Auditory–visual temporal integration measured by shifts in perceived temporal location

Philip M. Jaekl*, Laurence R. Harris

Centre for Vision Research, York University, Toronto, Ontario, Canada

Received 15 August 2006; received in revised form 21 January 2007; accepted 9 February 2007

Abstract

The perceived time of occurrence of a visual stimulus may be shifted towards the onset of an auditory stimulus occurring a short time later. The effect has been attributed to auditory–visual temporal integration although an unknown portion of the shift may be explained by the different processing times of visual and auditory stimuli. Here, perceived onset time is measured in a novel way that separates and compares the magnitude of these effects. Participants observed either a sequence consisting of a visual stimulus followed by an auditory stimulus and then another visual stimulus or the reverse. The temporal location of the second stimulus was varied systematically between the onset of the first and third stimuli, which were separated by a fixed duration. Two timescales were used: a short timescale that allowed for auditory–visual temporal integration to occur, and a long timescale that did not. Psychometric curves were fitted for both timescales, to the percentage the first interval was perceived is shortest, as a function of first interval duration. For the long timescale condition the point of subjective equality (PSE) of the two interval lengths was consistent with the different processing latencies. When visual and auditory stimuli occurred within 125 ms significant additional shifting of the PSE occurred. These results indicate that temporal integration shifts the perceived timing of a visual stimulus by an amount much larger than can be explained differential processing latencies.

Multisensory integration of auditory and visual stimuli has been demonstrated in the spatial dimension by using light flashes and brief noise bursts [5]. These experiments have established 'spatial ventriloquism' [11] in which the perceived location of the auditory component of an auditory/visual multisensory stimulus is displaced in a statistically optimal way towards the component with the least variability in localization estimates [1,9]. The dominance of vision in ventriloquism is compatible with vision usually being the least variable sense in specifying spatial location [22,33]. Conversely, audition has been shown to be more reliable than vision in determining the temporal aspects of a stimulus [22,32,33] and when the components of a multisensory stimulus are separated in time, the perceived timing of the visual component is shifted towards that of the sound [2,5,7,15,18,22,23,29–31].

When auditory and visual stimuli are spatially congruent, temporal integration is greatest when the onset of the visual stimulus occurs before the onset of the auditory stimulus [2,5,7,15,18,22,23,29–31]. The visual stimulus then appears shifted in time towards the temporal location of the auditory stimulus provided the SOA is less than about 220 ms [7,15,18,25,31]. The fact that temporal integration occurs more effectively if the visual stimulus comes first corresponds to the only possible natural temporal staggering of the visual and auditory components of a single event because of the slower speed of sound. That it takes longer for visual information to reach awareness (vision about 60 ms: [12,14]; auditory about 15 ms: [4] see also Ref. [13]) however confounds this observation [18]. When a visual stimulus is presented before an auditory stimulus, the longer processing time of vision means that it will actually be processed closer in time to the auditory stimulus. Conversely if the auditory stimulus is presented first, the faster processing time of audition results in the components being separated further in time.

The experiments described here distinguish shifts in the relative timing of stimuli due to temporal integration pulling them

* Corresponding author at: Department of Psychology, York University, Toronto, Ontario, Canada M3J 1P3. Tel.: +1 416 736 2100x40886; fax: +1 416 736 5857.

E-mail address: jaekl@yorku.ca (P.M. Jaekl).

together from incidental shifts due to differential latencies. Participants judged the relative lengths of the time intervals in a sequence of three auditory and/or visual stimuli. Two timescales were used: a long timescale (600 ms) in which the separations between the stimuli were too great for significant temporal integration [25], and a short timescale (125 ms) in which the stimuli were close enough together to facilitate temporal integration in addition to latency effects. By comparing data obtained with the two timescales differential latency effects and temporal integration effects could be distinguished.

Fifteen participants (5 female, mean age = 26 years, range 22–45 years) were paid $ 10/hr and signed an informed consent form. All participants had normal hearing and normal or corrected-to-normal vision. This study was conducted according to the procedures outlined in the York University ethics code for human participants.

Participants sat in the dark with their heads supported on a chinrest 40 cm from a horizontal array of three 0.5 cm square, orange LEDs separated by 14° in front of a speaker with a green fixation LED 2 cm below (Fig. 1).

Visual stimuli were LEDs switching on (luminance 0.03 cd/m^2), sound stimuli were 5 ms bursts of white noise (81 db; Tucker Davis Technologies RP2). MATLAB, in conjunction with the Psychophysics Toolbox extensions [3,21] and the Realtime Processor Virtual Design Studio, was used to program the sequences.

Two controls and two experimental conditions were presented at two timescales: long (600 ms) and short (125 ms) (Fig. 1b, experimental conditions). Participants were presented with a sequence of three stimuli with either a visual stimulus (V) between two temporally flanking auditory stimuli (A): the VAV condition (Fig. 1b, boxes i and iii); an auditory stimulus between two temporally flanking visual stimuli (Fig. 1b, boxes ii and iv): the AVA condition; or control conditions with three visual stimuli (VVV) or three sounds (AAA). The temporal location of the central member (probe) relative to the temporally flanking stimuli was varied using the method of constant stimuli over a ±50 ms range from the true midpoint. There were five pairs of intervals for each condition. The first interval ranged from 12.5 to 112.5 ms (short timescale) or from 250 to 350 ms (long timescale), in 25 ms steps. The duration of the second interval was co-varied to keep the total duration at 125 or 600 ms for the short and long timescales respectively. For each stimulus sequence participants indicated whether the first or second interval appeared shorter. Each set of intervals was repeated 30 times for a total of 150 trials for each of the four sequence types. Conditions were counterbalanced with the exception that the control conditions were run first.

The percentage of instances the first interval was chosen as shortest was plotted as a function of first interval duration. A cumulative Gaussian was fitted to the data for each participant in each condition using the equation: $f = 100/(1 + \exp(-(x - x_0)/b))$ where b is the standard deviation, x the length of the first interval (ms) and x_0 is the point of subjective equality (PSE), that is the timing at which the central stimulus was perceived as midway between the flanking stimuli. We define the just noticeable difference (JND) as the standard deviation of the curve (b), i.e. when there was an 84% chance detecting a difference in the two interval lengths. If the regression coefficient (r^2) value for a given participant was less than 0.5 the participant re-ran the condition. If, on this second run, the participant scored another r^2 value below criterion, the participant's data were completely omitted from analysis. Four individual-condition data sets (out of a total of 80: each participant completing 10 sets) were omit-
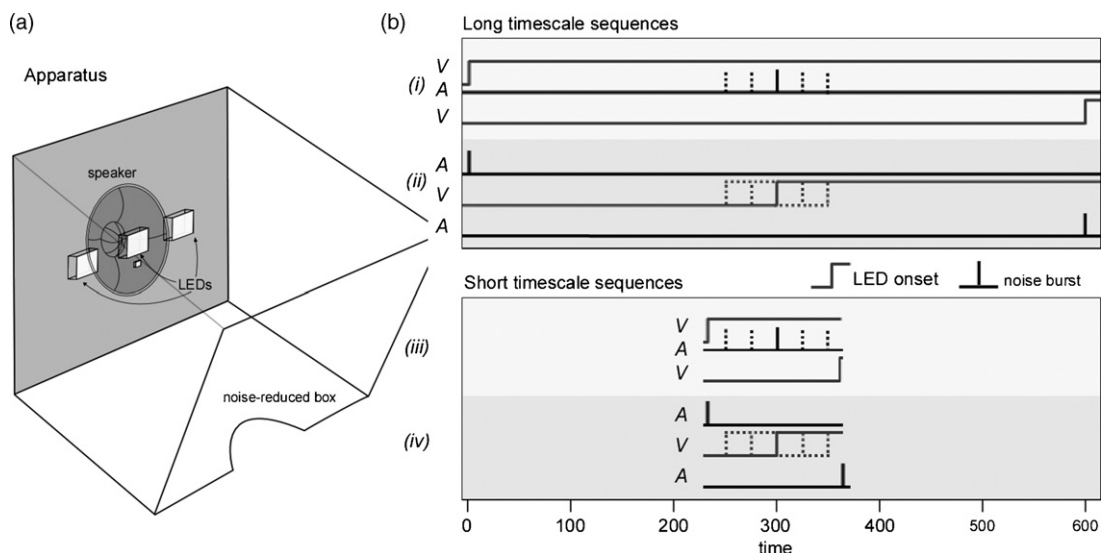

Fig. 1. (a) Experimental apparatus. Participants viewed a horizontal array of three square LEDs presented in a noise-reducing box. Behind the LEDs, a centrally positioned loudspeaker was covered with black material to keep it from view. (b) Stimulus sequences were either visual/auditory/visual (VAV) or auditory/visual/auditory (AVA) stimuli. The three rows in each panel show the timing of the first, second and third stimulus in each sequence, respectively. Visual stimuli are indicated with thick grey lines and auditory stimuli are shown with thick black lines. The configurations with the probe temporally centred between the first and third stimuli are shown with a solid line representing the probe. Other possible probe onset times are indicated with dashed lines. The possible onset times of the probe were within a ±50 ms range of the true midpoint between the onset of the first and third stimuli. (i and ii) Long timescale condition: intervals between stimulus onsets totaled 600 ms. (iii and iv) Short timescale condition: intervals totaled 125 ms.
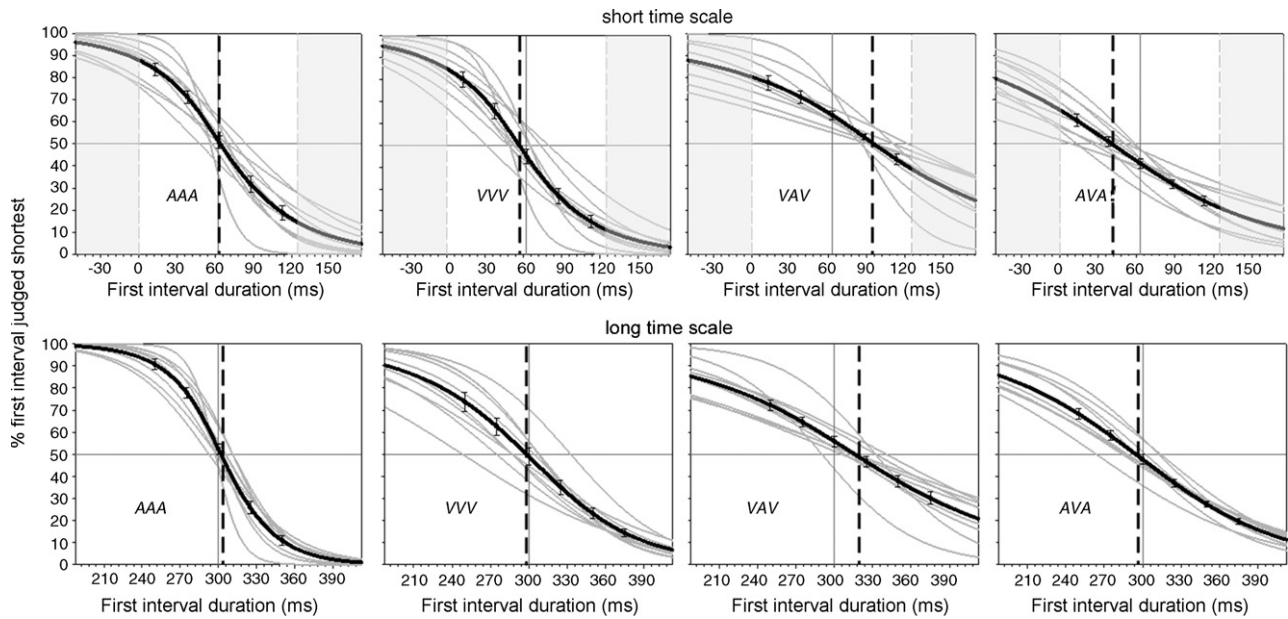
Fig. 2. Psychometric curves fit to the percentage of times that the first interval was judged shortest and plotted as a function of the length of the first interval for all conditions. Thin curves represent individual participants and thick curves are plotted through the mean of the individual curves. The data points are the average of all participants' responses for each interval with standard errors. The horizontal reference line indicates the 50% level where the participant was equally likely to choose the first or second interval as shortest. The psychometric curves cross this line, indicated by the vertical dashed lines, at the point of subjective equality. Solid vertical lines indicate the true midpoint between the flanking stimuli. *Top row*: Short timescale, midpoint = 62.5 ms. The temporal positions of the flanking stimuli are represented by the grey shaded areas. *Bottom row*: Long timescale, midpoint = 300 ms. Horizontal scales of these graphs have the same range.

ted from analysis because the r^2 values were below criterion. The PSEs for the four sigmoidal fits that were below the r^2 criterion were estimated according to equation 8.24 in [19]. The data were tested for outliers against a z-score criterion of 1.5. Five outliers were then removed (each point from a separate participant) and replaced using the same estimation procedure. 9/80 (11.25%) data points were estimated.

Fig. 2 illustrates for each condition (VVV, AAA, VAV, AVA) and each participant, psychometric curves fitted to the mean percentages at which the first interval appeared shortest, plotted as a function of the duration of the first interval for the long and short timescales.

Fig. 3 shows the displacement of the mean PSEs from the actual midpoints (0) for each of the four conditions. A positive displacement means the point of subjective equality occurred when the first interval was longer than the second (50% crossing to the right of the actual midpoint in Fig. 2).

For the VAV condition, the temporal position of the probe when it was perceived as central in the long timescale condition was shifted towards the trailing light by 16.8 ± 4.7 ms. This was increased to 35.6 ± 3.9 ms in the short timescale condition. For the AVA condition the temporal location of the light when it was perceived as central was displaced towards the leading sound by 7.0 ± 4.7 ms. This was increased to 24.9 ± 5.7 ms in the short timescale condition. The mean absolute displacement for the AVA and VAV conditions in the long timescale condition was 11.9 ms and this was increased by 18.4 to 30.3 ms in the short timescale condition.

Multiple one-sample t-tests were used to compare the mean shifts in the PSEs for each condition against a test value of 0 (no shift) (using Bonferroni type-1 error control; alpha = 0.05/4).

The mean shifts for the short timescale multisensory conditions were significant, with the mean for the VAV condition being significantly higher than 0 ($t_{vav} = 9.17$, df = 9, $p < 0.001$) and the mean for the AVA condition being significantly lower than 0 ($t_{ava} = -4.37$, df = 9, $p < 0.01$). For the long timescale condition, the mean PSE for the VAV combination was significantly shifted from 0 ($t_{vav} = 3.59$, df = 9, $p < 0.01$), but for the AVA condition was not ($t_{ava} = -1.49$, df = 9, $p = 0.84$). The JNDs for both the long and short timescales were significantly longer for the multisensory conditions (VAV and AVA) than for the unisen-
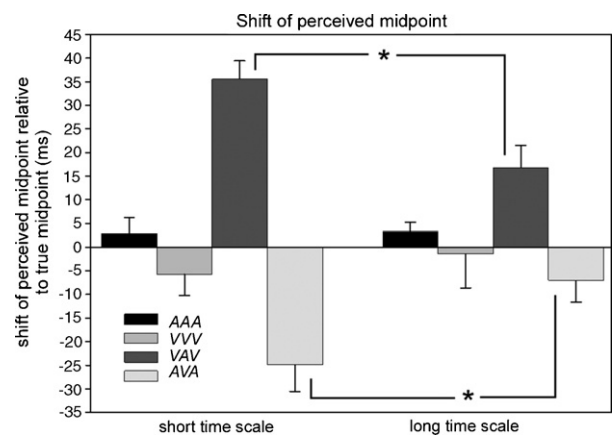

Fig. 3. The temporal position of a probe stimulus judged as centred for each condition. The mean perceived midpoints are the average of each participant's PSE relative to the true midpoints. Positive values indicate that intervals were judged equal when the first was longer. Standard error bars are shown. The asterisks indicate that the central probes of both the VAV and the AVA combinations were shifted significantly more in the short timescale condition than in the long timescale configuration.

sory conditions (AAA and VVV conditions) (t_{short} = 4.57, df = 9, $p_{onetail}$ < 0.001; t_{long} = 2.32, df = 9, $p_{onetail}$ < 0.05).

Paired-sample t-tests (alpha = 0.05/2) showed significant differences in the magnitudes of the shifts of the PSE between both the short and long timescale VAV conditions and the short and long timescale AVA conditions. For the VAV conditions the PSE was significantly more shifted towards the second visual stimulus in the short timescale condition compared to the long timescale condition (t_{vav} = 2.32, df = 9, $p_{onetail}$ < 0.025). For the AVA conditions the PSE was significantly more shifted towards the first auditory stimulus in the short timescale condition compared to the long timescale condition (t_{ava} = −3.37, df = 9, $p_{onetail}$ < 0.01).

This study has shown that when the temporal positions of stimuli are judged relative to each other, there are distortions due to the different processing times of light and sound and to sensory interactions. Our long timescale condition showed a shift attributable to sensory timing differences and our short timescale condition showed a significantly longer shift attributable to sensory interactions in addition to timing differences.

When the SOA between visual and auditory stimuli exceeds approximately 225 ms, stimuli are routinely perceived as separate in time [25]. Thus, at an SOA of between 250 and 300 ms (our long timescale), auditory/visual temporal integration is not likely to have contributed significantly to judgments of which interval was shortest (but see [10,16,20]). The shift of the central probe of about 12 ms can thus be attributed to the relatively longer processing time of visual as compared to auditory stimuli [15]. This value is less than the difference between previously estimated processing latencies for vision and audition [8,13,14,17,24] of around 40–55 ms. These studies used measurements obtained directly using electrophysiological methods [8,14,17,24] and reaction time differences [13]. Latencies also vary with characteristics of the stimulus (intensity, energy, etc.) and therefore it was necessary to ascertain the difference in delay for the particular stimuli used in this experiment. The value of 12 ms reported here represents the difference in processing times of these particular stimuli and with our particular task.

In the short timescale condition the central probe stimulus was always between 12.5 and 62.5 ms of a flanking stimulus and thus the probe and flanking stimuli were within the 225 ms window where auditory/visual integration is known to occur [25]. Both temporal integration and processing time differences would contribute to shifts in perceived temporal location in this condition. The average shift in the perceived midpoint was 26 ms. The long timescale condition demonstrated that about 12 ms of this could be accounted for by differential latency effects indicating a temporal shift due to sensory interactions of about 18 ms. The perceptual midpoints of the unisensory sequences (AAA, VVV) were not significantly different from the true midpoints for either timescale indicating that responses were uninfluenced by attentional [27,28], apparent motion, or other sensory bias (e.g. [25]).

Morein-Zamir et al. [18] found that a visual stimulus was temporally shifted towards an auditory stimulus under comparable circumstances. Visual stimuli that were integrated with auditory stimuli were separated by ±10° in that study and ±14° in the present study. This small spatial separation is unlikely to impair multisensory integration, especially in view of the poorly localizable, 5 ms sound burst. Indeed, Vroomen and Keetels [31] showed that separation of visual and auditory stimuli by as much
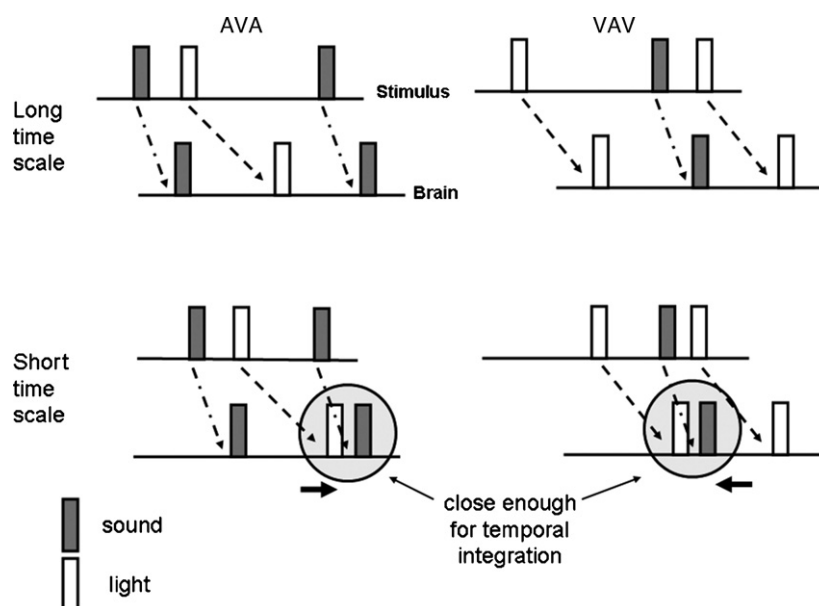

Fig. 4. The PSE for the AVA and VAV conditions could be expected to be displaced in the long timescale condition (upper panels) because of the difference in processing times for light and sound. The top trace in each panel indicates the timing of the stimuli; the lower trace indicates their representation in the brain. Sound is processed faster than light as shown by the connecting arrows – dashed for slower (vision), dots-and-dashes for faster (audition) – resulting in a shift of the stimuli as shown, resulting in the brain representations being equally spaced. In the short timescale condition (lower panels) this shift brings the mental representations of some pairs of stimuli together close enough (indicated by the circled areas, see text) for temporal integration to occur. The perceived temporal location of the central stimulus may then be shifted by an additional amount.

as 90° had no effect on temporal order judgements of visual stimuli.

Previous studies [2,7,25] measured relative shifts in the perceived time of sensory stimuli by measuring SOAs that were judged as simultaneous. A hypothetical explanation for these effects and the effects in the present study, is illustrated in Fig. 4 in which arrows are used to represent the processing times of individual stimuli. Auditory and visual stimuli reach the brain closer together when a visual stimulus precedes an auditory stimulus than they do when the same stimuli are presented in the reverse order allowing for more integration and contributing to asymmetries that have been reported (e.g. [18]).

The perceptual temporal shifts due to auditory/visual integration in the multisensory conditions may have resulted from either a shift of the auditory stimulus towards the initial visual stimulus or a shift of the visual stimulus towards the auditory stimulus. Statistically optimal integration [6] would predict that the sound, being more reliable [22], would be least likely of the two to be shifted. In fact, auditory stimuli consistently appear to be more temporally reliable and modulate the perceived timing of visual stimuli [2,5,7,15,18,22,23,29–31] rather than the reverse. Some examples of such modulation include [22] in which a temporally based visual aftereffect is induced by auditory stimuli and [24] in which auditory stimuli modulate a visual flash-lag illusion. These studies suggest that within the nervous system, the less accurate visual temporal information is guided by more accurate temporal coding from the auditory system. Thus, in the present study it is likely that the perceived timing of the visual stimulus moved towards the onset of the auditory stimulus.

As well as showing larger shifts, the psychometric functions for the multisensory stimulus combinations for the short timescale conditions were shallower than for the unisensory conditions. Participants anecdotally reported the multisensory conditions to be more difficult than the unisensory conditions. Spence and Driver [26] and Turatto et al. [29] have demonstrated that automatic, exogenous attention is allocated towards a stimulated sensory modality. When a sequence of stimuli such as VAV or AVA was presented, attention may have been drawn to the first-experienced modality thus making the comparisons harder than those for the VVV and AAA conditions.

The perception of the time at which a stimulus occurs is an important aspect of perception in general and is a critical aspect of identifying events in the real world and for locating moving objects. This study indicates that the perceived timing of events is not determined exclusively by the stimuli and processing times: when stimuli are close together in time their perceived timing is shifted in a way that reflects and encourages binding across sensory modalities.

Acknowledgement

References

[1] D. Alais, D.C. Burr, The ventriloquist effect results from near-optimal bimodal integration, Curr. Biol. 14 (2004) 257–262.

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

[3] D.H. Brainard, The psychophysics toolbox, Spat. Vis. 10 (1997) 433–436.

[4] G.G. Celesia, Organization of auditory cortical areas in man, Brain 99 (1976) 403–414.

[5] B. De Gelder, P. Bertelson, Multisensory integration, perception and ecological validity, Trends Cogn. Sci. 7 (2003) 460–467.

[6] M.O. Ernst, M.S. Banks, Humans integrate visual and haptic information in a statistically optimal fashion, Nature 415 (2002) 429–433.

[7] R. Fendrich, P.M. Corballis, The temporal cross-capture of audition and vision, Percept. Psychophys. 63 (2001) 719–725.

[8] J.J. Foxe, C.E. Shroeder, The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex, Cogn. Brain Res. 14 (1) (2002) 187–198.

[9] W.D. Hairston, M.T. Wallace, J.W. Vaughan, B.E. Stein, J.L. Norris, J.A. Schirillo, Visual localization ability influences cross-modal bias, J. Cogn. Neurosci. 15 (2003) 20–29.

[10] H. Helfrich, Time and Mind II: Information Processing Perspectives, Hogrefe & Huber, Gottingen, 2003.

[11] I.P. Howard, W. Templeton, Human Spatial Orientation, John Wiley and Sons, New York, 1966.

[12] D.A. Jeffreys, J.G. Axford, Source locations of pattern-specific components of human VEPs, Exp. Brain Res. 16 (1972) 1–21.

[13] A. Kopinska, L.R. Harris, Simultaneity constancy, Perception 33 (2004) 1049–1060.

[14] N. Lesevre, Chronotopographical analysis of the human evoked potential in relation to the visual field, Ann. N. Y. Acad. Sci. 388 (1982) 156–182.

[15] J. Lewald, R. Guski, Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli, Brain Res. Cogn. Brain Res. 16 (2003) 468–478.

[16] W.H. Meck, Functional and Neural Mechanisms of Interval Timing, CRC Press, Boca Raton, FL, 2003.

[17] S. Molholm, W. Ritter, M.M. Murray, D.C. Javitt, C.E. Shroeder, J.J. Foxe, Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study, Brain Res. Cogn. Brain Res. 14 (1) (2002) 115–128.

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

[19] J.L. Myers, A.D. Well, Research Design and Statistical Analysis, Lawrence Erlbaum Associates Inc., Hillsdale, 1995.

[20] J. Navarra, A. Vatakis, M. Zampini, S. Soto-Faraco, W. Humphreys, C. Spence, Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration, Brain Res. Cogn. Brain Res. 25 (2005) 499–507.

[21] D.G. Pelli, The VideoToolbox software for visual psychophysics: transforming numbers into movies, Spat. Vis. 10 (1997) 437–442.

[22] G.H. Recanzone, Auditory influences on visual temporal rate perception, J. Neurophysiol. 89 (2) (2003) 1078–1093.

[23] C.R. Scheier, R. Nijhawan, S. Shimojo, Sound alters visual temporal resolution, Invest. Ophthalmol. Vis. Sci. 40 (Suppl.) (1999) 4169.

[24] C.E. Shroeder, J.J. Foxe, The case for feedforward multisensory convergence during early cortical processing, Neuroreport 16 (5) (2005) 419–423.

[25] D.A. Slutsky, G.H. Recanzone, Temporal and spatial dependency of the ventriloquism effect, Neuroreport 12 (2001) 7–10.

[26] C. Spence, J. Driver, Audiovisual links in exogenous covert spatial orienting, Percept. Psychophys. 59 (1) (1997) 1–22.

[27] C. Spence, D.I. Shore, R.M. Klein, Multisensory prior entry, J. Exp. Psychol. Gen. 130 (2001) 799–832.

[28] E.B. Titchener, Lectures on the Elementary Psychology of Feeling and Attention, Macmillan, New York, 1908.

[29] M. Turatto, F. Benso, G. Galfano, C. Umilita, Nonspatial attentional shifts between audition and vision, J. Exp. Psychol. Hum. Percept. Perform. 28 (3) (2002) 1–22.

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

[31] J. Vroomen, M. Keetels, The spatial constraint in intersensory pairing: no role in temporal ventriloquism, J. Exp. Psychol. Hum. Percept. Perform., in press.

[32] Y. Wada, N. Kitagawa, K. Noguchi, Audio–visual integration in temporal perception, Int. J. Psychophysiol. 50 (2003) 117–124.

[33] R.B. Welch, D.H. Warren, Immediate perceptual response to intersensory discrepancy, Psychol. Bull. 88 (1980) 638–667.