Exp Brain Res (2003) 152:198–210
DOI 10.1007/s00221-003-1536-z

RESEARCH ARTICLE

Massimiliano Zampini · David I. Shore · Charles Spence

Audiovisual temporal order judgments

Received: 24 June 2002 / Accepted: 25 May 2003 / Published online: 22 July 2003

Abstract In two experiments, we examined the extent to which audiovisual temporal order judgments (TOJs) were affected by spatial factors and by the dimension along which TOJs were made. Pairs of auditory and visual stimuli were presented from either the left and/or right of fixation at varying stimulus onset asynchronies (SOAs), and participants made unspeeded TOJs regarding either "Which modality was presented first?" (experiment 1), or "Which side was presented first?" (experiment 2). Modality TOJs were more accurate (i.e. just-noticeable differences, JNDs, were smaller) when the auditory and visual stimuli were presented from *different* spatial positions rather than from the *same* position, highlighting an important potential confound inherent in previous research. By contrast, spatial TOJs were unaffected by whether or not the two stimuli were presented in different modalities. A between-experiments comparison revealed more accurate performance (i.e. smaller JNDs) when people reported which modality came first than when they reported which side came first for identical bimodal stimulus pairs. These results demonstrate that multisensory TOJs are critically dependent on both the relative spatial position from which stimuli are presented and on the particular dimension being judged.

Keywords Multisensory perception · Temporal order judgment · Auditory · Visual

M. Zampini (✉) · C. Spence
Department of Experimental Psychology,
University of Oxford,
South Parks Road, Oxford, OX1 3UD, UK
e-mail: massimiliano.zampini@psy.ox.ac.uk
Tel.: +44-1865-271380
Fax: +44-1865-310447

M. Zampini
University of Verona, Italy

D. I. Shore
Department of Psychology,
McMaster University,
Hamilton, Canada

Introduction

Previous studies have shown that spatial coincidence and temporal synchrony are two of the key factors determining whether or not multisensory integration takes place (see Driver and Spence 2000; Slutsky and Recanzone 2001; Stein and Meredith 1993). Relative to the extensive research devoted to trying to understand the role of *spatial* factors in modulating multisensory integration (see Bertelson and de Gelder 2003; Driver and Spence 1998 for reviews), far less research has been directed at understanding the role of *temporal* factors in multisensory binding.

One of the classic paradigms used by researchers to investigate temporal perception in humans is the temporal order judgment (TOJ) task (Bald et al. 1942; Dinnerstein and Zlotogura 1968; Hamlin 1895; Hirsh and Sherrick 1961; Rutschmann and Link 1964; Spence et al. 2001b; Sternberg et al. 1971; Teatini et al. 1976). In a typical TOJ experiment, participants are presented with a pair of stimuli at varying stimulus onset asynchronies (SOAs) and required to judge which stimulus was presented first (or second; see Shore et al. 2001). A second classic task is the simultaneous/successive task in which participants have simply to judge whether two stimuli appear to have been presented simultaneously or not (Bushara et al. 2001; Exner 1875; Raizada and Poldrack 2001; Slutsky and Recanzone 2001; Stone et al. 2001; Sugita and Suzuki 2003). Some researchers have also adopted a hybrid ternary response version of these tasks in which participants either judge which stimulus appeared first, or else respond that the stimuli were presented simultaneously (Allan 1975; Stone 1926; Ulrich 1987; Van de Par et al. 1999).

In their seminal study of multisensory TOJs, Hirsh and Sherrick (1961) reported that the smallest temporal interval between two stimuli needed for participants to be able to judge correctly which stimulus came first on 75% of trials (the so-called just noticeable difference, JND) was of the order of 20 ms. Surprisingly, Hirsh and Sherrick found that the JND remained relatively constant across a number of different intramodal and cross-modal combinations of

auditory, visual and tactile stimuli, and also across the various different dimensions along which TOJ responses were made. In Hirsh and Sherrick's intramodal TOJ studies, participants judged either "Which side came first?" for pairs of visual or auditory stimuli, or "Which frequency came first?" for pairs of auditory stimuli (see also Hirsh 1959); while, in their multisensory TOJ studies, participants judged "Which modality came first?".

Unfortunately, the appropriate interpretation of Hirsh and Sherrick's (1961) results is uncertain given the presence of a spatial confound in their multisensory TOJ experiments (see Spence et al. 2001b). In particular, the stimuli in each of the different modalities were always presented from *different* spatial locations: Auditory stimuli were presented over headphones, while visual stimuli were presented from a screen placed directly in front of the participants; and tactile stimuli were presented to the participant's hand, placed by their side. This means that participants in Hirsh and Sherrick's study may simply have based their TOJ responses on which *location* was stimulated first, rather than on the basis of which *modality* they perceived first (see Spence et al. 2001b). Alternatively, participants may simply have used redundant spatial cues to augment their modality discrimination responses. The existence of this spatial confound means that the JND of 20 ms reported by Hirsh and Sherrick may actually reflect a significant overestimation of peoples' ability to order correctly pairs of stimuli presented to different sensory modalities. This spatial confound is by no means limited to Hirsh and Sherrick's (1961) multisensory TOJ studies but, more critically, extends to *all* previously published studies of audiovisual temporal perception: The visual stimuli in these other studies were also placed in front of the participant, while auditory stimuli were presented either over headphones (Bushara et al. 2001; Jaskowski et al. 1990; Raizada and Poldrack 2001; Rutschmann and Link 1964) or else from some other external device (such a loudspeaker or "clicking" relay) producing the auditory stimuli (Bald et al. 1942; Drew 1896; Dinnerstein and Zlotogura 1968; Hamlin 1895; Smith 1933; Teatini et al. 1976; Whipple et al. 1899).

This means that, at present, we have no unambiguous estimate of the temporal precision with which people can judge the temporal ordering of stimuli presented to different sensory modalities that is free from this spatial confound (though see Gengel and Hirsh 1970, for a solitary exception). However, the derivation of robust and accurate estimates of people's sensitivity to audiovisual synchrony is important for a number of real-world applications, from the design of hearing aids (McGrath and Summerfield 1985; Pandev et al. 1986) to the derivation of guidelines for satellite telecommunications broadcasting (ITU-T1990; Reeves and Voelker 1993; Rihs 1995), as well as for the development of new virtual-conferencing technologies (Finger and Davis 2001; Mortlock et al. 1997).

But just how much of a problem does this spatial confound actually represent? Do people really make use of redundant spatial cues when judging which of two sensory modalities was presented first? If not, then we may be able to take these previous multisensory TOJ results at face value. Spence et al. (2001b; experiment 1) recently attempted to address these questions by assessing the influence of redundant spatial cues on visuotactile TOJs. Participants in their study were presented with pairs of visual and tactile stimuli from either the same or different locations by the left and/or right hand, and were required to judge which modality appeared to have been presented first. Spence et al. found that, while participants needed just 37 ms to judge the order of visual and tactile stimuli presented to different spatial locations accurately, they needed 68 ms in order to judge their order when the visual and tactile stimuli were presented from the same spatial location. These results provide empirical support for the view that previous researchers may have systematically overestimated the temporal precision with which people can correctly judge the order of pairs of stimuli presented to different sensory modalities.

However, given the numerous differences between the various senses and, in particular, the better temporal (and worse spatial) resolution of the auditory system as compared to either the visual or tactile modalities (Gebhardt and Mowbray 1959; Welch et al. 1986), we thought it possible that audiovisual temporal order judgments might be less affected by the relative spatial displacement of auditory and visual stimuli than was the case in Spence et al.'s (2001b) visuotactile study. Moreover, given that the majority of multisensory studies published to date have focused on audiovisual TOJs, rather than on the visuotactile pairing studied by Spence et al., it seemed both desirable and necessary to perform an analogous experiment for the pairing of auditory and visual stimuli.

The primary goal of the present study was therefore to examine whether a spatial modulation of multisensory TOJ performance could be demonstrated when people attempted to judge the order in which pairs of auditory and visual stimuli were presented. To this end, audiovisual stimulus pairs were presented from either the same or different locations to the left and/or right of fixation on each trial. The stimuli were separated by a variable SOA using the method of constant stimuli (Shore et al. 2001; Spence et al. 2001b). In the 1st experiment, participants were required to make unspeeded discrimination responses regarding "Which modality came first?".

We thought it possible that changing the dimension along which TOJ responses are made, such as from "Which side came first?" to "Which modality came first?" in Hirsh and Sherrick's (1961) previous studies, might affect the precision with which people can make multisensory temporal order judgments.[1] Indeed, McFar-

[1] In the Introduction to their paper, Hirsh and Sherrick (1961) point out that performance on simultaneous/successive judgment tasks will be affected by the dimension along which stimuli vary, despite the fact that they do not consider this point for the TOJ studies they report. However, this apparent inconsistency may simply reflect another of Hirsh and Sherrick's underlying assumptions, namely, that TOJs and simultaneous/successive judgment tasks rely on very different neural mechanisms.

land et al. (1998) recently reported that the temporal precision with which people make various intramodal auditory and visual TOJs varies significantly as a function of the dimension along which people responded (frequency or intensity judgments for auditory stimuli; and size, orientation or colour judgments for visual stimuli). Importantly, however, McFarland et al. always varied the nature of the stimuli presented to their participants whenever they changed the response dimension. This means that it is unclear whether the differences in temporal precision they reported reflect the effects of changes in the dimension along which participants responded versus simply differences in the temporal discriminability of the stimuli that were actually presented to participants in the various conditions. Therefore, in our 2nd experiment, participants were required to judge "Which side came first?" (instead of "Which modality came first?", as in experiment 1). Half of the stimulus pairs presented in experiment 2 were identical to those presented in experiment 1, thus enabling us to assess more directly (in between-experiments analyses) the effect of changing the dimension along which TOJs are made on the precision of audiovisual temporal processing.

Experiment 1

Methods

Participants

Ten right-handed participants (mean age 26 years) took part in the experiment as paid volunteers. Visual acuity was normal or corrected-to-normal, and all participants reported normal hearing. The participants were naïve as to the purpose of the experiment and varied in their previous experience of psychophysical testing procedures. The experiment took approximately 50 min to complete. All of the experiments reported in this study were non-invasive and had ethical approval from the Department of Experimental Psychology, University of Oxford, UK. The experiments were performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Apparatus and stimuli

The experiment was conducted in a completely dark, sound-attenuated room. A red light-emitting diode (LED; luminance of 64.3 cd/m^2) was placed on the table 62 cm in front of the participant and served as a fixation point. Similar LEDs were placed above and below this fixation point to provide post-response feedback to participants. Two identical loudspeaker cones (VE100AO) were positioned 26 cm to either side of the fixation light, and at the same distance from the participant. A red LED similar to the fixation light was centred directly in front of each one of these eccentric loudspeaker cones.

Auditory stimuli consisted of the presentation of a 9 ms white noise burst [82 dB(A) as measured from the participants' head position], and the visual stimuli consisted of the onset of either peripheral LED for 9 ms. No specific attempt was made to try and match the intensities of the stimuli, which were clearly suprathreshold (see Smith 1933; Spence et al. 2001b). White noise was presented continuously at 75 dB(A) throughout the experiment from a centrally positioned loudspeaker cone to mask any noises made by the participants.

Participants normally kept both of their thumbs on two separate keys placed vertically one above the other on a hand-held response pad. They were instructed to press the upper key to indicate that the auditory stimulus had been presented first, and the lower key whenever the visual stimulus appeared to have been presented first. Response feedback consisted of the illumination of one of the two LEDs placed directly above and below the fixation light immediately after a response had been detected. The lower light indicated a "vision-first" response, while the upper light indicated a "sound-first" response. Note that it was made clear to the participants that the feedback lights gave no indication as to whether the participants' response was correct or incorrect, but simply indicated which response they had made. Presentation of the stimuli and monitoring of the responses was controlled by an IBM 486-compatible microcomputer using a program written in Turbo Pascal 7.0. Timing was controlled by a custom timing routine that interfaced to the LEDs, loudspeaker cones and response keys.

Design

There were three within-participants factors: Stimulus position (same versus different), Side of vision (left versus right) and SOA (−200 ms, −90 ms, −55 ms, −30 ms, −20 ms, +20 ms, +30 ms, +55 ms, +90 ms, and +200 ms; Negative SOAs indicate that the auditory stimulus was presented first, whereas positive values indicate that the visual stimulus was presented first). The four possible stimulus configurations (sound left, vision left; sound right, vision right; sound left, vision right; and sound right, vision left) occurred with equal probability during each experimental session. The 40 possible conditions (10 SOAs ×4 possible stimulus configurations) were presented twice within each block of experimental trials. All participants completed 2 blocks of 30 practice trials, followed by 8 blocks of 80 experimental trials. The SOAs were doubled in the first practice block to facilitate the acquisition of the TOJ discrimination task by participants.

Procedure

The fixation light was illuminated at the beginning of each trial. Participants were instructed to maintain their fixation on this central red LED throughout each block of trials. The first stimulus was presented from either the left or right after a delay of 750 ms. The second stimulus was presented after the SOA specified for that particular condition. The task was unspeeded, and participants were informed that they should respond only when confident of their response (although within the 3,500 ms allowed before the termination of the trial). If participants responded prior to the onset of the first stimulus or failed to make a response before the trial was terminated, error feedback was presented. This consisted of the flickering of the fixation light for 1,000 ms. Such responses occurred on less than 1% of trials overall and were not analysed. Otherwise, the participant's response was indicated by the illumination of one of the central feedback lights for 500 ms after their response was detected. The fixation light was illuminated to indicate the start of the next trial 750 ms after the end of the preceding trial.

Results

The proportion of "vision first" responses was converted to its equivalent *Z*-score assuming a cumulative normal distribution (see Finney 1964; see Fig. 1). The intermediate eight SOAs were used to calculate a best-fitting

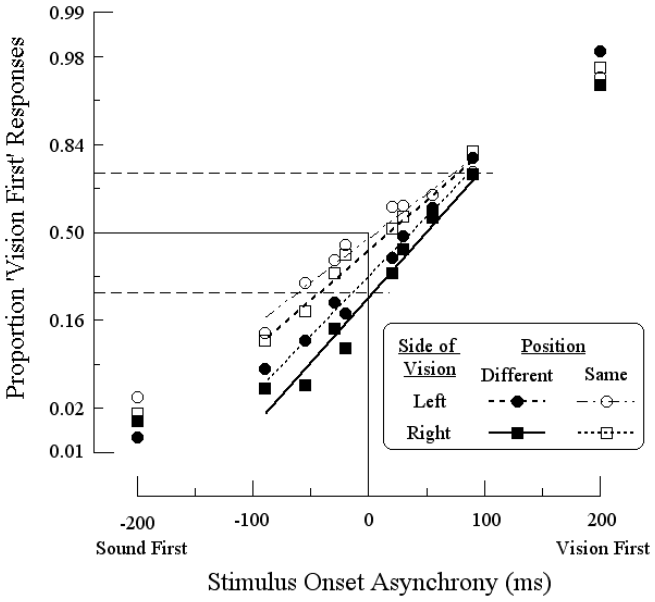

Fig. 1 Mean proportion of "vision first" responses plotted on a standard Gaussian *Z*-score scale as a function of the stimulus onset asynchrony (SOA) between the two stimuli for all four possible stimulus pairs presented in experiment 1. Participants performed an unspeeded "Which modality came first?" temporal order judgment (TOJ) task for bimodal pairs of auditory and visual stimuli presented to either the left and/or right of fixation. The best-fitting straight lines for the intermediate eight SOAs are indicated for each of the four stimulus pairs (see text for details)

straight line for each participant for each condition.[2] The slopes and intercepts from these best-fitting lines were used to calculate the JND and point of subjective simultaneity (PSS) for each of the four conditions for each participant (see Table 1 for a summary). The JND and PSS data were submitted to a repeated-measures analysis of variance (ANOVA) with the factors of Stimulus position (same versus different) and Side of vision (left versus right). One participant was removed from the analysis, because their PSS was larger than 250 ms, which was beyond the values tested (see Spence et al. 2001b for similar exclusion criteria), indicating their inability to perform the task.

Analysis of the JND data revealed a highly significant main effect of Spatial position ($F_{1,\ 8}$=17.11, P=0.003), reflecting the fact that participants found the TOJ task significantly easier when the two stimuli were presented from *different* positions (mean JND of 22 ms), rather than from the *same* position (mean JND of 32 ms). There was no significant main effect of Side of vision ($F_{1,\ 8}$=0.44, n.s.), nor any interaction between these two factors ($F_{1,\ 8}$=0.22, n.s.) (Fig. 2).

A similar analysis of the PSS data revealed a significant main effect of Spatial position ($F_{1,8}$=5.82, P=0.042), reflecting the fact that visual stimuli had to be presented further in advance of auditory stimuli to be perceived as occurring at the same time when the two stimuli came from different locations (mean PSS of 75 ms) than when presented from the same location (mean PSS of 60 ms). Once again, there was no significant effect of the Side of vision ($F_{1,8}$=1.91, n.s.), nor any interaction between these two factors ($F_{1,8}$=1.05, n.s.). For all four conditions, the PSS was significantly different from zero (M=67.4 ms; $t(8)$>9.9 for each comparison, P<0.0001 for all comparisons using a one

[2] The +200 ms points were excluded from this computation, because most participants performed nearly perfectly at this interval, so no additional variance was accounted for by these points. The inclusion of these points would actually have resulted in an artefactual reduction in slope.

Table 1 Mean and standard errors of the means for points of subjective simultaneity (*PSSs*), and just noticeaebale differences (*JNDs*) for experiment 1 and 2

Condition	PSSs		JNDs	
	Mean	SEM	Mean	SEM
Experiment 1: Which modality came first?				
Vision left, sound left	59.8	6.1	33.4	4.3
Vision right, sound right	60.2	4.9	29.7	3.4
Vision left, sound right	69.3	4.1	22.5	1.8
Vision right, sound left	80.4	7.8	21.9	2.1
Experiment 2: Which side came first?				
Sound left, sound right	−5.5	17.4	59.3	11.4
Vision left, vision right	−5.3	7.6	40.6	4.1
Vision left, sound right	−78.7	21.2	61.9	14.6
Sound left, vision right	72.5	10.4	62.2	10.1

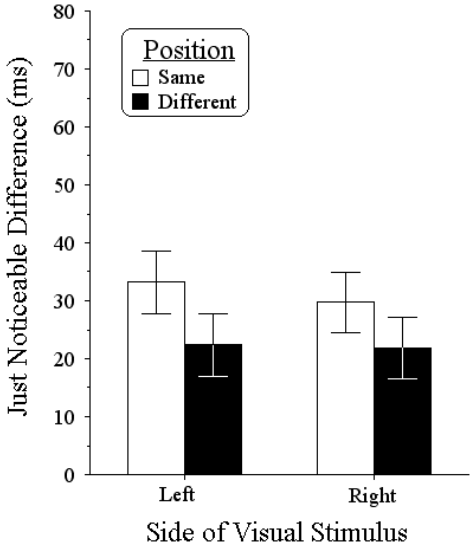

Fig. 2 Just noticeable differences for the four bimodal stimulus pairs in experiment 1. Participants performed an unspeeded "Which modality came first?" TOJ task for pairs of auditory and visual stimuli presented to either the left and/or right of fixation. *Error bars* represent the within-observer standard error of the mean based on the ANOVA reported in the text. The results highlight the improved temporal discrimination performance when bimodal pairs of auditory and visual stimuli were presented from different spatial positions rather than from the same position

sample two-tailed *t*-test), indicating that the visual stimulus had to be presented before the auditory stimulus.

Discussion

The major finding to emerge from the analysis of experiment 1 was the significant reduction in temporal precision observed when the auditory and visual stimuli were presented from the *same,* rather than from *different,* spatial locations. This result is consistent with Spence et al.'s (2001b) previous findings regarding the beneficial effect of presenting stimuli from different positions on visuotactile TOJs. Interestingly, however, the magnitude of the spatial modulation of TOJ performance reported in the present experiment 1 (a mean reduction in the JND of 10 ms when the stimuli were presented from different positions rather than from the same position) was somewhat smaller than that reported in Spence et al.'s previous visuotactile study (where a mean reduction in the JND of 31 ms was reported). Given that the magnitude of the spatial displacement involved in the different position conditions was similar across the two studies, it seems likely that this difference may reflect a more fundamental difference between the senses involved. One possibility is that the introduction of a spatial displacement may play a more significant role for pairs of modalities that both exhibit poor temporal resolution such as vision and touch, than for the pairings involving the auditory modality which demonstrates more accurate temporal (and less accurate spatial) resolution than the other senses (see Gebhardt and Mowbray 1959; Welch et al. 1986). It is possible that the increased temporal precision evident in the auditory modality may lead participants to place less emphasis on spatial factors such as the relative spatial location of the stimuli.

The results of experiment 1, when taken together with the results of Spence et al.'s (2001b) previous visuotactile study, clearly highlight the potential importance of the spatial confound inherent in all previous studies of multisensory temporal perception. Both studies unambiguously demonstrate that the presentation of stimuli in different sensory modalities from different spatial locations can lead to the facilitation of performance on TOJ tasks. This means that all of the previously published studies of multisensory temporal perception (with the exception of the solitary study by Gengel and Hirsh 1970) may have systematically overestimated the precision with which people can make multisensory TOJs by inadvertently presenting stimuli from different modalities from different spatial locations.

One possible account for these results is that participants in experiment 1 may have used the redundant spatial information available in the different position trials to facilitate their decision as to which modality had been presented first. According to this redundancy account (see Spence et al. 2001b), participants who were unsure as to which modality had been presented first on a particular trial may have used any information regarding which location they perceived to have been stimulated first to infer the correct response. No such redundant spatial information would have been available to facilitate performance on trials where the auditory and visual stimuli were presented from the same spatial location, hence potentially accounting for the worse performance seen in this condition.

However, an equally plausible alternative explanation for the improved JNDs reported in the different position condition of experiment 1 is in terms of the limited-capacity attentional resources thought to be present in each cerebral hemisphere (Banich 1988; Friedman and Polson 1981; Herdman and Freidman 1985; Kinsbourne and Cook 1971; Kinsbourne and Hicks 1978). When the auditory and visual stimuli were presented in different positions in experiment 1, they also projected, at least initially, to *different cerebral hemispheres*. By contrast, on trials where the two stimuli were presented from the same position, they projected to the *same cerebral hemisphere*. The presentation of both stimuli in rapid succession to the same cerebral hemisphere in the Same position condition may have increased the perceptual load on this limited capacity resource, as compared to when the two stimuli were presented to different hemispheres.[3] Therefore, presenting the two stimuli to different locations may have reduced the perceptual load on any given hemisphere and hence possibly lead to an improvement of temporal discrimination performance. This line of reasoning proposes a relative *cost* for having stimuli in the same location as opposed to a relative *benefit* for having stimuli at different locations, and it may be that both factors are involved.

A third possible explanation of the benefit for different locations (or cost for same locations) emerges when one considers the role of multisensory binding (see Stein and Meredith 1993). The likelihood that two stimuli from different sensory modalities will be bound into a single coherent object representation is greater when they come from the same location rather than from different locations. According to this account, binding the stimuli into the same object file (see Kahneman et al. 1992) results in the merging of the information concerning the different events associated with that object. This merging would tend to blur the relative onsets of the individual features. That is, perceiving a single object/event that both makes a noise and emits light would make it more difficult to say which of these events occurred first. Such temporal binding would not occur when the stimuli came from different locations, and thus a smaller interval between the stimuli would be needed to say reliably

[3] This hemispheric redundancy account can be considered as a specific example of a more general "spatial channel" account (see Regan 1982 for a review). That is, stimuli presented from the same spatial location may activate a common limited capacity spatial channel, whereas stimuli presented to different spatial positions may activate different spatial channels. However, we prefer the hemispheric redundancy account over the spatial channel account given the problems that are associated with the appropriate definition of a channel (see Regan 1982; Spence et al. 2001b).

which came first in this condition. We will return to a fuller evaluation of the relative merits of these three putative accounts of the spatial modulation of JNDs after the next experiment.

The results of experiment 1 also support previous claims that visual stimuli have to be presented before auditory stimuli for them to be perceived as simultaneous (Jaskowski et al. 1990; Sugita and Suzuki 2003; Van de Par et al. 1999; though see Bald et al. 1942; Rutschmann and Link 1964; Teatini et al. 1976). Importantly, visual stimuli had to lead auditory stimuli by a significantly smaller interval in the present study when they were presented from the same spatial position than from different positions. There are at least two plausible explanations for this finding: It may be that auditory and visual stimuli are more likely than other stimulus pairings to be drawn towards each other across time (i.e. a kind of temporal equivalent of the well-known spatial ventriloquism effect; e.g. Bertelson and de Gelder 2003), hence reducing the effective PSS when presented from the same spatial location than from different spatial locations (Morein-Zamir et al. 2003; Scheier et al. 1999; Slutsky and Recanzone 2001; though see Regan and Spekreijse 1977, for a possible exception to this view). This account is entirely consistent with the multisensory binding hypothesis outlined above for the difference in JND results. Alternatively, according to an attention-switching account of TOJs (Allan 1975; Kristofferson 1967; though see Shore et al. 2002), attention has to be shifted from the location of the first stimulus to the location of the second stimulus in order for participants to make a TOJ response. It will take more time for attention to shift to the second stimulus if it is presented from a different spatial location than the first stimulus, as that will require not only a shift of attention from one modality to another (Spence et al. 2001a), but also a shift of spatial attention from one location to another (see Spence 2001). Both the attention-switching and the temporal ventriloquism account predict that the amount of time by which vision has to lead for simultaneity to be achieved should be smaller when the auditory and visual stimuli are presented from the same rather than different positions. However, no matter what the cause of this reduction of the PSS turns out to be, the key point remains that our results demonstrate that the amount of time by which one modality (vision) has to lead another (audition) for them to be perceived as simultaneous is critically dependent on the relative spatial positioning of the stimuli. As far as we are aware, this spatial modulation of the PSS for multisensory stimulus pairs has not been considered in previous TOJ research. This spatial modulation may also help to explain why the PSSs reported in the present study were so much larger than those reported in many of the previous studies (where values of 20–40 ms are much more common; see Spence et al. 2001b for a review). Our presentation of visual stimuli from more eccentric positions than in the majority of previous studies (where visual stimuli were typically presented at fixation) may also help to explain this difference (see Rains 1963).

Overall, the results of experiment 1 demonstrate that both the precision with which people can make audiovisual TOJs (in terms of the JND) and also the PSS can be significantly affected by the relative spatial position from which auditory and visual stimuli are presented. In the next experiment, we assessed the effect of changing the dimension on which TOJs are made on measures of multisensory temporal perception.

Experiment 2

In the years since Hirsh and Sherrick's (1961) seminal study, few researchers have considered the possible impact of changes in the dimension along which TOJs are made on the accuracy of multisensory TOJs. Hirsh and Sherrick appear to have implicitly assumed that any such changes would have little impact on the magnitude of the JNDs they reported. They used a "Which side came first?" task to assess intramodal visual temporal perception, a "Which frequency came first?" task to assess intramodal auditory TOJs, and a "Which modality came first?" task to assess multisensory perception. At first glance, the fact that Hirsh and Sherrick found that participants needed only 20 ms to judge accurately the temporal order of pairs of stimuli, no matter which modalities the two stimuli were presented in, nor which dimension was used for responding, would appear to give some support for their view that the dimension along which responses were made has little effect on the precision of TOJs. Importantly, however, just as in McFarland et al.'s (1998) more recent studies, Hirsh and Sherrick never attempted to assess the effects of changing the response dimension while keeping the stimulus pairing the same to address this issue more convincingly: Intramodal temporal perception was always assessed using a spatial task, whereas multisensory perception was always assessed using an "explicitly" modality-first task. Therefore, in experiment 2, we decided to examine explicitly the effect of changing the dimension along which TOJs were made on multisensory JND and PSS response measures. Participants were now required to make an unspeeded TOJ regarding which side (left versus right) came first, regardless of the modality of the stimuli actually presented.

Methods

Participants

Ten participants recruited by advertisement took part in the experiment as paid volunteers. The participants were all right-handed and naïve as to the purpose of the experiment. The mean age for the participants was 25 years. All reported normal hearing and normal or corrected-to-normal vision.

Apparatus, materials, design and procedure

The apparatus, materials, design and procedure were as in experiment 1, with the following exceptions. Given the change in

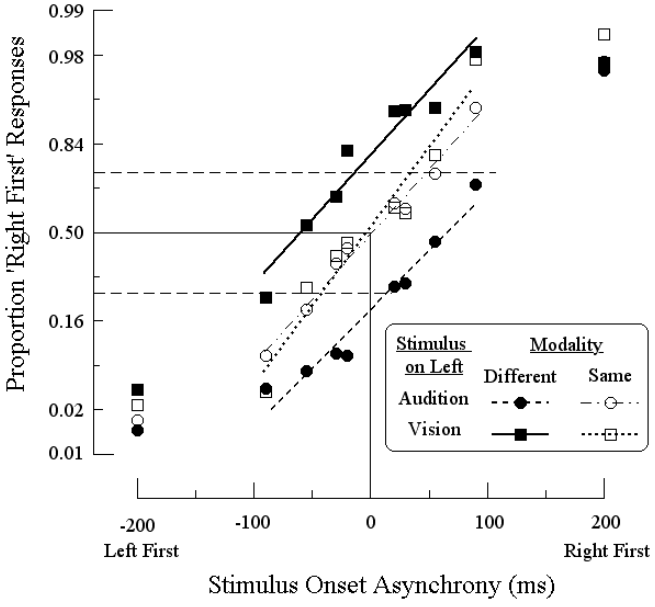

Fig. 3 Mean proportion of "right-first" responses plotted on a standard Gaussian *Z*-score scale as a function of the stimulus onset asynchrony (SOA) between the two stimuli for all four possible stimulus pairs in experiment 2. Participants performed an unspeeded "Which side came first?" temporal order judgment (TOJ) task for pairs of auditory and visual stimuli each presented to either side of fixation. The best-fitting straight lines for the intermediate eight SOAs are indicated for each of the four stimulus pairs (see text for details)

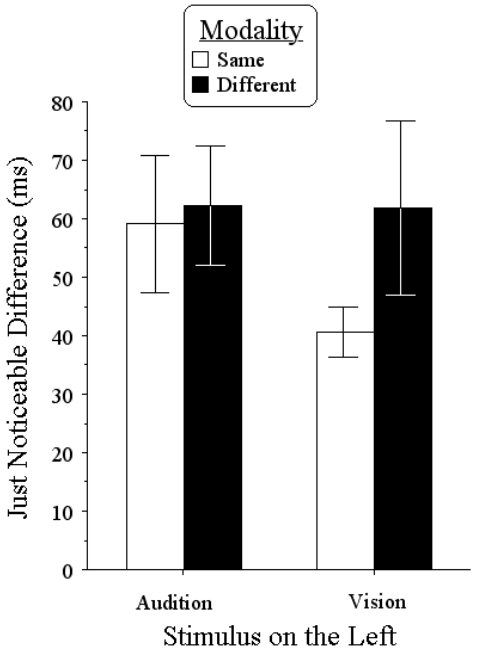

Fig. 4 Just noticeable differences for the four stimulus pairs in experiment 2. Participants performed an unspeeded "Which side came first?" TOJ task for pairs of auditory and visual stimuli presented one to either side of fixation. *Error bars* represent the within-observer standard error of the mean based on the ANOVA reported in the text. The results show that performance on the spatial TOJ task is not facilitated by the presentation of redundant modality cues (on the bimodal stimulus trials)

the dimension along which participants had to respond (i.e. from a modality discrimination task to a spatial discrimination task), the stimulus conditions were modified from those used in experiment 1. Specifically, participants were now presented with the following four possible stimulus combinations: sound left, vision right; sound right, vision left; sound left, sound right; and vision left, vision right. Participants were instructed to press the lower key whenever the left stimulus appeared to have been presented first, and the upper key whenever the right stimulus appeared to have been presented first. Response feedback consisted of the illumination of one of two LEDs placed directly above and below the fixation light immediately after a response had been detected. The lower light indicated a "left-first" response, and the upper light indicated a "right-first" response.

Results

The proportion of "right-first" responses was converted to its equivalent *Z*-score, and the intermediate eight SOAs were used to calculate a best-fitting straight line. Two participants were removed from the analysis because their calculated PSS was larger than 250 ms (see Spence et al. 2001b), indicating their inability to perform the task. The data from the remaining eight participants were subjected to a two-way within-participants ANOVA with the factors of Modality (unimodal versus bimodal stimulus pairs) and Stimulus on the left (vision versus audition; see Figs. 3 and 4, Table 1). JNDs were somewhat *smaller* for pairs of stimuli presented in the same modality (mean 50 ms; 41 ms for vision, 59 ms for audition) than for bimodal stimulus pairs (62 ms for both stimulus combinations), although the main effect of Modality failed to reach statistical significance ($F_{1,7}$=3.7, P=0.096). That is, in contrast to experiment 1, where redundant spatial information improved modality judgments, redundant modality information in the present experiment did not improve spatial TOJs and, if anything, *reduced* the accuracy of such judgments. The analysis of the JND data revealed no main effect of Stimulus on the left ($F_{1,7}$=1.61, P=0.25), nor any interaction between Stimulus on the left and Modality ($F_{1,7}$=1.31, P=0.29).

Analysis of the PSS data revealed a main effect of Stimulus on the left ($F_{1,7}$=28.1, P=0.001), and a two-way interaction between Modality and Stimulus on the left ($F_{1,7}$=15.7, P=0.006). This pattern of results can be attributed to the fact that the data for experiment 2 are coded in terms of "the amount by which the left stimulus had to lead", and the visual stimulus was on the left for one condition and on the right for the other condition. The interaction occurred because the left stimulus has to lead the right stimulus by 75 ms when vision is on the left and vice versa when vision is on the right. There was no significant difference between the amount by which the visual stimulus had to lead the auditory stimulus for these two conditions ($F_{1,7}$<1.0). If we were to convert the measure to "the amount by which the visual stimulus had to lead" for the bimodal stimulus pairs, then no interaction would be found. In comparing the PSS to the actual point

of simultaneity (i.e. to 0 ms), we found that the 75 ms by which visual stimuli had to be presented prior to the auditory stimuli to be significant [$t(7)=3.7$, $P=0.008$ for vision on the left; and $t(7)=7.0$, $P=0.0002$ for vision on the right]. For the unimodal stimuli, the left stimulus (auditory or visual) had to be presented 5 ms prior to the right stimulus for simultaneity to be achieved; however, this difference failed to reach statistical significance [$t(7)=0.32$ for audition, and $t(7)=0.70$ for vision; n.s.].

Comparison across experiments 1 and 2

Two of the stimulus pairings (sound left, vision right and sound right, vision left) were presented to participants in both experiments 1 and 2. We therefore conducted a between-experiments analysis to assess whether the dimension along which TOJs were made could affect the precision of multisensory temporal perception. A priori, we predicted that multisensory temporal precision would be unaffected by the dimension on which participants had to respond (see Hirsh and Sherrick 1961 for identical assumptions), especially if the stimulus pairs presented were identical. However, analysis of the JND data from the bimodal conditions in the two experiments (see Fig. 5) using a mixed between–within ANOVA with the factors of Task (modality versus spatial) and Stimulus on the left (vision or audition), revealed a significant main effect of Task ($F_{1,15}=14.6$, $P=0.002$). People were more precise when judging which modality was presented first (JND 22 ms) than when judging which side was presented first (JND 62 ms). There was no significant main effect of Stimulus on left ($F_{1,14}<1.0$, n.s.), nor any interaction between these two factors ($F_{1,14}<1.0$, n.s.).

A similar analysis of the PSS data revealed no main effects, nor interactions ($F_{1,15}=1.2$ for the two-way interaction; $F_{1,15}<1.0$ for both main effects). Note that for this analysis the PSS values for experiment 2 were converted from "the amount by which the right stimulus had to lead" to "the amount by which vision had to lead" to make the results compatible with those of experiment 1. This analysis showed that visual stimuli had to lead auditory stimuli by the same amount (i.e. 75 ms; range 70–80 ms) regardless of whether participants judged which modality came first, or which side came first.

Discussion

Presenting stimuli from different modalities in the *spatial* TOJ task of experiment 2 did not improve temporal performance. In fact, there was actually a marginal disadvantage for the bimodal stimulus pairs when compared with the unimodal stimulus pairs. This stands in contrast to the results of experiment 1, where we found that different *locations* helped in a *modality* TOJ task. In considering the three alternative accounts proposed in the Discussion to the previous experiment, it is clear that the redundant information account cannot be sustained. In experiment 2, participants were once again provided with redundant information on a subset of trials (i.e. on the bimodal trials), and yet we found no improvement in performance for these trials as compared to the unimodal trials. The other two remaining, potentially viable accounts—multisensory binding and limited capacity hemispheric resources—cannot be differentiated on the basis of the present data, since none of the stimuli were presented from the same location. That is, we can say from the results of experiment 2 that the contribution of redundant information cannot account for the relative cost for Same position trials in experiment 1, but cannot distinguish between the other two possible accounts of that effect.

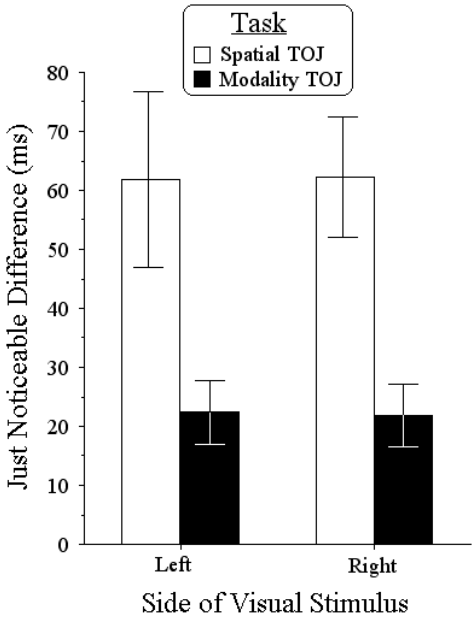

Fig. 5 Just noticeable differences for the two bimodal conditions in common across experiments 1 and 2, showing the effect of changing the dimension along which temporal order judgments (*TOJ*) are made on measures of multisensory temporal precision. *Error bars* represent the within-observer standard error of the mean based on the ANOVA reported in the text. The results show how changing the dimension on which TOJs are made can affect the accuracy of multisensory TOJs, even when the actual stimuli remain constant

The second observation to emerge from the analysis of experiment 2 relates to the relatively high JNDs observed in the cross-modal conditions. The between-experiments analysis revealed that JNDs for bimodal stimulus pairs are significantly affected by the dimension along which participants make their response: JNDs were significantly larger for spatial judgments as compared to modality judgments, even though the bimodal stimuli on which these TOJs were made were identical. One potential explanation for this effect comes from a consideration of the spatial ventriloquism effect—if the visual stimulus attracted the sound toward its position in the Different position condition, then observers might have found it more difficult to distinguish left from right, thus perhaps

making the spatial TOJ task harder than the modality judgment task.

The multisensory binding hypothesis outlined earlier may also help to account for this result. Even though we compared identical pairings of stimuli, they were presented in different contexts—in experiment 1, half of the stimuli were presented from the same position, whereas in experiment 2 all of the stimuli came from different locations. Given the relative difficulty of the Same position condition in experiment 1, observers may have enjoyed a relative benefit on the Different position conditions that was not present in experiment 2. Clearly this is a somewhat speculative account, but we believe that it is one that should be considered since, in many other domains, the context of presentation has been shown to have a significant influence on performance (see Sade and Spitzer 1998; Urbach and Spitzer 1995).

General discussion

Experiment 1 demonstrated a significant advantage for modality TOJs when the auditory and visual stimuli were presented from different spatial locations rather than from the same location. This raises a potentially important methodological confound in *all* previous multisensory TOJ research. In particular, these results indicate that people may be significantly worse at making multisensory TOJs than has been suggested previously. Interestingly, and in contrast to experiment 1, there was no advantage (in fact, there was actually a marginally significant disadvantage) when redundant modality cues were present in the spatial TOJ task of experiment 2. An additional finding to emerge from both experiments is that visual stimuli have to be presented prior to auditory stimuli for simultaneity to be achieved (at least for the particular stimulus configurations used in the present experiments). Finally, the dimension on which participants made their TOJs significantly affected the temporal precision for identical bimodal stimulus pairs.

Taken together, the results of the experiments reported in the present study suggest that the JND reported in any particular TOJ study may depend on a number of factors. In particular, the relative position of the stimuli and the dimension along which participants make their responses. In combination with a number of other factors previously shown to modulate the accuracy of multisensory TOJs, such as the experience of the participants on psychophysical tasks and the relative intensity of the stimuli (Bald et al. 1942; Dinnerstein and Zlotogura 1968; Spence et al. 2001b, 2003), it becomes increasingly clear how difficult it is to derive any meaningful conclusions from the comparison of TOJ performance across different studies of multisensory temporal perception where one or more of these factors vary.

Direct between-experiments comparison of performance on the bimodal trials (sound left, vision right and vision left, sound right) that were common to both experiment 1 and 2 revealed that TOJ accuracy was significantly better in the modality discrimination task of experiment 1 (mean JND of 22 ms) than in the spatial discrimination task of experiment 2 (M=62 ms). Given that exactly the same stimuli were presented in both tasks on these bimodal trials, it is rather surprising that performance was so much better when participants made modality discriminations than when they made spatial discriminations. Nevertheless, a similar pattern of results has also been demonstrated for visuotactile TOJs (Spence et al. 2001b; experiment 1) when participants had to switch between making "Which modality came first?" and "Which side came first?" judgments. These results therefore highlight another confound inherent in Hirsh and Sherrick's (1961) seminal TOJ studies, since they used different response dimensions when assessing intramodal versus multisensory TOJ performance.

In our discussion of the results of experiment 1, we presented three possible accounts for the improvement in TOJ performance (i.e. the reduction in JNDs) seen when auditory and visual stimuli were presented from different (rather than the same) spatial locations. According to the *redundancy account*, participants in experiment 1 may have used any redundant information regarding which location was stimulated first to facilitate their judgments regarding which modality had been presented first in the Different position condition. It was argued that the larger JNDs seen in the Same position condition might therefore reflect the fact that no such redundant spatial information was available to facilitate performance when the two stimuli were presented from the same spatial location. According to the redundancy account, participants should also have made more accurate TOJs when redundant modality information was provided in the bimodal trials of experiment 2. However, performance was, if anything, slightly worse on the bimodal trials than on the unimodal trials of this experiment, thus ruling out the redundancy account of our data.

The second possible account put forward to explain the facilitation of TOJ performance seen when the auditory and visual stimuli were presented from different positions in experiment 1 was in terms of the *multisensory binding* hypothesis. According to this account, auditory and visual stimuli are less likely to be bound together into a single multimodal percept when presented from different spatial locations than from the same location (Morein-Zamir et al. 2003; Scheier et al. 1999; Slutsky and Recanzone 2001; though see Regan and Spekreijse 1977). Consequently, participants in experiment 1 may have found it easier to distinguish between the temporal onset of the auditory and visual stimuli in the Different positions condition than in the Same position condition, because multisensory binding was less likely to have taken place in the former condition than in the latter condition. The multisensory binding account does not make any specific predictions regarding the results of experiment 2, since the stimuli in the bimodal trials were always presented from different positions.

The third possible account for the results of experiment 1 was in terms of *limited capacity hemispheric resources*.

According to this account, performance in the Different positions condition of experiment 1 may have been facilitated by the fact that each stimulus was processed, at least initially, by different cerebral hemispheres. Given that there appear to be to-some-degree independent attentional resources for each hemisphere (Banich 1988; Friedman and Polson 1981; Herdman and Freidman 1985; Kinsbourne and Cook 1971; Kinsbourne and Hicks 1978), one might argue that there may have been more attentional resources available to process the two stimuli in the Different positions condition than in the Same position condition of experiment 1. However, since the two stimuli were always presented from different sides (i.e. processed initially by different cerebral hemispheres) in experiment 2, the hemispheric resources account does not predict any difference between performance between the unimodal and bimodal trials of this experiment.

One might consider it unlikely that the hemispheric resource account could explain the difference in performance between the Same and Different position trials in experiment 1, given the generally very low perceptual demands of the task (i.e. participants were only required to make an unspeeded response to two briefly presented stimuli). However, it is worth noting that previous studies providing support for the role of hemispheric resources in human performance have typically used stimulus displays that were not much more complex or demanding than those used here (Banich 1988; Friedman and Polson 1981; Herdman and Freidman 1985; Kinsbourne and Hicks 1978).

Clearly, it will be important in future research to try and distinguish between the multisensory binding and hemispheric resources accounts of the relative spatial position effect demonstrated in experiment 1. One potential avenue for such research would be to vary the non-spatial attributes of the auditory and visual stimuli used, such as the degree of semantic congruency between them. For while varying the degree of congruency between the auditory and visual stimuli is known to influence the degree of multisensory integration (or cross-modal binding) that takes place (Calvert et al. 1998), it should have relatively little effect on the hemispheric distribution of resources required to process the stimuli. Alternatively, one could also vary the relative spatial position of the auditory and visual stimuli, while still presenting them both from within the same hemifield (such as by presenting the stimuli at different elevations on either the left or right), so that the two stimuli are always projected to the same cerebral hemisphere no matter whether they are presented from the same or different positions. We are currently conducting a series of experiments along these lines (see Zampini et al. 2003).

Another finding to emerge from the results of the present experiments is that no matter whether participants are making a modality or spatial TOJ, vision has to lead audition by 60–80 ms for the two stimuli to be perceived as simultaneous (i.e. for the PSS to be reached). There are several factors that help to account for this finding. First, neural transduction latencies at the peripheral sensory epithelia are very different for the two modalities; It takes less than 1 ms for auditory stimuli to be transduced at the cochlea, whereas it takes 30–40 ms for visual stimuli to be transduced at the retina (King and Palmer 1985; Pöppel et al. 1990; Spence and Squire 2003).[4]

A second factor that may also help to account for any differences in apparent processing latencies between the two modalities relates to differences in the baseline distribution of attention between the two modalities (see Spence et al. 2001b), which may in itself be linked to differences in the alerting properties for different sensory modalities (Posner et al. 1976). Additionally, the absolute stimulus intensity has also been shown to have an effect on the PSS (Neumann and Niepel 2003; Smith 1933), as has the absolute spatial direction from which the stimuli are presented (i.e. centrally presented stimuli tend to be perceived before peripherally presented stimuli; Smith 1933; Whipple et al. 1898; see also Posner et al. 1976).

So how does the brain cope with these asynchronies when trying to integrate stimuli from different sensory modalities? Neurophysiological studies of the response properties of neurons in multisensory brain structures such as the superior colliculus provide one possible solution. They suggest that many of the cells involved in multisensory integration solve the problem by having relatively wide temporal windows for the integration of sensory signals from different modalities (see King and Palmer 1985; Meredith and Stein 1983; Stein and Meredith 1993), thus overcoming any problems associated with asynchronous arrival times (at least within the ranges of asynchrony that are likely to occur naturally). Alternatively, other researchers have suggested that there may be cortical mechanisms for the resynchronization of sensory stimuli that have fallen out of synchrony due either to differences in transduction latencies or differences in more central cortical processing latencies (Bergenheim et al. 1996; Grossberg and Grunewald 1997; Moutoussis and Zeki 1997; see also Sugita and Suzuki 2003).

The neural basis of multisensory temporal perception

At present, little is known about the brain mechanisms involved in detecting temporal synchrony between multimodal sensory inputs (Bushara et al. 2001; Rao et al. 2001). In fact, only two neuroimaging studies have so far directly investigated the neural correlates of multisensory temporal perception (Bushara et al. 2001; Raizada and

[4] Because light travels through air far more rapidly than sound, these differences in transduction latencies can be offset, or even reversed, for pairs of audiovisual stimuli that occur very far from us (see Pöppel et al. 1990). For example, we often see distant lightening before hearing the associated thunder (Spence and Squire 2003). However, such differences are unlikely to have had any significant effect in the present studies since the auditory and visual stimuli were presented relatively close to the participants (distance=62 cm; see also Sugita and Suzuki 2003).

Poldrack 2001). Using a simultaneous/successive judgment task (rather than a TOJ task as used here), these researchers revealed that multisensory synchrony detection resulted in the activation of a large-scale dynamic network of cortical and sub-cortical areas, including the insula, posterior parietal and prefrontal cortex, and cerebellar areas. Unfortunately, the auditory and visual stimuli were presented from very different positions in these neuroimaging studies (i.e. the spatial confound was present), making it unclear whether the networks of activation they reported are specifically related to *multisensory synchrony detection* per se, or whether they might reflect neural processes associated with *spatial ventriloquism* (Bertelson 1998; Slutsky and Recanzone 2001; Spence 2001). It should be noted that the spatial ventriloquism effects that occur when auditory and visual stimuli are presented at approximately the same time, but from spatially disparate locations, are precisely the conditions that might have led to the perception of simultaneity in these neuroimaging studies.

Given the unknown contributions of spatial (i.e. ventriloquism) and temporal binding of multisensory information in both Bushara et al.'s (2001) and Raizada and Poldrack's (2001) results, it will be very interesting in future studies to use an improved methodology (i.e. avoiding the spatial confound) to investigate neural loci of specifically *temporal* binding. Furthermore, it will also be interesting to compare the pattern of neural activation across both the simultaneous/successive and TOJ tasks, to assess whether similar neural structures are recruited for both types of discrimination, given claims that they may reflect very different processes/mechanisms (i.e. one related to multisensory binding, and the other related to temporal discrimination instead; Shore et al. 2002; see also Allan 1975; Hirsh and Sherrick 1961; Mitrani et al. 1986).

Implications

The results of the present study add to a rapidly growing body of research highlighting the complex interplay of both spatial and temporal factors in modulating the multisensory integration of audiovisual stimuli (Fendrich and Corballis 2001; Lewald et al. 2001; Morein-Zamir et al. 2003; Roberson et al. 2001; Slutsky and Recanzone 2001). It is only by combining sophisticated behavioural methodologies with contemporary cognitive neuroimaging techniques such as functional magnetic resonance imaging (fMRI) and physiological measures such as event-related potentials and event-related fields that future studies will be able to provide a more comprehensive and accurate understanding of the neural underpinnings of multisensory temporal perception.

The present research findings may also have implications in a variety of real-world applications where the accurate estimation of people's sensitivity to audiovisual asynchrony is major importance. For example, researchers working on the design of hearing aids (McGrath and Summerfield 1985; Pandev et al. 1986) have to trade-off the potential benefits of increasing the quality of auditory signal processing provided by the hearing aid against the increased temporal delays (and consequent increase in the asynchrony between auditory and visual stimuli) that such improved signal processing typically requires. The modulatory role of spatial factors in the perception of simultaneity and temporal order identified in the present research have not, as yet, received consideration by researchers in these areas. Similarly, those working in a number of other areas, such as those working on the derivation of guidelines for satellite telecommunications broadcasting (ITU-T 1990; Rihs 1995), and in the development of new virtual-conferencing technologies (Finger and Davis 2001; Mortlock et al. 1997), are also very concerned with determining the limits of the perception of synchrony, given the desynchronization that such technologies often introduce into audiovisual broadcasts. Finally, it may also be possible that, by assessing the optimal asynchrony required for the perception of simultaneity (i.e. the PSS values derived in the present experiments), one may also be able to design more effective multisensory warning signals (see Spence and Driver 1999 for further discussion of this issue).

Acknowledgements Charles Spence and David I. Shore were funded by a Network Grant from the McDonnell-Pew Centre for Cognitive Neuroscience, University of Oxford. David I. Shore was also funded by an operating grant from the Natural Science and Engineering Research Council of Canada.

References

Allan LG (1975) The relationship between judgments of successiveness and judgments of order. Percept Psychophys 18:29–36

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

Banich MT (1998) The missing link: the role of interhemispheric interaction in attentional processing. Brain Cogn 36:128–157

Bergenheim M, Johansson H, Granlund B, Pedersen J (1996) Experimental evidence for a sensory synchronization of sensory information to conscious experience. In: Hameroff SR, Kaszniak AW, Scott AC (eds) Toward a science of consciousness: the first Tucson discussions and debates. MIT Press, London, UK, pp 303–310

Bertelson P (1998) Starting from the ventriloquist: the perception of multimodal events. In: Sabourin M, Craik FIM, Robert M (eds) Biological and cognitive aspects. (Advances in psychological science, Vol 2) Psychological, Hove, UK, pp 419–439

Bertelson P, Gelder B de (2003) The psychology of multimodal perception. In: Spence C, Driver J (eds) Crossmodal space and crossmodal attention. Oxford University Press, Oxford, UK

Bushara KO, Grafman J, Hallett M (2001) Neural correlates of auditory-visual stimulus onset asynchrony detection. J Neurosci 21:300–304

Calvert GA, Brammer MJ, Iversen SD (1998) Crossmodal identification. Trends Cognit Sci 2:247–253

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

Drew F (1896) Attention: experimental and critical. Am J Psychol 7:533–573

Driver J, Spence C (1998) Crossmodal links in spatial attention. Philos Trans R Soc B 353:1319–1331
Driver J, Spence C (2000) Multisensory perception: beyond modularity and convergence. Curr Biol 10:731–735
Exner S (1875) Experimentelle Untersuchung der einfachsten psychischen Processe (Experimental study of the most simple psychological processes). Pflugers Arch 11:403–432
Fendrich R, Corballis PM (2001) The temporal cross-capture of audition and vision. Percept Psychophys 63:719–725
Finger R, Davis AW (2001) Measuring video quality in videoconferencing systems. Technical Report SN187-D. Pixel Instrument Corporation, Los Gatos, CA
Finney DJ (1964) Probit analysis: statistical treatment of the sigmoid response curve. Cambridge University Press, London, UK
Friedman A, Polson MC (1981) Hemispheres as independent resource systems: limited-capacity processing and cerebral specialization. J Exp Psychol Hum Percept Perform 7:1031–1058
Gebhardt JW, Mowbray GH (1959) On discriminating the rate of visual flicker and auditory flutter. Am J Psychol 72:521–528
Gengel RW, Hirsh IJ (1970) Temporal order: the effect of single versus repeated presentations, practice, and verbal feedback. Percept Psychophys 7:209–211
Grossberg S, Grunewald A (1997) Cortical synchronization and perceptual framing. J Cognit Neurosci 9:117–132
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
Herdman CM, Friedman A (1985) Multiple resources in divided attention: a cross modal test of independence of hemispheric resources. J Exp Psychol Hum Percept Perform 11:40–49
Hirsh IJ (1959) Auditory perception of temporal order. J Acoust Soc Am 31:759–767
Hirsh IJ, Sherrick CE Jr (1961) Perceived order in different sense modalities. J Exp Psychol 62:423–432
ITU-T (1990) Television and sound transmission: tolerances for transmission time differences between the vision and sound components of a television signal. Telecommunication standardization sector of ITU, Recommendation J100, CMTT 717 in CCIR Recommendations 12. International Telecommunication Union, Düsseldorf, Germany
Jaskowski P, Jaroszyk F, Hojan-Jerierska D (1990) Temporal-order judgments and reaction time for stimuli of different modalities. Psychol Res 52:35–38
Kahneman D, Treisman A, Gibbs BJ (1992) The reviewing of object files: object-specific integration of information. Cognit Psychol 24:175–219
King AJ, Palmer AR (1985) Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. Exp Brain Res 60:492–500
Kinsbourne M, Cook J (1971) Generalized and lateralized effects of concurrent verbalization on a unimanual skill. Q J Exp Psychol 23:341–345
Kinsbourne M, Hicks RE (1978) Functional cerebral space: a model for overflow, transfer and interference effects in human performance: a tutorial review. In: Requin J (ed) Attention and performance, VII. Erlbaum, Hillsdale, NJ, pp 345–362
Kristofferson AB (1967) Attention and psychophysical time. Acta Psychol 27:93–100
Lewald J, Ehrenstein WH, Guski R (2001) Spatio-temporal constraints for auditory-visual integration. Behav Brain Res 121:69–79
McFarland DJ, Cacace AT, Setzen G (1998) Temporal-order discrimination for selected auditory and visual stimulus dimensions. J Speech, Lang Hearing Res 41:300–314
McGrath M, Summerfield Q (1985) Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. J Acoust Soc Am 77:678–685
Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. Science 221:389–391
Mitrani L, Shekerdjiski S, Yakimoff N (1986) Mechanisms and asymmetries in visual perception of simultaneity and temporal order. Biol Cybern 54:159–165
Morein-Zamir S, Soto-Faraco S, Kingstone AF (2003) Auditory capture of vision: examining temporal ventriloquism. Cogn Brain Res 17:154–163
Mortlock AN, Machin D, McConnell S, Sheppard P (1997) Virtual conferencing. BT Technol J 15:120–129
Moutoussis K, Zeki S (1997) A direct demonstration of perceptual asynchrony in vision. Proc R Soc Lond B Biol 264:393–399
Neumann O, Niepel M (2003) Timing of "perception" and perception of "time". In: Kaernbach C, Schröger E, Müller H (eds) Psychophysics beyond sensation: laws and invariants of human cognition. (Scientific Psychology Series) Erlbaum, Hove, UK
Pandev PC, Kunov H, Abel SM (1986) Disruptive effects of auditory signal delay on speech perception with lip-reading. J Aud Res 26:27–41
Pöppel E, Schill K, Steinbüchel N von (1990) Sensory integration within temporally neutral system states: a hypothesis. Naturwissenschaften 77:89–91
Posner MI, Nissen MJ, Klein RM (1976) Visual dominance: an information-processing account of its origins and significance. Psychol Rev 83:157–171
Rains JD (1963) Signal luminance and position effects in human reaction time. Vision Res 3:239–251
Raizada RDS, Poldrack RA (2001) Event-related fMRI of audio-visual simultaneity perception. Soc Neurosci Abstr 27:511.14
Rao SM, Mayer AR Harrington DL (2001) The evolution of brain activation during temporal processing. Nat Neurosci 4:317–323
Reeves B, Voelker D (1993) Effects of audio-video asynchrony on viewer's memory, evaluation of content and detection ability. Research report. Pixel Instruments, Los Gatos, CA
Regan D (1982) Visual information channeling in normal and disordered vision. Psychol Rev 89:407–444
Regan D, Spekreijse H (1977) Auditory-visual interactions and the correspondence between perceived auditory space and perceived visual space. Perception 6:133–138
Rihs S (1995) The influence of audio on perceived picture quality and subjective audio-visual delay tolerance. In: Hamberg R, Ridder H de (eds) Report 1071. Proc MOSAIC workshop, Eindhoven, 18–19 Sept: Advanced methods for the evaluation of television picture quality. Inst Percept Res, pp 133–137
Roberson GE, Hairston WD, Wallace MT, Stein BE, Laurienti PJ, Schirillo JA (2001) Unifying multisensory signals across time and space. Soc Neurosci Abstr 27:511.19
Rutschmann J, Link R (1964) Perception of temporal order of stimuli differing in sense mode and simple reaction time. Percept Mot Skills 18:345–352
Sade A, Spitzer H (1998) The effects of attentional spread and attentional effort on orientation discrimination. Spatial Vision 11:367–383
Scheier CR, Nijhawan R, Shimojo S (1999) Sound alters visual temporal resolution. Invest Ophthalmol Vis Sci 40:S792
Shore DI, Spence C, Klein RM (2001) Visual prior entry. Psychol Sci 12:205–212
Shore DI, Spry E, Spence C (2002) Confusing the mind by crossing the hands. Cogn Brain Res 14:153–163
Slutsky DA, Recanzone GH (2001) Temporal and spatial dependency of the ventriloquism effect. Neuroreport 12:7–10
Smith WF (1933) The relative quickness of visual and auditory perception. J Exp Psychol 16:239–257
Spence C (2001) Crossmodal attentional capture: a controversy resolved? In: Folk C, Gibson B (eds) Attention, distraction and action: multiple perspectives on attentional capture. Elsevier Science, Amsterdam, pp 231–262
Spence C, Driver J (1999) A new approach to the design of multimodal warning signals. In: Harris D (ed) Job design, product design and human-computer interaction. (Engineering psychology and cognitive ergonomics, Vol 4) Ashgate: Hampshire, pp 455-461

Spence C, Squire SB (2003) Multisensory integration: maintaining the perception of synchrony. Curr Biol 13:R519–R521
Spence C, Nicholls MER, Driver J (2001a) The cost of expecting events in the wrong sensory modality. Percept Psychophys 63:330–336
Spence C, Shore DI, Klein RM (2001b) Multisensory prior entry. J Exp Psychol Gen 130:799–832
Spence C, Baddeley R, Zampini M, James R, Shore DI (2003) Crossmodal temporal order judgments: when two locations are better than one. Percept Psychophys
Stein BE, Meredith MA (1993) The merging of the senses. MIT Press, Cambridge, MA
Sternberg S, Knoll RL, Gates BA (1971) Prior entry reexamined: effect of attentional bias on order perception. Presented, meeting Psychon Soc, November, St. Louis, MO
Stone JV, Hunkin NM, Porrill J, Wood R, Keeler V, Beanland M, Port M, Porter NR (2001) When is now? Perception of simultaneity. Proc R Soc B 268:31–38
Stone SA (1926) Prior entry in the auditory-tactual complication. Am J Psychol 37:284–287
Sugita Y, Suzuki Y (2003) Implicit estimation of sound-arrival time. Nature 421:911
Teatini G, Farnè M, Verzella F, Berruecos P Jr (1976) Perception of temporal order: visual and auditory stimuli. G Ital Psicol 3:157–164
Ulrich R (1987) Threshold models of temporal-order judgments evaluated by a ternary response task. Percept Psychophys 42:224–239
Urbach D, Spitzer H (1995) Attentional effort modulated by task difficulty. Vision Res 35:2169–2177
Van de Par S, Juola JF, Kohlrausch A (1999) Judged synchrony/asynchrony for light-tone pairs. Presented 40th Ann meeting Psychon Soc, Los Angeles, CA
Welch RB, DuttonHurt LD, Warren DH (1986) Contributions of audition and vision to temporal rate perception. Percept Psychophys 39:294–300
Whipple GM, Sanford EC, Colgrove FW (1899) Minor studies from the psychological laboratory of Clark University: on nearly simultaneous clicks and flashes: the time required for recognition: notes on mental standards of length. Am J Psychol 10:280–295
Zampini M, Shore DI, Spence C (2003) Multisensory temporal order judgments: the role of hemispheric redundancy. Int J Psychophysiol (in press)