

Auditory temporal modulation of the visual Ternus effect: the influence of time interval

Zhuanghua Shi · Lihan Chen · Hermann J. Müller

Received: 6 August 2009 / Accepted: 29 April 2010 / Published online: 16 May 2010
© Springer-Verlag 2010

Abstract Research on multisensory interactions has shown that the perceived timing of a visual event can be captured by a temporally proximal sound. This effect has been termed ‘temporal ventriloquism effect.’ Using the Ternus display, we systematically investigated how auditory configurations modulate the visual apparent-motion percepts. The Ternus display involves a multielement stimulus that can induce either of two different percepts of apparent motion: ‘element motion’ or ‘group motion’. We found that two sounds presented in temporal proximity to, or synchronously with, the two visual frames, respectively, can shift the transitional threshold for visual apparent motion (Experiments 1 and 3). However, such effects were not evident with single-sound configurations (Experiment 2). A further experiment (Experiment 4) provided evidence that time interval information is an important factor for crossmodal interaction of audiovisual Ternus effect. The auditory interval was perceived as longer than the same physical visual interval in the sub-second range. Furthermore, the perceived audiovisual interval could be predicted by optimal integration of the visual and auditory intervals.

Keywords Time perception · Vision · Audition · Temporal ventriloquism effect · Ternus display

Introduction

Most events we perceive in everyday life consist of simultaneous inputs from different sensory modalities. For example, when knocking at the door, we immediately perceive a sound coupled with a touch feedback. By integrating different sources of sensory information, our brain can achieve accurate representations of the environment (Calvert et al. [2004](#)

a visual event can be biased by an irrelevant and slightly asynchronous auditory stimulus (Fendrich and Corballis 2001; Morein-Zamir et al. 2003; Scheier et al. 1999; Shimojo et al. 2001). For example, Scheier et al. (1999); see also Morein-Zamir et al. 2003) demonstrated that when two sounds were presented, one slightly before a first flash and the other shortly after a second flash, the sounds attracted the temporal occurrence of the lights, thus improving the visual temporal resolution (i.e., the just noticeable difference, JND). In contrast, when two sounds were inserted in between two light flashes, the visual temporal resolution became worse (Scheier et al. 1999). This phenomenon, referred to as ‘temporal ventriloquism effect’ (Morein-Zamir et al. 2003), has elicited a great deal of interest in multisensory interaction research. It has been proposed that the temporal ventriloquism effect is related to the modality appropriateness or modality precision (Welch and Warren 1980, 1986). On this hypothesis, the sensory modality with the highest acuity may outweigh the others, so that, for example, audition with its high temporal resolution may dominate temporal perception.

The spatial and temporal constraints on the temporal ventriloquism effect were investigated in a number of follow-up studies (Bruns and Getzmann 2008; Freeman and Driver 2008; Getzmann 2007; Jaekl and Harris 2007; Keetels et al. 2007; Vroomen and Keetels 2006). As there was little influence of the relative spatial positions of the auditory and visual stimuli on the temporal ventriloquism effect (Bruns and Getzmann 2008; Vroomen and Keetels 2006), other studies turned to examining the influence of the temporal configuration on the effect. For example, Morein-Zamir et al. (2003) found that the temporal relationship of audiovisual events was important: with two sounds presented before and, respectively, after the two visual stimuli, the temporal ventriloquism effect occurred only when the second sound was trailing the second light within a range of 200 ms; and with two sounds presented in between the two visual stimuli, a temporal ventriloquism effect was observed only when the sounds were separated by 16 ms. Furthermore, several studies have shown that a single sound leaves temporal-order judgment (TOJ) performance uninfluenced (Morein-Zamir et al. 2003; Scheier et al. 1999; Shimojo et al. 2001). This has been taken to suggest that two sounds are required, one paired directly with each visual event, for the audiovisual stimuli to be perceived as a unitary event (Morein-Zamir et al. 2003; Welch 1999). Similar results have also been reported for the audiovisual apparent-motion paradigm (Bruns and Getzmann 2008; Freeman and Driver 2008; Getzmann 2007). For example, a temporal ventriloquism effect has been observed in an ambiguous bidirectional visual apparent-motion stream by introducing auditory beeps slightly lagging or leading the flashes (Freeman and Driver

2008). More recently, intramodal perceptual grouping has been shown to be an important factor in crossmodal integration (Sanabria et al. 2004; Vroomen and de Gelder 2004; Keetels et al. 2007; see reviews, Spence et al. 2007; Spence 2007). Intramodal grouping may segregate different sensory events from one another, thus weakening the crossmodal interaction. For example, using a sequence of sounds, Keetels et al. (2007) found the temporal ventriloquism effect to be diminished when the two target sounds (paired with visual targets) had the same frequency or rhythm as other flanker sounds.

Despite the empirical evidence reviewed above, the influence of sound structure on the temporal ventriloquism effect has thus far received only little attention in the literature. Arguably, however, to achieve an understanding of the mechanisms underlying the temporal ventriloquism effect, it is critical to examine crossmodal time perception—in particular, the perceived onset/offset time versus the perceived time interval of the events in the target and distractor modalities. Onset/offset time and time interval relate to the two fundamental concepts in time perception: succession and duration (Fraisse 1984). In most of the aforementioned studies, the temporal ventriloquism effects were implicitly assumed to be due to the visual events being captured by the accompanying auditory events at marked points (e.g., onsets or offsets) in time (temporal-marker hypothesis). To date, it is still unknown how the time interval of events in the distractor modality influences the temporal ventriloquism effect. If the onset or offset of a sound is the major factor determining the crossmodal interaction, a temporal ventriloquism effect should also be observable under single-sound conditions. Alternatively, as has been suggested by several authors (Morein-Zamir et al. 2003; Welch 1999), paired auditory events that correspond with paired visual events may be necessary to induce a temporal ventriloquism effect. Given this, the time intervals produced by the paired stimuli may also be an important factor in the crossmodal temporal interaction. Moreover, perceived intervals in the auditory modality have been shown to be longer than the same (physical) intervals in the visual modality (Goldstone and Lhamon 1974; Walker and Scott 1981). Thus, if time intervals are an influential factor and are perceived differently in the auditory and visual modalities, one might observe a temporal ventriloquism effect even when the auditory and the visual onsets occur simultaneously.

To investigate how sound onset and inter-sound interval influence the perception of visual apparent motion, in the present study, we adopted the Ternus apparent-motion paradigm (Harrar and Harris 2007; Ternus 1926). The typical Ternus apparent motion is produced by presenting two sequential visual frames; each frame consists of two horizontal dots, and the two frames, when overlaid, share

one common dot at the center. When the spatial configuration is fixed, observers typically report two distinct percepts dependent on the inter-stimulus onset interval (ISOI): ‘element motion’ and ‘group motion’ (Fig. 1). Short ISOIs usually give rise to the percept of element motion, that is: the outer dots are perceived as moving, while the center dot appears to remain static or flashing. By contrast, long ISOIs give rise to the perception of group motion: the two dots are perceived to move together as a group (Kramer and Yantis 1997; Pantle and Petersik 1980; Pantle and Picciano 1976). Numerous studies have shown that element and group motion are never perceived simultaneously, that is, the two types of percept are mutually exclusive (for a review, see Petersik and Rice 2006).

In Experiment 1, we combined dual sounds with the visual Ternus display, introducing three different audiovisual temporal arrangements. As temporal ventriloquism has been found in both temporal-order judgments (TOJ) and classical apparent-motion tasks, a similar interaction was expected to be observed with visual Ternus apparent motion. In Experiment 2, we examined the effects of single-sound audiovisual arrangements, with a sound presented in temporal proximity to either the first or the second visual frame. If sound onset captures the onset of visual stimuli, one would expect to obtain a similar ventriloquism effect in single-sound conditions as in dual-sound conditions. However, if other factors, such as inter-sound interval or auditory grouping, are important for producing a temporal modulation, the temporal ventriloquism effect with single sounds may be weak, if at all present. In Experiments 3 and 4, we further examined the importance of the interval for the audiovisual temporal interaction. If a given (physical) time interval is perceived differently in different modalities and the auditory interval influences the visual interval, one would expect a temporal interaction to occur even when the onsets of the visual and auditory events coincide in time. Additionally, the variances of the interval estimates for the auditory, the visual, and the combined audiovisual events were examined

further to quantitatively describe the interactions between the auditory and visual intervals.

Experiment 1

Two auditory clicks presented close in time to the visual events have been found to influence the sensitivity of (visual) temporal-order judgments (Morein-Zamir et al. 2003; Scheier et al. 1999) as well as the (visual) percept of apparent motion (Freeman and Driver 2008; Getzmann 2007). In Experiment 1, we examined whether a similar audiovisual temporal capture effect would also be found in the visual Ternus paradigm. In particular, we were interested in any temporal capture effect using an audiovisual interval of 30 ms, which is generally within the range of the audiovisual simultaneity window (Levitin et al. 2000; Stone et al. 2001).

Method

Participants

Ten participants (6 females and 4 males, mean age of 24.8) took part in Experiment 1. All had normal or corrected-to-normal vision and normal hearing. All were naïve as to the purpose of the experiment. Informed consent was obtained before the start of the experimental session.

Apparatus and stimuli

Visual stimuli were presented on a 17-inch CRT monitor (Viewsonic) with a refresh rate of 100 Hz and a resolution of 1024 × 768 pixels, which was controlled by a PC (HP AMD Athlon 64 Dual-Core Processor) with a Radeon 1700

were controlled via the monitor's vertical synchronization pulse.

Design and procedure

Prior to the experiment, participants were shown demos of 'element motion' and 'group motion' and then performed a block of trials for practice. A trial started with a fixation cross presented at display center for 300 ms. Next, a blank display was shown for a random duration of 500–700 ms. This was then followed by the first visual stimulus frame which was presented for 30 ms. After a variable ISOI (80, 110, 140, 170, 200, 230, 260, 290, or 320 ms), the second visual stimulus frame was presented, also for 30 ms. The two visual frames shared one common element, located at the center of the display. The location of the other, outer element of the first frame, either to the left or to the right of the shared element, was always opposite to the outer element of the second frame (Fig. 2a). Each visual frame was accompanied by one brief, 30-ms sound. There were three different conditions of audiovisual interval (Fig. 2b): In condition 1, the first sound preceded the first visual frame and the second sound trailed the second visual frame by 30 ms (hereafter, we will refer to these as 'outer sounds'). In condition 2, the two sounds were presented simultaneously with two visual frames ('simultaneous sounds'). And in condition 3, the first sound trailed the first visual frame and the second sound preceded the second visual frame by 30 ms ('inner sounds'). On each trial, stimuli of the same audiovisual configuration were repeated once more after a 1,000-ms blank interval, so as to provide participants with a strong-enough impression for a response decision. The display then disappeared, and after a random duration of 300–500 ms, participants were presented with a question mark which prompted them to make a two-forced choice (mouse button) response indicating whether they had perceived element or group motion. For half of the participants, a left mouse click corresponded to 'group motion' and a right click to 'element motion,' and vice versa for the other half. Thus, Experiment 1 implemented a 3 (audiovisual intervals) \times 9 (ISOIs) within-subjects design. Each configuration was presented 24 times, with the directions of apparent motion balanced across the 24 trial instances.

Results and discussion

The proportion of 'group motion' reports was plotted as a function of the ISOI and fitted by logistic regression for each participant. For each condition, the transition threshold between 'element motion' and 'group motion'—that is, the point at which 'group motion' and 'element motion' were reported equally frequently, which is also referred to

as the point of subjective equality (PSE)—was calculated by estimating the 50% performance point on the (fitted) $\text{PriTD}(\cdot)$

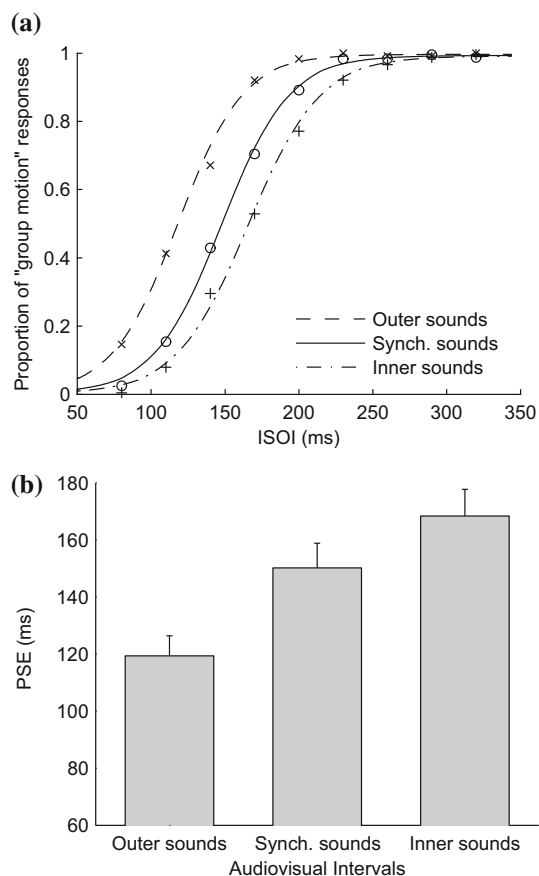


Fig. 3 **a** Psychometric curves fitted for the data from all participants in Experiment 1. The *solid curve* and *circles* represent the 'synchronous-sounds' condition, the *dashed curve* and *crosses* the 'outer-sounds' condition, and the *dash-dotted curve* and *pluses* the 'inner-sounds' condition. **b** Mean PSEs (and associated standard errors) for three conditions of audiovisual intervals

Table 1 Transition thresholds (and associated standard errors, ms) between 'element motion' and 'group motion' for three audiovisual (AV) intervals for Experiments 1 and 2

AV-interval (ms)	Transition thresholds (PSE \pm SE)		
	Experiment 1	Experiment 2a	Experiment 2b
0	150.1 (± 8.7)	149.8 (± 7.9)	153.5 (± 10.7)
30	119.3 (± 7.0)	146.1 (± 9.2)	157.1 (± 13.1)
-30	168.2 (± 9.4)	153.5 (± 7.5)	147.6 (± 10.7)

synchronous sounds and less likely with inner sounds compared to synchronous sounds. Both (outer and inner sounds) conditions produced the classical temporal ventriloquism effect (TVE) with a Ternus display: compared against the synchronous-sound baseline, the effect sizes were -30.8 (SE: 3.53) ms and 18.1 (SE: 3.05) ms for outer and inner sounds, respectively.

To our knowledge, with the Ternus display, this is the first demonstration of opposite TVEs on apparent motion

by outer and inner sounds, respectively. In this regard, the present results go beyond previous studies (Getzmann 2007; Morein-Zamir et al. 2003; Scheier et al. 1999). For example, Getzmann (2007) demonstrated the influence of inner clicks on a continuous apparent-motion percept, but failed to find a significant effect of outer clicks (though the latter showed some tendency toward an effect). In another study (Morein-Zamir et al. 2003), there was also one intervening-sounds condition (with a 40-ms inter-stimulus-interval between sounds) that yielded no effect in a TOJ task. The failure to find opposite effects in previous studies might be due to the short ISOIs used in TOJ tasks and participants' difficulties in classifying the multiple percepts in classical apparent-motion tasks (see above).

Possible accounts for the temporal interaction effect have been put forward in previous studies. One typical explanation is based on the modality precision hypothesis. When two sensory modalities provide discrepant temporal information about an event, this discrepancy is resolved by favoring the modality that is characterized by a higher precision in registering that event. The auditory system has higher temporal resolution than the visual system (Welch 1999; Welch and Warren 1980). Accordingly, auditory events are assigned high weights, and visual events low weights, in audiovisual integration (Morein-Zamir et al. 2003; Welch and Warren 1980, 1986). However, what remains unclear is whether the differential modality weighting is based on points in time (onsets/offsets) or time intervals. Experiment 2 was designed to produce evidence for deciding between these alternatives by examining how single sounds, which do not provide any auditory-interval information, would influence the visual Ternus apparent motion.

Experiment 2

In a previous study (Morein-Zamir et al. 2003), an influence of dual sounds on visual TOJs was observed only in conditions in which the second sound trailed the second visual event by 100 or 200 ms, while there was no temporal capture effect in a single-sound condition. In Experiment 2, we examined for any crossmodal temporal interaction effect of single sounds, which were temporally manipulated (i.e., preceding, synchronous, or trailing) with respect to either the first or the second visual event in visual Ternus display.

Method

The method was the same as in Experiment 1, with the following exceptions.

Participants

The same ten participants who had taken part in Experiment 1 also participated in Experiment 2. Five participants performed Experiment 2 on day 1 and Experiment 1 on day 2, and vice versa for the other five participants—thus counterbalancing potential practice effects across the two experiments.

Design and procedure

Experiment 2 consisted of two separate sessions, hitherto referred to as Experiments 2a and 2b, respectively. Half the participants performed Experiment 2a first and then Experiment 2b, and vice versa for the other half. In Experiment 2a, only one sound was presented, namely, close to the first visual frame. In Experiment 2b, only the second sound was presented. The settings were the same as in Experiment 1, except that either the second sound (Experiment 2a) or the first sound (Experiment 2b) was omitted. The three conditions of audiovisual interval were preceding sound (30 ms before the onset of the respective visual frame), synchronous sound, and trailing sound (30 ms after the onset of the respective visual frame).

Results and discussion

Individual PSEs for the three audiovisual intervals were computed as in Experiment 1. For Experiment 2a, in which the sound accompanied the first visual frame, the mean PSEs are presented in Table 1 (see also Fig. 4a). A repeated-measures ANOVA of the PSEs showed the main effect of audiovisual interval to be significant, $F(2,18) = 3.69$, $P < 0.05$. Separate paired t -tests (with the conservative Bonferroni correction) of the PSEs revealed only a (marginally) significant difference between the preceding-sound and trailing-sound conditions (difference of 7.3 ms, $P = 0.09$). However, the classical TVEs calculated relative to the synchronous-sound baseline were far from reaching significance (although they were in the right direction numerically): -3.7 ms, $P = 0.42$, and 3.7 ms, $P = 0.69$, for the preceding-sound and trailing-sound conditions, respectively.

Similar results were obtained in Experiment 2b, where the sound accompanied the second visual frame—see Table 1 for the mean PSEs (see also Fig. 4b). A repeated-measures ANOVA revealed the main effect of audiovisual interval to be significant, $F(2,18) = 5.49$, $P < 0.05$. Follow-on paired t -tests (with Bonferroni correction) showed the PSE for the trailing-sound condition to be significantly smaller compared to both the synchronous sound (classical TVE of -5.9 ms, $P < 0.05$) and the preceding sound (difference of -9.4 ms, $P < 0.05$) condition, while the

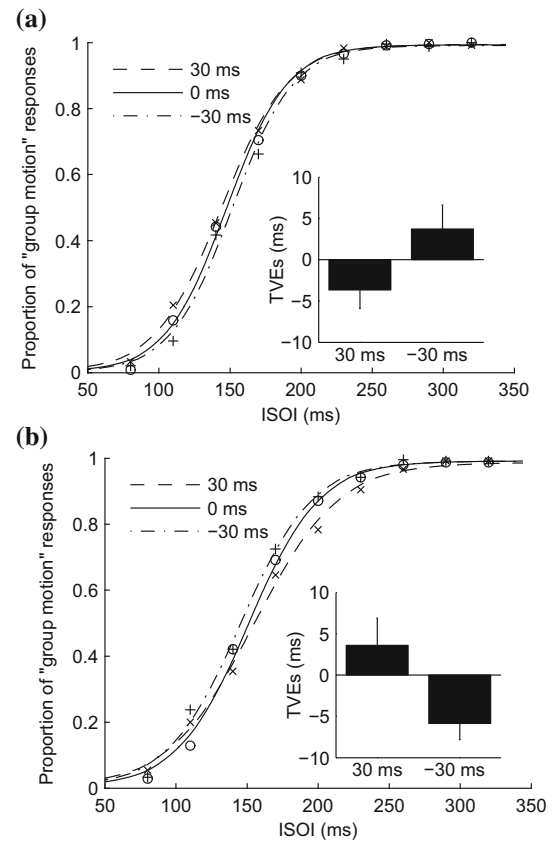


Fig. 4 **a** Psychometric curves fitted for the data from all participants in Experiment 2a. The *solid curve and circles* represent the 'synchronous-sound' condition (audiovisual interval 0 ms), the *dashed curve and crosses* the 'preceding-sound' condition (audiovisual interval 30 ms), and the *dash-dotted curve and pluses* the 'trailing-sound' condition (audiovisual interval -30 ms). The magnitude of the TVEs, calculated against the 'synchronous-sound' baseline, is presented in a subplot for the 'preceding-sound' (30 ms) and 'trailing-sound' conditions (-30 ms). **b** Psychometric curves fitted for the data from all participants in Experiment 2b. The magnitude of the TVEs, calculated against the 'synchronous-sound' baseline, is presented in a subplot for the 'preceding-sound' (30 ms) and 'trailing-sound' conditions (-30 ms)

PSEs did not differ reliably between the preceding-sound and synchronous-sound conditions (3.5 ms, $P = 0.9$).

The fact that there was no significant TVE with a sound preceding or trailing the first visual frame or a sound preceding the second visual frame in Experiment 2 is consistent with Bruns and Getzmann (2008), who reported a similar pattern with temporally proximal single sounds. However, a small TVE was found for trailing sound associated with the second visual frame in Experiment 2b. The asymmetry between preceding and trailing sounds is similar to the results obtained in a previous TOJ study with a dual-sound modulation (Morein-Zamir et al. 2003). The underlying mechanism is still unclear, though it has been suggested that this pattern relates to an asymmetry in audiovisual simultaneity (Dixon and Spitz 1980; Morein-Zamir et al. 2003).

In any case, compared to the TVEs obtained with the two-sound configurations in Experiment 1, TVEs by single sounds are relatively weak and significant only with a sound trailing the second visual frame. Even comparing the transition thresholds between the preceding- and trailing-sound conditions, the differences are merely 7.3 and 9.4 ms for Experiments 2a and 2b, respectively. Note that, compared with Experiment 1, all settings were the same except that one sound was removed. In addition, in Experiment 2, the additive (classical TVE) effect of the sound preceding the first visual frame plus that trailing the second visual frame was only -9.6 ms, which is much smaller than the TVE (of -30.8 ms) in the outer-sounds condition of Experiment 1. Similarly, the additive effect of two individual ‘inner sounds’ was 7.2 ms, which is also smaller than the TVE (of 18.1 ms) in the inner-sounds condition of Experiment 1. This suggests that the effects obtained with single sounds in Experiment 2 cannot fully explain the temporal ventriloquism effect, based only on the influence of the sounds’ onsets (i.e., temporal markers). Previous attempts to account for the weak or absent TVE with single-sound configurations have been in terms of a violation of the fundamental ‘assumption of unity’ (Morein-Zamir et al. 2003; Welch 1999). On this assumption, two separate multisensory events which share physical properties (e.g., both involving a visual component associated with an auditory component) are more likely considered as originating from the same source. Alternatively, if two sensory modalities are experienced by the observers as signaling disparate events (e.g., if one of the two events involves only one modality), the crossmodal interaction is less likely to happen (Welch 1999). Consequently, in single-sound conditions, two events with different physical properties (one involving audiovisual and the other visual stimuli only) weaken the attribution of a common cause and, therefore, the crossmodal temporal interaction. However, besides a violation of the ‘assumption of unity,’ an alternative account for the weak TVE in Experiment 2 could be the lack of an inter-sound interval (with single-sound configurations).

Experiment 3

One possible way to disentangle the role of temporal markers from that of time interval without violating the ‘assumption of unity’ is to keep the onsets of the audiovisual events physically the same, while making the perceived intervals different. To realize such a situation, in Experiment 3, synchronous audiovisual events (with sound onset occurring simultaneously with visual frame onset) were used. The rationale for this is as follows. There is ample evidence from studies of audiovisual time judgments that the interval between two auditory stimuli appears longer than that between two visual stimuli, even

if the auditory and visual stimuli are presented simultaneously (Goldstone and Lhamon 1974; Walker and Scott 1981; Wearden et al. 1998). If the inter-sound interval plays a (critical) role in audiovisual temporal interactions, one would expect that the transition threshold (from element to group motion) in visual Ternus displays would be lowered by the inter-sound interval in the synchronous audiovisual condition, compared with the unimodal (visual-display-only) condition. By contrast, the hypothesis of temporal-marker capture would predict that the transition threshold would remain the same for both conditions, as the auditory events are presented simultaneously with the visual events.

Method

The method was the same as in Experiment 2, with the following exceptions.

Participants

Ten participants (5 females, mean age of 26.5) took part in Experiment 3, all with normal or corrected-to-normal vision and normal hearing. Participants were naïve as to the purpose of the experiment. Informed consent was obtained before the start of testing. Participants were paid 8 Euros for their service.

Design and procedure

The settings were the same as in Experiment 2, except for the following differences. As most participants had made nearly 100% group motion judgments for long ISOIs in the previous experiments, the range of ISOIs was adjusted to 80–260 ms, with increasing step sizes of 30 ms. There were two experimental conditions: the audiovisual synchronous condition, in which the two sounds were presented synchronously with the two visual frame; and the unimodal condition, in which the visual stimuli were presented without sounds. In order to avoid inter-trial effects due to the presentation of sounds, the two conditions were presented blockwise, with block order counterbalanced across participants.

Results and discussion

Individual PSEs for two conditions were computed as in Experiment 1. The mean PSEs (and associated standard errors) were 164 (± 7.46) ms and 153 (± 6.78) ms for the unimodal (visual-only) and the audiovisual synchronous condition, respectively (see Fig. 5). A repeated-measures ANOVA of the TVEs revealed the main effect of condition to be significant, $F(1,9) = 10.67$, $P < 0.01$. That is,

perceptual reports were biased toward group motion in the audiovisual synchronous condition.

This finding cannot be explained by auditory temporal-marker capture, since the temporal onsets of the auditory and the visual stimuli were the same. One potentially critical factor for this finding in the audiovisual synchronous condition was the subjective auditory interval, which may have influenced the subjective visual interval. Studies of time perception have shown that auditory intervals or durations in the range of seconds are perceived as longer than the same intervals or durations in the visual modality (Goldstone and Lhamon 1974; Walker and Scott 1981). This may also be true for sub-second interval perception. Furthermore, auditory intervals may be assigned greater weights, so that they dominate the perception of audiovisual intervals (Goldstone and Goldfarb 1964a, b; Goldstone and Lhamon 1974; Wearden et al. 1998). To corroborate this hypothesis, Experiment 4 was designed to permit a direct comparison of the subjective visual intervals, auditory intervals, and audiovisual intervals in the visual Ternus paradigm.

Experiment 4

selected at random from the set [40, 70, 100, 130, 160, 190, 220] ms. The order of the two interval presentations (standard, audiovisual) was determined randomly and counterbalanced across trials. Following the presentation of both intervals, participants were prompted to make a

Table 2 Mean PSEs (\pm SE ms) and standard deviation (σ) of the estimation, the average auditory weights (w_a) as well as the predicted PSEs by the MLE models over all ten subjects for 5 interval conditions

Interval conditions	Parameters (Mean \pm SE)			
	PSE	σ	w_a	Predicted PSE
AA	129.1 (\pm 1.71)	38.8 (\pm 4.60)	–	–
VV	131.1 (\pm 2.84)	63.9 (\pm 9.53)	–	–
AV _{−40}	102.7 (\pm 1.39)	41.0 (\pm 3.89)	0.717 (\pm 0.03)	101.3
AV ₀	131.3 (\pm 1.54)	40.13 (\pm 2.71)	0.717 (\pm 0.03)	130
AV ₊₄₀	156.4 (\pm 1.67)	46.1 (\pm 5.80)	0.717 (\pm 0.03)	158.7

are listed in Table 2. The mean weight of the auditory interval (w_a) was 0.717, which was substantially higher than that for the visual interval ($w_v = 0.283$). More interestingly, the predicted PSEs were very close to the PSEs estimated from the empirical data.

The results of Experiment 4b go beyond the merely qualitative modality precision hypothesis: using the Bayesian model, the subjective audiovisual interval could be quantitatively predicted from the weighted integration of the auditory and visual intervals. Taken together with Experiment 3, the results provide further support for the hypothesis that synchronous auditory stimuli expand the (estimated) interval between visual stimuli. Consequently, one can conclude that the audiovisual synchronous presentation shifted the visual apparent-motion percept (in Experiment 3) toward group motion, thus decreasing the transition threshold. The results could also provide a coherent explanation for the findings of Experiment 1, that is, the influence of the auditory on the visual interval (which was demonstrated in Experiment 4) can also explain the shifts of the apparent-motion threshold in Experiment 1.

A second important prediction of optimal combination of auditory and visual information is that the variance of the combined estimate will be smaller than either individual estimate (Alais and Burr 2004; Ernst and Banks 2002):

$$\sigma_{av}^2 = \frac{1}{1/\sigma_a^2 + 1/\sigma_v^2} \leq \min(\sigma_a^2, \sigma_v^2) \quad (3)$$

where σ_{av}^2 , σ_v^2 , and σ_a^2 are the variances of the combined audiovisual, visual, and auditory information. In Experiment 4, the value of $\hat{\sigma}_{av}$ predicted from Equation (3) was 32.76. However, the mean standard deviation of the audiovisual interval was 42.4 (collapsed across the AV₊₄₀, AV₀, AV_{−40} conditions, see Table 2), which was not significantly smaller than the estimate for the auditory interval ($\hat{\sigma}_a = 38.8$), $P > 0.1$ (with Bonferroni correction). This suggests that the temporal discrimination of the audiovisual interval does not fully follow the prediction of the Bayesian model, as has been argued with respect to audiovisual spatial integration under low-noise conditions (Battaglia et al. 2003). In a recent study, Burr et al. (2009) also

showed that the Bayesian approach encounters a limitation when applied to a temporal-bisection task (an interval judgment task on a central stimulus appeared to be nearer in time to the first or last stimulus). However, in contrast to our results, Burr et al. found that the Bayesian model failed to predict the weights of the auditory and visual intervals, while it worked well in predicting the variances.

General discussion

The results of the present study illustrate that the perception of time intervals is important in audiovisual temporal integration. When a sound was presented slightly before the first visual frame and a second sound slightly after the second frame of the Ternus display, the apparent-motion percept was biased toward ‘group motion’—most likely because the two sounds attracted the temporal events of the two visual stimuli. A similar auditory capture effect, though in the opposite direction, was also observed when the first sound was presented after the first visual frame and the second sound before the second frame (Experiment 1). The opposite modulations of the inner and outer sounds on visual Ternus apparent motion resemble the pattern reported in previous studies of the temporal ventriloquism effect (Freeman and Driver 2008; Getzmann 2007; Morein-Zamir et al. 2003; Scheier et al. 1999). It has been suggested that the higher temporal acuity of audition captures the temporal markers of visual events, giving rise to the temporal ventriloquism effect (Freeman and Driver 2008; Getzmann 2007; Morein-Zamir et al. 2003; Welch and Warren 1980, 1986). However, the temporal capture effects were either not present or relatively weak in conditions with single (whether preceding or trailing) sounds (Experiment 2), which is also consistent with previous studies (Bruns and Getzmann 2008; Morein-Zamir et al. 2003; Scheier et al. 1999).

This dissociation between single- and dual-sound conditions raises an interesting question, namely: what is the critical temporal factor influencing audiovisual apparent motion, the temporal markers of the auditory stimuli or the auditory time interval? Some authors have argued that audiovisual pairing is a prerequisite for temporal capture:

by providing the same paired (audiovisual) features in all events (fulfilling the ‘assumption of unity’), crossmodal temporal integration becomes possible (Getzmann 2007; Morein-Zamir et al. 2003; Welch 1999). However, besides the imbalance in features, the auditory interval is also missing in single-sound conditions. Experiments 3 and 4 revealed that it is the audiovisual time interval, rather than the temporal event markers, that critically determines the audiovisual apparent motion. Since the onsets of the auditory and visual stimuli were the same in the audiovisual synchronous condition, the temporal-marker hypothesis and the notion of audiovisual pairing alone are not sufficient to explain the threshold shift of the apparent motion. Studies of crossmodal time perception have shown that perceived inter-stimulus intervals are not equal in the various modalities. For example, auditory intervals/durations are typically perceived as longer than visual intervals/durations with intervals/durations in the range of seconds (Goldstone and Lhamon 1974; Walker and Scott 1981). A recent study of vibro-tactile and visual asynchronies reported that empty tactile intervals were also perceived as longer than visual intervals (van Erp and Werkhoven 2004). To explain the asymmetric interval perception among sensory modalities, it has been proposed that the respective internal pacemakers run at different speeds (Wearden 2006): the internal clock runs faster in modalities with higher temporal precision, so that there are more ‘clock ticks’ accumulating for stimuli defined in these modalities compared to others (with the number of clicks determining the lengths of perceived intervals). Experiment 4a revealed a similar result, namely, that the auditory interval is perceived as longer than the same visual interval at the sub-second level. Experiment 4b further suggested that the perceived audiovisual interval is integrated from the optimal weights of the auditory and visual intervals. Quantitative probability-based models have previously been found to provide a good account of crossmodal spatial integration (Alais and Burr 2004; Battaglia et al. 2003; Ernst and Bühlhoff 2004; Ernst and Banks 2002). Yet it is less clear whether these models do also apply to crossmodal temporal integration. Requiring participants to make temporal-order judgments (TOJ), Ley et al. (2009) found performance in a bimodal TOJ task, with combined auditory-tactile stimuli, to be well explained by an optimal MLE model. By contrast, using a different, temporal-bisection task, Burr et al. (2009) found that the model of optimal combination fit only roughly with their pattern of results. In contrast to their study, using an apparent-motion paradigm, we provide new evidence that humans may be able to estimate crossmodal intervals in an optimal manner, though the variances of the estimates were not predicted accurately by the Bayesian model. The discrepancies among the relevant studies might arise from the different

paradigms used. However, it is also possible that prerequisites of Bayesian models, such as assumption of Gaussian noise, are not fully satisfied in crossmodal temporal judgments, as a result of which the predictions of Bayesian models are not quite accurate (Burr et al. 2009).

Besides the absence of an auditory interval, intramodal grouping was also absent in single-sound conditions (Experiment 2). A number of studies have shown that the strength of intramodal grouping can influence crossmodal integration—that is, stronger intramodal grouping may work as a segregation cue, thus reducing the crossmodal interaction (Sanabria et al. 2004; Vroomen and de Gelder 2004; Keetels et al. 2007; Spence et al. 2007; Spence 2007). Consistent with this, the present results (Experiment 2) revealed that absence of intramodal grouping can also weaken the crossmodal capture. This is in line with previous studies that have used audiovisual TOJ tasks, where there was no temporal ventriloquism effect in single-sound conditions (Morein-Zamir et al. 2003; Scheier et al. 1999). However, there is a nice exception to this generalization. A single sound, or a continuous sound, presented in between two visual events has been found to facilitate the percept of continuous motion (Bruns and Getzmann 2008; Getzmann 2007). Note that in the above studies, the single sound was located in between two visual events (the average audiovisual SOA was 88 ms), thus ‘splitting’ the visual interval into two short audiovisual intervals and making the visual events more likely to be perceived as a holistic movement. In contrast, when a single sound was presented in temporal proximity to the first or the second visual event, there was no evidence for a facilitation of the apparent-motion percept (Allen and Kolars 1981; Bruns and Getzmann 2008). In the present Experiment 2, a similar pattern was evident in the single-sound condition, with audiovisual SOAs of 30 and 0 ms. Taken together, this evidence suggests that both intramodal and crossmodal intervals play an important role in crossmodal integration.

In summary, the present study investigated auditory temporal modulation of the perception of visual Ternus displays. When two sounds were presented temporally slightly outside the two visual frames, the visual apparent motion was shifted to group motion. The opposite capture effect was seen with two sounds located in between the two visual frames. While two sounds produced larger (either positive or negative) temporal capture effects on the visual apparent-motion percept, capture effects by single sounds were relatively weak or nonexistent. Furthermore, the audiovisual synchronous Ternus display elicits more percepts of group motion compared to visual-only Ternus displays. The results from subjective interval estimation suggested that the visual *interval* was captured by the auditory *interval*. In addition, audiovisual interval estimation is well described by a Bayesian integration model with higher weights for auditory (compared to visual) intervals.

Acknowledgments This research was supported in part by a grant from the German Research Foundation (DFG) within the Collaborative Research Centre SFB 453. We thank Hans Strasburger for stimulating comments and the reviewers for insightful suggestions.

References

- Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14(3):257–262
- Allen PG, Kolars PA (1981) Sensory specificity of apparent motion. *J Exp Psychol Hum Percept Perform* 7(6):1318–1328
- Battaglia PW, Jacobs RA, Aslin RN (2003) Bayesian integration of visual and auditory signals for spatial localization. *J Opt Soc Am A Opt Image Sci Vis* 20(7):1391–1397
- Bermant RI, Welch RB (1976) Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition. *Percept Mot Skills* 42(43):487–493
- Bertelson P, Radeau M (1981) Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept Psychophys* 29(6):578–584
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10(4):433–436
- Bruns P, Getzmann S (2008) Audiovisual influences on the perception of visual apparent motion: exploring the effect of a single sound. *Acta Psychol* 129(2):273–283
- Burr D, Banks MS, Morrone MC (2009) Auditory dominance over vision in the perception of interval duration. *Exp Brain Res* 198(1):49–57
- Calvert G, Spence C, Stein BE (2004) The handbook of multisensory processes. MIT Press, Cambridge
- Dixon NF, Spitz L (1980) The detection of auditory visual desynchrony. *Perception* 9(6):719–721
- Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415(6870):429–433
- Ernst MO, Bühlhoff HH (2004) Merging the senses into a robust percept. *Trends Cogn Sci* 8(4):162–169
- Fendrich R, Corballis PM (2001) The temporal cross-capture of audition and vision. *Percept Psychophys* 63(4):719–725
- Fraisse P (1984) Perception and estimation of time. *Annu Rev Psychol* 35:1–36
- Freeman E, Driver J (2008) Direction of visual apparent motion driven by timing of a static sound. *Curr Biol* 18:1262–1266
- Getzmann S (2007) The effect of brief auditory stimuli on visual apparent motion. *Perception* 36(7):1089–1103
- Goldstone S, Goldfarb JL (1964a) Auditory and visual time judgment. *J Gen Psychol* 70:369–387
- Goldstone S, Goldfarb JL (1964b) Direct comparison of auditory and visual durations. *J Exp Psychol* 67:483–485
- Goldstone S, Lhamon WT (1974) Studies of auditory-visual differences in human time judgment. 1. Sounds are judged longer than lights. *Percept Mot Skills* 39(1):63–82
- Harrar V, Harris LR (2007) Multimodal Ternus: visual, tactile, and multisensory grouping in apparent motion. *Perception* 36:1455–1464
- Howard IP, Templeton WB (1966) Human spatial orientation. Wiley, New York
- Jaekl PM, Harris LR (2007) Auditory-visual temporal integration measured by shifts in perceived temporal location. *Neurosci Lett* 417(3):219–224
- Keetels M, Stekelenburg J, Vroomen J (2007) Auditory grouping occurs prior to intersensory pairing: evidence from temporal ventriloquism. *Exp Brain Res* 180(3):449–456
- Kramer P, Yantis S (1997) Perceptual grouping in space and time: evidence from the Ternus display. *Percept Psychophys* 59(1):87–99
- Levitin DJ, MacLean K, Mathews M, Chu L (2000) The perception of cross-modal simultaneity. *Int J Comput Anticip Syst* 323–329
- Ley I, Haggard P, Yarrow K (2009) Optimal integration of auditory and vibrotactile information for judgments of temporal order. *J Exp Psychol Hum Percept Perform* 35(4):1005–1019
- Morein-Zamir S, Soto-Faraco S, Kingstone A (2003) Auditory capture of vision: examining temporal ventriloquism. *Cogn Brain Res* 17(1):154–163
- Pantle AJ, Petersik JT (1980) Effects of spatial parameters on the perceptual organization of a bistable motion display. *Percept Psychophys* 27(4):307–312
- Pantle AJ, Picciano L (1976) A multistable movement display: evidence for two separate motion systems in human vision. *Science* 193(4252):500–502
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442
- Petersik JT, Rice CM (2006) The evolution of explanations of a perceptual phenomenon: a case history using the Ternus effect. *Perception* 35(6):807–821
- Pick HL, Warren DH, Hay JC (1969) Sensory conflict in judgments of spatial direction. *Percept Psychophys* 6(4):203–205
- Posner MI, Nissen MJ, Klein RM (1976) Visual dominance: an information-processing account of its origins and significance. *Psychol Rev* 83(2):157–171
- Radeau M, Bertelson P (1987) Auditory-visual interaction, the timing of inputs. Thomas (1941) revisited. *Psychol Res* 49(1):17–22
- Sanabria D, Soto-Faraco S, Chan JS, Spence C (2004) When does visual perceptual grouping affect multisensory integration? *Cogn Affect Behav Neurosci* 4(2):218–229
- Scheier C, Nijhawan R, Shimojo S (1999) Sound alters visual temporal resolution. *Invest Ophthalmol Vis Sci* 40(Suppl. 4):4169
- Shimojo S, Scheier C, Nijhawan R, Shams L, Kamitani Y, Watanabe K (2001) Beyond perceptual modality: auditory effects on visual perception. *Acoust Sci Technol* 22(2):61–67
- Soto-Faraco S, Kingstone A (2004) Multisensory integration of dynamic information. In: Calvert G, Spence C, Stein BE (eds) The handbook of multisensory processes. MIT Press, Cambridge
- Spence C (2007) Audiovisual multisensory integration. *Acoust Sci Technol* 28:61–70
- Spence C, Sanabria D, Soto-Faraco S (2007) Intersensory Gestalten and crossmodal scene perception. In: Noguchi K (ed) Psychology of beauty and Kansei: new horizons of Gestalt perception. Fuzanbo International, Tokyo, pp 519–579
- Stone JV, Hunkin NM, Porritt J, Wood R, Keeler V, Beanland M et al (2001) When is now? Perception of simultaneity. *Proc R Soc Lond B Biol Sci* 268(1462):31–38
- Ternus J (1926) Experimentelle Untersuchungen über phänomenale Identität. *Psychologische Forschung* 7:81–136
- Treutwein B, Strasburger H (1999) Fitting the psychometric function. *Percept Psychophys* 61(1):87–106
- van Erp JBF, Werkhoven PJ (2004) Vibro-tactile and visual asynchronies: sensitivity and consistency. *Perception* 33:103–111
- Vroomen J, de Gelder B (2004) Temporal ventriloquism: sound modulates the flash-lag effect. *J Exp Psychol Hum Percept Perform* 30:513–518
- Vroomen J, Keetels M (2006) The spatial constraint in intersensory pairing: no role in temporal ventriloquism. *J Exp Psychol Hum Percept Perform* 32(4):1063–1071
- Walker JT, Scott KJ (1981) Auditory-visual conflicts in the perceived duration of lights, tones and gaps. *J Exp Psychol Hum Percept Perform* 7(6):1327–1339

- Wearden JH (2006) When do auditory/visual differences in duration judgments occur? *Q J Exp Psychol* 59(10):1709–1724
- Wearden JH, Edwards H, Fakhri M, Percival A (1998) Why “sounds are judged longer than lights”: application of a model of the internal clock in humans. *Q J Exp Psychol* 51(2):97–120
- Welch RB (1999) Meaning, attention, and the ‘unity assumption’ in the intersensory bias of spatial and temporal perceptions. In: Aschersleben G, Bachmann T, Müsseler J (eds) *Cognitive contribution to the perception of spatial and temporal events*. Elsevier, Amsterdam, pp 317–387
- Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. *Psychol Bull* 88(3):638–667
- Welch RB, Warren DH (1986) Intersensory interactions. In: Boff KR, Kaufman L, Thomas JP (eds) *Handbook of perception and human performance: sensory processes and perception*, vol 1. Wiley, New York, pp 1–36
- Witten IB, Knudsen EI (2005) Why seeing is believing: merging auditory and visual worlds. *Neuron* 48(3):489–496