Lihan Chen

Crossmodal Temporal Capture in Visual and Tactile Apparent Motion:

Influences of temporal structure and crossmodal grouping

München 2009

Crossmodal Temporal Capture in Visual and Tactile Apparent Motion:

Influences of temporal structure and crossmodal grouping

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie
an der Ludwig-Maximilians-Universität
München

vorgelegt von

Lihan Chen

aus Zhejiang, China

im Mai 2009

Referent: Prof.Dr. Hermann Müller

Koreferent: Prof.Dr. Torsten Schubert

Tag der mündlichen Prüfung: 7. Juli, 2009

Table of Contents

Chapter 1 Introduction

1.1 Crossmodal integration

Sensory modalities are generally distinguished and organized on the basis of physical stimulation-light for vision, sound for hearing and skin pressure and friction for touch etc. Previous research has generally considered each sensory modality in isoaltion. However, most of our life experiences stem from acquiring information from different sensory modalties.The recent twenty years have witnessed a burgeon in the research of crossmodal interaction, in which the interpretation of data in one sensory modality are influenced by the data that acquired in another modality(Calvert, Spence et al. 2004; Spence, Senkowski et al. 2009). Today, there is a growing large body of research using behavioural, electrophysiological and neuroimaging techniques, as well as data from patients and mathematical modelling approaches to describe the principles of multisensory integration in both animals and humans (Driver and Noesselt 2008; Stein and Stanford 2008; Goebel and van Atteveldt 2009; Stein, Stanford et al. 2009). Studies on crossmodal integration have mainly focused on the following areas: principle mechanisms such as crossmodal grouping in crossmodal processing (Spence, Sanabria et al. 2007; Holmes 2009), the relations between crossmodal processing and perception and recent interests in the roles of attention as well as learning and memory issues in crossmodal integration(Shams, Kamitani et al. 2000). And, noticeably, there is an ongoing trend for the researchers to shift their attention from the spatial alignment and spatial representation across different coordinate frames (Hotting, Rosler et al. 2004; Spence and Driver 2004) to temporal alignment in crossmodal integration context (Getzmann 2007; Freeman and Driver 2008; Cook and Van Valkenburg 2009; Navarra, Hartcher-O'Brien et al. 2009).

Phenomenologically, crossmodal interaction has been demonstrated in a number of ways, such as the *perceived order of two events* (Scheier, Nijhawan et al. 1999; Morein-Zamir 2003; Getzmann 2007; Keetels 2007), *the subjective dislocalization* (Caclin, Supérieure et al. 2002), and the *perceived numbers of the events* (Shams 2000; Bresciani 2005; Bresciani 2007). Most of the studies have focused on spatial interactions of the intermodal conflict and on identity interactions. One classical and interesting phenomenon termed as the ‘ventriloquism effect’ has been extensively reported (Vroomen and de Gelder 2000; Aschersleben and Bertelson 2003; Bertelson and Aschersleben 2003; Vroomen and Keetels 2006) , for example, in a typical spatial ventriloquism effect, the apparent location of target sounds is displaced towards light flashes delivered simultaneously at some distance(Howard and Templeton 1966). In McGurk effect, what is being heard is influenced by what is being seen (for example, when hearing /ba/ but seeing the speaker say /ga/ the final perception may be /da/)(McGurk and MacDonald 1976). Ventriloquism has also been demonstrated with "crossmodal dynamic capture" in spatial apparent motion. A number of evidences have shown that the apparent motion in a certain modality can be influenced by static or dynamic events from another modality (Sekuler, Sekuler et al. 1997; Soto-Faraco, Lyons et al. 2002; Soto-Faraco, Spence et al. 2004). For example, the direction of auditory motion can be "captured" by the conflicting direction of visual motion, while the direction of visual motion is not affected by the incongruent auditory motion (Soto-Faraco and Kingstone 2004). Such phenomenon have been termed as "crossmodal dynamic capture", which has been well demonstrated across audition, touch and vision (Soto-Faraco, Lyons et al. 2002; Soto-Faraco and Kingstone 2004; Soto-Faraco, Spence et al. 2004; Sanabria, Soto-Faraco et al. 2005; Lyons, Sanabria et al. 2006; Soto-Faraco, Kingstone et al. 2006; Sanabria, Spence et al. 2007). Most of those studies have centered on the spatial interactions between two modalities with an "immediate response" approach, whereby participants were required to make a quick "motion direction" judgment on one of the conflicting motion streams in two modalities.

An essential point regarding the mechanism of the crossmodal interactions is that their various manifestations are strongly dependent on the *timing* of inputs. In most of the apparent motion dynamic capture paradigms adopted by Soto Faraco (2002, 2004), the asynchronous temporal distance between stimuli from two modalities was fixed at 500ms. Generally, this relative large temporal disparity imposed no capture effect in both 'congruent' and 'conflicting' conditions (Soto-Faraco, Lyons et al. 2002; Soto-Faraco, Spence et al. 2004)

In visual-audio interaction, a *window* of synchrony between auditory and visual events is crucial to spatial ventriloquism, as the effect disappear when the audio-visual asynchrony exceeds approximately 300ms (Slutsky and Recanzone 2001). The McGurk effect (a visual /ga/ combined with an audio /ba/ is often heard as /da/) also fails to occur when the audio-visual asynchrony exceeds 200-300ms (Massaro, Cohen et al. 1996; Munhull, Gribble et al. 1996). The temporal integration occurs more effectively if the visual stimulus comes first. Jaekl and Harris (2007) used two ranges of timescales to investigate the auditory-visual temporal integration and measured the shifts of the perceived temporal location (compared percentage of first interval duration and the shift of the perceived midpoint). One audio was inserted in between two visual events or one visual stimulus in between two audios to create two intervals. In the long timescale (600ms), the separations between the stimuli were too great for significant temporal integration, the point of subjective equality (PSE) of the two intervals lengths was consistent with the different processing latencies while in a short timescale (125ms), additional shifting of the PSE occurred, the stimuli were close enough together to facilitate temporal integration in addition to the latency effect. Jaekl and Harris (2007) results indicate that temporal integration shifts the perceived timing of a visual stimulus by an amount much larger than can be explained differential processing latencies(Jaekl and Harris 2007).

Actually, according to Wallace et al (1996), combinations of stimuli could have very different consequences in the same neuron in superior colliculus, depending on their temporal and spatial relationships, generally, multisensory interactions were evident when pairs of stimuli were separated from one another by <500ms, and the products of these interactions far exceeded the sum of their unimodal components(Wallace, Wilkinson et al. 1996).

Multisensory interactions between dynamic events appear to be *asymmetrical* phenomena. The fact that crossmodal dynamic capture occurs for a pair of modalities in one direction (e.g. vision captures audition) does not necessarily mean that the capture will occur in the other direction (e.g, in spatial domain, audition does not easily capture vision). According to Soto-Faraco et al (2003), with the dynamic spatial crossmodal capture, auditory apparent motion was strongly influenced by visual apparent motion (46%) and by tactile apparent motion (36%); tactile apparent motion was modulated by visual apparent motion (44%) and, to a smaller degree, by auditory apparent motion (15%); and finally, visual apparent motion was uninfluenced by auditory or tactile apparent motion. The capture effects always occurred in the synchronous condition but not in the control asynchronous condition, indicating that participants were not confused about which target modality demanded a response. Two complementary aspects have been offered to account this asymmetry. (1) Crossmodal weighting account: the perceputal system is organized so that input from one modality will receive a higher weighting in the integration process than input from another modality (Soto-Faraco, Kingstone et al. 2003). Also, the modality dominance may depend on what modality produces the most accurate information for the particular task being performed (Welch and Warren 1980). (2) Stimulus intensity variation within modalities. In one modality, if input presented at a high intensity or has a better 'quality' of information than input of weak intensity, the integration process may be biased by the more salient input. In this case, asymmetrical would be the consequence of a particular set of stimulus values rather than a built-in feature of the processing system.

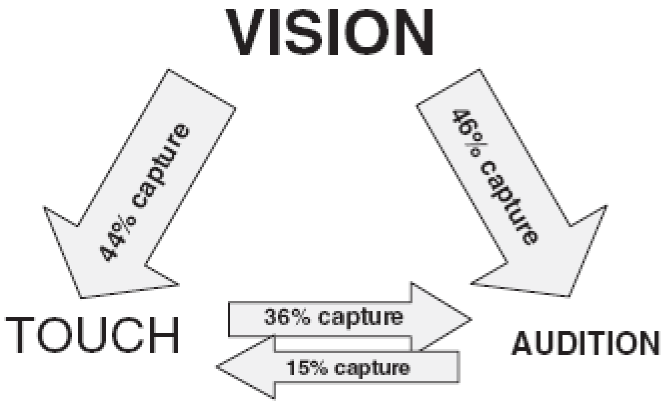


Figure 1.1 Schematic representation of the results of a study from Soto-Faraco, Kingstone,& Spence (2003), in which every possible pairing of target and distractor modality (across audition, vision and touch) was evaluated by asking participants to judge the direction of a moving stimulus in the target sensory modality while attempting to ignore a moving stimuli in the distractor modality.

In the mutisensory integration studies, several *hypotheses* have been put forward to account the findings and most of them are given to account the crossmodal capture in spatial domain. Three most commonly explanations of intersensory bias (capture) effect have been provided. (1) *Modality precision hypothesis* (Howard and Templeton 1966; Fisher 1968; Kaufman 1974; Choe, Welch et al. 1975). When two sensory modalities provide discrepant information about some characteristic of an event, the resolution of the discrepancy will favor the modality that is more precise of the two in registering that event. (2) *The modality appropriateness hypothesis* (Welch and Warren 1980; Welch, DuttonHurt et al. 1986). The different sensory modalities are not equivalently suited for the perception of a given event. For example, vision is suited to spatial and audition to temporal information processing. (3) *Bayesian integration models (Cue integration)* (Ernst and Banks 2002; Witten and Knudsen 2005) . Besides classical ventriloquism introduced above, in which vision ‘captures’ sound, the reverse effect has also been observed: when visual localization was made more difficult, sound became dominant and attracted the visual stimuli (Alais and Burr 2004). The results seem to be not well explained by one sense capturing the other, but by a simple model of optimal combination of visual and auditory information. Thus, irrespective of the sensory modality, the more reliable stimulus appears to dominate the less reliable stimulus. It was proposed that central nervous system integrates information in a statistically optimal fashion, this assumption was supported by recent audio-visual/visual-tactile integration experiments (Ernst and Banks 2002; Witten and Knudsen 2005). This optimal fashion is implemented by a Bayesian calculation procedure.

A hot debate as regards with the capture or multisensory integration was focused on the *mechanism of 'integration'*, genuine perceptual nature or postperceptual adjustment (response bias). Some investigators have stressed a decisional component as the underlying mechanism for crossmodal dynamic congruency effect (Meyer and Wuerger 2001) response interference (from the ignored modality) whereas others have argued that the effects can occur on earlier (perceptual) stages (Kitagawa and Ichihara 2002; Soto-Faraco, Navarra et al. 2004) . The crux to solve this debate is to create a non-transparent experimental condition, whereby the participants have no conscious experience of the seemingly disparity /conflict between crossmodal events, if in this case, crossmodal integration effect is still observed, then we may expect a dominant genuine perceptual processing (Bertelson and Radeau 1981; Vroomen and De Gelder 2004). Most of the ventriloquism studies have supported the 'perceptual' account, those studies based the perceptual integration on the following points: (1) sounds were spatially uninformative and neutral with respect to the response alternative; (2) randomized conditions and SOAs between stimuli; (3) context-dependent expectations were minimized, since the experiment only contains two simple events(Fendrich 2001; Bertelson 2003; Morein 2003; Vroomen 2004). Bertelson and Aschersleben (1998; also, Caclin et al,2002) demonstrated that the perceptual basis of ventriloquism can be addressed using a double psychophysical staircase methodology. The key manipulation in staircase procedure is the irrelevant stimulus dimensions are neutral with respect to the available responses and the intersensory conflict stimulation was not transparent. The staircase experiments demonstrated that cross-modal dynamic capture occurs in an experimental context that excludes response competition and potential top-down influences by an awareness of cross-modal conflict. These data provide strong support for the existence of early perceptual integration between auditory and visual motion signals.

Recent event-related potential technique and functional magnetic resonance imaging approaches have been employed to tackle the underlying processing in crossmodal integration. Due to their excellent temporal resolution, event-related potentials (ERPs) may provide a very helpful tool to assess the level at which the ventriloquism effect occurs. Mismatch negativity (MMN), which indexes the automatic detection of a deviant stimulus rarely occurring in a sequence of standard background stimuli, is indeed assumed to be elicited at a pre-attentive level (Carles 2007). The logic of using MMN in (spatial) ventriloquism is as follows: if ventriloquism effect prevents the spatial localization difference between the deviant and the standard sounds from being perceived, the generation of an MMN could be inhibited. On the contrary, if the spatial separation between the deviant sound and the visual signal is too large to elicit a ventriloquism effect, the MMN evoked by the auditory source discrepancy should be preserved. With the recording of MMN, Colin et al (2002) found the MMN evoked by spatial separation contrasts (20 and 60 degrees) in the auditory alone condition was suppressed by the corresponding audiovisual condition only when the latter yielded a ventriloquism effect, suggesting that this effect occurs at an early perceptual level (Colin, Radeau et al. 2002). Stekelenburg, Vroomen and de Gelder (2004) used an ‘oddball’ paradigm whereby simultaneously presented sounds and flashes coming from the same location served as standard. The deviant consisted of a sound originating from the same source as the standard together with a flash at 20 degrees spatial separation, which evoked an illusory sound shift. This illusory sound shift evoked an MMN closely resembling the MMN evoked by an actual sound shift. And visual-only control condition ruled out that the illusory-evoked MMN was confounded by the visual part of the audiovisual deviant. These results indicate that the crossmodal interaction on which the ventriloquist illusion is based takes place automatically at an early processing stage, within 200 ms after stimulus onset (Stekelenburg, Vroomen et al. 2004). With the similar reasoning, Stekelenburg and Vroomen (2009) used apparent visual motion paradigm, in an audiovisual condition, auditory standards and deviants were synchronized with a visual stimulus that moved in the same direction as the auditory standards, the audiovisual deviants did not evoke an MMN, the ventriloquism from visual motion reduced the perceptual difference between sound motion of standards and deviants. In this case, the auditory and visual motion signals are integrated at early sensory processing stages (Stekelenburg and Vroomen 2009).

Alink et al (2008) used fMRI to investigate whether early visual and auditory motion areas are involved in the generation of the cross-modal dynamic capture (CDC) illusion, to do this they functionally defined ROIs (regions of interest) for the visual motion area hMT/V5+ and the auditory motion complex (AMC) and compared the trials in which CDC occurred to those in which it did not, and to see whether AMC and the visual motion area hMT/V5+ were affected by this illusion. The results show that the CDC illusion is preceded by an enhanced activation that is most dominantly present in the ventral intraparietal sulcus, the motion coherency was found to enhance activation in bilateral hMT/V5+ as well as in an area adjacent to the right

AMC. In the early visual motion area hMT/V5+, motion information is integrated across senses and vision seems to be winning the competition between the two senses. With full-brain analysis, when motion is conflicting, it activates several areas in frontal and parietal cortices more strongly compared with coherent motion, those effects in frontoparietal cortex is related to response selection, which reflects an increased demand on decisional processes involved in preparing the motor response when motion is conflicting across senses. The results indicate audiovisual integration occurs in early motion areas but the cognitive state of subjects before stimulus onset plays a role in this illusion. Thus, fMRI study of Alink et al (2008) provides evidence for the co-existence of both perceptual and decisional components involved in audiovisual motion processing(Alink, Singer et al. 2008).

1.2 Crossmodal temporal integration-temporal ventriloquism

The perceived time of occurrence of a visual stimulus can be biased by the presentation of an irrelevant and slightly asynchronous auditory stimulus. This phenomenon has been labeled the temporal ventriloquism effect (Morein-Zamir, Soto-Faraco et al. 2003). Several recent studies have been carried out to deal with temporal ventriloquism effect in auditory/visual modality and adopted such tasks as temporal order judgment or the categorizations of apparent motion percepts, determination of the perceived apparent motion percept in the target modality-'capture' effect by distractor modality (Morein-Zamir, Soto-Faraco et al. 2003; Vroomen and Keetels 2006; Getzmann 2007; Bruns and Getzmann 2008; Freeman and Driver 2008). For the temporal order judgment task, the claim was that the sensitivity of a participant's judgments concerning the temporal order in which a pair of visual stimuli were presented is enhanced (i.e., the just noticeable difference, JND, is lower) when two auditory stimuli are presented, one shortly before the first visual stimulus and the other shortly after the second visual stimulus. The sensitivity was reduced when the two auditory stimuli are presented in between the two stimuli, with each one was in the vicinity of the two visual stimuli respectively (Shimojo, Scheier et al. 2001; Morein-Zamir, Soto-Faraco et al. 2003). Freeman and Driver (2008) found, that in a repeated two flashed visual apparent motion stream with equal inter-flash interval, auditory beeps slightly lagging or leading the flashes strongly influenced the perceived visual motion direction, although the beeps provided no spatial information. They argued that the accompanied beeps influence the perceived visual timing, due to higher temporal acuity in auditory modality than in visual modality.

Over the years, studies have shown the perceived rate of stimulation in one sensory modality can be modulated by the rate of stimulus presentation in another modality. The rate at which a rapidly alternating visual stimulus appears to be flickering can be modulated quite dramatically by changes in the rate at which a simultaneously-presented auditory stimulus is made to flutter(Gebhard and Mowbray 1959) , and the auditory flutter typically has a much greater influence on judgments of perceived visual flicker than vice versa (Shipley 1964). The auditory driving effect on tactile stimuli have also been found (Bresciani, Ernst et al. 2005; Bresciani and Ernst 2007). Since the stimuli presented in the two aforementioned audiovisual studies were close to the flick/flutter-fusion threshold, the results may just reflect the perception of a given stimulus attribute, rather than the perceptual organization itself. The convincing evidence for the existence of genuinely perceptual organization comes from the research on the two-flash illusion (Shams, Kamitani et al. 2000; Shams, Kamitani et al. 2002), whereby a typical behavioural finding is: in one-flash trials, participants reported to see two lights whenever two or more beeps were auditorily presented. This line of research appears to highlight the importance of stimulus structure in contraining the nature of crossmodal interactions, with the modality carrying more discontinuous and salient signal being the modulating modality in crossmodal interactions (Shimojo, Scheier et al. 2001) .

1.3 Perceptual grouping in crossmodal integration

To date, only a small body of research has attempted to investigate the extent to which the perceptual organization of stimuli taking place in one sensory modality can influence the perceptual organization of stimuli in another sensory modality, or the nature of any crossmodal interactions between stimuli presented in different sensory modalities. The studies we cited below clearly highlight the necessity of considering intramodal perceptual grouping when investigating the crossmodal grouping of sensory information. The problem was boiled down to the following: to what extent does human use information from one sensory modality in order to impose a certain organization on the perceptual stimuli (array) present in another sensory modality (Spence, Sanabria et al. 2007).

In cross-modal spatial capture studies, some evidence have shown intramodal perceptual grouping constrains the crossmodal binding of apparent motion from two modalities. For example, the modulation of visual motion on the direction of auditory apparent motion is reduced when the visual stimuli are embedded in a more spatiotemporally extended visual motion stream (Sanabria, Soto-Faraco et al. 2004; Sanabria, Soto-Faraco et al. 2005). Similar results have been confirmed by visual-tactile apparent motion study, in which crossmodal dynamic capture effects of the tactile apparent motion by the conflicting visual apparent motion was stronger in the 2-lights condition than in the 6-lights condition (Lyons, Sanabria et al. 2006). One common explanation is that when the intramodal perceptual grouping was set up *prior to* the crossmodal grouping, crossmodal interaction would become weaker (Spence, Sanabria et al. 2007).

In crossmodal temporal integration, similar evidences of intramodal grouping influence on crossmodal interaction have been found in temporal ventriloquism effect (Keetels, Stekelenburg et al. 2007) and in crossmodal apparent motion (Bruns and Getzmann 2008). In Keetels et al (2007) study, they examined how principles of auditory grouping relate to intersensory pairing. Two sounds that normally enhance sensitivity on a visual temporal order judgement task (temporal ventriloquism) were embedded in a sequence of flanker sounds which either had the same or different frequency, rhythm or location. In all experiments they found that temporal ventriloquism only occurred when the two capture sounds differed from the flankers, demonstrating that grouping of the sound in the auditory stream took priority over intersensory pairing (Keetels, Stekelenburg et al. 2007). Bruns and Getzmann recently used a visual motion categorization task to examine the auditory grouping and sound temporal structure on the perceived visual motion. Interestingly, they found that either a continuous sound filling the gap of two lights or a short sound intervening two lights enhanced the continuous motion, while such enhancement was absent when sound was part of a tone sequence that allowed for intramodal perceptual grouping *prior to* the multisensory integration of the audiovisual stimuli. The question therefore arises as to whether the multisensory integration would still be affected when intramodal perceptual grouping was removed or absent.

Cross-modal interactions between auditory and visual stimuli take place not only in stimuli pairs but also between auditory and visual sequences (O'Leary and Rhodes 1984; Rahne, Deike et al. 2008). O’Leary and Rhodes (1984) investigated the influence of perceptual grouping (or organization) within one sensory modality on the perceived grouping of stimuli presented in another modality. In the interactions of visual and auditory streams, they observed that participant’s subjective reports of whether they perceived one or two streams in a given modality could be influenced by whether they perceived one or two streams in the other modality at the same time. Hence, their study represents a seminal one of empirical research on the nature of crossmodal perceptual organization. Cook and Valkenburg (2009) conducted series of experiments to determine whether cross-modal interaction precedes unimodal grouping and whether the temporal ventriloquism effect could be found between grouped auditory and visual sequences. Using a repeating four tone auditory sequence (high-middle-middle- low) where the middle tones could group with either the high tone or the low tone, paired with a sequence of light flashes with a single flash one side and three the other, they demonstrated that the temporal ventriloquism effect occurs between grouped auditory and visual sequences, where participants reported that the isolated light and tone from grouped visual and auditory sequences seemed to synchronize when they were between 120 and 240ms apart, this suggests that unisensory grouping can occur prior to cross-modal interaction.(Cook and Van Valkenburg 2009)

Gilbert (1938, 1941) introduced the term 'inter-sensory Gestalten' to drescribe the crossmodal interactions. In Gilbert's initial interpretation, perceptual organization is believed to take place within each modality individually, but it does not exclude the possibility that the perceptual organization taking place in one modality can influence the perceptual organization occurring in another sensory modality(Gilbert 1938; Gilbert 1941). Allen & Kolers (1981) proposed a 'common or suprasensory organizing principle' in multiensory organization (Allen and Kolers 1981). However, empirical evidence that can be taken to support the existence of genuine intersensory Gestalten is weak (Spence, Sanabria et al. 2007) , in Spence and Sanabria et al's view, the possible existence of crossmodal (intermodal) apparent motion would be a good support of this genuine intersensory Gestalten, i.e, in crossmodal apparent motion, the patterns of crossmodal perceptual organization rely for their existence on stimulation in more than one sensory modality, but those patterns are absent in their constituent sensory modalities. Particularly, in the research outlook section, they pointed out that it might be good for investigating any crossmodal interactions between the auditory and tactile modalities, given the greater similarity in the nature of the physical signals that are transduced by these two sensory systems (Von Bekesy 1959; Hotting and Roder 2004) .

1.4 Apparent motion as a research paradigm

Apparent motion (AM) is a particularly useful probe of motion mechanisms because it describes controllable experimental situations in which nothing physically moves, yet a compelling percept of motion is generated. Also, AM is easy to create and AM is considered perceptually equivalent to real motion (Kolers 1963).

Korte published what have been called the 'laws' of apparent motion, which describes the relationship between the “primary” variable that affects the visual illusion (Korte 1915). These variables include the exposure time of each stimulus, as well as the temporal and spatial separation between the stimuli. Korte’s third law states that as this spatial separation increases, the interstimulus interval (ISI) must also increase if apparent motion is to be maintained (with the exposure time held constant). This ‘law’ has received modifications since it has been proposed. For example, Neuhaus (Neuhaus 1930) reported that apparent motion can be seen for a range of ISIs,and that as the spatial separation increases, the range of the ISIs that produce apparent motion decrease. As an extension of the third law, when the stimulus duration is lengthened, the interstimulus onset interval must also be increased in order to produce optimal conditions for apparent movement. Sherrick demonstrated that this relation can be observed between ISOI(Interstimulus onset interval) and stimulus duration with two successively presented vibrotactile stimuli(Sherrick 1968). Not only stimulus duration and ISIs determine the quality of apparent motion, the number of stimulators also play a role. In tactile AM, it was found that the frequency of judgments of continuous movement increased monotonically as the number of stimulators increased from two to eight. Longer duration(100ms) also increased the frequency of apparent movement judgment, and when long durations were combined with a large number of stimulators, good apparent movement was obtained over a entire range of ISOIs (20ms,30ms,50ms,70ms,90ms,110ms) (Kirman 1974).

Apparent motion has been observed in different sensory modalities with different corresponding physical stimuli. Early research which based on subjective reports have suggested the existence of crossmodal apparent motion between all possible combinations of auditory, visual and tactile stimuli (Galli 1932; Zapparoli 1969). Recently, Harrar et al (2008) compared the properties of apparent motion among three combinations: light-light, light-touch and touch-touch and revealed different saliency of ‘apparent motion’ percepts in the above combinations. In different modalities, apparent motion demonstrated different saliencies. Harrar et al (2008) compared the properties of apparent motion between a light and a touch with apparent motion between either two lights or two touches. Subjects rated the quality of apparent motion between each stimulus combination for a range of stimulus onset asynchronies (SOAs). Subjects reported perceiving apparent motion between all three stimulus combinations. For light-light visual apparent motion, it was consistent with Korte's third law. Touch-touch apparent motion also obeyed Korte's third law, but over a smaller range of distances, showing that proprioceptive information concerning the position of the fingers was integrated into the tactile motion system. The threshold and preferred SOAs for visuotactile apparent motion did not vary with distance, suggesting a different mechanism for multimodal apparent motion. The above studies revealed apparent motion saliency is dependent on different modalities

Of the various apparent motion paradigms, perhaps the one associated with the Ternus display is the most fascinating (Ternus 1926; Pantle and Picciano 1976) and for the underlying mechanism it is extensively explored (Pantle and Picciano 1976; Petersik and Pantle 1979; Braddick 1980; Breitmeyer and Ritter 1986; Breitmeyer and Ritter 1986; Grossberg 1991; Grossberg and Rudd 1992; Petersik and Rice 2008). The Typical Ternus phenomenon occurs in a fairly simple and straightforward apparent-movement display consisting of two elements arranged horizontally on each of two stimulus frames, often presented along with interstimulus interval (ISI) frames. The first frame of the display contains two elements (for example, dots, lines, squares or bars) equally spaced in the horizontal dimension. The second frame contains the same two elements displaced laterally by the distance separating neighboring elements, such that the positions of two of the original elements are now occupied by two different elements. Two common 'apparent motion' percepts can be obtained under different conditions.-'group motion' and 'element motion'. In group motion, the two elements shift back and forth together as a group, while in 'element motion', a percept of stationary (or flashing) 'central' elements with a single 'outside' element moving from end to end is perceived. The two movement impressions do not reflect the frank physical stimulation that impinged on the visual retina, but the consequence of the visual system's 'act' upon the external stimulus (He and Ooi 1999).

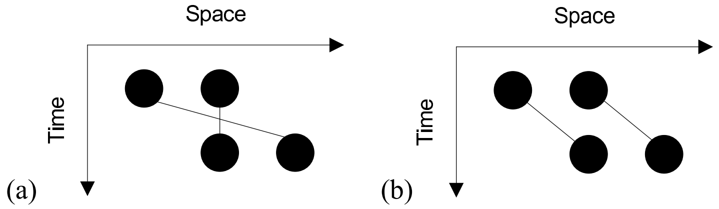


Figure 1.2. Schematic representation of the classical Ternus effect. (a) 'Element motion' percept. As illustrated, the dot which occupies the same position in two frames ('center' dot) is perceived to remain in the same location, while the 'outer' dots (the remaining two dots) are perceived to move from one location to the other. (b) 'Group motion' percept. Two dots are perceived to move together in a manner consistent with the physical displacement.

Two temporal durations of the visual frames and the interval between those two frames determine the two percepts: duration of the stimulus frames and the ISI (such as a blank dark or gray background in the monitor) between them. In general, the ISI for the transition from 'element motion' to 'group motion' is in the range of 100 to 300 ms, dependent on stimulus duration, element distance, element contrast, etc(Petersik and Pantle 1979; Kramer and Yantis

1997). Intermediate ISIs yield different percentages of element and group movement, a steady viewing of the alternating frames leads to a bistable percept (periodic changeovers from one percept to the other) but never both simultaneously.

When Ternus first put forward the classical ‘visual Ternus effect’ (Ternus 1926), He was concerned with what today ‘the correspondence problem’ (Ullman 1979) or the question how the visual system determines which points in one image correspond with their partners in another, different, image. As a Gestaltist, Ternus expressed his idea of ‘phenomenal identity’- elements in his displays rest upon the ‘meaning or role which the [part] has in the two figures’… “Phenomenal identity depends upon Gestalt identity; homologous parts in two Gestalten will exhibit phenomenal identity; phenomenal identity is a characteristic not of piecewise relationship but of whole-phenomena” (Ternus 1938) (page154). Ternus further introduced the concepts of i-points (those in the same locations across frames) versus d-points (those points that appear in different locations across frames), identity retention (when the same points appear perceptually in the same locations across frames) versus identity exchange. With a similar categorization, the idea that meaningful groups of elements are identified in individual stimulus frames has been called ‘spatial grouping effects’ by Kramer and Yantis (1997) and ‘within-frame grouping’ by He and Ooi (1999). Karmer and Yantis (1997) made clear that spatial grouping effects favor the production of group motion, whereas temporal grouping effects (‘across-frame grouping’) favor element motion (Kramer and Yantis 1997). He and Ooi conducted several experiments in which the perceived organization of the elements undergoing Ternus motion was manipulated to favor within-frame grouping or across-frame grouping (He and Ooi 1999). As an extention of the visual Ternus, Harrar and Harris (2007) first demonstrated an explicit Gestalt grouping in the sense of touch with tactile Ternus paradigm (Harrar and Harris 2007). Petersik and Rice (2008) differentiated two correspondences: spatial correspondence, which emphasizes the maintenance of similar stimulus elements in given locations over time. The other principle is relation correspondence, which emphasizes the maintenance of the inter-organization of stimulus elements across frames. An emphasis upon spatial correspondence resulted in an increase in reports of element motion compared to a featureless control condition. Emphasis on relation correspondence resulted in an increase in reports of group motion(Petersik and Rice 2008). The advantages of taking the Ternus paradigm to address the crossmodal temporal capture lie in: the two percepts in Ternus display are mutually exclusive and the ISOI for the transition from ‘element motion’ to ‘group motion’ is in a relatively broad range (Petersik and Pantle 1979; Kramer and Yantis 1997). Also, the the

Ternus paradigm is good to reveal the correspondence process between the two tokens in the Ternus display. These three points make the Ternus display a promising candidate for investigating crossmodal temporal-capture phenomena.

Chapter 2 Auditory temporal modulation of the visual Ternus effect: the influence of time interval

2.1 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 multi-element 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 1.1 and 1.3). However, such effects were not evident with single-sound configurations (Experiment 1.2). A further experiment (Experiment 1.4) 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

2.2 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, Spence et al. 2004) . However, when different modalities convey inconsistent information, perception is usually biased towards one modality – this has been referred to as modality dominance. For example, a sound source is often localized (or 'captured') towards the location of a light flash which is presented simultaneously at some distance to the sound. This effect is known as 'ventriloquism effect' (Howard and Templeton 1966; Bermant and Welch 1976; Bertelson and Radeau 1981) A wealth of literature on multisensory integration has demonstrated phenomena of modality dominance in the spatial domain, with vision being dominant in most cases (Pick, Warren et al. 1969; McGurk and MacDonald 1976; Posner, Nissen et al. 1976; Welch and Warren 1980; Welch, DuttonHurt et al. 1986; Soto-Faraco, Spence et al. 2004). . Besides the classical ventriloquism effect of vision 'capturing' sound, the reversed ventriloquism effect has also been observed when the visual location information was poor (Ernst and Banks 2002; Alais and Burr 2004). Recently, probability-based models, such as Bayesian integration models, have been proposed to provide quantitative accounts of the spatial ventriloquism effect (Ernst and Banks 2002; Battaglia, Jacobs et al. 2003; Alais and Burr 2004).

Interestingly, multisensory interactions also occur in the temporal domain. For instance, the perceived occurrence of a visual event can be biased by an irrelevant and slightly asynchronous auditory stimulus (Scheier, Nijhawan et al. 1999; Fendrich and Corballis 2001; Shimojo, Scheier et al. 2001; Morein-Zamir, Soto-Faraco et al. 2003). For example, Scheier et al. (1999) 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, Nijhawan et al. 1999). This phenomenon, referred to as 'temporal ventriloquism effect' (Morein-Zamir, Soto-Faraco 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; Welch and Warren 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 (Vroomen and Keetels 2006; Getzmann 2007; Jaekl and Harris 2007; Keetels, Stekelenburg et al. 2007; Bruns and Getzmann 2008). As there was little influence of the relative spatial positions of the auditory and visual stimuli on the temporal ventriloquism effect (Vroomen and Keetels 2006; Bruns and Getzmann 2008), 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 (Scheier, Nijhawan et al. 1999; Shimojo, Scheier et al. 2001; Morein-Zamir, Soto-Faraco et al. 2003). 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 (Welch 1999; Morein-Zamir, Soto-Faraco et al. 2003) . Similar results have also been reported for the audiovisual apparent-motion paradigm(Getzmann 2007; Bruns and Getzmann 2008) .Paired sounds induced temporal ventriloquism effects, whereas a single sound preceding or trailing the visual stimulus had no influence (Bruns and Getzmann 2008) . However, a single sound presented temporally midway between two visual stimuli can influence visual apparent motion. More recently, auditory grouping has been shown to be an important factor in audiovisual temporal interactions (Keetels, Stekelenburg et al. 2007). Using a sequence of flanker sounds which had the same or a different frequency, rhythm, or location as the two ‘capture sounds’ paired with the visual targets, Keetels et al. (2007) found the temporal ventriloquism effect to become manifest only when the paired target sounds were odd-one-out (or ‘pop-out’) stimuli.

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. However, arguably, 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, namely, 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 (Welch 1999; Morein-Zamir, Soto-Faraco et al. 2003) 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 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 the temporal ventriloquism effect even when the auditory and the visual onsets occur simultaneously.

Most studies of the temporal ventriloquism effect have used temporal order judgment (TOJ) or apparent motion tasks (Scheier, Nijhawan et al. 1999; Shimojo, Scheier et al. 2001; Morein-Zamir, Soto-Faraco et al. 2003; Getzmann 2007; Keetels, Stekelenburg et al. 2007; Bruns and Getzmann 2008; Parise and Spence 2008). Both tasks are similar, for example, in that they both present two visual events and use a short timescale of inter-stimulus onset intervals (ISOI). In TOJ tasks, the threshold for (perceived) synchrony/asynchrony between the two visual events is usually small, less than 60 ms (Brecher 1932; Elliott, Shi et al. 2007). Given this short time range, it is hard to present and manipulate sounds between two visual stimuli. Perhaps owing to this, the results obtained with such auditory intervening conditions have been mixed; for example, a temporal interaction has been reported for intervening sounds with an (inter-sound) interval of 16 ms, but not of 40 ms (Morein-Zamir, Soto-Faraco et al. 2003). Classical two-dot apparent-motion tasks, by contrast, allow for a wider range of ISOIs between the visual stimuli. Yet, there are multiple percepts of two-dot apparent motion, ranging from simultaneity, continuous apparent motion, to successive motion, etc., and participants often find it difficult to make a multi-alternative forced-choice judgment (Ekroll, Faul et al. 2008). However, there is a visual illusion associated with apparent motion known as the ‘Ternus effect’ (Ternus 1926). The typical Ternus effect 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. Dependent on the length of the ISOI (and other factors; see below), observers typically report two distinct percepts: ‘element motion’ and ‘group motion’. In element motion, the outer dots are perceived as moving, while the center dot appears to remain static or flashing; in group motion, the two dots are perceived to move together as a group (Figure 2.1). Thus, the Ternus paradigm provides a convenient means for examining the processes of establishing correspondence between the elements in the two visual frames(Ullman 1979; Dawson, Nevin-Meadows et al. 1994). Numerous studies have shown that element and group motion are never perceived simultaneously, that is, the two types of percept are mutually exclusive (Petersik and Rice 2006). In addition, when the (spatial) element configuration is fixed, the percept is mainly dependent on the ISOI: short ISOIs usually give rise to the percept of element motion, longer ISOIs to that of group motion (Pantle and Picciano 1976; Pantle and Petersik 1980; Kramer and Yantis 1997) . In general, the ISOI for the transition from element to group motion is of the order of 100 to 300 ms, dependent on stimulus duration, element distance, element contrast, etc (Petersik and Pantle 1979; Kramer and Yantis 1997). Given this, the Ternus paradigm readily lends itself to investigating crossmodal temporal interactions, because (i) the percepts associated with it are mutually exclusive and (ii) the transition threshold between the two percepts is larger than the visual synchrony/asynchrony threshold, which permits sounds to be presented in-between the two visual events.

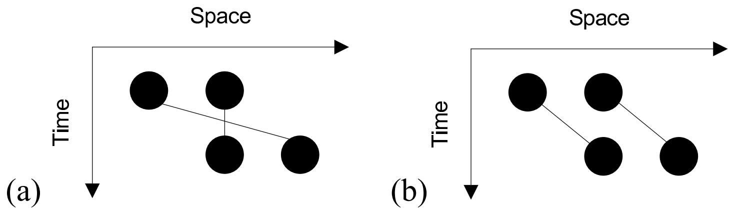


Figure 2.1. Schematic representation of the classical Ternus effect. (a) ‘Element motion’ percept. As illustrated, the dot which occupies the same position in two frames (‘center’ dot) is perceived to remain in the same location, while the ‘outer’ dots (the remaining two dots) are perceived to move from one location to the other. (b) ‘Group motion’ percept. Two dots are perceived to move together in a manner consistent with the physical displacement.

For these reasons, the classical Ternus display was employed in the present study to investigate how sound onset and inter-sound interval influence the perception of visual apparent motion. In Experiment 1.1, dual sounds were presented in three different audiovisual temporal arrangements. As temporal ventriloquism has been found in both TOJ and classical apparent-motion tasks, a similar interaction was expected to be observed with the visual Ternus illusion. Experiment 1.2 was designed to examine 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 1.3 and 1.4, 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.

2.3 Experiment 1.1 Dual sounds on visual Ternus apparent motion

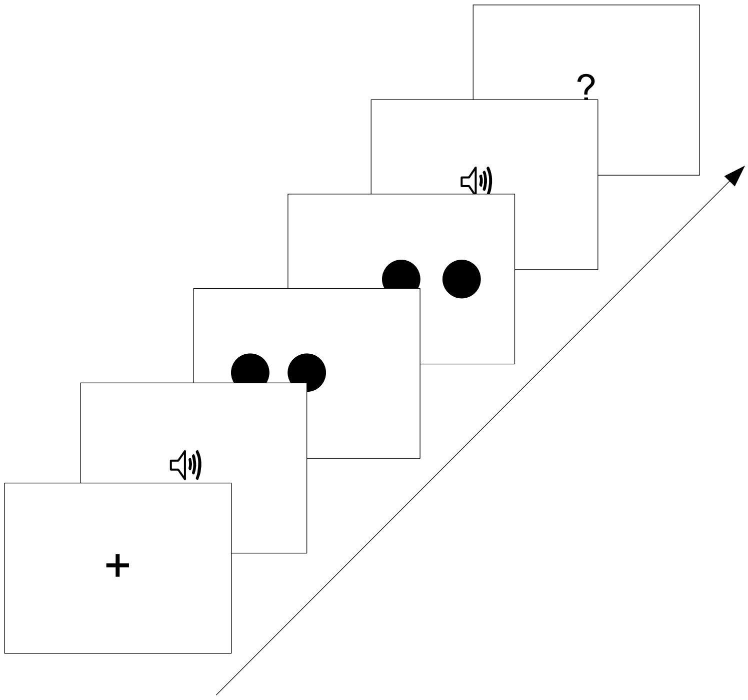
Two auditory clicks presented close in time to the visual events have been found to influence the sensitivity of (visual) temporal-order judgments (Scheier, Nijhawan et al. 1999; Morein-Zamir, Soto-Faraco et al. 2003) as well as the (visual) percept of continuous apparent motion (Getzmann 2007). In Experiment 1.1, 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, MacLean et al. 2000; Stone, Hunkin et al. 2001) .

2.3.1 Method

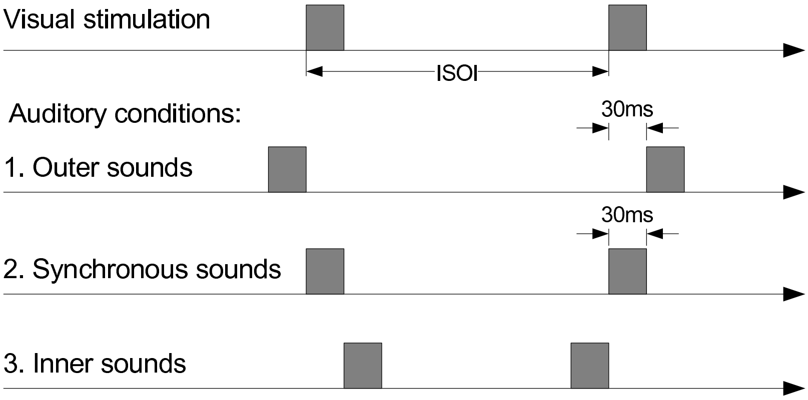
Participants. Ten participants (6 females and 4 males, mean age of 24.8 years) took part in Experiment 1.1, All had normal or corrected-to-normal vision and normal hearing. All were naïve as to the purpose of the experiment. Informed consent was obtained before the start of the experimental session.

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 1024x768 pixels, which was controlled by a PC (HP AMD Athlon 64 Dual-Core Processor) with a Radeon 1700 FSC graphics card. The computer program for controlling of the experiment was developed with Matlab (Mathworks Inc.) and the Psychophysics Toolbox (Brainard 1997; Pelli 1997). The testing cabin was dimly lit with an average ambient luminance of 0.09 cd/m2. The viewing distance was set to 57 cm, maintained by using a chin-rest. Visual stimuli consisted of two ‘stimulus frames’, each containing two black disks (1.3° of visual angle in diameter, 0.24 cd/m2 luminance) presented on a gray background (10.6 cd/m2). The separation between the two disks was 2° of visual angle. The two frames shared one element location at the center of the monitor, but contained two other elements located at horizontally opposite positions relative to the center (see Figure 1). Mono sounds (65 dB, 1000 Hz) were generated and delivered via an M-Audio card (Delta 1010) to a headset (RT-788V, RAPTOXX). To ensure accurate timing of the auditory and visual stimuli, the durations of the visual stimuli and the synchronization of the auditory and visual stimuli 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 to 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 the right of the shared element, was always opposite to the outer element of the second frame (Figure 2.2a). Each visual frame was accompanied by one brief, 30-ms sound. There were three different conditions of audiovisual interval (Figure 2.2b): 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 1000-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 to 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 1.1 implemented a 3 (audiovisual intervals) x 9 (ISOIs) within-subjects design. Each configuration was presented 24 times, with the directions of apparent motion balanced across the 24 trial instances.



(a)

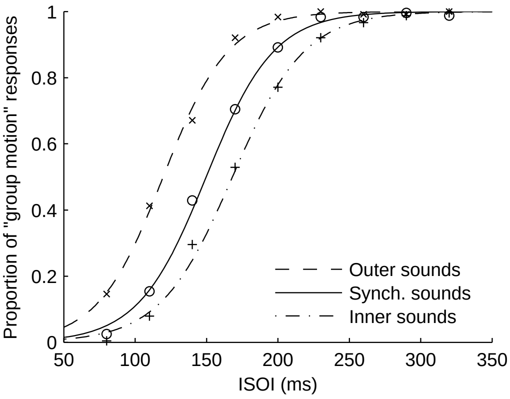


(b)

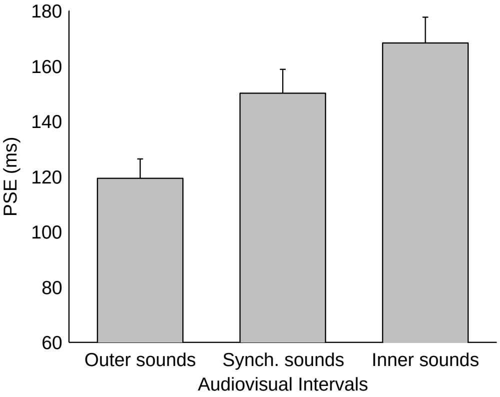
Figure 2.2 (a) Schematic illustration of the events presented on one trial. The example illustrates the condition of 'outer sounds', with the first sound presented before the first visual frame and the second sound after the second visual frame. (b) Illustration of the temporal succession of events for the three different conditions. In the 'outer-sounds' condition, the first sound was presented 30 ms before the onset of the first visual frame and the second sound 30 ms after the onset of the second visual frame. In the 'synchronous-sounds' condition, two sounds were presented simultaneously with the visual frames. In the 'inner-sounds' condition, the first sound was presented 30 ms after the onset of the first visual frame and the second sound 30 ms before the onset of the second visual frame.

2.3.2 Results and discussion

The proportion of 'group motion' reports were 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% threshold from the (fitted) logistic function (Treutwein and Strasburger 1999).



(a)



(b)

Figure 2.3. (a) Psychometric curves fitted for the data from all participants in Experiment 1.1. The solid curve and circles represent the synchronous-sounds condition, the dashed curve and crosses the outer-sounds condition, and the dotted curve and pluses the inner-sounds condition. (b) Mean PSEs (and associated standard errors) for three conditions of audiovisual intervals.

Figure 2.3 presents the results of Experiment 1.1 The mean PSEs are given in Table 1.1 A repeated-measures analysis of variance (ANOVA) of the estimated PSEs, with audiovisual interval (outer sounds, synchronous sounds, inner sounds) as factor, revealed the main effect to be significant, F(2, 18)=91.2, p<0.001 – that is, the temporal configuration of two sounds influenced the visual Ternus effect (see Figure 1.3b). Separate paired t-tests (with Bonferroni correction) confirmed the PSE to be significantly smaller for outer sounds compared to synchronous sounds, p<0.01, and to be significantly larger for inner sounds compared to synchronous sounds, p<0.01. In other words, participants were more likely to make a ‘group motion' judgment with outer sounds compared to synchronous sounds, and less likely with inner sounds compared to synchronous sounds. Both (outer- and inner-sound) 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. These effects were observed with a short audiovisual SOA (30 ms), which was within the range of simultaneity perception (Levitin, MacLean et al. 2000; Stone, Hunkin et al. 2001) . This may be taken to suggest that the TVE effects occur at a perceptual level of processing.

Table 1.1 Transition thresholds (PSE ± SE ms) between 'element motion' and 'group motion' as a function of the audiovisual (AV) interval for Experiments 1.1 and 1.2.

AV-Interval (ms)	Transition thresholds (PSE ± 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)

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 (Scheier, Nijhawan et al. 1999; Morein-Zamir, Soto-Faraco et al. 2003; Getzmann 2007). 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 towards an effect). In another study (Morein-Zamir, Soto-Faraco 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 and Warren 1980; Welch 1999). Accordingly, auditory events are assigned high weights, and visual events low weights, in audiovisual integration (Welch and Warren 1980; Welch and Warren 1986; Morein-Zamir, Soto-Faraco et al. 2003). However, what remains unclear is whether the differential modality weighting is based on points in time (onsets / offsets) or time intervals. Experiment 1.2 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.

2.4 Experiment 1.2 Single sounds on visual Ternus apparent motion

In a previous study (Morein-Zamir, Soto-Faraco 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 1.2, we examined for any crossmodal temporal interaction effect of single sounds, which were temporally manipulated (i.e., proceeding, synchronous, or trailing) with respect to either the first or the second visual event in visual Ternus display.

2.4.1 Method

The method was the same as in Experiment 1.1, with the following exceptions.

Participants. The same ten participants who had taken part in Experiment 1.1 also participated in Experiment 1.2. Five participants performed Experiment 1.2 on day 1 and Experiment 1.1 on day 2, and vice versa for the other five participants – thus counterbalancing potential practice effects across the two experiments.

Design and procedure. Experiment 1.2 consisted of two separate sessions, hitherto referred to as Experiments 1.2a and 1.2b, respectively. Half the participants performed Experiment 1.2a first and then Experiment 1.2b, and vice versa for the other half. In Experiment 1.2a, only one sound was presented, namely, close to the first visual frame. In Experiment 1.2b, only the second sound was presented. The settings were the same as in Experiment 1.1, except that either the second sound (Experiment 1.2a) or the first sound (Experiment 1.2b) 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).

2.4.2 Results and discussion

Individual PSEs for the three audiovisual intervals were computed as in Experiment 1.1.

For Experiment 1.2a, in which the sound accompanied the first visual frame, the mean PSEs are presented in Table 1 (see also Figure 2.4a). 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 1.2b, where the sound accompanied the second visual frame – see Table 1.1 for the mean PSEs (see also Figure 2.4b). 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 PSEs did not differ reliably between the preceding-sound and synchronous-sound conditions (3.5 ms, p=0.9).

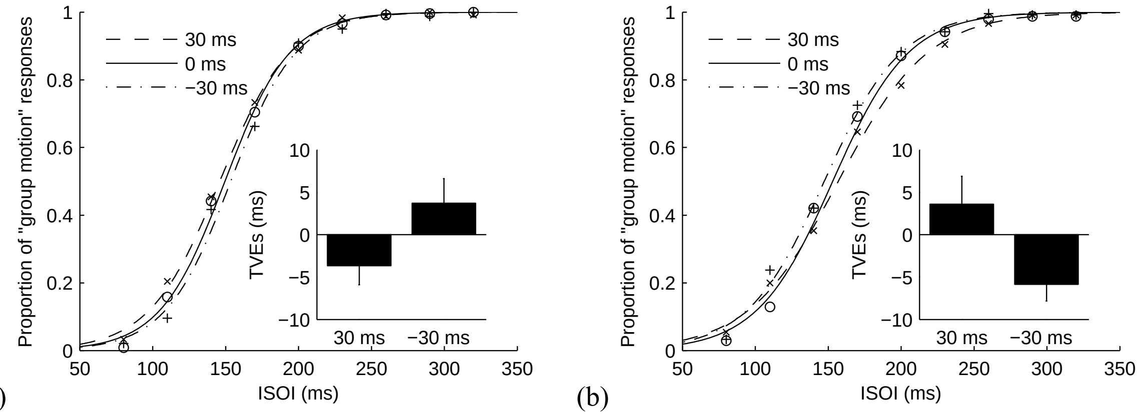


Figure 2.4. (a) Psychometric curves fitted for the data from all participants in Experiment 1.2a. The solid curve and circles represent the synchronous-sound condition (audiovisual interval 0 ms), the dashed curve and crosses the preceding-sound condition (audiovisual interval 30 ms), and the dash-dotted curve and pluses the trailing-sound condition (audiovisual interval -30 ms). The magnitude of the TVEs, calculated against the synchronous-sound baseline, are 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 1.2b. The magnitude of the TVEs, calculated against the synchronous-sound baseline, are presented in a subplot for the preceding-sound (30 ms) and trailing-sound conditions (-30 ms).

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 1.2 is consistent with Bruns and Getzmann (2008), who reported a similar pattern with temporally proximal single sounds. However, the small TVE was found for trailing sound associated with the second visual frame in Experiment 1.2b. This may rather relate to an asymmetry in audiovisual simultaneity (Dixon and Spitz 1980), namely: a sound occurring immediately after a visual stimulus is more likely to be perceived as a synchronous event than a visual stimulus occurring after a sound. Thus, in Experiment 1.2b, the sound trailing the second visual event may have biased the perceived time of this visual event.

In any case, compared to the TVEs obtained with the two-sound configurations in Experiment 1.1, 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 1.1, all settings were the same except that one sound was removed. In addition, in Experiment 1.2, 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 1.1. 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 1.1. This suggests that the effects obtained with single sounds in Experiment 1.2 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’ (Welch 1999; Morein-Zamir, Soto-Faraco et al. 2003). 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 considered by the observers to be signalling 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 1.2 could be the lack of an inter-sound interval (with single-sound configurations).

2.5 Experiment 1.3 Synchronous sounds on visual Ternus apparent motion

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 1.3, synchronous audiovisual events (with sound onset occurring simultaneously with visual frame onset) were used. The rational 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, Edwards 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.

2.5.1 Method

The method was the same as in Experiment 1.2, with the following exceptions.

Participants. Ten participants (5 females, mean age of 26.5 years) took part in Experiment 1.3, 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 1.2, 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.

2.5.2 Results and discussion

Individual PSEs for two conditions were computed as in Experiment 1.1. 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 Figure 2.5). 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 towards 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; Goldstone and Goldfarb 1964b; Goldstone and Lhamon 1974; Walker and Scott 1981; Wearden, Edwards et al. 1998). – To corroborate this hypothesis, Experiment 1.4 was designed to permit a direct comparison of the subjective visual intervals, auditory intervals, and audiovisual intervals in the visual Ternus paradigm.

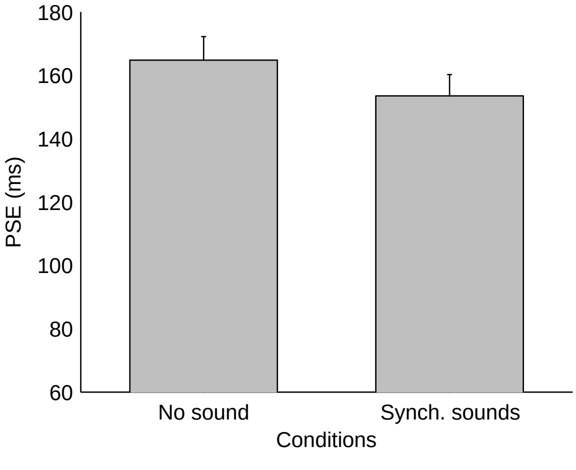


Figure 2.5. Mean PSEs (and associated standard errors) for the visual-only (i.e., no sound) condition and the audiovisual synchronous condition.

2.6 Experiment 1.4 Auditory and visual interval estimations

In Experiment 1.4, we estimated and compared the subjective intervals (near the thresholds obtained in the preceding experiments) between two Ternus display frames with and without synchronous auditory stimuli (Experiment 1.4a). In addition, we examined the variances of the interval estimates for the auditory, the visual, and the audio-visual intervals, so as to be able to quantitatively examine the influence of the auditory interval on the perception of the audio-visual interval (Experiment 1.4b). If audiovisual temporal integration, like the spatial integration, operates by optimally integrating the independent interval estimates from each modality , the audio-visual interval would be predictable by Bayesian maximum-likelihood estimation (MLE) ((Ernst and Banks 2002; Battaglia, Jacobs et al. 2003; Alais and Burr 2004; Burr, Banks et al. 2009). On the MLE model, the ‘optimal’ weighting scheme for crossmodal integration is based on the variability of the unimodal estimates:

$$\hat{I}_{av} = \omega_a \hat{I}_a + \omega_v \hat{I}_v, \qquad (1)$$

where $\hat{I}_a$, $\hat{I}_v$, and $\hat{I}_{av}$ are the perceived auditory, visual, and audio-visual intervals, ωa and ωv are the weights of unimodal auditory and visual intervals, respectively. The weights are inversely proportional to the variances σ_a^2 and σ_v^2 of the auditory and visual intervals:

$$\omega_a = \frac{1/\sigma_a^2}{1/\sigma_a^2 + 1/\sigma_v^2} \text{ and } \omega_v = 1 - \omega_a \qquad (2)$$

The estimate of $\hat{I}_{av}$ is optimal since it has the greatest reliability

2.6.1 Method

The method was the same as in Experiment 1.1, with the following exceptions:

Participants. Ten participants took part in Experiment 1.4a (5 females, mean age of 25.4) and ten in Experiment 1.4b (6 females, mean age of 25.9), all with normal or corrected-to-normal vision and normal hearing. Participants were naïve as to the purpose of the study. Payment for participating was 8 Euros.

Design and procedure. In Experiment 1.4a, the visual Ternus display was adopted for visual interval presentation. Prior to the experiment, participants performed one block of trials for practice. A trial started with a fixation cross presented at the center of the display for 300–500 ms. After that, a blank display was shown for a random duration of 300 to 500 ms. This was then followed by the successive presentation of two types of interval (with a random gap of 1–1.2 s between the presentations): (i) One was the standard interval in-between the two successive visual frames (presented without auditory stimuli). The inter-stimulus interval (ISI) was fixed at 130 ms, and both frames were presented for 30 ms. (ii) The other was the comparison interval which was created by two 30-ms ‘frames’ of audio-visual synchronous stimuli. The ISI between the frames was 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 counter-balanced across trials. Following the presentation of both intervals, participants were prompt to make a judgment of which of the two intervals was longer. Participants pressed the left or right buttons to indicate the first or, respectively, the second interval was the longer one. Overall, the (within-subject design) experiment consisted of 336 trials: comparison interval (7 levels) x 48 trials with counterbalanced order of interval presentation (2 levels) and direction of (apparent) motion (2 levels).

In Experiment 1.4b, the variances associated with three different interval estimates were examined: (i) visual interval (V), realized in terms of the classical visual Ternus display (without sounds; see Experiment 1.4a above); (ii) auditory interval (A), created by two 500-Hz tones of 30-ms duration; (3) audio-visual interval (AV), formed by the audio-visual Ternus display (see Experiment 4a above). Unlike Experiment 1.4a, for each condition, the same type of interval (i.e., visual or auditory or audio-visual) was used for both the standard stimulus and the comparison stimulus. Similar to Experiment 1.4a, in the comparison stimulus, a random ISI of 40, 70, 100, 130, 160, 190, or 220 ms was inserted between two visual frames (V condition), two sounds (A condition), and two audiovisual synchronous stimuli (AV condition), respectively. The standard stimuli for the auditory and visual conditions involved the same interval, of 130 ms. In the audiovisual condition, there were three standard audiovisual stimuli which had a similar temporal structure to those illustrated in Figure 2.2(b) for Experiment 1.1: two brief sounds flanking (outer-sounds, inner-sounds) or synchronizing with the two visual frames. The ISI between visual frames was fixed to 130 ms, but the auditory intervals were 170, 130, and 90 ms for the outer-sound, synchronous-sound and inner-sound conditions, respectively. These conditions will hitherto be abbreviated as AV+40 , AV0 , AV-40 (where the number subscript gives the extension/contraction of the interval time relative to the standard interval of 130 ms). In all other respects, the procedure was the same as in Experiment 1.4a. Experiment 1.4b implemented a within-subject design: the five types of interval (i.e., A, V, AV+40 , AV0 , AV-40) were presented in separate trial blocks (with randomized block order), with 7 levels of comparison interval presented randomly within blocks, with 24 repetitions for each condition.

2.6.2 Results and discussion

In Experiment 1.4a, the PSEs were calculated individually from the estimated psychometric curves. The mean PSE (±SE) was 103 (±10.9) ms, which means that an interval of

103 ms in the audiovisual synchronous condition was perceived as equivalent to the visual interval of 130 ms. A t-test comparing the mean PSE to 130 ms revealed the difference to be significant, $t(9) = -2.59$, $p<0.05$. Thus, with audiovisual synchronous presentation, the interval was perceived as longer than the same physical visual-only interval, due to the auditory interval expanding the (synchronous) visual interval. This result is consistent with previous comparisons of audio-visual intervals (Goldstone and Lhamon 1974; Walker and Scott 1981; Wearden, Edwards et al. 1998).

For Experiment 1.4b, a psychometric function based on the cumulative normal distribution was fitted for each interval condition (i.e., A, V, AV+40 , AV0 , AV-40) within each participant, and the PSEs and the standard deviations of the distribution (σ) were calculated from the thresholds of 50% and 84.1% (see details in Alais & Burr, 2004; Battaglia et al., 2003; Ernst & Banks, 2002; Witten & Knudsen, 2005). The overall psychometric curves are presented in Figure 2.6. The mean PSEs and mean standard deviations of the distributions for the five different intervals are listed in Table 1.2.

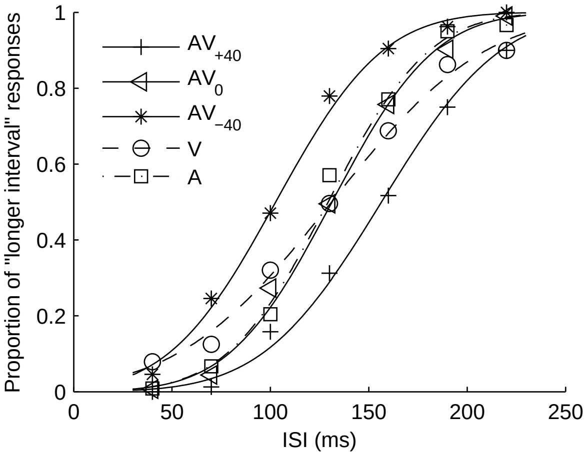


Figure 2.6. Psychometric curves fitted for the data from all participants in Experiment 4b. The solid curves with stars, triangles, and pluses represent the AV-40, AV0, and AV+40 audiovisual interval conditions, respectively; the dashed curve and circles the visual interval condition (V); and the dash-dotted curve and squares the auditory interval condition (A).

Separate t-tests comparing the PSEs relative to standard interval of 130 ms revealed the PSEs to differ significantly from 130 ms only in the AV-40 and the AV+40 conditions, $t(9)=-20.67$, $p<0.01$, and $t(9)=16.72$, $p<0.01$, respectively. This means that the audiovisual subjective interval was influenced by the auditory interval. To further quantify this interaction, the weights associated with the auditory interval (modality) were calculated according to Equation (2) and the PSEs predicted by the MLE model were calculated according to Equation (1). These values are listed in Table 2. The mean weight of the auditory interval (ωa) was 0.717, which was substantially higher than that for the visual interval (ωv = 0.283). More interestingly, the predicted PSEs were very close to the PSEs estimated from the empirical data.

The results of Experiment 1.4b 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 1.3, 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 1.3) towards group motion, thus decreasing the transition threshold. The results could also provide a coherent explanation for the findings of Experiment 1.1, that is, the influence of the auditory on the visual interval (which was demonstrated in Experiment 1.4) can also explain the shifts of the apparent-motion threshold in Experiment 1.1.

Table 1.2. Mean PSEs (± SE ms) and standard deviation (σ) of the estimation, the average auditory weights (wa) as well as the predicted PSEs by the MLE models over all ten subjects for 5 interval conditions.

Interval Conditions	Parameters (Mean ± SE)			
	PSE	σ	w_a	Predicted PSE
AA	129.1 (±1.71)	38.8 (±4.60)	-	-
VV	131.1 (±2.84)	63.9 (±9.53)	-	-
AV_{-40}	102.7 (±1.39)	41.0 (±3.89)	0.717 (±0.03)	101.3
AV_0	131.3 (±1.54)	40.13 (±2.71)	0.717 (±0.03)	130
AV_{+40}	156.4 (±1.67)	46.1 (±5.80)	0.717 (±0.03)	158.7

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 (Ernst and Banks 2002; Alais and Burr 2004).

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

Where σ_{av}^2, σ_v^2 and σ_a^2 are the variances of the combined audiovisual, visual and auditory information. In Experiment 1.4, 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, Jacobs 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(Burr, Banks et al. 2009).

2.7 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 towards ‘group motion’ – most likely because the two sounds attracted the temporal events of 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 1.1). 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 ((Scheier, Nijhawan et al. 1999; Morein-Zamir, Soto-Faraco et al. 2003; Getzmann 2007). 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 (Welch and Warren 1980; Welch and Warren 1986; Morein-Zamir, Soto-Faraco et al. 2003; Getzmann 2007). However, the temporal capture effects were either not present or relatively weak in conditions with single (whether preceding or trailing) sounds (Experiment 1.2), which is also consistent with previous studies (Scheier, Nijhawan et al. 1999; Morein-Zamir, Soto-Faraco et al. 2003; Bruns and Getzmann 2008).

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 (Welch 1999; Morein-Zamir, Soto-Faraco et al. 2003; Getzmann 2007). However, besides the imbalance in features, the auditory interval is also missing in single-sound conditions. Experiments 1.3 and 1.4 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 1.4a revealed a similar result, namely, that the auditory interval is perceived as longer than the same visual interval at the sub-second level. Experiment 1.4b 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 (Ernst and Banks 2002; Battaglia, Jacobs et al. 2003; Alais and Burr 2004; Ernst and Bülthoff 2004). Yet it is less clear whether these models do also apply to crossmodal temporal integration. Using a temporal bisection task, Burr et al. (2009) found that the model of optimal combination fit only roughly with their pattern of results (Burr, Banks et al. 2009). In contrast to their study, using an apparent-motion paradigm, we provide new evidence that Bayesian model can well predict the weights for crossmodal intervals, though the predictions of the variances were not accurate. This discrepancy might arise from the different paradigms used (temporal-bisection task vs. apparent-motion judgment task). However, it is also possible that prerequisites of Bayesian models, such as assumption of Gaussian noise, are not fully satisfied in crossmoal temporal judgments, as a result of which the predictions of Bayesian models are not quite accurate (Burr, Banks et al. 2009) .

The auditory interval has a higher weight in audiovisual interval perception, which would explain the different results between dual-sound (Experiment 1.1) and single-sound (Experiment 1.2) configurations. The major difference is the absence of an auditory interval in single-sound conditions. This could explain that there is no evidence of a temporal capture effect of single sounds on dual visual events, such as in the TOJ task (Scheier, Nijhawan et al. 1999; Morein-Zamir, Soto-Faraco et al. 2003) Although a single sound, or a continuous sound, presented in-between two visual events has been found to facilitate the percept of continuous motion (Getzmann 2007; Bruns and Getzmann 2008), this phenomenon may relate to the short audiovisual interval and perceptual fusion of three events (Bruns and Getzmann 2008). Note that in the above studies, the single sound was located in-between two visual events (the average audio-visual 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 Kolers 1981; Bruns and Getzmann 2008). In the present Experiment 1.2, a similar pattern was evident in the single-sound condition, with audio-visual SOAs of 30 and 0 ms.

In summary, the present study investigated auditory temporal modulation of the perception of the visual Ternus display. 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 non-existent. Furthermore, the audiovisual synchronous Ternus display elicits more percepts of group motion compared 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.

Chapter 3 Influences of intra- and crossmodal grouping on visual and tactile Ternus apparent motion

3.1 Abstract

Previous studies of dynamic cross-modal integration have revealed that the direction of apparent motion in a target modality can be influenced by a spatially incongruent motion stream in another, distractor modality. However, it remains controversial whether non-motion-type intra- and crossmodal perceptual grouping can affect apparent motion in a given target modality. To address this question, we examined visual and tactile Ternus apparent motion and quantified the intra- and crossmodal influences of spatiotemporal grouping on two apparent-motion states: 'element motion' and 'group motion'. Flashing (in order to induce grouping of) the middle visual element prior to the Ternus display enhanced 'element motion' percepts in visual Ternus apparent motion, whereas an analogous stimulation (priming) of the middle tactile element did not influence tactile Ternus apparent motion. Furthermore, under crossmodal conditions, prime stimuli (middle element) in the distractor modality (e.g., touch) did not influence apparent motion in the target modality (e.g., vision). By contrast, tactile stimuli (middle element) presented synchronously with the visual Ternus display frames biased the apparent-motion percept towards 'group motion', while tactile Ternus apparent motion was unaffected by synchronous visual stimuli. We propose that the differential roles of spatiotemporal grouping in vision and touch could explain the asymmetric intra- and crossmodal effects in visual and tactile Ternus apparent motion.

3.2 Introduction

Investigating crossmodal interactions is essential for a comprehensive understanding of the perceptual system (Welch and Warren 1986). A number of studies on crossmodal spatial attention have shown that stimulation in one modality (e.g., tactile motion or touch stimulation) can enhance perceptual sensitivity for spatially congruent stimulus locations in another modality (e.g., visual signal detection) (Macaluso, Frith et al. 2000; Kennett, Eimer et al. 2001; Eimer, van Velzen et al. 2002; Gray and Tan 2002). Crossmodal interactions have also been shown to have asymmetrical influences on the (perceived) direction of apparent motion between different modalities (Soto-Faraco, Lyons et al. 2002; Sanabria, Soto-Faraco et al. 2004; Strybel and Vatakis 2004; Craig 2006; Sanabria, Lupianez et al. 2007; Occelli, Spence et al. 2009).

In a typical crossmodal apparent-motion capture study (e.g. Soto-Faraco, Lyons et al. 2002), two crossmodal apparent-motion streams are presented synchronously or asynchronously in collocated positions, with either congruent or incongruent directions of motion. Observers are asked to indicate the movement direction of the stimuli in the target modality, while ignoring that of the stimuli in the distractor modality. Observers usually make highly accurate judgments of motion direction with congruent motion directions, while their performance is relatively poor with incongruent directions in the 'synchronous' condition. This phenomenon has been termed 'dynamic-capture effect' (Soto-Faraco, Lyons et al. 2002; Soto-Faraco, Spence et al. 2004; Soto-Faraco, Spence et al. 2004). However, if the distractor stimuli are presented asynchronously, the capture effects become weak or even disappear (Soto-Faraco, Spence et al. 2004; Soto-Faraco, Spence et al. 2004). Performance in the incongruent condition also depends on which modality has been selected as (irrelevant) distractor modality. For example, direction judgments of auditory apparent motion have been shown to be reduced to chance level by incongruent visual apparent motion, while the direction of visual apparent motion is rather unaffected by incongruent auditory apparent motion (Soto-Faraco, Lyons et al. 2002; Sanabria, Soto-Faraco et al. 2004; Sanabria, Lupianez et al. 2007). Similarly, direction judgments of tactile apparent motion have been shown to be influenced by incongruent visual and incongruent auditory apparent motion (Craig 2006; Lyons, Sanabria et al. 2006). Similar asymmetrical dynamic capture effects have also been found between touch and audition (Sanabria, Soto-Faraco et al. 2004; Occelli, Spence et al. 2009). Furthermore, dynamic capture can be influenced by intramodal grouping. For example, Sanabria and colleagues (Sanabria, Soto-Faraco et al.

2004; Sanabria, Soto-Faraco et al. 2005) manipulated the strength of task-irrelevant visual apparent motion (distractor modality) by increasing the number of visual stimuli and extending their presentation from before to after the presentation of auditory apparent motion (target modality). They found that enhanced intramodal visual apparent motion reduced the crossmodal capture effect. Moreover, dynamic capture was significantly reduced when the visual stimuli were presented prior to the combined audiovisual display, compared to when the audiovisual display was presented first (Sanabria, Soto-Faraco et al. 2005); this reduction was not attributable to a temporal warning effect. From this, Sanabria et al. concluded that intramodal visual grouping would improve perceptual segregation of the auditory (target) from the visual (distractor) events.

Hitherto, demonstrations of crossmodal interactions in apparent motion have mostly focused on conflict of crossmodal motion and judgments of motion direction. Recently, however, Getzmann and colleagues examined the influence of non-motion-type temporal auditory grouping on visual apparent motion (Getzmann 2007; Bruns and Getzmann 2008). They found the presentation of short sounds (at a fixed location) temporally intervening between the visual stimuli facilitated the impression of continuous visual motion relative to the baseline (visual stimuli without sounds), whereas sounds presented before the first or after the second visual stimulus as well as simultaneously presented sounds reduced the continuous-motion impression. Bruns and Getzmann (2008) argued that crossmodal temporal grouping, which is thought to give rise to a temporal ventriloquism effect (Morein-Zamir, Soto-Faraco et al. 2003), is the main factor influencing the visual motion impression. However, despite the recent focus of grouping mechanisms, there is as yet little understanding of how motion perception per se in one modality is influenced by intra- and crossmodal spatiotemporal non-motion-type grouping.

In the present study, we adopted the two-state Ternus apparent-motion paradigm (Ternus 1926; Harrar and Harris 2007) to examine the influences of intra- and cross-modal non-motion-type grouping on visual and tactile apparent motion (see next paragraph for details). We chose vision and touch to investigate crossmodal apparent motion for the following reasons. First, motion in touch and vision activate at least some common brain areas, such as the middle temporal cortex (hMT), an area known to encode movement velocity (Zeki 1974; Reisbeck and Gegenfurtner 1999; Hagen, Franzen et al. 2002). Second, vision and touch have different temporal properties. Vision has a lower temporal resolution than touch and exhibits transient persistence beyond stimulus offset (Breitmeyer 1984); by contrast, there is essentially no persisting neural activity in peripheral tactile nerve fibers after the offset of a vibratory stimulus (Gescheider, Wright et al. 2009). Accordingly, the role of spatio-temporal grouping for crossmodal apparent motion may differ between vision and touch. With the same spatiotemporal configuration, we expected intramodal visual stimuli presented in close temporal proximity to be grouped more readily compared to analogous tactile stimuli, due to the low temporal resolution and large temporal-integration window in vision. By contrast, the higher temporal resolution in touch may promote perceptual segregation, rather than temporal grouping. In the present study, we introduced stimuli in the distractor modality (collocated with the 'middle elements' in successive Ternus display frames; see below for details) either shortly before ('priming') or synchronously with the presentation of a Ternus apparent-motion display in the target modality. It was expected that the differential propensities for intramodal temporal grouping in vision versus segregation in touch with regard to the middle element of the Ternus display would result in differential capture effects in Ternus apparent motion.

Ternus apparent motion arises from a typical Ternus display (see Figure 2.1), which consists of two sequential visual frames, each presenting two horizontal dots (with the same inter-dot distance in the two frames), where the two frames, when overlaid, share one common dot at the center (the 'middle' element). With different inter-stimulus (i.e., inter-frame) intervals (ISIs), there are often two distinct percepts: 'element motion' and 'group motion'. In element motion, the outer dots are perceived as moving, while the center dot appears to remain static or flashing; in group motion, the two dots are perceived to move together as a group. It has been proposed that in Ternus apparent motion, temporal and spatial grouping processes are in competition (Kramer and Yantis 1997). At short ISIs between the two frames, temporal grouping prevails, that is, the element in the 'overlapping' position of the first frame is likely to be grouped with the element appearing at the same location in the second frame, leading to the percept of 'element motion'. By contrast, at long ISIs, temporal grouping weakens and spatial grouping within a frame becomes more prominent, giving rise to a dominant percept of 'group motion'. A recent study of multimodal Ternus displays indicated that the Gestalt grouping principles are similar in both touch and vision (Harrar and Harris 2007). Thus, Ternus apparent motion is a promising apparent-motion paradigm for examining intra- and crossmodal spatiotemporal grouping effects on crossmodal motion perception.

Specifically, in Experiment 2.1 we enhanced temporal grouping of the middle elements (either the middle dot in the visual Ternus display or the middle tactile tap in the tactile Ternus display) within a given modality by priming the middle element twice (intramodal prime stimuli) prior to the onset of the Ternus display, and comparing the intramodal grouping effects on Ternus apparent motion between the visual and tactile modalities (see Figure 3.2A). If temporal grouping is stronger in the visual modality than in tactile modality, one would expect to obtain a strong priming effect for visual apparent motion, but a lesser effect for tactile apparent motion. In Experiments 2.2 and 2.3, we introduced crossmodal stimulation of the middle elements either asynchronously (visual and tactile middle elements presented sequentially) or synchronously (visual and tactile middle elements presented simultaneously), so as to be able to compare intra- and crossmodal grouping effects on visual and tactile Ternus apparent motion (see schematic illustrations in Figures 3.3A and 3.3B).

3.3 Experimental procedures

3.3.1 Participants

A total of twenty-four naïve observers participated in the study experiments for payment: seven in Experiment 2.1 (four females; average age 25.8 years), eight in Experiment 2.2 (three females; average age 23.4 years), and nine in Experiment 2.3 (4 females; mean age 23.6 years). All had normal or corrected-to-normal vision and none of them reported any history of somato-sensory disorders. They all gave informed consent prior to the experiment.

3.3.2 Stimuli and Apparatus

Visual stimuli were generated using three green LEDs which were arranged horizontally on a sponge with an inter-LED distance of 2 cm. The tactile stimuli were produced using solenoid actuators with embedded cylinder metal tips which would tap the fingers to induce indentation taps when the solenoid coils were magnetized (Heijo Box, Heijo Research Electronics, UK). The three solenoids were embedded near the three LEDs in the same sponge, so that the three visual and the three tactile stimuli were effectively collocated (Figure 3.1). The LEDs and tactile solenoids were controlled via the parallel port of a PC (HP AMD Athlon 64 Dual-Core Processor, Radeon 1700 FSC graphics card). The ON- (luminance of 90 cd/m2) and OFF-states of the LEDs and the tactile taps were controlled by a Matlab program using the Psychophysics Toolbox (Brainard 1997; Pelli 1997). The duration of each ON-state was 30 ms for the LED and 5 ms for the tactile (tap) stimuli. The testing cabin was dimly lit, with an average ambient luminance of 0.09 cd/m2.

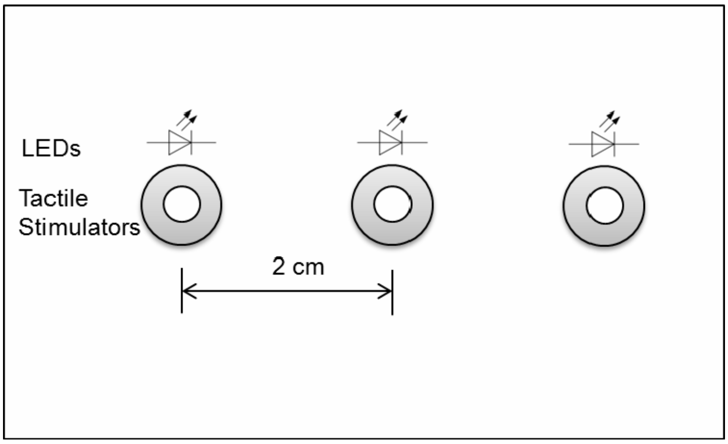


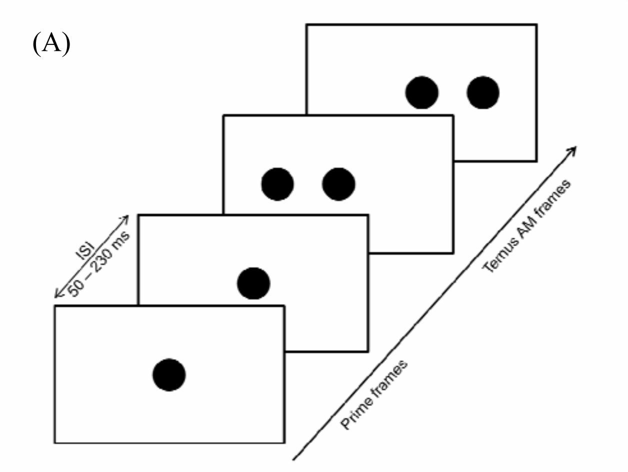
Figure 3.1 Schematic diagram of the experimental setup. Observers were asked to touch three solenoids (indicated by the rings in the figure) with three left-hand (ring, middle, and index) finger tips. Three LEDs were visible near finger tips.

3.4 Experiment 2.1 Influence of intra-modal priming on Ternus apparent motion

Experiment 2.1 consisted of two separate sessions, one dedicated to the visual and the other to the tactile Ternus apparent-motion tasks (with task order counter-balanced across participants). Prior to the experiment, participants were shown demos of ‘element motion’ and ‘group motion’ in visual and tactile Ternus displays, respectively, and they performed a block of trials to become familiar with these two alternative percepts.

In each (visual or tactile) Ternus apparent-motion task, there were two conditions: one with and one without prime. In the visual Ternus apparent-motion condition, a trial started with a beep, followed by a random blank interval of 300 to 500 ms. In the prime condition, the stimulus sequence was as follows (see Figure 3.2A): the middle LED was briefly illuminated twice, each time for 30 ms, with a variable inter-stimulus interval (ISI). The ISI was randomly selected from seven different lengths equally spaced between 50 and 230 ms. After a blank interval of the same ISI, the first visual frame of the Ternus display (the left or the right two LEDs) was turned on for 30 ms. It was then again followed by a blank of the same ISI. Immediately after that, the second visual frame of the Ternus display (the right or the left two LEDs, opposite to the first frame) was turned on for 30 ms. Finally, after a duration of 600 ms, participants were prompted by a ‘beep’ to make a two-alternative forced choice (2AFC) judgment indicating whether they had perceived ‘element motion’ or ‘group motion’. In the baseline condition without prime, only the standard visual Ternus frames were presented, with the same stimulus durations and ISI settings as in the prime condition. In the session with tactile Ternus configurations, the stimuli were tactile taps in place of the LED flashes. All other details (such as the ISI settings etc.) were the same as in the visual Ternus apparent-motion task. For the tactile stimulation, participants had to place the tips of their left-hand ring, middle, and fore fingers on the left, middle, and right solenoid actuators, respectively. Choice responses were collected via the mouse buttons; using their right hand, participants pressed the left button for element motion and the right button for group motion. The inter-trial interval (ITI) varied randomly between 1200 and 1700 ms. In summary, Experiment 2.1 adopted a full-factorial within-subjects design, with 2 (visual/tactile) stimulus modalities × 2 (with/without) prime conditions × 7 ISIs. Each condition was repeated 32 times, with counter-balanced directions of motion.

We measured priming effects on both visual and tactile apparent motion in a paradigm where the middle element (visual dot or tactile tap) was temporally grouped before the Ternus frames (we refer these as priming frames, see Figure 3.2A). Participants' judgments of apparent motion were collected using a 2-alternative forced-choice (2AFC) procedure (i.e., 'group motion' vs. 'element motion'). For each observer, the proportions of 'group motion' were calculated for both baseline and priming conditions; then the psychometric functions were fitted using a logistic function (Treutwein and Strasburger 1999). Figures 3.2B and 3.2C present the psychometric functions (averaged across observers) for the visual and the tactile Ternus configurations.



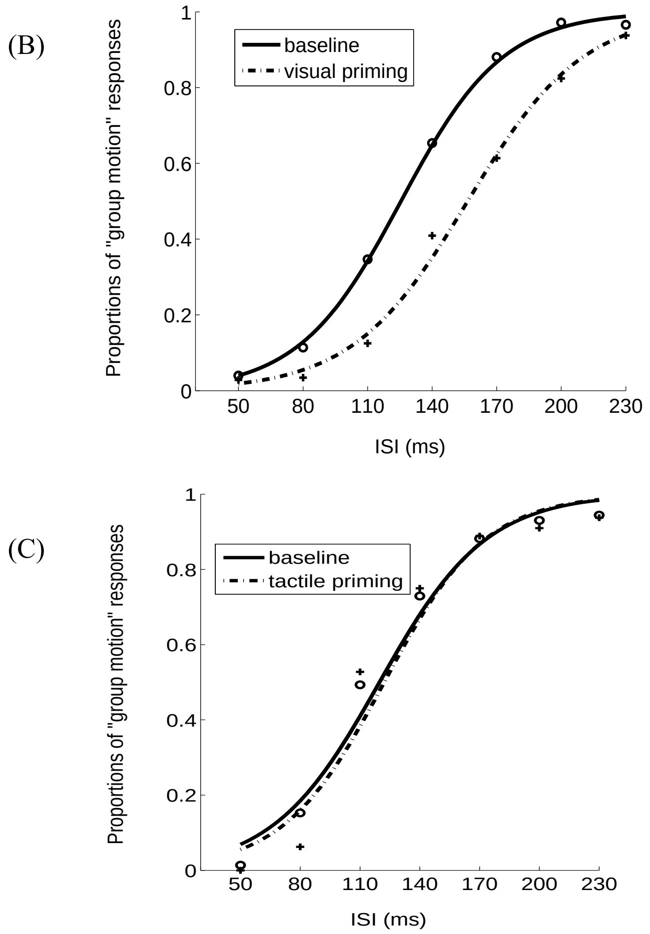


Figure 3.2 Intra-modal priming effects on Ternus apparent motion. (A) Experimental paradigm. Illustrated is the priming condition, in which the middle visual (or tactile) stimulus was presented twice (30 ms for visual stimuli, 5 ms for tactile taps), separated by an ISI, prior to the presentation of the standard visual (or tactile) Ternus display. In the baseline condition, the two prime frames were omitted and only the standard Ternus display was presented (in either the visual or the tactile modality). (B) Responses of ‘group motion’ to the visual Ternus display. The circles and solid line represent the proportions of ‘group motion’ responses in the baseline condition; the dash-dot line and crosses depict the proportions of ‘group motion’ reports in the priming condition. (C) Responses of ‘group motion’ to the tactile Ternus display.

Subsequently, the transition thresholds (the points of subjective equality, PSEs) between the two apparent-motion percepts and the just noticeable differences (JNDs) were calculated individually. With visual Ternus configurations, the average PSEs (±SE) were 159.0±29.2 ms and 130.2±17.2 ms for the priming and baseline conditions, respectively. A repeated-measures analysis of variance (ANOVA) revealed the PSE to be significantly higher in the priming condition, $F(1,6)=6.307$, $p<0.05$. This can be taken to indicate that priming (i.e., flashing) the middle LED prior to the Ternus display enhanced the temporal grouping of the middle visual elements across frames (instead of the middle element moving towards the left or the right; see illustration in Figure 3.1), thus shifting the visual apparent-motion percept towards ‘element motion’. In contrast, with tactile configurations, the PSEs (±SE) in the (tactile) priming and standard conditions – 131.4±27.9 ms and 127.9±23.4 ms, respectively –and did not differ significantly from each other, F(1, 6)=0.156, p>0.1; that is, the priming of the middle tactile stimulus did not influence the subsequent tactile apparent-motion percept.

Further analyses showed that the JNDs were not influenced by the prime stimuli, neither for visual nor for tactile Ternus configurations, F(1,6)=1.294, p>0.1, and F(1,6)=1.683, p>0.1, respectively. Thus, the primes neither enhanced nor degraded the sensitivity for the two motion states in the Ternus apparent-motion display.

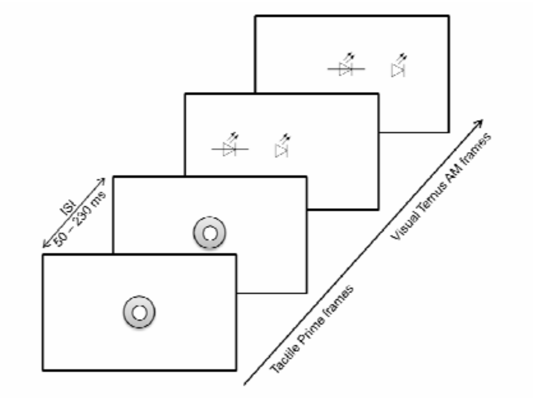
3.5 Experiment 2.2 – Influence of tactile priming on visual Ternus apparent motion

To examine the influence of tactile stimuli on visual Ternus apparent motion, three different conditions were introduced: (a) A baseline condition, in which the standard visual Ternus display was presented without tactile taps; (b) a tactile priming condition, in which the middle solenoid tapped the middle finger twice prior to the presentation of the standard visual Ternus display (Figure 3.3A); and (c) a tactile synchronous condition, in which the onsets of the two taps applied to the middle finger were synchronized with the onsets of the two visual Ternus frames (Figure 3.3B). The ISI configuration and the sequence of events on a trial were the same as in Experiment 2.1. And again, participants were asked to make a 2AFC judgment on the visual motion percept (‘element motion’ vs. ‘group motion’). In summary, Experiment 2.2 used a full-factorial within-subjects design, with 3 conditions x 7 ISIs; each condition was repeated 32 times, with counter-balanced directions of motion.

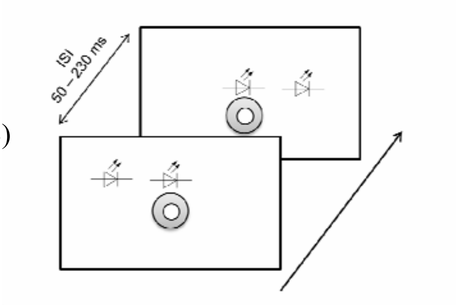
Experiment 2.1 revealed intramodal temporal-grouping effects for visual Ternus configurations, but not for tactile configurations. In Experiment 2.2, we went on to examine how crossmodal temporal grouping would affect Ternus apparent motion. In particular, we asked whether tactile asynchronous (‘priming’) or, respectively, synchronous grouping of the middle elements would influence visual Ternus apparent motion. Accordingly, we compared three different conditions: (a) ‘baseline’ condition, in which pure visual Ternus apparent motion was tested; (b) ‘priming’ condition, which examined the effect of crossmodal asynchronous grouping induced by tapping the middle finger twice before the visual Ternus frames (Figure 3.3A); and (c) ‘synchronous’ grouping condition, in which the tactile taps were presented synchronously with the onset of the two visual Ternus frames (Figure 3.3B).

The proportions of ‘group motion’ responses were calculated for each condition, individually for each observer, followed by the fitting of the psychometric functions (using the logistic function). Figure 3.3C presents the average psychometric functions for the three conditions. Subsequently, from the individual functions, the PSEs and JNDs were determined. On average, the PSEs (±SE) were 144.9±13.4 ms, 135.4±12.1 ms, and 114.9±28.1 ms for the baseline, priming, and synchronous conditions, respectively. A repeated-measures ANOVA revealed the main effect of condition to be significant, F(2,14) = 9.79, p<0.01. Bonferroni-corrected pairwise comparisons showed that the PSE was significantly lower, by 29.9 ms, in the synchronous compared to the baseline condition, p <0.05. That is, in the ‘synchronous’ condition, the concurrent tactile events systematically increased reports of ‘group motion’ in the visual Ternus display. By contrast, compared to the baseline condition, PSE in the priming was lower 9.5 ms, p >0.1, cross-modal interaction or ‘capture’ effects were not evident in the tactile priming condition.

(A)



(B)



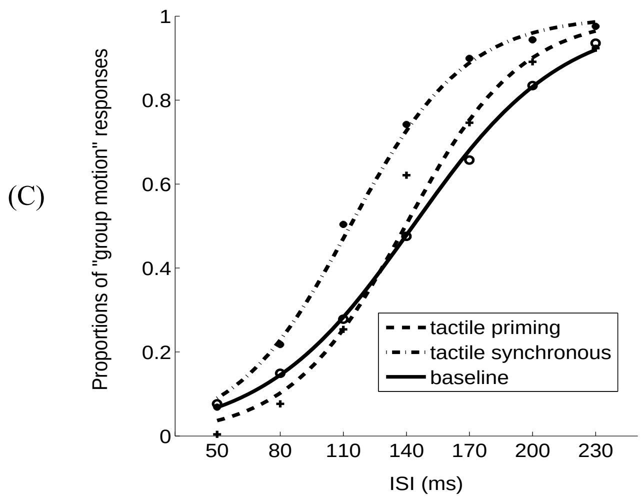


Figure 3.3 (A) Crossmodal priming condition, in which middle solenoid metal tip (indicated by the ring in the figure) tapped the middle finger twice prior to the presentation of the standard visual Ternus frames (composed of LEDs). (B) 'Synchronous' condition, in which the onsets of the middle-finger taps were synchronized with the onsets of two visual frames. (C) Proportions of 'group motion' responses as a function of ISI and fitted psychometric curves in Experiment 2. The proportion of 'group motion' responses are represented by the solid line and circles for the baseline condition, the dash-dot line and stars for the 'synchronous-tap' condition, and the dashed line and crosses for the tactile-priming condition.

The average JNDs in Experiment 2.2 were 25.9±2.3 ms, 19.8±1.9 ms, and 21.1±1.8 ms for the baseline, the priming, and the synchronous conditions, respectively. A repeated-measures ANOVA failed to reveal significant differences among the three conditions (main effect of condition: F(2,14) = 2.796, p=0.095), suggesting that tactile stimulation does not change the sensitivity for visual apparent motion.

3.6 Experiment 2.3 –Influence of visual priming on tactile Ternus apparent motion

The stimulus settings were the same as in Experiment 2.2, just with the roles of target and priming modality reversed. That is, the tactile modality was the target modality in which the Ternus display was presented, and the visual modality was the priming modality, with visual primes and, respectively, synchronous visual stimuli generated from the middle LED. Similar to Experiment 2.2, there were three conditions: (a) baseline condition: tactile Ternus apparent-motion display without visual events; (b) 'synchronous' condition: middle LED illuminated synchronously with the onset of the first and the second tactile frame; (c) visual priming condition: middle LED flashed twice prior to the tactile events. The procedure was the same as in Experiment 2.2, except that the task was changed to a 2AFC judgment on the tactile apparent-motion percept (‘element motion’ vs. ‘group motion’).

In Experiment 2.2, temporal grouping induced by tactile priming had no influence on the subsequent visual apparent motion, whereas grouping by concurrent taps did alter the visual apparent-motion percept, giving rise to an increased proportion of ‘group motion’ reports. To further explore the pattern of crossmodal interactions, in Experiment 2.3, we examined the influence of visual priming and synchronous grouping on tactile Ternus apparent motion.

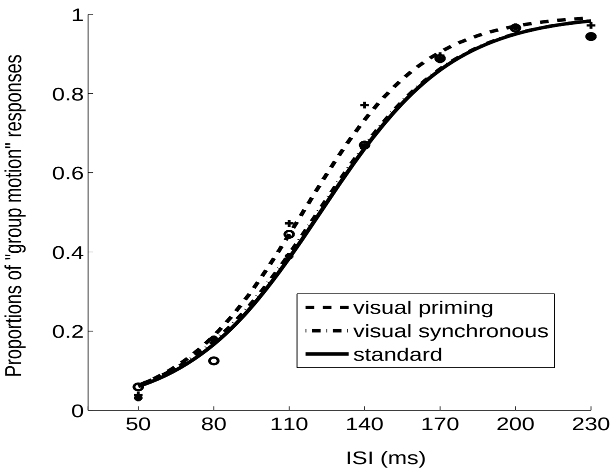


Figure 3.4. Proportions of ‘group motion’ responses as a function of ISI and the fitted psychometric curves in Experiment 2.3. The proportion of ‘group motion’ responses are represented by the solid line and circles for the baseline condition, the dash-dot line and stars for the ‘synchronous-flash’ condition, and the dashed line and crosses for the visual-priming condition.

The overall psychometric functions are shown in Figure 2.4. The mean PSEs (±SE) were 122.4 ±25.8 ms, 121.0±25.9 ms, and 115.4±13.7 ms for the baseline, priming, and synchronous conditions, respectively. A repeated-measures ANOVA failed to reveal any significant differences among the PSEs in the three conditions, F(2,16)=1.551, p=0.242. Similarly, a repeated-measures ANOVA of the JNDs revealed no significant condition effect, F(2,16)=0.972, p=0.399. Thus, in contrast to Experiment 2.2, neither visual priming nor synchronous visual events modulated the perception of tactile Ternus apparent motion.

3.7 General discussion

With non-motion-type intramodal (asynchronous) temporal grouping, differential effects emerged between visual and tactile Ternus apparent motion: Priming of the middle flash shifted the visual Ternus apparent-motion percept towards ‘element motion’; by contrast, there was no effect of tactile priming (of the middle tap) on the tactile Ternus apparent-motion percept. The visual modality is known to have lower temporal resolution than the tactile modality. For example, the highest flicker fusion rate is approximately 60 Hz in peripheral vision (Hartmann, Lachenmayr et al. 1979), which is substantially lower than the vibrotactile upper limit 1000 Hz (Verrillo, Fraioli et al. 1969). When visual stimuli are presented in rapid succession at the same location, visual persistence and masking effects come to the fore (Breitmeyer and Ogmen 2006). These properties help visual stimuli to become temporally grouped. Accordingly, in the intramodal visual-grouping condition of Experiment 2.1, the flashes of the middle element prior to the visual Ternus frames may have helped the middle elements in the Ternus frames to temporally group, as a result of which the two flanker elements (i.e., that in the first and that in the second frame) become more likely to be linked, forming the percept of ‘element motion’. Indeed, similar results have been reported by He and Ooi (1999). They also observed a dominant ‘element motion’ percept using two-element priming in a four-element Ternus display, which they attributed to a perceptual bias in the apparent-motion percept owing to visual persistence. By contrast, the tactile modality has a higher temporal resolution and exhibits less persistent neural activity (Gescheider, Wright et al. 2009). Furthermore, vibrotactile masking studies have shown that backward masking is more effective than forward masking (Snyder 1977; Evans 1987; Craig 2006), that is, later tactile stimuli are more potent in disrupting the perception of earlier stimuli (than vice versa). Applied to the present paradigm, backward tactile masking may have overridden the temporal grouping effect, so that the priming of the middle tactile stimuli would have had little influence on the subsequent tactile Ternus apparent motion.

For the synchronous crossmodal stimulation conditions examined in Experiments 2.2 and 2.3, the results showed that synchronous (non-motion-type) tactile stimuli presented at the middle position could readily shift the visual Ternus apparent-motion percept towards of ‘group motion’ (Exp. 2.2), whereas synchronous visual stimuli had no influence on tactile Ternus apparent motion (Exp. 2.3). There are several possible explanations for this dissociation. First, the sense of touch has high temporal resolution, and it may thus promote segregation between two temporally proximate visual stimuli. Indeed, there is evidence that temporally proximate tactile stimuli improve visual temporal discrimination (Keetels and Vroomen 2008); also, tactile intervals are perceived as longer than physically equivalent visual intervals (van Erp and Werkhoven 2004). Accordingly, the longer (perceived) interval between the two tactile stimuli may have ‘captured’ the shorter visual interval, which in turn may have biased the visual Ternus apparent-motion percept towards ‘group motion’. (Recall that reports of ‘group motion’ percepts are increased at longer ISIs.). Second, the sense of touch monitors biologically vital features of the environment by means of direct contact, and may therefore capture attention more efficiently than other, more ‘remote’ sensory modalities such as vision (Spence, Nicholls et al. 2001; Spence 2002; Gallace and Spence 2008). Applied to our experimental setting, with synchronous visual-tactile stimuli, the tactile stimuli may have captured attention (rather than the visual stimuli) and impose a temporal segregation of the two visual frames, which in turn led to a shift of visual Ternus apparent-motion percept towards ‘group motion’. 1 By contrast, in the tactile apparent-motion experiment (Exp. 2.3), due to tactile taps being highly salient (indentations), attention may be allocated mostly to the (on-body) touch stimuli. Thus, synchronous visual stimuli would have received less attention, diminishing their influence on tactile-motion perception. Interestingly, recent neuroimaging evidence has shown that tactile motion/tactile stimuli, even in the absence of visual stimuli, can activate hMT/V5 (Macaluso, Frith et al. 2000; Hagen, Franzen et al. 2002), a brain area known to be specialized for visual-motion processing (Zeki 1974; Maunsell and Van Essen 1983a; Maunsell and Van Essen 1983b). Such cortical connections could provide a neural mechanism by which synchronous tactile stimuli might influence visual motion perception.

Unlike the synchronous crossmodal grouping effects, crossmodal priming (intended as an asynchronous-grouping manipulation) did not influence the perception of Ternus apparent motion in either the visual or the tactile modality. The failure to find such influences may relate to ‘intersensory Gestalten’ being weak perceptual formations. The notion of an ‘intersensory Gestalt’ refers to the structure of stimuli in one sensory modality influencing the perceived organization of stimuli in another modality (Gilbert 1938; Gallace and Spence 2008). Generally, intersensory Gestalten have been found to be relatively weak (see review Gallace and Spence 2008). In our experimental setting, presenting tactile and visual stimuli temporally separated in the middle position may not yield a uniform grouping (i.e., an intersensory Gestalt), but rather two independent tactile and visual objects (events). Consistent with this, weak or ineffective interactions of asynchronous crossmodal stimuli have also been found in previous studies of

---

[1] The two aspects discussed here are not necessarily mutually exclusive. Due to its higher temporal resolution, the tactile interval between successive stimuli will capture the visual interval (according to the ‘assumption of unity’, Welch and Warren, 1986). As the tactile interval is relatively longer perceptually (e.g., van Erp and Werkhoven, 2004), it moves apart the two (temporally captured) visual Ternus frames. And interval capture may be enhanced by the greater power of tactile stimuli to attract attention.

motion direction capture, where dynamic-capture effects were diminished when the (apparent) motion in the distractor modality was presented asynchronously with the (apparent) motion in the target modality (Soto-Faraco, Lyons et al. 2002; Soto-Faraco, Spence et al. 2004; Soto-Faraco, Spence et al. 2004).

In summary, the findings of the current study suggest that the effects intra- and crossmodal temporal grouping within and between the visual and tactile modalities on Ternus apparent motion are subject to the temporal properties and biological constraints in the two sensory systems. We propose that the differential temporal resolution in the two systems could explain the asymmetric intramodal grouping effects in visual and tactile Ternus apparent motion. Furthermore, on-body touch may modulate the subjective temporal interval (i.e., prolong it in the tactile modality) and bias the allocation of attention towards the tactile modality, which in turn exerts asymmetric influences on crossmodal Ternus apparent motion (prolongation of the visual interval, biasing the visual motion percept towards 'group motion').

Chapter 4 The influences of auditory timing and temporal structure on tactile apparent motion

4.1 Abstract

Previous studies have demonstrated that the temporal interactions between different sensory modalities on the perceived apparent motion, however, the contributions of timing and event structure remain unclear. To address this question, we investigated the influences of auditory timing and auditory temporal structure on the perceptual rivalry of tactile apparent motion in a long audiotactile stream (lasted about 90 seconds). With one-to-one mapping (“full-pairing”) audiotactile stimuli, auditory timing was crucial to systematically influence the tactile apparent motion. Surprisingly, the influence of the auditory timing was diminished in the “half-pairing” configuration and was even reversed in temporal shifted “full-pairing” configuration. Our findings suggest that auditory timing could influence the tactile apparent motion, yet it is not the determining factor. Auditory temporal structure, influencing on the saliency and attentional redistribution of crossmodal events, also contribute to the ultimate multisensory percept.

4.2 Introduction

Apparent motion is a common perceptual phenomenon in daily life. For example, two brief flashes of light, when separated in both time and space, create an illusion of movement from the location of the first flash to that of the second when the spatiotemporal parameters of the display are within the appropriate range (Exner 1875). Apparent motion occurs in different sensory modalities, given the respective physical stimuli. Early research, based on subjective reports, has revealed the existence of crossmodal apparent motion between all possible combinations of auditory, visual, and tactile stimuli (Korte 1915; Neuhaus 1930; Sherrick 1968; Kirman 1974; Kirman 1974; Allen and Kolers 1981; Allen and Kolers 1981; Ohmura 1987; Getzmann 2007; Harrar, Winter et al. 2008). Furthermore, a number of studies have shown that apparent motion in a particular modality may be influenced by static or dynamic events in another modality (Sekuler, Sekuler et al. 1997; Soto-Faraco, Lyons et al. 2002; Soto-Faraco, Spence et al. 2004). For example, the direction of auditory motion can be captured by visual motion in a conflicting direction, whereas the direction of visual motion is not affected by incongruent auditory motion (Soto-Faraco and Kingstone 2004). Such interactions, referred to as "crossmodal dynamic capture", have been well demonstrated across audition, touch, and vision (Soto-Faraco, Lyons et al. 2002; Soto-Faraco and Kingstone 2004; Soto-Faraco, Spence et al. 2004; Sanabria, Soto-Faraco et al. 2005; Lyons, Sanabria et al. 2006; Soto-Faraco, Kingstone et al. 2006; Sanabria, Spence et al. 2007). Most of these studies have examined spatial interactions between two modalities using an "immediate-response" approach, where participants were required to make a speeded motion direction judgment on one of the (potentially conflicting) motion streams in two modalities. Recent work on crossmodal interaction has also shown that apparent motion in one modality may be influenced solely by temporal modulations in another modality. For example, visual apparent motion can be modulated by spatially uninformative auditory events (Getzmann 2007; Bruns and Getzmann 2008; Freeman and Driver 2008). Using a visual apparent-motion paradigm, Freeman and Driver (2008) found that, in a repeated two-flash visual apparent-motion stream with equal inter-flash interval (for which, when presented alone, the perceived motion direction would be ambiguous), auditory beeps slightly lagging or leading the flashes strongly influenced the perceived visual motion direction – even though the beeps provided no spatial information. Freeman and Driver argued that the accompanying beeps influenced the perceived visual timing, owing to the higher temporal acuity in the auditory, compared to the visual, modality. Similar influences of auditory timing on visual timing have been found in temporal-order judgment tasks. As demonstrated in a number of other studies, such influences give rise to what has been referred to as “temporal-ventriloquism” effect, that is: when auditory and visual stimuli occur slightly asynchronously, one stimulus is pulled into temporal alignment with the other stimulus, forming a bound or unitary percept (Bertelson and Aschersleben 2003; Morein-Zamir, Soto-Faraco et al. 2003; Vroomen and de Gelder 2004; Getzmann 2007).

Some studies of the auditory influence on visual apparent motion have attempted to show that the superior precision of auditory timing (the standard theoretical assumption) is not the only determining factor. Studies of crossmodal spatial capture have provided evidence that intramodal perceptual grouping can constrain the ‘binding’ of apparent-motion signals from two modalities. For example, the modulatory effect of visual motion on the direction of auditory apparent motion is reduced when the intramodal visual apparent-motion stream is presented prior to the audiovisual apparent-motion stimuli (Sanabria, Soto-Faraco et al. 2004; Sanabria, Soto-Faraco et al. 2005). Similar results have also been found in a visual-tactile apparent-motion study in which crossmodal dynamic capture of the tactile apparent motion by a conflicting visual apparent motion was stronger in a two-flash condition than in a six-flash condition (Lyons, Sanabria et al. 2006). One common explanation is that when the intramodal perceptual grouping is set up prior to the crossmodal grouping, crossmodal interactions become weaker (Spence, Sanabria et al. 2007). With regard to crossmodal temporal integration, analogous evidence – of intramodal grouping influencing crossmodal interactions – has been revealed in studies of the temporal-ventriloquism effect (Keetels, Stekelenburg et al. 2007) and of crossmodal apparent motion (Bruns and Getzmann 2008; Shi, Chen et al. 2010). For example, Bruns and Getzmann used a visual-motion categorization task to examine the effects of auditory grouping and sound temporal structure on the perceived visual motion. Interestingly, they found that either a continuous sound filling the gap between two light flashes or a short sound intervening between two flashes enhanced reports of continuous motion, while there was no such enhancement when the sound was part of a tone sequence that allowed for intramodal perceptual grouping prior to the multisensory integration of the audiovisual stimuli. This raises the question whether multisensory integration would still be affected when intramodal perceptual grouping is removed or absent. In a more recent study, using visual Ternus apparent motion, we demonstrated that removing intramodal auditory grouping (e.g., by merely presenting a single sound click) can also diminish the temporal-ventriloquism effect, while the effect was observed in two-sound grouping conditions (Shi, Chen et al. 2010). These findings suggest that crossmodal interactions are constrained by an “assumption of unity”, that is, the notion that the interaction of different sensory inputs depends on observers’ inherent assumption that physically separate events are attributable to a common cause (Welch and Warren 1980; Welch 1999). However, how the “reasoning” underlying the assumption of unity is implemented and what factors influence this assumption are still under debate.

Crossmodal interactions take place among different combinations of modalities, with audiovisual interactions having been investigated most extensively. Compared to the visual modality, both the auditory and tactile modalities are superior in temporal acuity, and sense vibratory stimuli with the basilar membrane in the cochlea and, respectively, the Meissner and Pacinian corpuscles in the skin (Von Bekesy 1959; Soto-Faraco and Deco 2009). Given the similarities between hearing and touch, one would naturally expect to find a close interplay between the two senses in crossmodal perception. Mutual influences between audition and touch have been explored in several recent studies. For example, sound can drive the tactile perceived rate (Bresciani, Ernst et al. 2005; Bresciani and Ernst 2007) and capture tactile apparent motion (Soto-Faraco, Spence et al. 2004). Irrelevant tactile stimuli presented together with task-relevant sounds can improve auditory detection (Gillmeister and Eimer 2007). Moreover, it has been demonstrated that a centrally located tactile stimulus attracts a peripheral sound towards the middle (Caclin, Supérieure et al. 2002). Of note, however, the above mentioned studies on audiotactile interactions focused mainly on spatial interactions using the immediate-response approach, with observers asked to make a speeded response to the target stimuli involving location discrimination. By contrast, with regard to the temporal domain, there are almost no reported studies of the influence of auditory temporal grouping and structure on tactile apparent motion, especially on the perceptual rivalry between different directions of apparent motion (which is optimal with constant SOA between alternate stimulus events) in a long-running (> 1 minute) tactile apparent-motion steam. Given that apparent-motion rivalry is different from more transient apparent-motion phenomena, in the former the different percepts such as the motion directions are fluctuated, the perceptual decision making of a certain percept is implemented by an activity accumulation of a population of neurons which win the competitions in the rivalry.(Carter et al, 2008; Deco, Scarano & Soto-Faraco, 2007) For the latter, the apparent motion percept is relatively clear, subjects usually made an immdiate response after the presentation of motion stimuli.(Soto-Faraco, Spence & Kingstone, 2004) and given the greater similarity in the nature of the physical signals that are transduced by both auditory and tactile sensory systems, it might be good grounds for investigating any crossmodal interactions between the auditory and tactile modalities(Spence, Sanabria & Soto-Faraco,2007). Particularly, it is theoretically interesting to investigate the crossmodal interaction between auditory and tactile modalities, i.e, the influence of auditory events on the tactile apparent motion rivalry.

On this background, the present study was designed to investigate how the crossmodal temporal-ventriloquism effect and unimodal temporal structure operate in the audiotactile modalities. Specifically, we examined whether and to which degree the temporal-ventriloquism effect induced by static sounds influences tactile apparent motion in a long-running motion rivalry stream and how the intramodal sound structure influences the crossmodal interaction. We implemented streams of audiotactile stimuli in two conditions: "full" and "half audiotactile stimulus mapping", respectively. In both conditions, a train of beeps was pairing with a stream of tactile taps: in the "full-pairing" condition, the even-numbered beeps were synchronous with the onsets of the tactile taps on one side (e.g., middle finger of left hand) and the odd-numbered beeps asynchronous with the tactile taps on the other side (e.g., middle finger of right hand) by a given SOA. In the "half-pairing" condition, by contrast, the even-numbered synchronous sounds were absent, leaving only the odd-numbered beeps (therefore "half-pairing"). If the crossmodal temporal-ventriloquism effect takes priority over intramodal temporal structure, one would expect that sound can bias the tactile apparent motion in both situations, with the same audiotactile stimulus onset asynchronies (SOAs) in both conditions. Alternatively, if auditory temporal structure dominates in the audiotactile interaction, one would envisage differential outcomes: the full-pairing audiotactile stream would be subject to a crossmodal temporal-ventriloquism effect; by contrast, the half-pairing condition would show little influence of the auditory timing due to the incomplete grouping of the auditory with the tactile events, analogously to the results of our audiovisual temporal-ventriloquism study (Shi, Chen et al. 2010). Furthermore, if auditory temporal structure does indeed play a critical role in audiotactile crossmodal interactions, one would expect different auditory temporal structures lead to corresponding different 'capture' effect on the tactile apparent motion, with respective dominant percept of motion direction. These predictions were tested using full-pairing, half-pairing, and temporally shifted (see below) full-pairing audiotactile stimuli in a long tactile apparent-motion rivalry stream. Experiments 3.1 and 3.2 were designed to investigate whether spatially uninformative sound can bias tactile apparent motion; and, if so, whether a crossmodal temporal-ventriloquism effect based on the asynchrony between audiotactile events would take priority over the intramodal (auditory) grouping in this bias. In contrast with Experiment 3.2, Experiment 3.3 used an audiotactile streams of full auditory-tactile pairing but with shifted temporal locations of audiotactile stimuli, to examine the influence of auditory temporal structure on the perception of tactile apparent motion.

4.3 Experiment 3.1 Tactile apparent motion in full-pairing audiotactile stream

In Experiment 3.1, we use full-paired audiotactile streams. The audiotactile stimulus-onset asynchrony (SOA between the odd-numbered auditory stimulus and the paired tap) was systematically varied from -75 ms to 75 ms, where negative SOAs indicate auditory stimuli leading tactile stimuli, and vice versa for the positive SOAs. The short SOAs serve two purposes: to prevent a warning-signal effect evoked by non-target auditory events, and to trigger a genuine crossmodal interaction when audiotactile events fall into the short temporal window within which observers perceive the audiotactile events to originate from a common source. The purpose of Experiment 1 was to examine for an influence of spatially uninformative sounds on tactile apparent motion, that is, whether there would be a similar auditory temporal-ventriloquism effect on tactile apparent motion. as sounds did on visual apparent motion (Freeman & Driver, 2008).

4.3.1 Method

Participants. Seven observers (6 females, average age 26.6 years) participated in Experiment 3.1 for payment. None of them reported any history of somatosensory disorders. They were naïve as to the purpose of the study and were paid after the experiment.

Apparatus and stimuli. A customized tactile stimulus generator (Heijo Research Electronics) was connected to a HP PC (AMD Athlon 64 Dual-Core processor) via the LPT port. The two solenoid actuators, which were embedded in a sponge with a fixed center-to-center distance of 10 cm, alternately produced “indentation” taps to two fingers (see Figure 4.1A). The duration of a single tap was 10 ms and the stimulus onset asynchrony (SOA) between two successive taps was 400 ms. Mono-beeps (60 dB, 1000 Hz, 30 ms) were generated by an embedded high-precision M-AUDIO Delta 1010 Sound Card and delivered through a headset (RT-788V, RAPTOXX) to both ears. Participants’ responses were acquired via two foot pedals. The experimental program was developed using Matlab (Mathworks Inc.) and Psychophysics Toolbox (Brainard 1997).

Design and procedure. Prior to the formal experiment, observers received a practice session to become familiarized with the procedure and the experimental task. They were asked to place the tips of their left and right middle fingers such as to cover the surface of the left and right tactile actuators. A trial started with a fixation cross in the center of the monitor in front of the participants, which observers were instructed to fixate throughout the trial. After a random interval of 500–1000 ms, the two tactile actuators produced alternating (finger indentation) taps with a fixed SOA of 400 ms (2.5 Hz), repeated for 90 seconds. The initial tap occurred randomly on either the left or the right middle finger. Experiment 3.1 comprised of seven audiotactile conditions (SOAs) and one baseline condition, which were randomized across trials. In the audiotactile conditions, a train of beeps was pairing with a train of tactile taps, where even-numbered beeps were synchronous with the onsets of the tactile taps on one side and odd-numbered beeps were asynchronous, by a given SOA (-75, -50, -25, 0, 25, 50, 75 ms), with the onsets of the tactile taps on the other side (see Figure 1B). After an initial presentation of these events for 4 seconds, a visual-cue word (“begin”) was presented in the center of the screen prompting observers to initiate their responses, that is, indicate the perceived direction of the tactile apparent motion, irrespective of the accompanying sounds. Observers were asked to hold one foot pedal pressed to indicate the perceived direction of tactile apparent motion (left foot pedal for leftward motion, right pedal for rightward motion) and to switch the foot pedal immediately when the perceived direction changed. In the experiment, eight conditions were repeated four times with counter-balancing of the initial motion direction.

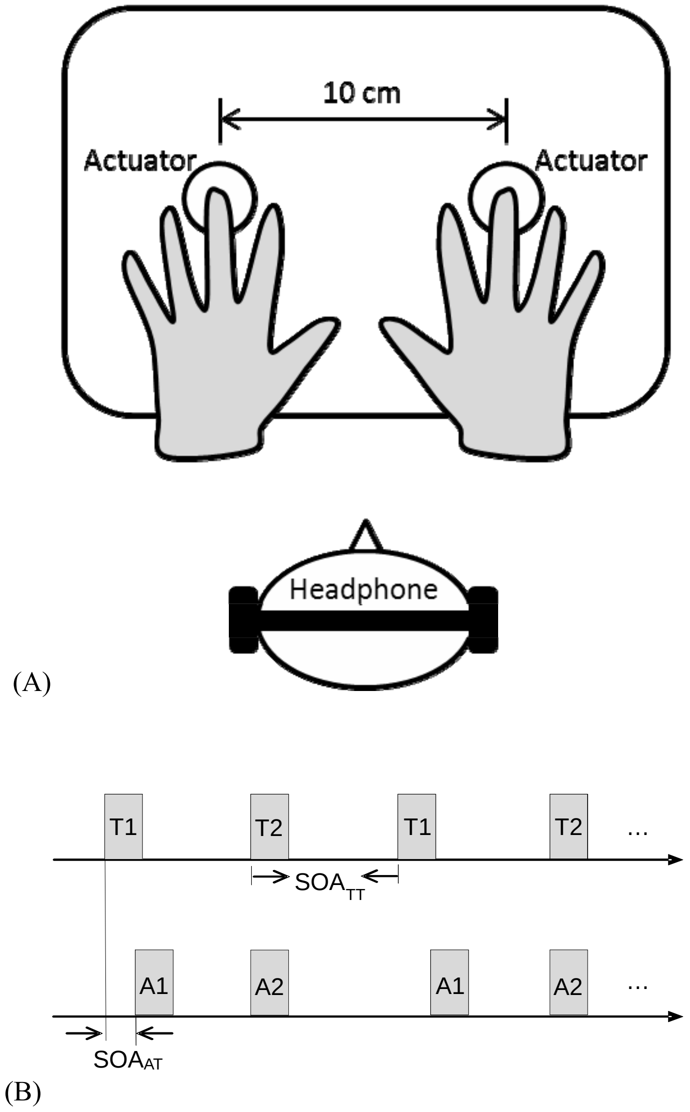


Figure 4.1 (A) Illustration of the experimental setup. (B) Schematic representation of the audiotactile timing conditions in Experiment 1. Asynchronous and synchronous audiotactile pairs were alternated in a train of 90 seconds audiotactile stream. The SOA between tactile stimuli (SOA_{TT}) was 400 ms, and the SOA between asynchronous audiotactile stimulus pairs (SOA_{AT}) was varied across trials (see text for details).

Data Analysis. Bistable perception has often a transient initial preference which is biased towards the initial presentation (Mamassian and Goutcher 2005). To reduce such initial biases in the present experiment, response recording commenced only four seconds after the initiation of the audiotactile streams. To further disassociate any transient preference from an influence of auditory timing, the responses of left- and rightward tactile apparent-motion directions were recoded in terms of “initial direction” (i.e., perceived direction congruent with the direction indicated by the first two taps) and “reverse direction” (opposite to the “initial direction”) and, accordingly, the pedal press times (i.e., phase durations) were collected and calculated separately for the “initial” and the “reverse directions” in each audiotactile condition. As the phase durations (of perceiving motion in one or the other direction) varied substantially among participants, while they can described with same distribution (Mamassian and Goutcher 2005; Pastukhov and Braun 2007), the phase durations were normalized for each observer relative to their respective mean.

4.3.2 Results and discussion

Figure 4.2 shows the mean normalized phase duration for the two types of responses as a function of the audiotactile SOA. Pairwise t-test showed that the responses of “initial direction” and “reverse direction” were not significantly different in the baseline condition (without sounds), t(6)=-1.098, p=0.314, indicating that initial bias virtually disappeared when tactile apparent-motion direction responses were collected only after four seconds. This was also true for the synchronous-sound condition (i.e., audiotactile SOA = 0 ms), t(6)=-1.968, p=0.097. Figure 2 shows clear trends of perceptual dominance across the different audiotactile SOAs. Individual pairwise t-tests showed that the two motion percepts (“initial direction” vs. “reverse direction”) were significantly different at the SOAs of -75,-50, 25, 50, and 75 ms, p<0.05. This suggests that asynchronous auditory stimuli indeed influence tactile apparent motion. We selected the phase durations of the “initial direction” for further analysis of the influence of auditory timing (the results would be analogous for the “reverse direction”). A repeated-measures ANOVA revealed a significant main of auditory timing, F(6, 36)=19.712, p<0.01, and a follow-on linear contrast test showed that the phase duration increased linearly with increasing audiotactile SOA, F(1,6) = 42.8, p<0.01. This indicates that the influence of the auditory stimuli on tactile apparent motion was temporally systematic and bidirectional. For example, an audiotactile SOA of 50 ms (i.e., with the first beep lagging the first tap by 50 ms) produced a dominant percept of “initial direction”, while an audiotactile SOA of -50 ms (i.e., with the first beep leading the first tap by 50 ms) gave rise to a dominant percept of “reverse direction”. Thus, the results are consistent with a temporal ventriloquism effect (Morein-Zamir, Soto-Faraco et al. 2003) in audiotactile apparent motion, with auditory timing influencing the perceived timing of temporally close tactile events. In the long audiotactile streams, lagging odd-numbered beeps pulled timings of corresponding odd-numbered taps closer to the even-numbered taps, thus leading to dominant responses of “initial direction”. Similarly, leading odd-numbered beeps pulled timings of corresponding odd-numbered taps faraway from the even-numbered taps,shifted responses to the opposite, “reverse direction”. The results are analogous to those of

(Freeman and Driver 2008), who found auditory timing to influence visual apparent motion.

Besides the temporal-ventriloquism account, one might argue there are alternative explanations for the auditory modulation of tactile apparent-motion direction observed in Experiment 3.1. In particular, auditory grouping or auditory structure may play a role in the modulation of tactile apparent motion; that is, interval variations between beeps may promote auditory perceptual grouping and thereby exert an influence on crossmodal grouping (Keetels, Stekelenburg et al. 2007; Spence, Sanabria et al. 2007). In order to disentangle this alternative explanation from the temporal-ventriloquism account, we further examined a half-pairing condition in Experiment 3.2, where the audiotactile temporal asynchronies were maintained but (potential) auditory grouping by interval variations was removed.

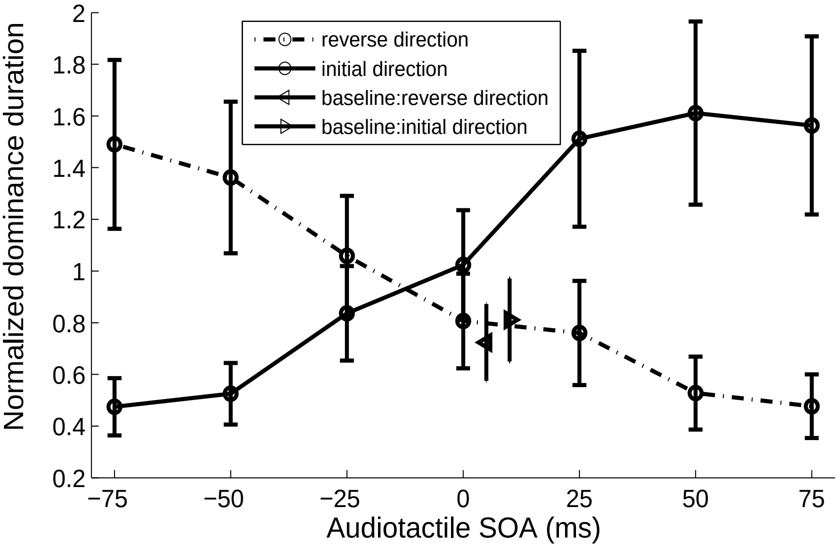


Figure 4.2. Normalized phase durations of tactile apparent motion as a function of audiotactile SOA (full-pairing condition). Normalized phase durations as a function of the audiotactile SOA. The solid line represents mean phase durations of the "initial direction", the dotted line of the "reverse direction". The rightward-pointing triangle denotes responses of "initial direction", and the rightward-pointing triangle responses of "reverse direction", for the baseline condition. Error bars indicate 95% confidence intervals (n=7).

4.4 Experiment 3.2 Tactile apparent motion with a half-paired audiotactile stream

To disassociate possible influences of auditory timing and auditory grouping, in Experiment

2, we only presented "asynchronous" beeps (as in Experiment 3.1), while omitting the "synchronous" beeps – in order to remove the potential for (intramodal) auditory grouping. With this manipulation, auditory beeps were paired only with one side (either the left or the right) of tactile taps, and the auditory-auditory SOA was 800 ms, while the (variable) audiotactile SOAs remained the same as in Experiment 3.1. If one, nevertheless, obtains the same pattern of effects as in Experiment 3.1 under these conditions (i.e., in the absence of auditory grouping), one would conclude that auditory timing is the main factor determining the audiotactile interaction.

4.4.1 Method

The same seven observers who had taken part in Experiment 3.1 also participated in Experiment 3.2 (for payment). The method was essentially the same as in Experiment 1, except that the "synchronous" beeps were removed in the audiotactile stream. That is, only every first tactile stimulus was paired with a sound (Figure 4.3).

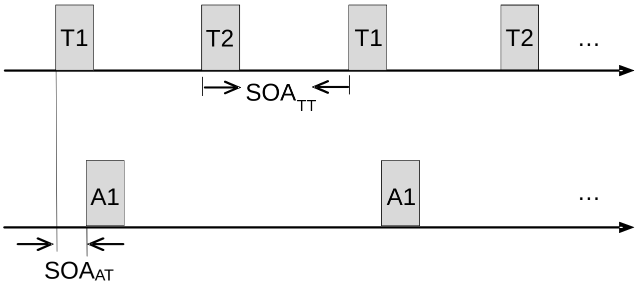


Figure 4.3. Schematic representation of the audiotactile timing conditions in Experiment 2. Auditory beeps were only paired with the taps from the initial side (either the left or the right). The tactile-tactile SOA was 400 ms and the auditory-auditory SOA 800 ms. The SOA of asynchronous audiotactile pairs (SOA$_{AT}$) was varied across trials.

4.4.2 Results and discussion

The mean normalized phase durations are shown in Figure 4.4. A pairwise t-test between the two perceived directions in the baseline condition (without beep) revealed no difference, t(6)=0.635, p=0.549. A repeated-measures ANOVA of the phase durations for "initial-direction" responses, with the single factor audiotactile SOA, failed to reveal a significant SOA effect, F(7,42) = 1.789, p=0.115. Similarly, there were no significant differences among audiotactile SOAs in the phase durations of "reverse-direction" responses, F(7,42)=1.965, p=0.083. However, with phase durations collapsed across all SOAs, a further ANOVA revealed a significant main effect of response ("initial direction" vs. "reverse direction"), F(1,6) = 21.113,

$p<0.01$. In contrast to Experiment 3.1, in Experiment 3.2, “initial-direction” responses were dominant across all seven audiotactile SOAs, regardless of auditory timing (the audiotactile SOA varied from -75 ms to 75ms). This suggests that the half-pairing auditory beeps created a “globally” dominant percept of motion direction from the side of the audiotactile stimuli to the side of the tactile-only stimuli.

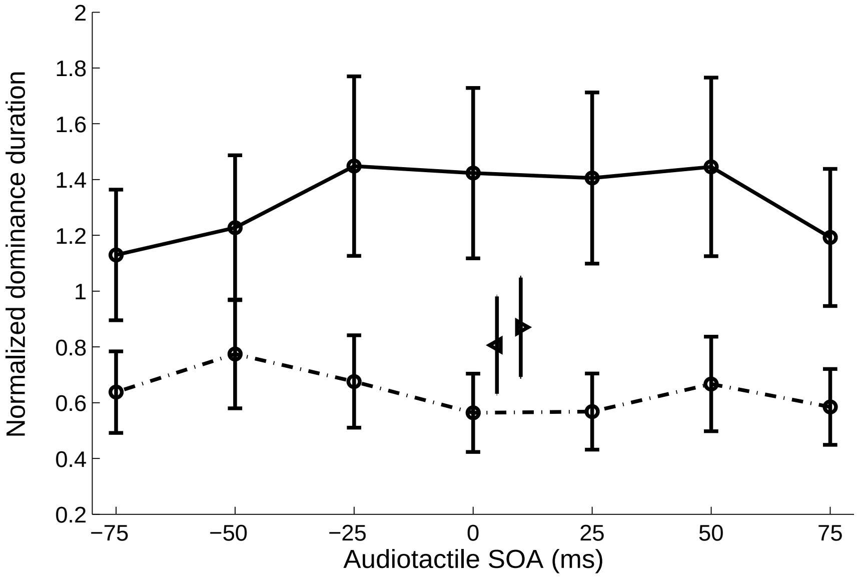


Figure 4.4 Normalized phase durations of tactile apparent motion as a function of audiotactile SOA (half-pairing condition). The solid line represents mean phase durations of the “initial direction”, the dotted line of the “reverse direction”. The rightward-pointing triangle denotes responses of “initial direction”, and the rightward-pointing triangle responses of “reverse direction”, for the baseline condition. Error bars indicate 95% confidence intervals (n=7).

At first glance, the failure to find a temporal-ventriloquism effect (of auditory timing) on tactile apparent motion in the half-pairing manipulation of Experiment 3.2 may relate to the “assumption of unity”, an internal estimation of the degree of concordance of the sensory inputs with a unitary source (Radeau and Bertelson 1978; Bertelson and Radeau 1981; Radeau and Bertelson 1987). Essentially, the more amodal properties are shared by events from different modalities, in particular space and time, the more likely that they originate from a common source or object (Welch and Warren 1980; Radeau 1994; Bertelson 1999; Welch 1999). The half-pairing configuration in Experiment 3.2 made the left and right events unbalanced (audiotactile event vs. tactile event), thus weakening the assumption of unity of audiotactile events, as well as the auditory influence on tactile apparent motion. However, as such, this could not explain the observed unanimous dominance of one direction of apparent motion across all audiotactile SOAs. Instead, we propose that, in the half-pairing condition, the audiotactile stimuli create salient events which may capture attention and thus serve as the starting point (cue) in motion direction, contributing to the "global" shift in tactile apparent-motion direction. This dominance of one direction of apparent motion is similar to the crossmodal line motion illusion (Shimojo, Miyauchi et al. 1997). In the demonstration of this illusion, a beep sound or an electric pulse ("cue") is presented on one side and followed by a visual line in proximity after this cue; this line is perceived to grow rapidly from the cued side even though both the cue and line are presented at the same time (the "line motion effect"). The crossmodal line motion effect has been attributed to a spatial-attentional bias induced by the auditory or tactile "cues". In our half-pairing condition, the auditory signals may have similarly served as a "cue" (though the auditory beep carried no spatial information), inducing one dominant motion direction, namely that starting from the tap paired with the cue. Such an attentional bias may have prevented a crossmodal temporal-ventriloquism effect from emerging.

4.5 Experiment 3. 3. Tactile apparent motion with a "shifted full-pairing" audiotactile stream

The results of Experiment 3.2 suggested that half-pairing beeps introduced an attentional bias on the tactile apparent motion, and unbalanced audiotactile sensory inputs prevented a temporal-ventriloquism effect (as seen in the full-pairing condition of Experiment 3.1) from occurring. However, it is still unknown how sensory weights or spatial-attentional shifts interact with auditory timing in the full-pairing condition. To examine the potential interactions, in Experiment 3.3, we manipulated the auditory temporal structure by shifting the auditory stimuli towards one tactile event with a full-pairing audiotactile stream, that is, we kept the even-numbered beeps and even-numbered taps synchronized, while varied the temporal distance between odd-numbered beeps and odd-numbered taps.The asynchronous audiotactile events were either closer (SOA=75ms) or faraway (SOA=325ms). Compared with the condition of one-to-one synchronized audiotactile events (control condition), we would expect to have opposite responses towards the perceived directions of tactile apparent motion in the two asynchronous conditions.

4.5.1 Method

The method was the same as in Experiment 3.1, with the following exceptions. 10 paid participants (6 females, average age 25.7 years) took part in the experiment (three of them had already taken part in the previous experiments). The experimental procedure was similar to Experiment 3.1, though the following three audiotactile SOAs were compared: 0, 75, and 325 ms (see Figures 4.5a and 4.5b). Note that with an audiotactile SOA of 325 ms, the first beep led the second tactile tap by 75 ms, while the second beep occurred synchronized with the second tap (thus, the auditory-auditory SOA was shorter, namely: 75 ms, in this condition). The audiotactile SOAs were randomized as in Experiment 3.1 and Experiment 3.2.

4.5. 2 Results and discussion

Figure 4.5c presents the mean phase durations for “initial-direction” and “reverse-direction” responses as a function of the (variable) audiotactile SOA. A repeated-measures ANOVA with the factors audiotactile SOA and response (“initial direction” vs. “reverse direction”) revealed the main effect of SOA to be significant, F(2, 18) = 4.677, $p<0.05$, while the main effect of response (just) failed to reach significance, F(1,9)=4.715, p=0.06. Thus, while auditory timing influenced the tactile apparent motion, responses exhibited no differences when the data were collapsed across all audiotactile SOAs; this was further confirmed by a non-significant difference between responses in the synchronous condition (audiotactile SOA of 0 ms), t(9)=1.566, p=0.152. However, the SOA x response interaction was significant, F(2,18)=12.784, $p<0.01$ – owing to opposite effects in the perceived directions of tactile motion with the 75-ms SOA (initial direction dominant; t(9)=4.212, $p<0.01$) versus the 325-ms SOA (reverse direction dominant; t(9)=-2.654, $p<0.05$) (see Figure 4.5c).

The reversal in the dominance of the perceived tactile-motion direction between the audiotactile SOAs of 75 ms and 325 ms is the most interesting result of Experiment 3.3. With an audiotactile SOA of 75 ms, the dominance pattern was the same as in Experiment 3.1 (initial direction dominant), which can be explained in terms of a temporal-ventriloquism effect. That is, the auditory timing shifted the tactile tap T1 close to T2, thus exerting an influence on the tactile timing – which, in turn, influenced the tactile apparent motion, yielding a dominant initial direction percept. By contrast, the opposite result with the audiotactile 325-ms SOA (reverse direction dominant) cannot be explained by auditory timing or auditory grouping. If the asynchronous beeps captured either the first or the second tactile tap, the auditory signal at the 325-ms SOA would still enhance the “initial-direction” percept, since the sound would also pull the second tap close to the first one, just as the sound at the 75-ms SOA pull the first tap close to the second tap. Similarly, merely auditory grouping would also lead to a dominance of “initial direction”, since, in a long-running temporal audio stream, the patterns of auditory interval variations and auditory groupings are similar in both the 75- and 325-ms conditions, that is, both audio streams were composed by numbers of audio pairs of alternated short and long intervals. However, the results showed the opposite, “reverse direction” dominance. A possible explanation of this is that at 325-ms SOA condition, two beeps, which are both close to the even taps, enhance the sensory inputs of the even audiotactile events. Similar to the conditions in Experiment 3.2, unbalanced weights on sensory inputs give rise to an attentional bias influencing the tactile apparent motion – that is, the tactile stimuli which are more strongly weighted attentionaly by the auditory events (T2 in Figure 4.5b) are likely to be perceived as the starting position of the apparent motion, making the percept of “reverse direction” dominant.

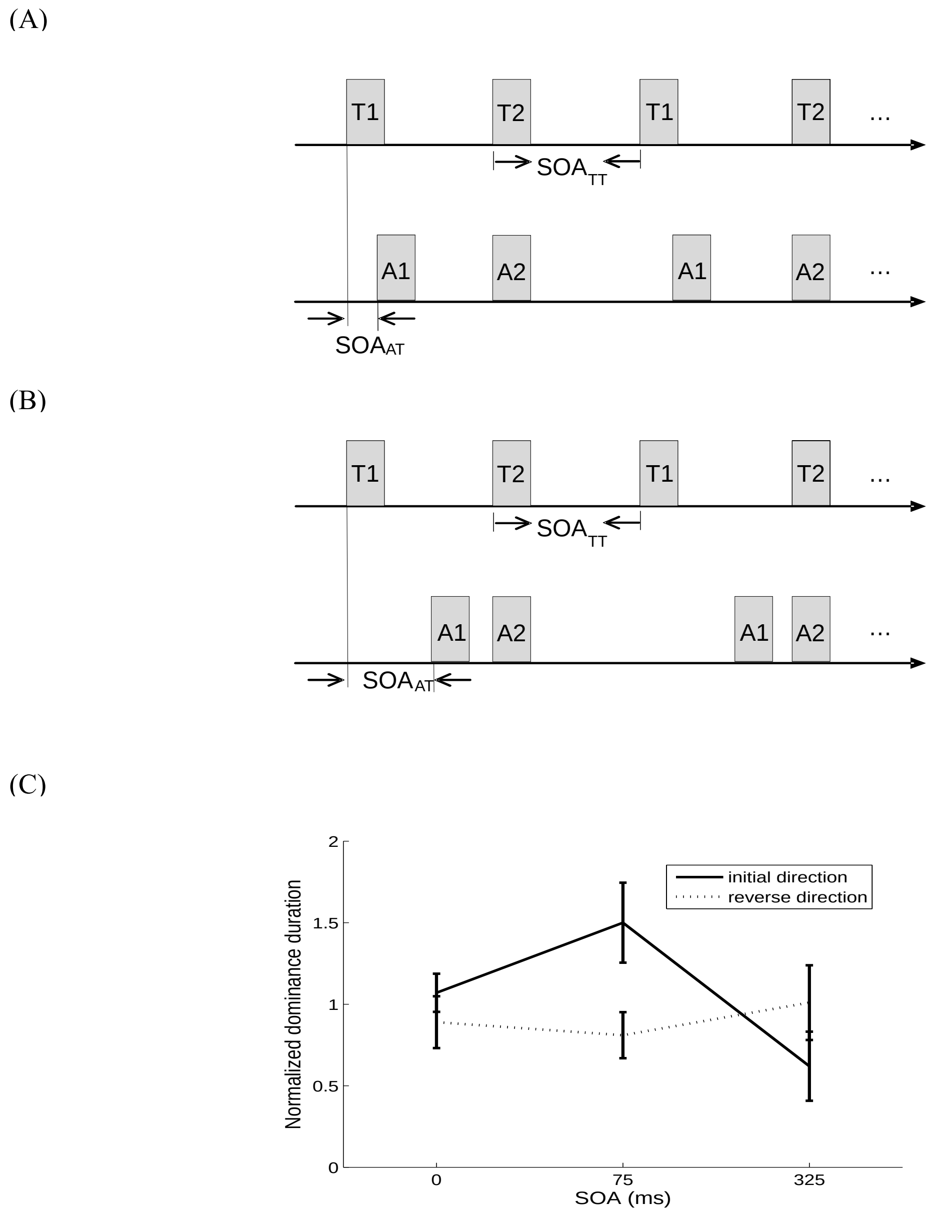


Figure 4.5. (a) Full-pairing audiotactile stream with SOA_{AT} = 75 ms. (b) Reversed full-pairing audiotactile stream with SOA_{AT} = 325 ms. (c) Normalized phase duration of tactile apparent motion as a function of audiotactile SOA. The error bars indicate 95%

confidence intervals (n=10).

4.6 General Discussion

With a full-pairing audiotactile configuration (Experiment 3.1), we modulated audiotactile asynchronies from -75 ms (beep leading tap) to 75 ms (beep trailing tap) in one pair, while keeping the second pair synchronous. We observed ambiguous rivalry of bistable tactile apparent motion to be systematically influenced by the audiotactile asynchrony. Interestingly, this systematic modulation disappeared when the synchronous beeps were removed (Experiment 3.2). Instead, with this half-pairing configuration, a global bias of apparent motion from the side of audiotactile stimuli to the side of tactile-only stimuli was observed. This suggests that differential weights associated with sensory inputs induce an attentional bias which prevents crossmodal interaction in motion perception. In a long-running audiotactile stream with a shifted full-pairing audiotactile configuration (325-ms SOA, Experiment 3.3), we found the reverse auditory capture effect to that in the "full-paring" condition of Experiment 3.1 or 3.3 (75-ms SOA). We have ruled out the independent roles of audiotactile temporal ventriloquism and merely the auditory groupings in inducing these capture effects. In full-paring audiotactile conditions (Experiment 3.1 and Experiment 3.3), the manipulations of auditory temporal structure lead to different perceived directions of tactile apparent motion. This indicates that auditory temporal structure plays an important role in audiotactile apparent motion interactions.

The "assumption of unity" is often thought to play an crucial role in crossmodal interaction and integration, since integration across two modalities makes sense only when the perceptual system has evidence that the events from two modalities originated from a common source (Welch and Warren 1980). In terms of the assumption of unity, with the half-pairing configuration, one would expect the perceived motion starts from one side of tap and goes across the linking sound "emitted" from that tap to the other side. Thus, depending on which tap is the 'source' emitting the sound, two opposite dominant tactile motion directions should be observed when varying the audiotactile asynchrony. However, this account failed to predict the results actually observed with the half-pairing configuration, that is, the global dominance of tactile motion from the side of the audiotactile stimuli to the side of tactile-only stimuli, which was independent of the audiotactile SOA. Besides the assumption of unity, the modality precision hypothesis has been widely used for explaining multisensory interaction in crossmodal conflict or inconsistent situations (Welch and Warren 1980; Welch, DuttonHurt et al. 1986), that is, perception is dominated by the information from the modality with the higher precision. This notion has been formally developed over the last decade drawing on Bayesian integration models (Ernst and Banks 2002; Alais and Burr 2004), which do provide a good quantitative account of crossmodal ventriloquism effects. However, only based on the modality precision hypothesis, it is not possible to explain the global dominance of apparent motion in one direction for the half-pairing configuration

We surmised that half-pairing beeps on one temporal location may induce an attention shift towards stimuli on that location (although observers had been instructed not to pay attention to the sounds), which leads to a systematic auditory capture effect on tactile apparent motion. Such an attention shift under conditions of ongoing tactile motion rivalry may well inhibit any modulation by auditory timing (i.e., the temporal-ventriloquism effect). Experiment 3.3 further demonstrated the attentional shift by varying auditory temporal structure in tactile apparent motion. It is known that crossmodal integration takes place within a certain, limited temporal and spatial range (Slutsky and Recanzone 2001; Alais and Burr 2004; Gepshtein, Burge et al. 2005; Bresciani, Dammeier et al. 2006; Shi, Chen et al. 2010). Accordingly, we shifted two paired beeps to one side of taps, weakening the full-pairing crossmodal integration and increasing the attentional bias. With an audiotactile SOA of 75 ms (Figure 4.5a), beeps were temporally closely aligned with alternatively separate tactile taps. Thus, sensory inputs from both sides were similar and a temporal-ventriloquism effect was triggered with asynchronous auditory timing. With an audiotactile SOA of 325 ms (Figure 4.5b), two beeps were temporally aligned with taps from one side, yet the audiotactile SOA between the closest audiotactile stimuli was preserved at 75 ms while the asynchronous beep still occurred in the interval between T1 and T2. However, in contrast to the condition of the audiotactile SOA of 75 ms, an opposite capture pattern was attained. We suggested that, similar to the half-pairing configuration, the shifted full-pairing condition induced an attentional bias towards the audiotactile side, as a result of which a dominant apparent motion starting from this (attended) side was observed. This is consistent with previous findings of audiovisual or tactile-visual "line motion effects" (Shimojo, Miyauchi et al. 1997), where a robust line motion from the auditory or tactile cued side was observed. Similar attentional modulations of apparent motion have been reported in the literature (Cavanagh 1992; Lu and Sperling 1995; Shimojo, Miyauchi et al. 1997; Berman and Colby 2002). Consistent with the results of Experiment 3.2, differential weighting of sensory inputs on the two sides with the audiotactile SOA of 325 ms may have prevented the temporal-ventriloquism effect from occurrimng. Taking these findings together, we propose that with long-running ambiguous / bistable tactile apparent motion, the unbalanced audiotactile input conditions lead to a variation of attentional saliency, and subjects' tracking of this saliency, resulted in the different auditory capturing tactile apparent motion effect, rather than the contributions from the tactile events themselves (since the tactile motion is bistable in the control condition) or merely the audiotactile asynchronies.

In summary, using audiotactile configurations and tactile rivalry apparent motion, we found a systematic influence of auditory timing on tactile apparent motion in a balanced full-pairing condition. A global bias on tactile apparent motion was observed and the temporal-ventriloquism effect was abolished in unbalanced (half-pairing) configurations and diminished in temporally shifted full-pairing conditions. The "assumption of unity" or the modality precision hypothesis alone cannot well explain this pattern of audiotactile temporal interactions. We suggest that auditory structure may induce an attentional-bias influence on tactile apparent motion rivalry. It would be interesting to compare the reversed interaction, that is, the influence of touch modulations on auditory apparent motion rivalry, to test and generalize our hypothesis of an attentional saliency modulation in crossmodal temporal capture in the apparent-motion paradigm.

Chapter 5 Deutsche Zusammenfassung

Der Grossteil unserer Erfahrungen entsteht aus der Interaktion unterschiedlicher Sinnes-Modalitäten. In solcher kreuzmodaler Interaktion scheinen zwei Bedingungen eine zentrale Rolle für die Wahrnehmung zu spielen: die Identität der Ereignisse und die räumlichen und zeitlichen Beziehungen zwischen den Ereigessen.

Von Interaktionen zwischen dem visuellen, auditorischen und taktilen Sinn wird in der Literatur vielfach berichtet. Diese Interaktionen wurden auf verschiedenste Weisen, wie zum Beispiel durch Manipulation der scheinbaren Reihenfolge zweier Ereignisse , deren subjektive Dislokalisierung , und die Anzahl wahrgenommener Ereignisse nachgewiesen. Ein klassisches und interessantes Phänomen ist der Bauchredner- oder Ventriloquismus-Effekt (Aschersleben, 2003; Bertelson, 2003; Vroomen & de Gelder, 2000; Vroomen, 2006). Beispielsweise wird der wahrgenommene Ursprungsort eines Tones durch die gleichzeitige Darbietung eines Lichtreizes in Richtung der Lichtquelle verschoben (Howard and Templeton, 1966).

Den Ventriloquismus-Effekt kann man auch für zeitliche Verschiebungen nachweisen. Eine Reihe von Studien zeigten den Ventriloquismus-Effekt in der Zeit, etwa bei Reihenfolgebestimmungen auditiver und visueller Reize oder beim Erfassen einer Distraktormodalität während der Richtungsbestimmung einer kurzzeitigen Scheinbewegung in der Zielmodalität (Morein-Zamir, 2003; Soto-Faraco et al.,2004; Vroomen & Keetels, 2006; Getzmann,2007; Bruns & Getzmann,2008). Jedoch ist der Erfassungseffekt in der Distraktormodalität in den meisten Studien räumlich durch die Methode der unmittelbaren Wirkung eingegrenzt. Beim zeitlichen Ventriloquismus-Effekt, der kürzlich in der audio-visuellen Variante erforscht wurde, ist die Zeitspanne zwischen Reizdarbietungen normalerweise zu groß um eine wahre kreuzmodale Integration zu erwirken, während die Wahrnehmung bei Bewegungsstatusbeurteilungen von Natur aus nicht eindeutig ist.

In dieser Arbeit werde ich versuchen zwei Fragen über die zeitliche Erfassung zwischen zwei verschiedenen Sinnesmodalitäten zu beantworten, nämlich wann und wo diese Erfassung stattfindet. Diese Frage wurde mit Hilfe zweier Paradigmen der Scheinbewegung untersucht: das eine ist die visuell-taktile Ternus-Illusion über kurze Zeitspannen (Harrar & Harris, 2007; Pantle & Picciano, 1976; Ternus, 1926) und das andere eine taktile Scheinbewegung übere längere Zeitspannen.

Der typische Ternus-Effekt entsteht durch die Darbietung zweier rasch aufeinanderfolgender visueller (taktiler) Fenster, wobei ein Fenster jeweils aus zwei horizontal benachbarten visuellen (taktilen) Reizen besteht. Das erste und das zweite Fenster sind räumlich leicht gegeneinander verschoben, teilen sich aber durch teilweise Überlagerung den mittleren Reiz. Je nach zeitlichem Abstand zwischen den Darbietungen (Interstimulusintervall, ISOI) enstehen meist zwei unterschiedliche Wahrnehmungen: das Springen des aussenliegenden Elements von einem Fenster in das nächste (‘Einzelbewegung’) oder die parallele Verschiebung beider Elemente von einem Fenster in das nächste (‘Gruppenbewegung’). Diese zwei möglichen Wahrnehmungen des Ternusdisplays schliessen sich gegenseitig aus und die Spanne der ISOI für den Übergang von der Wahrnehmung einer ‘Einzelbewegung’ zu der einer ‘Gruppenbewegung’ ist relativ breit. Aufgrund dieser Eigenschaften bietet sich das Ternus-Display als eine vielversprechende Methode an, das Phänomen der kreuzmodalen zeitlichen Erfassung zu untersuchen. Das zweite Paradigma nützt längerfristige (90 Sekunden) taktile Scheinbewegungen, welche durch das scheinbare Wandern - ohne physikalische Bewegung - zweier taktiler Finger-Tipp-Reize von einer Seite zur anderen Seit entstehen.

1 Auditorisches Erfassen des visuell-taktilen Ternus-Effekts

Das klassische Ternus-Display besteht aus einem mehrteiligen Reiz, der die Wahrnehmung einer von zwei Scheinbewegungen auslösen kann: "Einzelbewegung" oder "Gruppenbewegung."

In Experiment 1 wurden auditive Konfigurationen systematisch moduliert, um den auditiven Erfassungseffekt beim visuell-taktilen Ternus-Effekt zu untersuchen. Die Studie zeigte einen Tongruppierungseffekt, was bedeutet dass zwei Töne, die in zeitlicher Nähe (30ms) zu den beiden visuell-taktilen Fenstern des Ternus-Displays gezeigt wurden, die Übergangsschwelle zur Scheinbewegungswahrnehmung verschoben. Zum Beispiel entstand bei der Paarung von “äußeren“ Tönen mit den beiden visuellen Fenstern, also einer Tondarbietung kurz vor dem ersten und kurz nach dem zweiten visuellen Reiz, der Eindruck einer Gruppenbewegung. "Innen" präsentierte Töne hingegen erweckten den Eindruck einer Einzelbewegung. Die Darbietung eines einzelnen Tons in zeitlicher Nähe zu einem der beiden visuellen Fenster hatte keinen Einfluss wahrgenommene Art der Scheinbewegung. In einer weiteren Konfiguration wurden die beiden Töne in einer Reihe anderer Töne eingebettet, wobei die zwei Töne wieder in zeitlicher Nähe zu den visuellen Reizen lagen. Verglichen mit der Darbietung ohne zusätzliche Töne war hier der zeitliche Ventriloquismus-Effekt (ZVE) signfikant kleiner. Dies legt nahe, dass nur effektiv gepaarte Töne erfasst werden können, wohingegen das bloße gleichzeitige Darbieten von Tönen und visuellen Reizen nicht zur Gänze den ZVE erklären kann. Ein weiteres Experiment zeigte dass das subjektive Intervall zwischen zwei Tönen länger ist als ein physikalisch gleiches Inter-Ton-Intervall in einer Reihe von Tönen. Dies deutet darauf hin, dass das subjektive auditive Intervall eine kritische Rolle beim kreuzmodalen zeitlichen Erfassen spielen kann.

2 Asymmetrien in der kreuzmodalen Interaktion und Priming der Ternus illusion

Es wurde bereits gezeigt, dass die Wahrnehmung visueller Scheinbewegung von visuellen 'priming probes' und gleichzeitigen auditiven Reizen beeinflusst wird. In diesem Kapitel untersuchte ich die Priming-Effekte und kreuzmodale Interaktionen mithilfe eines 'three-token' Ternusdisplays.

Um den Priming-Effekt hervorzurufen wurde der mittlere visuelle Reiz, der also in beiden Fenstern des Ternusdisplays auftaucht, zweimal vor dem normalen Ternusdisplay angezeigt. Das Inter-Stimulus-Intervall (ISI) zwischen Prime und Probe war dasselbe wie das des Ternusdisplays. Dieses visuelle 'Token'-Priming verursachte in der Ternus-Illusion eine starke Wahrnehmung von ’Einzelbewegung’, bei dem der Punkt der subjektiven Gleichheit (point of subjective equality, PSE) um 35 ms anstieg, wohingegen taktiles Priming die Diskriminierung der beiden visuellen Bewegungswahrnehmungen verschlechterte, und den kleinsten bemerkbaren Unterschied (just noticeable difference, JND) um 5 ms erhöhte. Zudem wurde gezeigt, dass synchrone visuell-taktile Ereignisse in der Ternus-Illusion zur vermehrten Wahrnehmung von ‘Gruppenbewegung’ führten indem sie den PSE um 34 ms verlangsamten und die Sensitivität für visuelle Bewegungen erhöhten (die JND verringerte sich um 6 ms). Auf die taktile Ternusillusion zeigten weder das visuelle Priming noch synchrone visuelle Ereignisse einen Einfluss. Diese Asymmetrie in kreuzmodalem Priming und kreuzmodaler Interaktion lässt sich über die Erregbarkeit einer Sinnesmodaliät wie auch ihre Widerstandfähigkeit gegenüber den Einflüssen anderer Modalitäten, die auf biologisch begründeten Beschränkungen der zwei sensorischen Systeme beruhen erklären.

3 Das Hören beeinflusst taktile Scheinbewegungen

In diesem Kapitel habe ich gezeigt, dass in über längere Spannen präsentierten taktilen Scheinbewegungen eine Wahrnehmungsrivalität der Bewegungsrichtung durch sukzessive taktile Scheinbewegung induziert werden kann, die sich wieder auflösen lässt, indem man zeitnah (also mit einer 'stimulus-onset-asynchrony', oder SOA, von weniger als 75 ms) zu den taktilen Reizen Töne präsentiert. Diese kurze SOA-Spannweite dient als zeitliches Integrationsfenster und wurde in einfachen audiotaktilen Reihenfolgebestimmungs- und Frequenzdiskriminierungsaufgaben gemessen. Die gruppierenden Töne enthalten dabei keinerlei Information über die räumliche Lokalisation. Die tonerfassende taktile Scheinbewegung ist robust und nimmt eine eigene Stufe in der Wahrnehmung ein.

Zusammenfassend legen die Ergebnisse dieser Arbeit nahe, dass die temporale kreuzmodale Erfassung ein robustes Phänomen ist, und dass es auf einer perzeptuellen Ebene und innerhalb eines kurzen zeitlichen Integrationsfensters stattfindet. Eine weiterführende elektrophysiologische oder bildgebende Herangehensweise mag der Weg sein, die entsprechende Gehirndynamik aufzudecken und aufzuklären, wie diese temporale Kreuzerfassung entsteht und sich entwickelt.

References

Alais, D. and D. Burr (2004). "The ventriloquist effect results from near-optimal bimodal integration." Current Biology 14(3): 257-262.

Alink, A., W. Singer, et al. (2008). "Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex." Journal of Neuroscience 28(11): 2690-2697.

Allen, P. G. and P. A. Kolers (1981). "Sensory specificity of apparent motion." Journal of Experimental Psychology: Human Perception and Performance 7(6): 1318-1328.

Aschersleben, G. and P. Bertelson (2003). "Temporal ventriloquism: crossmodal interaction on the time dimension. 2. Evidence from sensorimotor synchronization." International Journal of Psychophysiology 50(1-2): 157-163.

Battaglia, P. W., R. A. Jacobs, et al. (2003). "Bayesian integration of visual and auditory signals for spatial localization." Journal of the Optical Society of America A:optics, image science, and vision 20(7): 1391-1397.

Berman, R. A. and C. L. Colby (2002). "Auditory and visual attention modulate motion processing in area MT+." Cognitive Brain Research 14(1): 64-74.

Bermant, R. I. and R. B. Welch (1976). "Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition." Perceptual and Motor Skills 42(43): 487-493.

Bertelson, P. (1999). Ventriloquism: A case of crossmodal perceptual groupings. Cognitive Contributions to the Perception of Spatial and Temporal Events Eds G Aschersleben, T Bachmann,JMusseler (Amsterdam: Elsevier Science). G. Aschersleben, T. Bachmann and J. Müssler. Amsterdam, Elsevier Science: 347 -362.

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

Bertelson, P. and M. Radeau (1981). "Cross-modal bias and perceptual fusion with auditory-visual spatial discordance." Perception & Psychophysics 29(6): 578-584.

Braddick, O. J. (1980). ""Low-level and high-level processes in apparent motion"." Philosophical Transactions of the Royal Society of London,Series B 290: 137-151.

Brainard, D. H. (1997). "The Psychophysics Toolbox." Spatial Vision 10(4): 433-436.

Brecher, G. A. (1932). "Die Enstehung und biologische Bedeutung der subjectiven Zeiteinheit--des Momentes The origin and biological significance of the subjective temporal unit: the "moment." " Zeitschrift für Vergleichende Physiologie 18: 204-243.

Breitmeyer, B. (1984). Visual Masking: An Integrative Approach. New York, Oxford University Press.

Breitmeyer, B. and H. Ogmen (2006). Visual Masking: Time slices through conscious and unconscious vision. New York, Oxford University Press.

Breitmeyer, B. G. and A. Ritter (1986). "The role of visual pattern persistence in bistable stroboscopic motion." Vision Research 26(11): 1801-1806.

Breitmeyer, B. G. and A. Ritter (1986). "Visual persistence and the effect of eccentric viewing, element size, and frame duration on bistable stroboscopic motion percepts." Perception & Psychophysics 39(4): 275-

280.

Bresciani, J. P., F. Dammeier, et al. (2006). "Vision and touch are automatically integrated for the perception of sequences of events." Journal of Vision 6(5): 554-564.

Bresciani, J. P. and M. O. Ernst (2007). "Signal reliability modulates auditory-tactile integration for event counting." Neuroreport 18(11): 1157-1161.

Bresciani, J. P., M. O. Ernst, et al. (2005). "Feeling what you hear: auditory signals can modulate tactile tap perception." Experimental Brain Research 162: 172-180.

Bresciani, J. P., Ernst,M.O (2007). "Signal reliability modulates auditory-tactile integration for event counting." Neuroreport 18(11): 1157-1161.

Bresciani, J. P., Ernst,M.O., Drewing,K., Bouyer,G.,Maury,V.,& Kheddar,A (2005). "Feeling what you hear: auditory signals can modulate tactile tap perception." Experimental Brain Research 162: 172-180.

Bruns, P. and S. Getzmann (2008). "Audiovisual influences on the perception of visual apparent motion: exploring the effect of a single sound." Acta psychologica 129(2): 273-283.

Burr, D., M. S. Banks, et al. (2009). "Auditory dominance over vision in the perception of interval duration." Experimental Brain Research 198(1): 49-57.

Caclin, A., E. N. Supérieure, et al. (2002). "Tactile “capture” of audition." Perception & Psychophysics 64(4): 616-630.

Calvert, G. A., C. Spence, et al. (2004). "The Handbook of multisensory processes, MIT Press., Cambridge, MA."

Carles, E. (2007). "The mismatch negativity 30 years later: How far have we come?" Journal of Psychophysiology 21(3-4): 129-132.

Carter, O. Konkle, T., Wang, Q, Hayward, V., & Moore, C. Tactile Rivalry Demonstrated with an Ambiguous Apparent-Motion Quartet. Current Biology, 2008;18(14),1050-1054

Cavanagh, P. (1992). "Attention-based motion perception." Science 257(5076): 1563-1565.

Choe, C. S., R. B. Welch, et al. (1975). "The "ventriloquist effect": Visual dominance or response bias?" Perception & Psychophysics 18: 55-60.

Colin, C., M. Radeau, et al. (2002). "Electrophysiology of spatial scene analysis: the mismatch negativity (MMN) is sensitive to the ventriloquism illusion." Clinical Neurophysiology 113(4): 507-518.

Cook, L. A. and D. L. Van Valkenburg (2009). "Audio-visual organisation and the temporal ventriloquism effect between grouped sequences: evidence that unimodal grouping precedes cross-modal integration." Perception 38(8): 1220-1233.

Craig, J. C. (2006). "Visual motion interferes with tactile motion perception." Perception 35(3): 351-367.

Dawson, M. R., N. Nevin-Meadows, et al. (1994). "Polarity matching in the Ternus configuration." Vision Research 34(24): 3347-3359.

Deco G.,Scarano L., Soto-Faraco S. Weber’s Law in decision making: integrating behavioral data in humans with a neurophysiological model. Journal of Neuroscience 2007;27:11192-11200

Dixon, N. F. and L. Spitz (1980). "The detection of auditory visual desynchrony." Perception 9: 719-721.

Driver, J. and T. Noesselt (2008). "Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments." Neuron 57(1): 11-23.

Eimer, M., J. van Velzen, et al. (2002). "Cross-modal interactions between audition, touch, and vision in endogenous spatial attention: ERP evidence on preparatory states and sensory modulations." Journal of Cognitive Neuroscience 14(2): 254-271.

Ekroll, V., F. Faul, et al. (2008). "Classification of apparent motion percepts based on temporal factors." Journal of Vision 8(4): 1-22.

Elliott, M. A., Z. Shi, et al. (2007). "The effects of subthreshold synchrony on the perception of simultaneity." Psychological Research 71(6): 687-693.

Ernst, M. O. and H. H. Bülthoff (2004). "Merging the senses into a robust percept " Trends in Cognitive Sciences 8(4): 162-169.

Ernst, M. O. and M. S. Banks (2002). "Humans integrate visual and haptic information in a statistically optimal fashion." Nature 415(6870): 429-433.

Evans, P. M. (1987). "Vibrotactile masking: temporal integration, persistence, and strengths of representations." Perception & Psychophysics 42(6): 515-525.

Fendrich, R., & Corballis,P.M (2001). "The temporal cross-capture of audition and vision." Perception & Psychophysics 63: 719–725.

Fisher, G. H. (1968). "Agreement between the spatial senses." Perceptual and Motor Skills, 126: 849-850.

Fraisse, P. (1984). "Perception and estimation of time" Annual Review of Psychology 35: 1-36.

Freeman, E. and J. Driver (2008). "Direction of visual apparent motion driven solely by timing of a static sound." Curr Biol 18(16): 1262-1266.

Gallace, A. and C. Spence (2008). "The cognitive and neural correlates of "tactile consciousness": a multisensory perspective." Consciousness and Cognition 17(1): 370-407.

Galli, P. A. (1932). "Uber mittelst verschiedener Sinnesreize erweckte Wahrnehmung von Scheinbewegungen [On the perception ofapparent motion elicited by different sensory stimuli]." Archiv fur die gesamte Psychologie 85: 137-180.

Gebhard, J. W. and G. H. Mowbray (1959). "On discriminating the rate of visual flicker and auditory flutter." American Journal of Psychology 72: 521-529.

Gepshtein, S., J. Burge, et al. (2005). "The combination of vision and touch depends on spatial proximity." Journal of Vision 5(11): 1013-1023.

Gescheider, G. A., J. H. Wright, et al. (2009). Information-processing channels in the tactile sensory system: A psychophysical and physiological analysis New York, Taylor & Francis Group, LLC.

Getzmann, S. (2007). "The effect of brief auditory stimuli on visual apparent motion." Perception 36(7): 1089-1103.

Gilbert, G. M. (1938). "A study in inter-sensory Gestalten." Psychological Bulletin 35: 698.

Gilbert, G. M. (1941). "Inter-sensory facilitation and inhibition." Journal of General Psychology 24: 381-407.

Gillmeister, H. and M. Eimer (2007). "Tactile enhancement of auditory detection and perceived loudness." Brain Research 1160: 58-68.

Goebel, R. and N. van Atteveldt (2009). "Multisensory functional magnetic resonance imaging: a future perspective." Experimental Brain Research 198(2-3): 153-164.

Goldstone, S. and J. L. Goldfarb (1964a). "Auditory and Visual Time Judgment " Journal of General Psychology 70: 369-387.

Goldstone, S. and J. L. Goldfarb (1964b). "Direct Comparison of Auditory and Visual Durations." Journal of Experimental Psychology 67: 483-485.

Goldstone, S. and W. T. Lhamon (1974). "Studies of auditory-visual differences in human time judgment. 1. Sounds are judged longer than lights." Perceptual and motor skills 39(1): 63-82.

Gray, R. and H. Z. Tan (2002). "Dynamic and predictive links between touch and vision." Experimental Brain Research 145(1): 50-55.

Grossberg, S. (1991). "Why do parallel cortical systems exist for the perception of static form and moving form?" Perception & Psychophysics 49(2): 117-141.

Grossberg, S. and M. E. Rudd (1992). "Cortical dynamics of visual motion perception: short-range and long-range apparent motion." Psychological Review 99(1): 78-121.

Hagen, M. C., O. Franzen, et al. (2002). "Tactile motion activates the human middle temporal/V5 (MT/V5) complex." European Journal of Neuroscience 16(5): 957-964.

Harrar, V. and L. R. Harris (2007). "Multimodal Ternus: visual, tactile, and visuo-tactile grouping in apparent motion." Perception 36(10): 1455-1464.

Hartmann, E., B. Lachenmayr, et al. (1979). "The peripheral critical flicker frequency." Vision Research 19(9): 1019-1023.

He, Z. J. and T. L. Ooi (1999). "Perceptual organization of apparent motion in the Ternus display." Perception 28(7): 877-892.

Holmes, N. P. (2009). "The principle of inverse effectiveness in multisensory integration: some statistical considerations." Brain Topography 21(3-4): 168-176.

Hotting, K. and B. Roder (2004). "Hearing cheats touch, but less in congenitally blind than in sighted individuals." Psychological Science 15(1): 60-64.

Hotting, K., F. Rosler, et al. (2004). "Altered auditory-tactile interactions in congenitally blind humans: an event-related potential study." Experimental Brain Research 159(3): 370-81.

Howard, I. P. and W. B. Templeton (1966). "Human Spatial Orientation.Wiley, New York, NY."

Jack, C. E. and W. R. Thurlow (1973). "Effects of degree of visual association and angle of displacement on the "ventriloquism" effect." Perceptual and motor skills 37(3): 967-79.

Jaekl, P. M. and L. R. Harris (2007). "Auditory-visual temporal integration measured by shifts in perceived temporal location." Neuroscience Letters 417(3): 219-224.

Jousmaki, V. and R. Hari (1998). "Parchment-skin illusion: sound-biased touch." Current Biology 8(6): R190.

Kandel, E..(2000). Principle of neural science, edited by Eric, R; Schwartz, J; Jessell, T. New York:McGraw-Hill, Health Professions Division.

Kaufman, L. (1974). "Sight and mind. New York: Oxford University Press."

Keetels and J. Vroomen (2008). "Tactile-visual temporal ventriloquism:No effect of spatial disparity."

Perception & Psychophysics 70(5): 765-771.

Keetels, M., J. Stekelenburg, et al. (2007). "Auditory grouping occurs prior to intersensory pairing: evidence from temporal ventriloquism." Experimental Brain Research 180(3): 449-456.

Kennett, S., M. Eimer, et al. (2001). "Tactile-visual links in exogenous spatial attention under different postures: convergent evidence from psychophysics and ERPs." Journal of Cognitive Neuroscience 13(4): 462-478.

Kirman, J. H. (1974). "Tactile apparent movement: the effects of number of stimulators." Journal of Experimental Psychology 103(6): 1175-1180.

Kitagawa, N. and S. Ichihara (2002). "Hearing visual motion in depth." Nature 416(6877):172-174.

Kolers, P. A. (1963). "Some differences between real and apparent visual movement." Vision Research 3:191-206.

Korte, A. (1915). "Kinematoskopische Untersuchungen." Zeitschrift fur Psychologie 72: 194- 296.

Kramer, P. and S. Yantis (1997). "Perceptual grouping in space and time: evidence from the Ternus display." Perception & Psychophysics 59(1): 87-99.

Levitin, D. J., K. MacLean, et al. (2000). "The perception of cross-modal simultaneity." International Journal of Computing and Anticipatory Systems: 323-329.

Lu, Z. L. and G. Sperling (1995). "Attention-generated apparent motion." Nature 377(6546): 237-9.

Lyons, G., D. Sanabria, et al. (2006). "The modulation of crossmodal integration by unimodal perceptual grouping: a visuotactile apparent motion study." Experimental Brain Research 174(3): 510-516.

Macaluso, E., C. D. Frith, et al. (2000). "Modulation of human visual cortex by crossmodal spatial attention." Science 289(5482): 1206-1208.

Mamassian, P. and R. Goutcher (2005). "Temporal dynamics in bistable perception." Journal of Vision 5(4): 361-375.

Massaro, D. W., M. M. Cohen, et al. (1996). "Perception of asynchronous and conflicting visual and auditory speech. ." The Journal of the Acoustical Society of America 100(3): 1777-1786.

Maunsell, J. H. R. and D. C. Van Essen (1983a). "Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation." Journal of Neurophysiology 49: 1127-1147.

Maunsell, J. H. R. and D. C. Van Essen (1983b). "Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity." Journal of Neurophysiology 49: 1148-1167.

McGurk, H. and J. MacDonald (1976). "Hearing lips and seeing voices." Nature 264(5588): 746-748.

Meyer, G. F. and S. M. Wuerger (2001). "Cross-modal integration of auditory and visual motion signals." Neuroreport 12(11): 2557-2560.

Morein-Zamir, S., S. Soto-Faraco, et al. (2003). "Auditory capture of vision: examining temporal ventriloquism." Cognitive Brain Research 17(1): 154-163.

Munhall, K. G., P. Gribble, et al. (1996). "Temporal constraints on the McGurk effect." Perception & Psychophysics 58(3): 351-362.

Navarra, J., J. Hartcher-O'Brien, et al. (2009). "Adaptation to audiovisual asynchrony modulates the speeded detection of sound." Proceedings of the National Academy of Sciences USA 106(23): 9169-9173.

Neuhaus, W. (1930). " Experimentelle Untersuchung der Scheinbewegung." Archiv für die gesamte Psychologie 75: 315-458.

O'Leary, A. and G. Rhodes (1984). "Cross-modal effects on visual and auditory object perception." Perception & Psychophysics 35(6): 565-569.

Occelli, V., C. Spence, et al. (2009). "The effect of sound intensity on the audiotactile crossmodal dynamic capture effect." Experimental Brain Research 193(3): 409-419.

Pantle, A. J. and J. T. Petersik (1980). "Effects of spatial parameters on the perceptual organization of a bistable motion display " Perception & Psychophysics 27(4): 307-312.

Pantle, A. J. and L. Picciano (1976). "A multistable movement display: evidence for two separate motion systems in human vision." Science 193(4252): 500-502.

Parise, C. and C. Spence (2008). "Synesthetic congruency modulates the temporal ventriloquism effect." Neuroscience Letters 442(3): 257-261.

Pastukhov, A. and J. Braun (2007). "Perceptual reversals need no prompting by attention." Journal of Vision 7(10): 1-17.

Pelli, D. G. (1997). "The VideoToolbox software for visual psychophysics: Transforming numbers into movies." Spatial Vision 10: 437-442.

Petersik, J. T. and A. J. Pantle (1979). "Factors controlling the competing sensations produced by a bistable stroboscopic motion display." Vision Research 19(2): 143-154.

Petersik, J. T. and C. M. Rice (2006). "The evolution of explanations of a perceptual phenomenon: a case history using the Ternus effect." Perception 35(6): 807-821.

Petersik, J. T. and C. M. Rice (2008). "Spatial correspondence and relation correspondence: grouping factors that influence perception of the Ternus display." Perception 37(5): 725-739.

Pick, H. L., D. H. Warren, et al. (1969). "Sensory conflict in judgments of spatial direction." Perception & Psychophysics 6: 203-205.

Posner, M. I., M. J. Nissen, et al. (1976). "Visual dominance: An information-processing account of its origins and significance. ." Psychological Review 83(2): 157-171.

Radeau, M. (1994). "Auditory-visual spatial interaction and modularity." Cahiers de Psychologie Cognitive 13: 3-51.

Radeau, M. and P. Bertelson (1978). "Cognitive factors and adaptation to auditory-visual discordance." Perception & Psychophysics 23(4): 341-343.

Radeau, M. and P. Bertelson (1987). "Auditory-visual interaction and the timing of inputs. Thomas (1941) revisited." Psychological Research 49(1): 17-22.

Rahne, T., S. Deike, et al. (2008). "A multilevel and cross-modal approach towards neuronal mechanisms of auditory streaming." Brain Research 1220: 118-131.

Reisbeck, T. E. and K. R. Gegenfurtner (1999). "Velocity tuned mechanisms in human motion processing." Vision Research 39: 3267-3285.

Sanabria, D., J. Lupianez, et al. (2007). "Auditory motion affects visual motion perception in a speeded discrimination task." Experimental Brain Research 178(3): 415-421.

Sanabria, D., S. Soto-Faraco, et al. (2005). "Intramodal perceptual grouping modulates multisensory integration: evidence from the crossmodal dynamic capture task." Neuroscience Letters 377(1): 59-64.

Sanabria, D., S. Soto-Faraco, et al. (2004). "When does visual perceptual grouping affect multisensory integration?" Cognitive, Affective, & Behavioral Neuroscience 4(2): 218-229.

Sanabria, D., S. Soto-Faraco, et al. (2005). "Intramodal perceptual grouping modulates multisensory integration: evidence from the crossmodal dynamic capture task." Neuroscience Letters 377(1): 59-64.

Sanabria, D., C. Spence, et al. (2007). "Perceptual and decisional contributions to audiovisual interactions in the perception of apparent motion: a signal detection study." Cognition 102(2): 299-310.

Scheier, C., R. Nijhawan, et al. (1999). "Sound alters visual temporal resolution." Investigative Ophthalmology & Visual Science 40(Suppl. 4): 4169.

Schutz, M. and M. Kubovy (2009). "Causality and cross-modal integration." Journal of Experimental Psychology: Human Perception and Performance 35(6): 1791-810.

Sekuler, R., A. B. Sekuler, et al. (1997). "Sound alters visual motion perception." Nature 385(6614): 308.

Shams, L., Y. Kamitani, et al. (2000). "Illusions. What you see is what you hear." Nature 408(6814): 788.

Shams, L., Y. Kamitani, et al. (2002). "Visual illusion induced by sound." Cognitive Brain Research 14(1): 147-152.

Shams, L., W. J. Ma, et al. (2005). "Sound-induced flash illusion as an optimal percept." Neuroreport 16(17): 1923-1927.

Sherrick, C. E. (1968). "Studies of apparent tactual. movement." In: The Skin Senses, edited by D.R.. Kenshalo. Springfield, C.C. Thomas: 331-345.

Shimojo, S., S. Miyauchi, et al. (1997). "Visual motion sensation yielded by non-visually driven attention." Vision Research 37(12): 1575-1580.

Shimojo, S., C. Scheier, et al. (2001). "Beyond perceptual modality: Auditory effects on visual perception " Acoustical Science and Technology, 22(2): 61-67.

Shipley, T. (1964). "Auditory Flutter-Driving of Visual Flicker." Science 145: 1328-1330.

Slutsky, D. A. and G. H. Recanzone (2001). "Temporal and spatial dependency of the ventriloquism effect." Neuroreport 12(1): 7-10.

Snyder, R. E. (1977). "Vibrotactile masking: A comparison of psychophysical procedures." Perception & Psychophysics 22(5): 471-475.

Soto-Faraco, S. and G. Deco (2009). "Multisensory contributions to the perception of vibrotactile events." Behavioural Brain Research 196(2): 145-154.

Soto-Faraco, S. and A. Kingstone (2004). Multisensory integration of dynamic information. The Handbook of Multisensory Processes. G. Calvert, C. Spence and B. E. Stein. Cambridge, MA, MIT Press: 49 - 67.

Soto-Faraco, S., A. Kingstone, et al. (2003). "Multisensory contributions to the perception of motion 41(13)." Neuropsychologia 41(13): 1847-1862.

Soto-Faraco, S., A. Kingstone, et al. (2006). "Integrating motion information across sensory modalities: the role of top-down factors." Progress in Brain Research 155: 273-86.

Soto-Faraco, S., J. Lyons, et al. (2002). "The ventriloquist in motion: illusory capture of dynamic information across sensory modalities." Cognitive Brain Research 14(1): 139-146.

Soto-Faraco, S., J. Navarra, et al. (2004). "Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task." Cognition 92(3): B13-23.

Soto-Faraco, S., C. Spence, et al. (2004). "Congruency effects between auditory and tactile motion: extending the phenomenon of cross-modal dynamic capture." Cognitive, affective & behavioral neuroscience 4(2): 208-217.

Soto-Faraco, S., C. Spence, et al. (2004). "Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities." Journal of Experimental Psychology: Human Perception and Performance 30(2): 330-45.

Soto-Faraco, S., C. Spence, et al. (2004). "Cross-modal dynamic capture: Congruency effects in the perception of motion across sensory modalities." Journal of Experimental Psychology:Human Perception and Performance 30(2): 330 -345.

Spence, C. (2002). "Multisensory attention and tactile information-processing." Behavioural Brain Research 135(1-2): 57-64.

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

Spence, C., M. E. Nicholls, et al. (2001). "The cost of expecting events in the wrong sensory modality." Perception & Psychophysics 63(2): 330-336.

Spence, C., D. Sanabria, et al. (2007). Intersensory Gestalten and crossmodal scene perception. Psychology of beauty and Kansei: New horizons of Gestalt perception. K. Noguchi. Tokyo, Fuzanbo International: 519-579.

Spence, C., D. Senkowski, et al. (2009). "Crossmodal processing." Experimental Brain Research 198(2-3): 107-11.

Stein, B. E. and T. R. Stanford (2008). "Multisensory integration: current issues from the perspective of the single neuron." Nature Review Neuroscience 9(4): 255-266.

Stein, B. E., T. R. Stanford, et al. (2009). "Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness." Experimental Brain Research 198(2-3): 113-126.

Stekelenburg, J. J. and J. Vroomen (2009). "Neural correlates of audiovisual motion capture." Experimental Brain Research 198(2-3): 383-90.

Stekelenburg, J. J., J. Vroomen, et al. (2004). "Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity." Neuroscience Letters 357(3): 163-6.

Stone, J. V., N. M. Hunkin, et al. (2001). "When is now? Perception of simultaneity." Proceedings of the Royal Society of London. Series B, Biological sciences 268(1462): 31-38.

Strybel, T. Z. and A. Vatakis (2004). "A comparison of auditory and visual apparent motion presented individually and with crossmodal moving distractors." Perception 33(9): 1033-1048.

Ternus, J. (1926). "Experimentelle Untersuchung über phänomenale Identität." Psychologische Forschung 7:

81-135.

Ternus, J. (1938). "The problem of phenomenal identity." in A Source Book of Gestalt Psychology. E.d, W D. Ellis (1938, London: Routledge and Kegan Paul).

Treutwein, B. and H. Strasburger (1999). "Fitting the psychometric function." Perception & Psychophysics 61(1): 87-106.

Ullman, S. (1979). "The Intepretation of Visual Motion." The MIT Press.

van Erp, J. and P. Werkhoven (2004). "Vibro-tactile and visual asynchronies: Sensitivity and consistency." Perception 33: 103-111.

Verrillo, R. T., A. J. Fraioli, et al. (1969). "Sensory magnitude of vibrotactile stimuli." Perception & Psychophysics 6: 366-372.

Von Bekesy, G. (1959). "Similarities between hearing and skin sensations." Psychological Review 66(1): 1-22.

Vroomen, J. and B. de Gelder (2000). "Sound enhances visual perception: cross-modal effects of auditory organization on vision." Journal of Experimental Psychology: Human Perception and Performance 26(5): 1583-1590.

Vroomen, J. and B. De Gelder (2004). "Perceptual Effects of Cross-modal Stimulation: Ventriloquism and the Freezing Phenomenon." In The Handbook of multisensory processes, MIT Press., Cambridge, MA.

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

Vroomen, J. and M. Keetels (2006). "The spatial constraint in intersensory pairing: no role in temporal ventriloquism." Journal of Experimental Psychology: Human Perception and Performance 32(4): 1063-1071.

Walker, J. T. and K. J. Scott (1981). "Auditory-visual conflicts in the perceived duration of lights, tones and gaps " Journal of Experimental Psychology: Human Perception and Performance 7(6): 1327-1339.

Wallace, M. T., L. K. Wilkinson, et al. (1996). "Representation and integration of multiple sensory inputs in primate superior colliculus." Journal of Neurophysiology 76(2): 1246-66.

Wearden, J. H., H. Edwards, et al. (1998). "Why "sounds are judged longer than lights": application of a model of the internal clock in humans " The Quarterly Journal of Experimental Psychology B,51(2): 97-120.

Wearden, J. H. e. a. (2006). "When do auditory/visual differences in duration judgments occur? ." The Quarterly Journel of Experimental psychology 59(10): 1709-1724.

Welch, R. B. (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 contributions to the perception of spatial and temporal events. Elsevier, Amsterdam,pp 371-387.

Welch, R. B. (1999). "Meaning,attention,and the "unity assumption" in the intersensory bias of spatial and temporal perceptions." In: Aschersleben G,Bachmann T,Musseler J (eds) Cognitive contributions to the perception of spatial and temporal events.Elsevier,Amsterdam,pp 371-387: .

Welch, R. B., L. D. DuttonHurt, et al. (1986). "Contributions of audition and vision to temporal rate perception." Perceptual Psychophysiology 39: 294-300.

Welch, R. B. and D. H. Warren (1980). "Immediate Perceptual Response to Intersensory Discrepancy." Psychological Bulletin 88(3): 638-667.

Welch, R. B. and D. H. Warren (1986). Intersensory interactions. Handbook of perception and human performance: Sensory processes and perception K. R. Boff, L. Kaufman and J. P. Thomas. New York, Wiley. 1: 1-36.

Witten, I. B. and E. I. Knudsen (2005). "Why seeing is believing: merging auditory and visual worlds." Neuron 48(3): 489-96.

Zapparoli, G. C., & Reatto, L. L. (1969). "The apparent movementbetween visual and acoustic stimulus and the problem of intermodal relations." Acta Psychologica 29: 256-267.

Zeki, S. M. (1974). "Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey." The Journal of Physiology 236: 549-573.

Acknowledgements

This dissertation was written at the Ludiwig-Maximilians University in Muncih. My PhD study and research in Munich were supported by SFB 453 Project M9 from the German Research Foundation (DFG) and Scholarship from LMU-Chinese Scholarship Council (CSC) joint program.

I am lucky that many people have helped me during the process of PhD work. I am extremely grateful to Prof Hermann Müller that he has given the whole and warm-hearted support throughout my research work. My sincere thanks go to Dr. Zhuanghua Shi, it is because of his supervision that I gain concrete experience in the experimental work and have exercised basic programming as well as data analysis skills. I am grateful to Prof. Zhijun Zhang and Prof Mowei Shen from my home university who have also offered their warmhearted supports of my PhD work here. As a member of CSC students, my gratitude also goes to Dr Matthias Hadesbeck, Mrs Monique Esnouf and Mrs Qiqi from the Foreign Affairs office at LMU who have also given the consistent care and help during my study in Munich. Thank Mr. Dieter Moser and his family's regular invitation and I really appreciate the support from them.

I extend my sincere thanks to my colleagues. I enjoy the friendship and collaboration with my lab partner Heng Zou and appreciate the help from Yiquan Shi, both helped me a lot when I first arrived in Munich. Dr. Thomas Geyer has offered his great help in my PhD study and offered kind suggestions on my PhD work. Dr. Michael Zehetleitner originally gave me the first demonstration on how to use the Heijo Device. Prof Heiner Deubel kindly lent his oscilloscope and provided valuable suggestions during my research. Prof. Kathrin Finke offered kind guidance of making presentations during PhD seminar. I thank Mr. Gilbers for his kindness and help during my work. I would sincerely thank Birgitt Aßfalg for her kind help and her cooking tea to make me refreshed in dealing with the afternoon work in cabin. Native speakers Nicholas Myers and Patricia Graf have also offered kind help in the German Summary.

I would thank Prof Torsten Schubert and PD Dr. Lutz Wiegrebe from Neurobiology Department who were kindly willing to be my readers and examiners. My gratitude also goes to Mrs. Doris Jacobson from Promotion office for her patience in collecting my documents.

Finally, my special gratitude goes to my beloved wife Jing Lu for her consistent encouragement and everlasting support. Her considerate understanding and wholehearted support are crucial to me, especially during the late stage of my thesis writing.

I have become an open vessel to accept much assistance from all sides. Two years' experience in Munich becomes a life treasure. May God bless my colleagues and friends in Munich!

Curriculum Vitae

Lihan Chen

Born on 02. January 1978 in Zhejiang Province, China.

Education

2006 – 2009 Ph.D. in Psychology
2002 – 2005 Master in Psychology (Zhejiang University)
1994 – 1999 Bachelor in Biomedical Engineering (Zhejiang University)
1991 – 1994 Middle School

Professional Experience

2007 – 2009
Research Fellow (DFG SFB 453 Project M9, Experimental psychology, LMU)

2005 – 2006
Lecturer (Department of Psychology, WenZhou Medical College, China)

1999 – 2002
Assistant Engineer, Instrument and Equipment Department, Wenzhou 3rd hospital