuli is equal to a constant proportion of the SOA of about 10% (at least within the range of SOAs from −150 to 300 ms). However, this interpretation of a constant influence of the auditory offset is not in line with general expectations about temporal ventriloquism; it could be a good description for the range of SOAs used here, but for increasingly large (negative and positive) SOAs, the effect should return to zero, since temporal ventriloquism, like spatial ventriloquism, is not expected to reveal equal cross-modal influence for all values of the relevant physical parameters (SOA or difference in source directions, respectively). Such a breakdown of cross-modal interaction in the time domain has recently been demonstrated by Roach et al. (2006) for temporal rate discrepancies.

A more likely explanation is that the auditory influence on the perceived position of the visual markers is highest when the SOAs are small and lie within the apparent synchrony range. When the SOAs become too large and are at the boundaries or outside the synchrony window, the net effect of the auditory component on the perceived temporal position of the flash should decrease and finally disappear. Within the range of perceived synchrony, the effect size of temporal ventriloquism, relative to the SOAs, might then be predicted from the synchrony judgment response curve. This means that the relative influence should be strongest at SOAs near the PSS and uniformly decrease back to zero for smaller and larger SOAs.

To test this view, we designed the second experiment, in which we extended the second part of Experiment 1 by increasing the range and the number of SOAs in the markers.

Experiment 2

The goal of the second experiment was to measure the influence of AV marker SOA on the adjusted temporal position of the visual target over a larger range and for intermediate values of SOAs. These delays were chosen on the basis of the results of Experiment 1 to give an answer to the following questions: Does the adjusted value return to zero for more negative SOAs than those tested in Experiment 1 (< −150 ms)? And is there a maximum shift for SOAs in the range between +50 and +300 ms?

The data should allow a joint evaluation together with those from Experiment 1; therefore, the same stimuli and the same design were used as before for the second experiment. Due to practical reasons, only some of the original participants were available also for Experiment 2.

Method

Stimuli and apparatus

The stimuli and apparatus used were identical to those in Experiment 1.

Design

Every participant performed two sessions of the adjustment task with AV markers and a visual target. Within each session, the relative SOA between flash and click of the markers was varied. There were eight values (−300, −225, −75, −25, +25, +75, +150, and +225 ms), with negative values indicating audio first and positive values indicating video first. Physical synchrony was defined as an SOA of 0 ms.

Participants

Eighteen participants took part in the experiment (10 females). Three of the participants were experienced in this research area (the authors) and voluntarily joined the experiment. The other participants were naïve about the experiment and received a payment of 30 Euros. Seven of the participants had also participated in Experiment 1. All participants reported (corrected-to-) normal vision and normal hearing. The participants varied in age from 18 to 64 years, with a mean of 25 years.

Procedure

The participants received written instructions about the response options and the use of the keyboard. Participants had to align the target stimulus (flash) with the rhythmic sequence of the marker stimuli (flashes). The audio components were distractors. At the start of a trial, the target was placed at a random position between the second and fourth (marker) positions. After perceiving this start position, the participant adjusted the temporal position of the target by pressing keys on the keyboard (see the "Procedure" section, Experiment 1).

In the two sessions, eight different SOAs in the markers were used, and each SOA was presented 10 times per session, which resulted in 80 trials per session. The SOAs were presented in random order. The sessions lasted around 45–60 min. In each session, participants could take a small break

after each trial, as needed. The two sessions were run on 2 different days within a 2-week period. After the second session, the participants were asked some general questions about their age and whether they played a musical instrument, and they were asked to describe their tactics in the AV sessions.

Results

The mean adjusted values and standard deviations were calculated for each participant for both sessions. Initial tests (a repeated measures ANOVA, 2 [session] × 8 [SOA]) showed no difference in the adjusted values between the two sessions, $F(1, 14) = 3.1$, $p = .10$; therefore, the data were combined (Table 3; Fig. 3, where stars indicate the data from Experiment 1). As compared with Experiment 1, we found here a smaller negative correlation between the number of adjustment steps (mean = 6.0) and the standard deviation of the adjusted values ($R^2 = .29$).

To test the overall effect of audio delay, the mean adjusted values for each participant were analyzed in a one-way (audio delay) repeated measures ANOVA. The analysis found a significant main effect for audio delay, $F(7, 119) = 5.1$, $p < .01$, partial $\eta^2 = .23$, as in Experiment 1. In the analysis, the effect of audio delay shows significant linear and cubic relations (both ps < .01, partial $\eta^2 = .38$ and .60, respectively). A closer look at the data from Experiment 2 (Fig. 3) suggests a need for an analysis in terms of higher-order trends. The results for the eight SOAs were therefore analyzed with a cubic regression. The data could be well described by a cubic function of the form $y = -10.43 + 0.17x - 2.3 \cdot 10^{-6}x^3$, $R^2 = .95$ (see short-dashed curve in Fig. 4a).

In the next analysis, all the data obtained from the 7 participants common to Experiments 1 and 2 were analyzed in the same way, resulting in the following function: $y = -14.31 + 0.19x + 8.1 \cdot 10^{-5}x^2 - 1.9 \cdot 10^{-6}x^3$, $R^2 = .88$ (see long-dashed curve in Fig. 4a). Comparing the two fits in Fig. 4a reveals highly similar parameter values; thus, for the further analysis, we have combined all data obtained in Experiments 1 and 2.

Repeating the same analysis with all data from the two experiments combined gives an overall significant cubic function: $y = -16.42 + 0.17x + 8.5\ 10^{-5}x^2 - 1.6 \cdot 10^{-6}x^3$, $R^2 = .80$ (see solid curve in Fig. 4a).

For the following discussion, we will define the term auditory *influence* to refer to the slope of the relation between marker SOAs and adjusted visual target position. The results reveal that the relation between marker SOA and shift

Table 3 Results of Experiment 2: Mean adjusted positions and standard deviations for each condition for each participant

Audio delay	−300 ms		−225 ms		−75 ms		−25 m		+25 ms		+75 ms		+150 ms		+225 ms	
Participant	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
P1	−1.2	70.1	4.1	62.6	−27.1	59.3	15.9	49.1	−2.4	62.0	18.9	51.4	9.4	53.8	−7.7	72.3
P2	−41.3	117.4	−46.0	117.0	−36.0	118.0	−28.3	122.5	−10.6	79.1	−3.5	92.8	−64.3	100.3	−37.2	98.2
P3	−0.6	87.8	−44.3	63.4	−38.9	73.7	−26.1	98.5	−23.0	80.7	14.2	85.7	−18.9	98.6	−25.8	101.5
P4	−30.7	46.0	−1.2	41.2	1.8	37.0	1.8	46.1	0.6	38.8	3.5	44.3	−5.9	38.0	−17.7	38.9
P5	15.9	68.9	−0.6	77.8	−7.1	77.0	−30.7	76.0	0.6	86.2	−8.9	77.5	22.4	72.0	0.6	104.1
P6	−23.6	69.2	−43.1	72.5	−18.9	80.5	−24.7	79.7	6.2	94.9	31.9	87.7	29.2	73.6	−2.5	80.7
P7	−35.4	27.1	−32.5	24.2	−8.3	25.7	3.0	41.0	−7.7	29.0	3.5	39.1	10.0	27.7	4.1	25.2
P8	−65.5	48.0	−74.9	42.9	−42.5	44.2	−16.5	42.5	20.7	59.3	−37.2	56.6	23.0	50.1	7.1	46.6
P9	−31.9	81.6	−40.7	86.2	−5.9	96.2	−16.5	120.1	27.1	104.9	−26.6	86.4	40.7	80.9	44.8	82.7
P10	−6.5	46.8	−5.9	52.7	−24.8	35.3	−45.4	44.9	−7.7	41.1	−6.5	33.9	−7.1	45.0	−13.6	51.8
P11	65.5	117.8	−31.9	130.2	28.3	117.0	64.3	127.8	−23.0	138.3	5.9	144.3	7.7	135.6	−44.8	124.9
P12	−37.8	70.8	−54.9	75.2	14.2	87.0	−26.0	95.9	21.2	92.6	30.1	109.5	9.4	86.2	−31.9	118.4
P13	−27.7	47.1	6.5	53.9	−67.3	51.8	−59.0	57.0	−17.1	75.7	−17.7	81.5	−24.8	56.4	−20.1	76.8
P14	−10.0	59.2	−30.7	58.5	−74.3	54.3	−10.0	58.6	−6.5	44.7	−3.5	55.2	−3.0	60.0	6.5	63.7
P15	−36.0	45.5	−58.4	37.2	−62.0	32.7	−37.2	27.4	−26.6	34.4	−26.0	28.0	−14.8	35.9	−11.8	31.1
P16	−2.4	41.0	−10.0	45.5	−44.8	42.0	−16.5	36.2	−3.0	62.3	4.1	38.7	8.3	39.6	8.3	34.9
P17	26.0	47.0	2.4	40.1	−49.6	54.2	−23.6	55.2	27.1	54.8	15.9	57.2	30.7	63.9	17.7	56.8
P18	−7.7	71.4	−34.2	66.8	−11.2	63.3	−20.1	71.8	−1.2	72.8	0.6	78.5	0.0	83.0	20.7	79.8
Mean	−13.9		−27.6		−26.4		−16.4		−1.4		−0.1		2.9		−5.7	
SE	7.0		5.8		6.6		6.3		3.9		4.4		5.8		5.3	

Note. All conditions had audiovisual markers and a visual target. "Audio delay" indicates the temporal delay of the audio component of the markers, relative to the visual component. The values are in milliseconds.

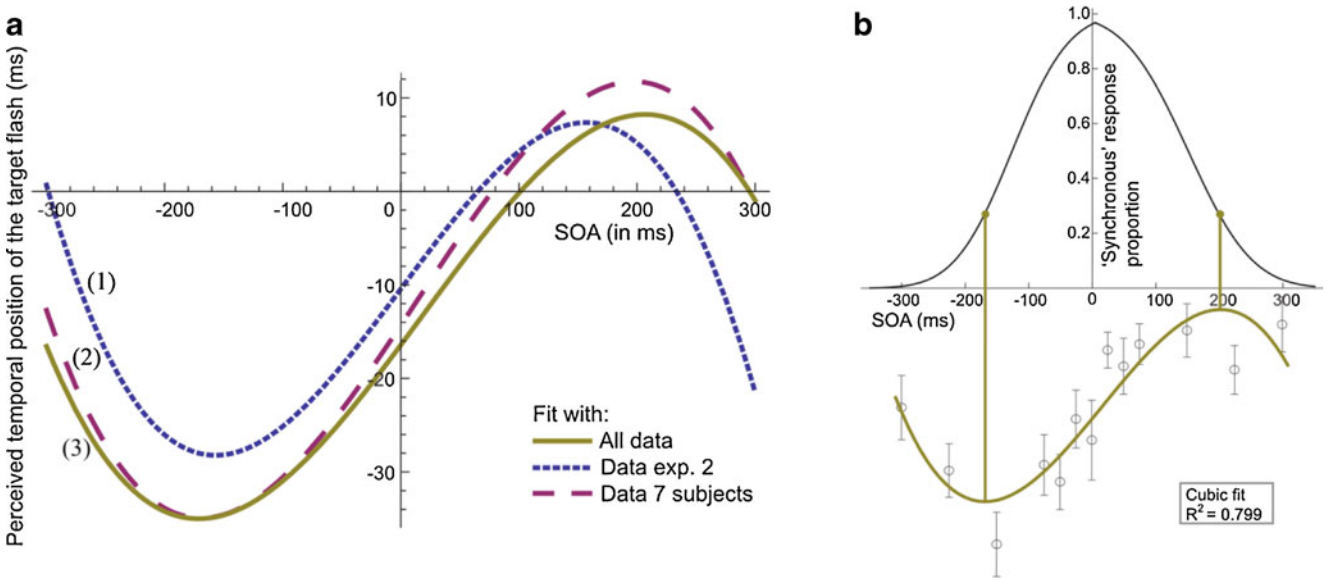

Fig. 4 **a** Three cubic best fits based on all data from Experiment 2 (short-dashed curve), all data of the 7 participants who participated in both Experiments 1 and 2 (long-dashed curve), and all data from Experiments 1 and 2 combined (continuous curve). **b** Fitted "synchronous" response judgment curve of the condition from Kuling, Van Eijk, Juola, and Kohlrausch (2012) that used stimuli similar to those used in the present studies, and the cubic fit of the combined audiovisual data (solid curve replotted from left panel, together with experimental results). The vertical lines indicate the temporal positions of the maximum and minimum values of the cubic function. On the left side, the minimum value corresponds to an SOA leading to a synchrony response probability of 27.0%; on the right side, the maximum value corresponds to a synchrony response probability of 27.9%. This observation suggests a relation between maximum overall shift observed in temporal ventriloquism and synchrony judgment probabilities for the same stimuli

in the visual target position cannot be expressed as a constant proportion of the SOA in the markers but that it is better described by a cubic relation. This means that the auditory influence is strongest for SOAs close to the PSS and decreases for both more positive and more negative SOAs.

In a more detailed analysis, we compared the SOA leading to the maximum and the minimum values of the overall cubic function with the synchrony curves from earlier experiments with the same short auditory and visual stimuli (Kuling et al., 2012). As Fig. 4b shows, the minimum of the cubic function corresponds to a percentage value of the (left side of the) synchrony curve of 27.0%, and the maximum of the cubic function to a percentage value of 27.9%. This observation indicates that the SOAs leading to the strongest positive and negative shifts in perceived position of the visual component coincide with the outer boundaries of the synchrony curve. At these two SOAs, participants perceive, on average, the AV stimuli as synchronous with a probability of about 25%. Furthermore, the influence of the audio component on the visual component in the marker is strongest for those SOAs that correspond to the central portion of the synchrony curve, with a maximum slope of about 0.16 ms/ms. For this range, we see a monotonically increasing relation between SOA and perceived temporal position of the visual component, leading to maximum shifts of about −35 and + 10 ms. Only for SOAs beyond this range does the induced temporal shift decrease again.

The observation of a potentially close relation between synchrony perception and induced temporal shifts suggested an alternative theoretical approach, based on the model described by Maij, Brenner, and Smeets (2009). This model combines the influence of the relative temporal position of the sound on the judged temporal position of the flash (linear component) with the probability that the flash and sound are considered to arise from the same event (Gaussian component). Multiplying these two factors gives the prediction of the perceived temporal shift for each SOA:

$$y = (ax + d)\cdot e^{-\frac{(x-b)^2}{2\cdot c^2}}. \quad (1)$$

In this formula, y is the observed shift of the visual target, x indicates the SOA in milliseconds, and a through d are four free parameters. We applied this theoretical curve for a different number of free parameters and used all data obtained in Experiments 1 and 2 as input. The first fit (continuous curve in Fig. 5) was obtained when all four parameters a–d were optimized (R^2 = .91; fit parameters: $a = 0.200$, $b = -0.088$, $c = 0.112$, $d = -0.018$). For the second fit (dashed curve in Fig. 5), the probability function (parameters b and c, $b = 0.013 + d/a$, $c = 0.117$) was derived from the synchrony curves in a synchrony judgment task with the same stimuli (from Kuling et al., 2012), so that only two parameters, a and d, were optimized from the present data set (R^2 = .91; fit parameters: $a = 0.184$, $d = -0.018$). Very similar

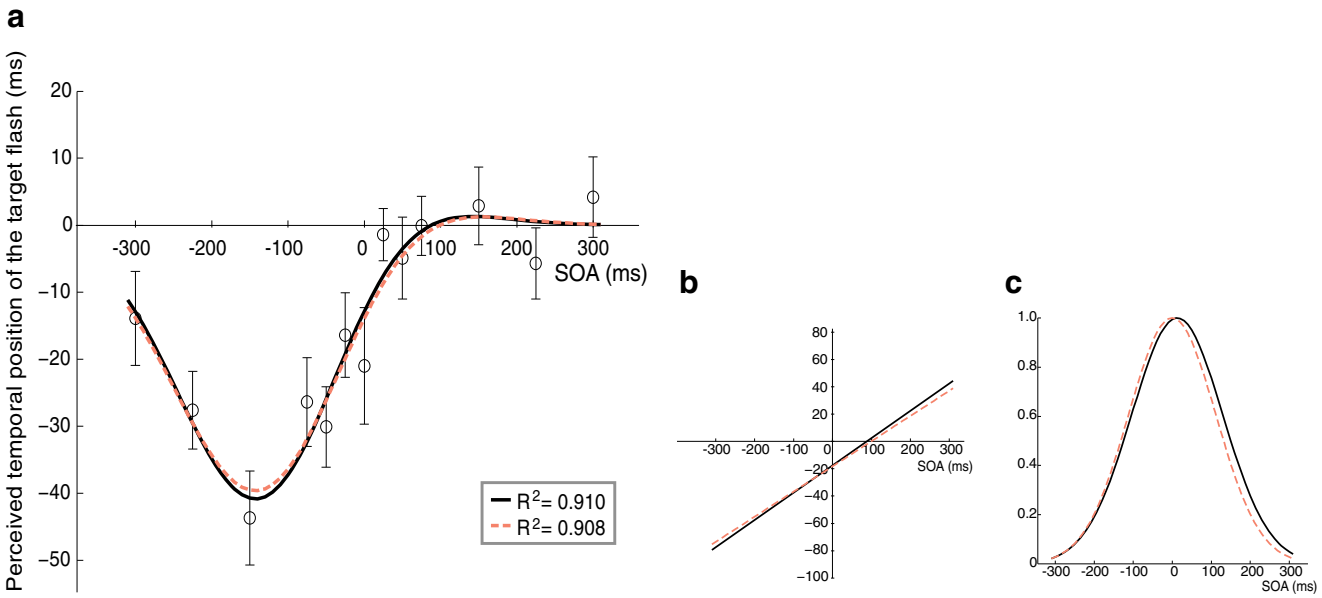

Fig. 5 Fits of the combined experimental data computed with (an adapted version of) the model used by Maij, Brenner, and Smeets (2009) (**a**). This model combines the influence of the temporal position of the tone on the judged temporal position of the flash (**b**) with the probability of the flash and tone being considered to arise from the same event (**c**). The continuous black curve is obtained when all four parameters a–d are optimized. For the dashed curve, the probability function (parameters b and c) is derived from the synchrony curves in a synchrony judgment task with the same stimuli (from Kuling, Van Eijk, Juola, & Kohlrausch, 2012), so that only two parameters, a and d, are optimized

results were obtained when the analysis was limited to the 7 participants who participated in both experiments (R^2 = .87, a = 0.207, d = −0.017).

A comparison of the fits with two and four free parameters shows that the general shape of the synchrony judgment curve is of relevance for the quality of the fit, but the detailed temporal position of the maximum of this curve is not so important. In fact, the maxima of the two Gaussian functions in Fig. 5c differ by about 30 ms, but the quality of the overall fit result is practically unaffected. Also, the linear component of the fits, indicating the influence of the temporal position of the tone on the judged temporal position of the flash, is quite robust between the fits, with a slope value of about 0.2.

With this model, the effect of SOA might be predicted from the synchrony curve as obtained in a synchrony judgment task. However, the model (and method) is not sensitive enough to accurately predict PSS values and the synchrony curve from the data of the temporal adjustment task.

Discussion

The results of Experiment 2 show that the auditory influence on the perceived moment of occurrence of the visual component depends on the AV SOA. The influence—that is, the change in position per change in SOA—is largest for SOAs in a range around the PSS for these AV stimuli (about 15 ms). This means that when the sensory inputs are perceived as synchronous, the auditory component strongly influences the perceived temporal position of the flash. This relative influence decreases for larger positive and negative SOAs, but the absolute shift value increases on the left side up to an SOA of about −168 ms, inducing a maximum temporal displacement of −37 ms, and on the right side to an SOA of +201 ms, leading to a maximum temporal displacement of +9 ms.

When we compare these results with results of single-event experiments, we find some very interesting analogies. For example, the asynchronies found in synchrony judgment tasks in single-event experiments are comparable to the asymmetry of the cubic function used to describe our results. Comparing our data with the mean synchrony perception curves (Kuling et al., 2012) shows synchrony response probabilities of about 27% for the two SOAs corresponding to the two maxima of the absolute shift. This would suggest that the influence of the auditory distractors exceeds the typical synchrony range (i.e., the range between the 50% points) used in synchrony judgment tasks. Furthermore, the auditory influence found in Experiment 2 was largest around the typical PSS found with the same stimuli in earlier studies (Kuling et al., 2012; Van Eijk et al., 2008, 2010).

The data on the outer left (−300 ms) and the outer right (+300 ms) sides are not as close to zero as we would expect from the unimodal conditions in Experiment 1. A closer look at the response processes showed that participants used more adjustment steps for the conditions with the largest (positive and negative) SOAs. They also reported them as being the most difficult trials. However, no effect of this difficulty can

be seen in the precision of the adjustments. This suggests that for these trials, other distracting factors might influence perception. For example, when the auditory and visual components are not temporally close enough to be perceived as coming from the same source, the five flashes and four noise bursts might be perceived as two different streams. For different rhythms, an auditory rhythm distracts from focusing on a visual rhythm, but most of the times, the visual rhythm can be correctly identified (Frings & Spence, 2010). This distracting effect could explain why the results for the outer SOAs are still noisy and do not perfectly return to zero.

A quantitative comparison of the present results with the data obtained by Burr et al. (2009) is not straightforward, due to some methodological differences. In our procedure, the perceived temporal position of the *visual* component in the marker is measured by adjusting a visual target. In contrast, in the experiment by Burr et al. (2009), the perceived temporal position of the AV marker was established by an AV target presented in physical synchrony. That is, the latter authors determined the contribution (weights) of auditory and visual components on the temporal perception of the total AV stimulus, while we directly measured the induced shift in perceived timing of the visual component. Although these two paradigms certainly address related problems, it is not obvious that they are identical and that the derived weights can be compared directly.

In the following discussion, we will attempt to apply the model approach from the present study to the results on temporal ventriloquism found by others. Morein-Zamir et al. (2003) measured the effect of one or two sounds on the JND between two visual stimuli. They started with a baseline condition in which the onsets of two lights were both accompanied by a simultaneous click (the lights stayed on, so there was no offset). In this condition, they found a JND of about 60 ms. With a lag of 75 ms between clicks and onsets (one click preceding the first onset and the other click following the second onset), Morein-Zamir et al. found a decrease in the JND of about 19 ms. For this condition, the model developed in the previous section predicts a decrease in the JND of 31ms (see Table 4). This decrease is mainly caused by an influence of the first sound on the first visual onset (induced temporal shift for an SOA of −75 ms is equal to −32 ms) and only to a small extent by the effect of the second sound on the second visual onset (SOA +75 ms leads to a shift of −1 ms). Also, for the data obtained for SOAs of 150 and 225 ms, the model predicts a stronger effect on the JND than was observed experimentally. The model predicts decreases of 42 and 31 ms, respectively, while the decreases in the experimental results amounted to 20 and 11 ms.

These results found in the literature are thus smaller than what is predicted using the model parameters that described our data very well. A source for this difference could be a difference in tasks. For example, Vroomen and de Gelder (2004) and Vroomen and Keetels (2009) found a relative influence of about 5% of the SOAs. In the present experiment, an *influence* of 20% was found, which might suggest that perception of rhythmic sequences induces a stronger temporal ventriloquism effect than does "single event" perception. If we want to realize a weaker influence in our model, we can do this by changing the slope of the linear component, which had a value of 0.2 in our best fit. Changing the slope of the linear component in our model to 0.05, shows quite good predictions of the results found by Morein-Zamir et al. (2003). In Table 4, comparisons can be found between the predictions with two versions of this model (slopes of the linear component of 0.2 and 0.05) and results found by Morein-Zamir et al.

The literature reviewed here for comparison with our model predictions reveals a large range of values for, for example, baseline JNDs. Morein-Zamir et al. (2003) used only an onset stimulus, whereas Shimojo et al. (2001) used equal-duration auditory and visual components. The JNDs found in these two studies differ by a factor of 2 (60 vs. 30 ms). Even within the study by Morein-Zamir et al., a large difference in JND occurs between their Experiments 1 and 2. They explain this difference by a learning effect, but other factors (e.g., brightness [Boenke, Deliano, & Ohl, 2009] and stimulus duration [Kuling et al., 2012]) might play a role in the differences as well.

The differences in overall shift size between rhythmic sequences and single-event perception suggest that some kind of temporal recalibration might occur during the rhythmic sequences. Both Vroomen, Keetels, de Gelder, and Bertelson (2004) and Fujisaki, Shimojo, Kashino, and Nishida (2004) found that after a few minutes exposure to a train of sounds and flashes with a constant time lag, the PSS values for different simultaneity judgments and TOJs were changed in the direction of the time lag. Despite the fact that adaptation and recalibration are supposed to take several minutes, they might have influenced our results, which could explain some of the differences between the shift sizes induced by temporal ventriloquism in multi- and single-event perception.

General discussion

In this article, we described a set of experiments in which we tested a method, known from auditory and speech perception research, for quantifying temporal ventriloquism in AV perception. In line with the original publications on this cross-modal effect (e.g., Aschersleben & Bertelson, 2003; Morein-Zamir et al., 2003), we focused on induced temporal shifts in the perceived moment of occurrence of the visual component caused by the auditory stimulus component. From our data, we were able to derive an equation that describes the shift in perceived temporal position of the visual component as a

Table 4 Comparison between the results of Morein-Zamir, Soto-Faraco, and Kingstone (2003) and predictions from our model for two different slopes of the linear model component

	Results Found in Morein-Zamir et al., 2003		Model Predictions for Change in JND	
SOA	Baseline JND	Change in JND	Slope 0.2	Slope 0.05
Simultaneous first sound + trailing second +100 ms	72 ms	−3 ms	−24 ms	−11 ms
Preceding first sound + simultaneous second−100 ms	72 ms	−13 ms	−14 ms	−16 ms
Preceding first sound + trailing second, ±100 ms	72 ms	−11 ms	−38 ms	−26 ms
Preceding first sound + trailing second, ±75 ms	62 ms	−19 ms	−31 ms	−16 ms
Preceding first sound + trailing second, ±150 ms	62 ms	−20 ms	−42 ms	−21 ms
Preceding first sound + trailing second, ±225 ms	62 ms	−11 ms	−31 ms	−14 ms

Note. The slope of 0.2 represents the best fit of our model to the data of Experiments 1 and 2, while the slope of 0.05 is an indication of the effect size of auditory distracters on visual stimuli in previous experiments (Vroomen & de Gelder, 2004; Vroomen & Keetels, 2009).

function of the temporal distance (positive or negative) between the visual and the auditory stimulus components. Such a quantitative relation has not been previously developed.

As was mentioned in the introduction, the term *temporal ventriloquism* has been used in the recent literature for two related but clearly different experimental phenomena. In the original publications from 2003, ventriloquism was analyzed as the shift in the perceived moment of perception. For example, Morein-Zamir et al. (2003) tested the hypothesis "that auditory events can alter the perceived *timing* of target lights" (p. 155), and they related this hypothesis to the spatial equivalent where the presence of a visual stimulus affects the perceived *location* of an auditory stimulus. In contrast to this experimental paradigm, Burr et al. (2009) and Hartcher-O'Brien and Alais (2011) interpreted the term *ventriloquism* as indication of the dominance of one modality on the perceived moment of occurrence for AV stimuli with intermodal delays, resulting in weights for the auditory and visual contributions to the perceived moment of occurrence of the AV stimulus.

We want to emphasize that these two paradigms measure different aspects of cross-model temporal influences, and they also require different theoretical analyses. When studying the shift in perceived occurrence (as in Morein-Zamir et al., 2003), a wide range of intermodal delays can be tested, because it is likely that for delays that are too large, any intermodal influence will disappear and the visual component will be perceived unaltered. In contrast, in the paradigms used in the two recent papers (Burr et al., 2009; Hartcher-O'Brien & Alais, 2011), it is necessary that the auditory and visual components are perceived together at one point in time, demanding that the relative delays be within the window of synchronicity.

In the terminology from the study by Roach et al. (2006), the latter paradigm requires multisensory integration based on a weighted average of two sensory estimates, whereas the former paradigm is rather immune to mandatory cross-modal integration according to a maximum-likelihood estimate. In their own experiments, Roach et al. investigated the effects of temporal rate conflicts in terms of finding a balance between costs and benefits of multisensory integration. Participants had to judge the perceived rate of a modulated stimulus (either auditory or visual) using the rate of an AV stimulus with synchronous rate modulations as a comparison (this is a baseline experiment similar to our unimodal conditions in Experiment 1). Then the task-relevant test modulation was combined with a task-irrelevant modulation in the opposite modality, and the influence of this irrelevant modulation rate on the perceived modulation of the test modulation was established (this corresponds to our conditions with bimodal markers). For small discrepancies in the relevant and irrelevant modulation rates, the judged rate could be modeled as a weighted combination, where the weights were derived from the unimodal rate discrimination sensitivity. A particular outcome was that in conditions with equal unimodal sensitivities (in which the usually high auditory sensitivity was decreased by reducing the modulations depth), both components had about equal weights, indicating a clear example that also in the temporal domain, the usually less appropriate modality can get a high weight (see Alais & Burr, 2004, for such a demonstration in the spatial domain). When the rate discrepancy increased further, the cross-modal influence disappeared, and the rate estimates returned to their veridical values (as we tend to see for rather large positive and negative SOAs in our marker stimuli).

The present experiments allowed deriving an equation that described the observed temporal shift in perceiving the visual onset as a function of the SOA in the AV marker stimuli. The equation is made up of two components: a linear one, which we interpreted as a measure of the strength of the cross-modal attraction, and a Gaussian component reflecting the likelihood of perceived synchrony between the auditory and visual components. The weaker the perception of synchrony, the smaller the weights and overall effect of cross-modal timing shifts. Thus, at large SOAs, the effect must disappear, and there is a parallel with the breakdown of cross-modal rate discrepancy effects shown by Roach et al. (2006).

Our quantitative approach also allows a reinterpretation of the experiment by Morein-Zamir et al. (2003). These authors interpreted interactions among their stimuli such that the first sound affected the (perceived) onset of only the first visual stimulus, and the second sound the perceived onset of only the second visual stimulus. In our analysis of these data, we computed the expected shifts in the same way. But there is no direct evidence given by Morein-Zamir et al. that this close link between the individual sounds and the individual visual onsets happens in exactly this way. Let us take their condition with the sounds and closest visual onsets being 75 ms apart from each other and the two visual onsets being separated by an SOA of 45 ms. For this specific condition, the first sound is presented 75 ms *before* the first visual onset and 120 ms before the second onset. Accordingly, sound 2 is presented 75 and 120 ms *after* the two visual onsets. According to our theoretical model shown in Fig. 5, the first sound should have an influence *both on the first and also on the second visual onsets*. Just on the basis of the temporal distances, our model predicts that the first sound should have a stronger overall shift impact on the second visual onset (about 38 ms), as compared with a shift of about 33 ms for the first visual onset. That is, in order to interpret the data in the way of the original article by Morein-Zamir et al., one needs to make the additional assumption that each sound only affects one, but not both, of the visual onsets.

In conclusion, the experimental method proposed in the present article, based on rhythm perception, has proven to be useful for measuring temporal ventriloquism over a large range of AV SOAs. The model based on the results seems to be promising for quantifying and predicting the net result of temporal ventriloquism in different experiments. However, to find the exact details, possibilities, and limitations of the model, future work is necessary. We hope that by presenting a quantitative, testable prediction for the induced time shift in visual stimuli, more conditions will be created to broaden the experimental data base for the phenomenon of temporal ventriloquism. A particular goal would be to better understand, and possibly integrate, the two sides of temporal ventriloquism emphasized in the introduction and the "General discussion" section: the cross-modally induced shift in perceived timing of individual components of an AV stimulus and the weighted contributions of the individual components to the perceived moment of occurrence of an AV stimulus.

Acknowledgments We would like to thank Bjorn Vlaskamp, Raymond van Ee, and two anonymous reviewers for constructive comments on an earlier version of the manuscript.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14,* 257–262.

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., & Bertelson, P. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension 2. Evidence from sensorimotor synchronization. *International Journal of Psychophysiology, 50,* 157–163.

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension 1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology, 50,* 147–155.

Boenke, L. T., Deliano, M., & Ohl, F. W. (2009). Stimulus duration influences perceived simultaneity in audiovisual temporal-order judgment. *Experimental Brain Research, 198,* 233–244.

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research, 198,* 49–57.

Frings, C., & Spence, C. (2010). Crossmodal congruency effects based on stimulus identity. *Brain Research, 1354,* 113–122.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience, 7,* 773–778.

Gebhard, J. W., & Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *American Journal of Psychology, 72,* 521–528.

Glenberg, A. M., & Jona, M. (1991). Temporal coding in rhythm tasks revealed by modality effects. *Memory & Cognition, 19,* 514–522.

Grahn, J. A. (2012). See what I hear? Beat perception in auditory and visual rhythms. *Experimental Brain Research, 220,* 51–61.

Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see: Auditory encoding of visual temporal sequences. *Psychological Science, 16,* 228–235.

Hartcher-O'Brien, J., & Alais, D. (2011). Temporal ventriloquism in a purely temporal context. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 1383–1395.

King, A. J. (2005). Multisensory integration: Strategies for synchronization. *Current Biology, 15,* 339–341.

Kuling, I. A., Van Eijk, R. J. L., Juola, J. F., & Kohlrausch, A. (2012). Effects of stimulus duration on audio-visual synchrony perception. *Experimental Brain Research, 221,* 403–412.

Maij, F., Brenner, E., & Smeets, J. B. J. (2009). Temporal information can influence spatial localization. *Journal of Neurophysiology, 102,* 490–495.

Marcus, S. M. (1976). Perceptual centers. Doctoral dissertation Cambridge, UK.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17,* 154–163.

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1085–1099.

Riggs, L. A. (1971). Vision. In J. W. King & L. A. Riggs (Eds.), *Woodworth & Schlossberg's experimental psychology* (3rd ed.). New York: Holt, Rinehart, and Winston.

Roach, N. W., Heron, J., & McGraw, P. W. (2006). Resolving multisensory conflict: A strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society B, 273,* 2159–2168.

Rose, D., & Summers, J. (1995). Duration illusions in a train of visual stimuli. *Perception, 24,* 1177–1187.

Schimmel, O., & Kohlrausch, A. (2006). On the influence of interaural differences on temporal perception of masked noise bursts. *Journal of the Acoustical Society of America, 120,* 2818–2829.

Schimmel, O., & Kohlrausch, A. (2008). On the influence of interaural differences on temporal perception of noise bursts of different

durations. *Journal of the Acoustical Society of America, 123,* 986–997.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature, 408,* 2000–2000.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research, 14,* 147–152.

Shi, Z., Chen, L., & Müller, H. J. (2010). Auditory temporal modulation of the visual Ternus effect: The influence of time interval. *Experimental Brain Research., 203,* 723–735.

Shimojo, S., Scheier, C., Nijhawan, R., Shams, L., Kamitani, Y., & Watanabe, K. (2001). Beyond perceptual modality: Auditory effects on visual perception. *Acoustical Science and Technology, 22,* 61–67.

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.

Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. *Perception & Psychophysics, 66,* 1171–1189.

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.

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.

Vroomen, J., & de Gelder, B. (2004). Temporal ventriloquism: Sound modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 513–518.

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.

Vroomen, J., & Keetels, M. (2009). Sounds change four-dot masking. *Acta Psychologica, 130,* 58–63.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception & Psychophysics, 72,* 871–884.

Yeung, N., & Monsell, S. (2003). Switching between tasks of unequal familiarity: The role of stimulus-attribute and response-set selection. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 455–469.