Exp Brain Res (2005) 167: 635–640
DOI 10.1007/s00221-005-0067-1

RESEARCH ARTICLE

Mirjam Keetels · Jean Vroomen

The role of spatial disparity and hemifields in audio-visual temporal order judgments

Received: 14 January 2005 / Accepted: 9 May 2005 / Published online: 21 September 2005

Abstract We explored whether sensitivity to audio-visual temporal order judgments (TOJs) was affected by the amount of spatial separation between a sound and light, and by whether the sound and light were presented in the same or in different hemifields. Participants made TOJs about noise bursts and light flashes, and judged whether the stimuli came from the same location or not. Flashes were presented either in the left or right hemifield (at ±10° from central fixation), and sounds either came from the same location as the lights, or at small or large disparities (20 or 40° from the light, respectively), thereby crossing the hemifields or not. TOJs became more accurate (i.e., the just noticeable difference, JND, became smaller) when spatial disparity increased and when hemifields were crossed. Location discrimination of the sound and light was affected similarly. These results demonstrate that audio-visual TOJs are critically dependent on both the relative position from which stimuli are presented and on whether stimuli cross hemifields or not.

Keywords Multisensory perception · Audio-visual temporal order judgment · Spatial disparity · Hemifields

Introduction

Temporal synchrony, along with spatial coincidence, may provide one of the most salient cues regarding whether information from different sensory modalities refer to a single multimodal percept, or rather should be treated as separate and independent perceptual objects or events (cf. Radeau and Bertelson 1987; Radeau 1994). Yet, the perceptual system can correct for some amount of spatial and temporal discrepancy between modalities and still provide the observer with a unified perceptual experience. For investigating sensitivity to temporal asynchronies, a majority of research in the temporal domain has used the temporal order judgment (TOJ) task. In a typical multisensory TOJ task, participants are presented with pairs of target stimuli in different sensory modalities at various stimulus onset asynchronies (SOAs) and are asked to judge in which modality a stimulus appeared first. Analysis of the responses across the range of SOAs allows one to calculate the just noticeable difference (JND), the minimal interval a participant needs for correctly judging which of the two stimuli had been presented first.

In many crossmodal TOJ studies, pairs of stimuli were presented from different spatial locations (e.g., auditory stimuli via headphones, visual stimuli from somewhere in front of the participant, and tactile stimuli somewhere at the participant's skin; see Bald et al. 1942; Dinnerstein and Zlotogura 1968; Hamlin 1895; Jaskowski et al. 1990; Rutschmann and Link 1964; Smith 1933; Teatini et al. 1976; Whipple et al. 1899). Recently, though, it has been shown that temporal precision in these studies might have been overestimated, as spatial separation between stimuli of which temporal order has to be judged can improve TOJ accuracy (Bertelson and Aschersleben 2003; Spence et al. 2001). For example, Bertelson and Aschersleben (2003) showed that audio-visual TOJ accuracy improved when sound and light were presented from different locations rather than from a common central location. Spence et al.(2003) also demonstrated enhanced performance for spatially separated visual-tactile stimulus pairs, and similar findings were reported for the audio-visual case (Spence et al. 2003; Zampini et al. 2003a), though not for the audio-tactile one (Zampini et al. 2005, in press).

Two major accounts have been put forward to explain improved audio-visual temporal precision when stimuli are spatially separated (Spence et al. 2003). First, it may be that when bimodal stimulus pairs are pre-

M. Keetels · J. Vroomen (✉)
Department of Psychology, Tilburg University,
Warandelaan 2, Tilburg, The Netherlands
E-mail: j.vroomen@uvt.nl
Tel.: +31-13-4662394
Fax: +31-13-4662370

sented from different locations, participants may actually have extra spatial information on which to base their responses. Participants may not know initially which modality had been presented first, but still know on which side the first stimulus appeared, and what the relative positions of the modalities were. So, because of this spatial redundancy, participants may infer which modality had been presented first. Second, it may be that multimodal pairing makes TOJs less accurate when stimuli are presented at the same location. Possibly, same location stimuli presented close in time are more likely paired together as a single multimodal event when compared to stimuli presented far apart (see e.g., Radeau 1994). Any such tendency to pair stimuli could make the relative temporal onsets of the components lost, and paired stimuli would thus be 'ventriloquized' in time (see e.g., Bertelson and Aschersleben 2003; Morein-Zamir et al. 2003; Vroomen and de Gelder 2004; Vroomen et al. 2004). Spatial redundancy and multimodal pairing therefore both predict better TOJ performance when audio-visual stimulus locations differ.

The finding that spatial discrepancy enhances audio-visual TOJ accuracy has, however, recently been qualified by Zampini et al. (2003b). Like others, they observed that TOJ accuracy improved when sounds and lights were presented from different locations, but only so when the locations crossed hemifields. Thus, when a sound and light were presented at the left and right of fixation (at −24 and +24°), JNDs improved from 73 to 58 ms for same versus different locations, respectively. In contrast, when the stimuli were presented, with the same spatial separation, vertically (at +24 and −24°) or within one hemifield (at −14 and −62°, or at +14 and +62°), no improvement was found (JNDs were 88 and 92 ms for same vs. different locations in the vertical arrangement, and 68 and 85 ms for same vs. different locations in the horizontal arrangement in the same hemifield). This made Zampini et al. (2003b) conclude that the critical factor for the TOJ improvement was that the individual components of an audio-visual stimulus were presented in different hemifields. They hypothesized that whenever stimuli are initially processed by different cerebral hemispheres, more resources would be available for processing these stimuli, leading to better TOJ performance (i.e., lower JNDs). On this account, it is thus not spatial separation as such that affects TOJ performance, but it is the initial projection to the different cerebral hemispheres that matters.

We considered the possibility, though, that the results of Zampini et al. (2003b) were confounded by the fact that the visual stimuli were presented at different eccentricities across the experiments. It is known that audio-visual fusion areas are larger in the periphery than at central locations, and since they also depend on whether the direction of the disparity is vertical or horizontal (Godfroy et al. 2003), it is conceivable that there was more bimodal fusion when stimuli were presented in the periphery vertically, rather than centrally. Furthermore, in conditions in which the hemifields were crossed, visual stimuli were always presented at 24° from fixation, while this was 14 and 62° for visual stimuli that did not cross hemifields. Given that visual resolution (Wandell 1995), accuracy of visual stimulus localization (Hairston et al. 2003), and speed of visual processing varies with retinal eccentricity (Carrasco et al. 2003; Rutschmann 1966), it seems conceivable that spatial and/or temporal cues of the visual stimuli were not comparable across conditions (in defence, though, in a post-hoc test Zampini et al. (2003b) did not find a difference between TOJs for visual stimuli presented at 14° vs. 62°). Finally, the hemispheric effect reported by Zampini et al. (2003b) was essentially based on a comparison between different experiments and participants using a blocked design. We felt it necessary to explore the phenomenon in a potentially more sensitive randomized within-subjects design.

In the present study, participants performed a TOJ task about auditory-visual stimulus pairs presented either at the same location, or at 20 of 40° disparity, in which case the sound and light either crossed the hemifields or not (see Fig. 1). The amount of spatial disparity between the sound and the light was varied as a critical test for the hemispheric account, as it predicts that TOJ accuracy will only improve when sounds and lights are presented in different hemifields, and thus irrespective of their spatial disparity. As an alternative, though, we considered the possibility that, all other things being equal, the perceived distance between sound and light might increase when hemifields were crossed. If so, then an effect of hemifield in the TOJ task might be explained by a difference in perceived spatial separation (smaller in the same hemifield than in different hemifields). To obtain an independent measure of perceived spatial disparity, participants judged, in a different part of the experiment, in a location discrimination task whether the sound and light were presented from the same or from different locations.

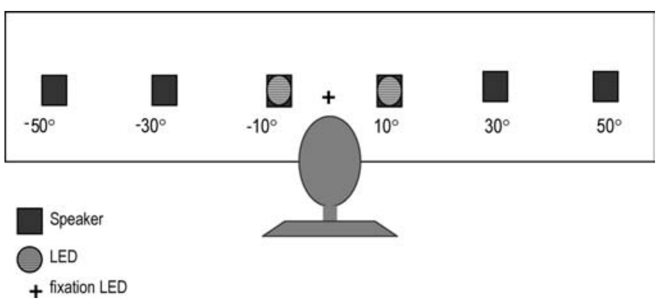

Fig. 1 The experimental setup. Participants were presented with two flashes on the left or right (10° from central fixation) and two sound bursts from one of six hidden loudspeakers. There was a variable stimulus onset asynchronies between the sound and light. The sound either came from the same location as the light, or it was presented at 20 or 40° separation, either crossing the hemifields or not. In the temporal order judgments task, participants judged whether the sound or the light was presented first. In the location discrimination task, participants judged whether the sound and light came from the same location or not

Method

Participants

Fifteen students from Tilburg University participated. All reported normal hearing and normal or corrected-to-normal seeing. They were tested individually and were naive as to the purpose of the experiment. They gave informed consent to participate in the study according to the Declaration of Helsinki and the ethics committee.

Stimuli

The auditory stimuli consisted of two white noise bursts of 20 ms duration, each with a 5 ms linear fade-in and fade-out (ISI = 1,000 ms) presented at 73 dB(A) by either one of six hidden loudspeakers. The speakers (FRWS 5 8 OHM, peak power 5 W, with a diameter of 4.5 cm) were placed at eye-level 10, 30, and 50° to the left and right of fixation. The visual stimuli consisted of two 20 ms flashes of a red LED (diameter of 1 cm, luminance of 40 cd/m^2, ISI = 1,000 ms), placed directly in front of the loudspeakers at 10° on the left and right. A small green LED was placed at eye-level, at central location, 57 cm in front of the participant and served as a fixation point.

Procedure

Participants sat at a table in a dark sound-proof booth. Head movements were precluded by a chin- and fore-head-rest. The fixation light was illuminated at the beginning of the experiment, and participants were instructed to maintain fixation on this central green LED during testing. Each trial consisted of the presentation of two noise bursts and two light flashes with a variable Stimulus Onset Asynchrony (SOA) between the sounds and flashes. The flashes were presented at 10° from central fixation unpredictably on the left or right, and the sounds were presented, unpredictably, either from the same location as the lights, or at 20 or 40° separation, thereby crossing the hemifields or not.

The study consisted of two parts. In the main part of the experiment, TOJs were made on which modality was presented first (sound or light). In the other part, participants judged whether the sounds and lights were presented from the same location or not. In both tasks, participants made an unspeeded response by pressing one of the two designated keys on a response box at the table. The next trial started 2,000 ms after a response was made.

Design

In the TOJ task, three within-subject factors were used: spatial disparity (five levels, sound and light either from the same location, or at 20 or 40° spatial separation in the same or in different hemifields), Side of light (left or right) and SOA (−240, −120, −90, −60, −30, 0, 30, 60, 90, 120, or 240 ms; negative values indicate that the sound was presented first). These factors yielded 110 equi-probable conditions. Each condition was presented 16 times for a total of 1,760 trials, presented in 16 blocks of 110 trials each. Within blocks, all combinations of SOA, side of light, and location were varied randomly.

In the location discrimination task, stimulus presentation was as in the TOJ task, except that, in order to reduce testing time, only SOAs of −120, −60, 0, 60, and 120 ms were used, and the number of trials per condition was reduced to eight instead of sixteen. Total testing lasted approximately 2 h, divided over 3 days.

To acquaint participants with TOJ task, experimental blocks were preceded by a training session in which 40 trials were presented with SOAs of −240, −120, 120, or 240 ms, presented at the ten possible combinations of location and side of light. During training, participants received corrective feedback (the green fixation LED flickering three times) whenever they made an erroneous response. Blocks continued until the proportion of correct responses was 85% or above. For the location discrimination task, 20 practice trials were given.

Results

TOJ task

Data of the TOJ task were analyzed as in Zampini et al. 2003a, b). Trials of the training session and trials with an SOA of −240 ms and +240 ms were excluded from further analyses, because most participants performed nearly perfect at this interval, and therefore no additional variance was accounted for by these measurements (see also Zampini et al. 2003b). The percentages of 'vision-first' responses were calculated for each participant and were converted into equivalent Z-scores assuming a cumulative normal distribution. For each condition, the best-fitting straight line was then calculated over the nine SOAs. The lines' slopes and intercepts were used to determine the JND (JND = 0.675/slope) and the point of subjective simultaneity (PSS). The JND represents the smallest interval between two stimuli needed by participants to judge correctly which stimulus came first on 75% of the trials. The PSS represents the average interval by which one stimulus had to lead the other for being perceived as simultaneous. An overall 5(spatial disparity) × 2 (side of light) ANOVA on the JNDs and on the PSSs showed that there was no effect of side of light, $F(1,14) = 0.43$, $P = 0.52$, and $F(1,14) = 0.17$, $P = 0.68$, respectively, nor did it interact with spatial disparity, $F(4,56) = 2.12$, $P = 0.90$, and $F(4,56) = 1.12$, $P = 0.36$, respectively. The proportions of 'vision-first' responses were therefore pooled over side of light. In an ANOVA on the PSSs, the effect of spatial disparity was not significant, $F(4, 56) = 1.47$, $P = 0.22$

(average PSS = +7.71 ms). Further analyses were therefore restricted to the JNDs (see Table 1). To obtain a measure of the effect of hemifield and distance on the JNDs, different scores were calculated for each participant with the JND of 'same location' as baseline. In a 2 (hemifield) × 2 (distance) ANOVA on the JND difference scores, a significant overall effect was found, $F(1, 14) = 11.19$, $P < 0.005$, indicating that JNDs were indeed smaller (i.e., sensitivity increased) whenever the stimulus pairs were spatially separated. There was also a significant main effect of hemifield, $F(1, 14) = 4.97$, $P < 0.05$, as JNDs were lower when sounds and lights were presented in different hemifields rather than in the same (a 9.1 ms vs. 5.5 ms improvement, respectively). Importantly, there was an effect of distance, $F(1, 14) = 5.70$, $P < 0.05$, indicating that JNDs for stimuli at 40° separation were lower than at 20° (a 9.9 ms vs. 4.7 ms improvement, respectively). The interaction between hemifield and distance was not significant, $F(1,14) = 1.24$, $P = 0.29$.

Location discrimination task

To examine the extent to which the spatial disparity between the sounds and lights was perceived, we calculated the percentage of 'same location' responses for each disparity. The overall 5 (spatial disparity) × 2 (side of light) × 5 (SOA) ANOVA showed that there was no effect of side of light, $F(1,14) = 0.11$, $P = 0.746$, an effect of spatial disparity, $F(4,56) = 58.57$, $P < 0.001$ (less same responses when pairs were spatially separated), and an effect of SOA, $F(4,56) = 3.01$, $P < 0.05$ (on average, 3% more 'same' location responses at 0 ms SOA than at the largest SOAs). To obtain an overall measure of perceived distance that could be compared with performance on the TOJ task, different scores were computed and pooled over side of light and SOA with the 'same location' condition as baseline (Table 1, right panel). In the 2 (hemifield) × 2 (distance) ANOVA on these difference scores, there was a main effect of distance, $F(1,14) = 18.40$, $P < 0.001$, because the perceived distance at 20° separation was smaller than at 40° (after correction for baseline, 73% vs. 88% of the trials were judged to be at different locations, respectively). There was no main effect of hemifield, $F(1,14) = 0.89$, $P = 0.36$ (77% vs. 84% for same vs. different hemifields, respectively), but the interaction between hemifield and distance was marginally significant, $F(1,14) = 4.38$, $P = 0.055$. The interaction indicated that the spatial disparity at 40° was perceived more reliably when sound and light were presented in different hemifields rather than the same (93% vs. 82%, respectively).

To compare performance on the two tasks, and to test whether location discrimination and TOJs were affected similarly by hemifield and distance, difference scores of both tasks were converted into Z-scores. The normalized scores were then entered into a 2 (task) × 2 (hemifield) × 2 (distance) MANOVA. A main effect of hemifield was found, $F(1,14) = 5.86$, $P < 0.05$, showing that in both the TOJ task and the location discrimination task the effect of spatial separation was largest when stimulus pairs were presented in different hemifields. There was also a main effect of distance, $F(1,14) = 14.79$, $P < 0.01$, showing that when distance increased, JNDs were lowered and less 'same location' responses were given. All other interactions were not significant (task and hemifield, $F(1,14) = 0.103$, $P = 0.75$; task and distance, $F(1,14) = 0.041$, $P = 0.84$, hemifield and distance, $F(1,14) = 0.04$, $P = 0.85$, task, hemifield and distance, $F(1,14) = 3.029$, $P = 0.104$), indicating that the effects of distance and hemifield were essentially the same for both tasks.

General discussion

We explored whether sensitivity to audio-visual TOJs was affected by the amount of spatial separation between sound and light, and by whether the stimuli were presented in the same or in different hemifields. As observed before, sensitivity improved (i.e., lower JNDs) when the individual components of an audio-visual stimulus were presented in different locations rather than the same location (see also Bertelson and Aschersleben 2003; Spence et al. 2001, 2003; Zampini et al. 2003a, b). Moreover, and in line with the hemispheric redundancy hypothesis proposed by Zampini et al. (2003b), JNDs were lower when sounds and lights were presented in different hemifields rather than in the same

Table 1 Mean performance on the temporal order judgment task (TOJ) and the location discrimination task as a function of the locations of the audio-visual stimulus pairs

	Separation (°)	TOJ task		Location discrimination task	
		JND[a]	Difference[c]	P(same)[b]	Difference[c]
Same location	0	47.6		0.97	
Same hemifield	20	45.5	2.1	0.25	0.72
	40	38.8	8.8	0.15	0.82
Different hemifield	20	40.4	7.2	0.23	0.74
	40	36.7	10.9	0.03	0.93

[a] *JND* mean just noticable difference (in ms)
[b] *P(same)* mean proportion of 'same location' responses
[c] *Difference* different scores compared to same location

hemifield. Importantly, JNDs also improved when the distance between the sound and the light increased from 20 to 40°, independent of whether the stimuli were presented in different hemifields or not. The latter is in clear contradiction with a strict interpretation of the hemispheric account, and it thus requires further elaboration.

Here we considered the possibility that what underlies the hemifield effect is that spatial disparity is more salient when sounds and lights are presented in different hemifields rather than in the same hemifield. On this view, the crucial feature is the extent to which sounds and lights are actually perceived to emerge from different locations. To check on that, and to obtain an independent measure of perceived distance, participants performed a location discrimination task. The results showed that even though the TOJ task and the location discrimination task were affected about equally by distance and hemifield, there was no strict one-to-one relation between the two (as there was a main effect of hemifield in the TOJ task, and an interaction between hemifield and distance in the location discrimination task). Most likely, then, two mechanisms underlie TOJ improvements: the perceived distance between sound and light, and whether or not the stimuli cross the hemifields (and possibly, hemispheres, although it is not clear to which extent sound processing is lateralized in a free-field listening situation, e.g., Woldorff et al. 1999).

The question remains why in the study of Zampini et al. (2003b), no improvement was found when stimuli were presented at different locations within the same hemisphere, as they used a spatial separation that was even larger than the one used here. One potential critical difference is that a trial in our setup consisted of a train of two sounds and two flashes instead of just one. We observed earlier that participants become more sensitive (i.e., lower JNDs) when the audio-visual stimuli are presented more than once (see also Morein-Zamir et al. 2004). More importantly, we also found that location discrimination of the sound and flash becomes more accurate. For example, performance in the spatial discrimination task was only 59.4% correct (chance level being 50%) when the audio-visual stimulus pair was presented once (Keetels and Vroomen 2004), while it was 83.6% correct in the present case where each stimulus pair was presented twice. It may, thus, well be that in the study of Zampini et al. (2003a), spatial separation between the sound and light was not perceived reliably because their stimuli were presented only once, which on its turn might explain the absence of a spatial separation effect. It should be noted, though, that with two stimulus presentations, participants had a chance to move their eyes towards the (second) flash. However, if it were indeed the case that participants, despite instructions, moved their eyes toward the visual stimuli, flashes would become central and the difference between 'same' versus 'different' hemifields should disappear. Since we found the effect of hemifield, we presume that participants followed instructions and fixated centrally.

The finding that TOJs improve and that spatial separation is perceived more accurately when stimuli are presented at large compared to small separations, is in line with both the spatial redundancy and the multimodal pairing account. Thus, whenever stimuli are presented at large separations, redundant spatial cues might be more distinctive, and/or multimodal pairing might occur less, compared to when pairs are presented with small separations. Both accounts are for the time being possible, and at present no distinction can be made between multimodal pairing or spatial redundancy. The two accounts differ in that multimodal pairing is critically dependent on that sound and light are presented in close temporal proximity of each other (say ±200 ms), while this is, in principle, not the case for spatial redundancy. However, given that TOJ accuracy is typically examined within the temporal fusion window of multimodal pairing, it is at present difficult to disentangle the two accounts. This puzzle, though, might be solved if, for example, multimodal pairing but not spatial discrimination could be prevented.

Acknowledgements We thank the anonymous reviewers for comments on a previous version of this paper.

References

Bald L, Berrien FK, Price JB, Sprague RO (1942) Errors in perceiving the temporal order of auditory and visual stimuli. J Appl Psychol 26:382–388

Bertelson P, Aschersleben G (2003) Temporal ventriloquism: crossmodal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. Int J Psychophysiol 50:147–155

Carrasco M, McElree B, Denisova K, Giordano AM (2003) Speed of visual processing increases with eccentricity. Nat Neurosci 6:699–700

Dinnerstein AJ, Zlotogura P (1968) Intermodal perception of temporal order and motor skills: effects of age. Percept Motor Skill 26:987–1000

Godfroy M, Roumes C, Dauchy P (2003) Spatial variations of visual-auditory fusion areas. Perception 32:1233–1245

Hairston WD, Wallace MT, Vaughan JW, Stein BE, Norris JL, Schirillo JA (2003) Visual localization ability influences cross-modal bias. J Cogn Neurosci 15:20–29

Hamlin AJ (1895) On the least observable interval between stimuli addressed to disparate senses and to different organs of the same sense. Am J Psychol 6:564–575

Jaskowski P, Jaroszyk F, Hojan-Jerierska D (1990) Temporal-order judgments and reaction time for stimuli of different modalities. Psychol Res 52:35–38

Keetels M, Vroomen J (2004) The role of space in audio-visual temporal order judgments. Poster presented at the 5th meeting of the International Multisensory Research Forum. Barcelona, Spain (June 2–5), http://www.imrf.info/2004/103

Morein-Zamir S, Soto-Faraco S, Kingstone A (2003) Auditory capture of vision: examining temporal ventriloquism. Cogn Brain Res 17:154–163

Morein-Zamir S, Li K, Kingstone A (2004) Looking at the window of perceived simultaneity: repetition, rate, and crossmodal asymmetries. Paper presented at the 5th meeting of the International Multisensory Research Forum. Barcelona, Spain (June 2–5), http://www.imrf.info/2004/131

Radeau M (1994) Auditory-visual interaction and modularity. Curr Psych Cogn 13:3–51

Radeau M, Bertelson P (1987) Auditory-visual interaction and the timing of inputs: Thomas (1941) revisited. Psychol Res 49:17–22

Rutschmann R (1966) Perception of temporal order and relative visual latency. Science 152:1099–1101

Rutschmann J, Link R (1964) Perception of temporal order of stimuli differing in sense mode and simple reaction time. Percept Motor Skill 18:345–352

Smith WF (1933) The relative quickness of visual and auditory perception. J Exp Psychol 16:239–257

Spence C, Shore DI, Klein RM (2001) Multisensory prior entry. J Exp Psychol Gen 130:799–832

Spence C, Baddeley R, Zampini M, James R, Shore DI (2003) Multisensory temporal order judgments: when two locations are better than one. Percept Psychophys 65:318–328

Teatini G, Farnè M, Verzella F, Berruecos P Jr (1976) Perception of temporal order: visual and auditory stimuli. Giorn Ital di Psicologia 3:157–164

Vroomen J, de Gelder B (2004) Temporal ventriloquism: sound modulates the flash-lag effect. J Exp Psychol Hum Percept Perform 30:513–518

Vroomen J, Keetels M, de Gelder B, Bertelson P (2004) Recalibration of temporal order perception by exposure to audio-visual asynchrony. Cogn Brain Res 22:32–35

Wandell BA (1995) Foundations of vision. Sinauer, Sunderland

Whipple GM, Sanford EC, Colgrove FW (1899) Minor studies from the psychological laboratory of dark University: on nearly simultaneous clicks and flashes: the time required for recognition: notes on mental standards of length. Am J Psychol 10:280–295

Woldorff MG, Tempelmann C, Fell J, Tegeler C, Gaschler-Markefski B, Hinrichs H, Heinz HJ, Scheich H (1999) Lateralized auditory spatial perception and the contralaterality of cortical processing as studied with functional magnetic resonance imaging and magnetoencephalography. Hum Brain Mapp 7:49–66

Zampini M, Shore DI, Spence C (2003a) Audiovisual temporal order judgments. Exp Brain Res 152:198–210

Zampini M, Shore DI, Spence C (2003b) Multisensory temporal order judgments: the role of hemispheric redundancy. Int J Psychophysiol 50:165–180

Zampini M, Brown T, Shore DI, Maravita A, Roder B, Spence C (2005) Audiotactile temporal order judgments. Acta Psychol (in press)