Available online at www.sciencedirect.com
SCIENCE DIRECT
Cognitive Brain Research 22 (2004) 32–35
COGNITIVE BRAIN RESEARCH
www.elsevier.com/locate/cogbrainres

Research report

Recalibration of temporal order perception by exposure to audio-visual asynchrony

Jean Vroomen[a,*], Mirjam Keetels[a], Beatrice de Gelder[a,b], Paul Bertelson[a,b]

[a]*Department of Psychology, Tilburg University, Warandelaan 2, Tilburg 500 LE, The Netherlands*
[b]*Université Libre de Bruxelles, Brussels, Belgium*

Accepted 6 July 2004
Available online 26 August 2004

Abstract

The perception of simultaneity between auditory and visual information is of crucial importance for maintaining a coordinated representation of a multisensory event. Here we show that the perceptual system is able to adaptively recalibrate itself to audio-visual temporal asynchronies. Participants were exposed to a train of sounds and light flashes with a constant time lag ranging from −200 (sound first) to +200 ms (light first). Following this exposure, a temporal order judgement (TOJ) task was performed in which a sound and light were presented with a stimulus onset asynchrony (SOA) chosen from 11 values between −240 and +240 ms. Participants either judged whether the sound or the light was presented first, or whether the sound and light were presented simultaneously or successively. The point of subjective simultaneity (PSS) was, in both cases, shifted in the direction of the exposure lag, indicative of recalibration.
© 2004 Elsevier B.V. All rights reserved.

Theme: Sensory systems
Topic: Visual psychophysics and behavior

Keywords: Intersensory perception; Recalibration; Aftereffects; Temporal order judgement

1. Introduction

Most natural events are processed by a number of different neural mechanisms. For example, seeing and hearing a talker provides multisensory information that is processed by specialized visual and auditory neural pathways. Several behavioural and neurophysiological studies have now highlighted the crucial role that temporal synchrony plays in binding such intersensory information so that a coherent representation of an event is obtained [1,2,5]. If temporal co-occurrence is indeed of crucial importance, the question arises how synchronization in the brain is achieved, as there are not only differences in physical transmission time between sound and light, but neural pathways also often differ considerably in processing speed. Given that neural architectures change over lifetime and that experience or attention can alter the neural response time to preferred or attended stimuli (e.g., the law of prior entry states that attended objects are perceived more rapidly than unattended objects [8,9]), it seems that any synchronization mechanism would need to be flexible in order to properly perform its function.

Here, we explore whether the perceptual system does indeed adapt to changes in the timing of intersensory events. The experiments build explicitly upon our work on adaptation to audio-visual *spatial* conflict [2,10]. The logic of the spatial conflict situation is that two sets of data, delivered in the auditory and visual modality, specify different locations but that other parameters, in particular synchronized timing, are those normally associated with a single event, and thus favour pairing of the two conflicting data. Provided that the conflicting data are indeed paired, their disagreement about location may be considered a misalignment of the sensory systems. Such misalignment manifests itself as an *immediate bias* of the perceived

* Corresponding author. Tel.: +31 13 4662394; fax: +31 13 4662067.
E-mail address: j.vroomen@uvt.nl (J. Vroomen).

0926-6410/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogbrainres.2004.07.003

auditory location towards the visual distracter, and following prolonged exposure to intersensory discrepancy, leads to adaptation or *recalibration* that is observable as an aftereffect [6]. Recalibration, in essence, reduces the conflict between vision and audition by shifting the least reliable modality—for spatial information the auditory one—towards the more precise modality, in this case vision [12].

Following the same logic, we predicted that if *temporally* misaligned but spatially co-localized auditory and visual data are paired, then the intersensory temporal misalignment might also manifest itself both as an immediate bias and as an aftereffect. Immediate temporal bias has indeed been demonstrated, as for example the perceived occurrence of a flash is attracted towards a temporally misaligned sound burst [5,11] (note that in the temporal domain, sound attracts vision as time is more accurately coded in audition). Temporal *aftereffects*, though, have not been explored. Here we therefore examined whether aftereffects indicative of temporal recalibration might be observed as well. Participants were exposed to a series of sounds and light flashes with a fixed temporal offset for some time. Following this exposure phase, the point of subjective simultaneity of a sound and light flash was measured by obtaining psychometric functions on temporal order judgements (TOJ) in two different tasks. Participants either judged on test trials whether a sound or a light had appeared first, or they judged whether a sound and light were presented simultaneously or successively. Two different tasks were used instead of one to check whether strategic effects rather than adaptive sensory changes influenced the results. Strategic adjustments are likely to be different in one or the other task, but not so for sensory changes. True temporal recalibration should therefore manifest itself in that in both tasks, the point of subjective simultaneity is shifted towards the previously experienced temporal conflict.

2. Method

2.1. Participants

Twenty students from Tilburg University participated. Half of them judged which modality appeared first (sound or light), the other half judged whether sound or light were presented simultaneously or successively. Participants reported normal hearing and normal or corrected-to-normal seeing. They were tested individually and were unaware of the purpose of the experiment.

2.2. Stimuli

The auditory stimulus consisted of a 2000 Hz tone of 20 ms duration (5 ms fade-in and fade-out) presented at 82 dB(A). Sounds were presented via a hidden loudspeaker placed at eye-level, 50 cm in front of the participant at a central location. The visual stimulus consisted of a 20 ms flash of a red LED (diameter of 1 cm, luminance of 40 cd/m^2), placed directly in front of the loudspeaker.

2.3. Design

The experiment had two within-subjects factors: Exposure lag during the exposure phase (−200, −100, 0, +100 and +200 ms with negative values indicating that the sound was presented first) and stimulus onset asynchrony (SOA) between the sound and light of the test stimulus (−240, −120, −90, −60, −30, 0, +30, +60, +90, +120 and +240 ms). These factors yielded 55 equi-probable conditions for the experimental trials. Each condition was presented 12 times for a total of 660 trials. Trials were presented in 15 blocks of 44 trials each (four repetitions for each SOA), preceded by two warm-up trials. Exposure lag was constant within a block and SOA varied randomly. Exposure lag varied between successive blocks with order counterbalanced across participants.

2.4. Procedure

Participants sat at a table in a dark and sound-proof booth. Head movements were precluded by a chin- and forehead-rest. Each block of experimental trials started with an exposure phase of 240 repetitions (3 min) of the sound–light stimulus pair (ISI=750 ms) with a constant time lag between the sound and the light. To ensure that participants were looking at the light during exposure, they had to attend the position where the light was presented. Unpredictably, a small green LED (a catch trial) could at that position be flashed once for 50 ms, and this occurred two, three, or four times during the exposure phase. Participants had to count and report at the end of exposure the number of catch trials. After a 10-s delay, the first trial then started.

Each trial consisted of two parts: an audio-visual re-exposure phase followed by the presentation of a sound and light whose temporal order had to be judged. The re-exposure phase consisted of a train of eight sounds and lights with the same time lag as was used during the exposure phase. After 1000 ms, the sound and light of the test stimulus were presented (with a variable SOA between them). The participants' task was to judge either whether the sound or the light of the test stimulus was presented first, or to judge whether the sound and light were presented simultaneously or successively. Participants made an unspeeded response by pressing one of two designated keys on a response box. The next trial started 2000 ms after a response.

Training was given prior to testing. Trials of the training session were not preceded by an exposure or re-exposure phase. Participants were either trained to distinguish sound-first from light-first trials (one block of −240 versus +240 ms SOA, and one block of −120 versus +120 ms SOA), or to distinguish simultaneous from successive trials (one

block of −240 and +240 ms versus 0 ms SOA, and one block of +120 and −120 versus 0 ms SOA). Each block consisted of 32 trials where each of the SOAs was presented equally often in random order. Whenever participants made an erroneous response, they received corrective feedback (a green LED flickering three times). Training continued until fewer than three erroneous responses were made within a block. Following this initial training, participants were exposed to the full range of SOAs without feedback in a single block of 154 trials (14 times the 11 SOAs used in the experiment proper). Testing lasted about 4 h, and was run on two consecutive days.

3. Results

Trials of the training session and warm-up trials were excluded from further analyses. Performance on catch trials was flawless, indicating that participants were indeed looking at the light during exposure. For participants who judged whether the sound or the light appeared first, the percentage of 'light-first' responses was calculated for each participant and for each combination of exposure lag (−200, −100, 0, +100 and +200 ms) and SOA (from −240 to +240 ms). For each of the thus obtained distribution of responses, an individually determined psychometric function was calculated by fitting a cumulative normal distribution using maximum likelihood estimation. The mean of the resulting distribution (the interpolated 50% crossover point) is the point of subjective simultaneity (henceforth the PSS), and the slope is a measure of the sharpness with which stimuli are distinguished from one another. For participants who judged whether the sound and the light were presented simultaneously or successively, the percentage of 'simultaneous' responses was calculated for each participant and for each combination of exposure lag and SOA. As the SOA varied from −240 to +240 ms, the probability of responding 'simultaneous' increased and then decreased. The resultant distribution of responses were fitted with a Gaussian function using maximum-likelihood estimation. Three defining parameters were estimated for each participant: the mean of the distribution providing a measure of the PSS, the peak height and the standard deviation.

Results showed, as expected, that in both tasks exposure to an audio-visual asynchrony shifted the PSS in the direction of the lag (Fig. 1), while there was no effect of exposure lag on the other estimated parameters (all F's<1). Thus, following exposure to sound-first adapters (−200 or −100 ms exposure), a sound of the test stimulus had to be presented earlier to the light in order to be perceived as simultaneous if compared with exposure to light-first adapters (+100 and +200 ms). In the overall ANOVA on the PSS, there was a highly significant effect of exposure lag, $F(4,72)=24.72$, $p<0.001$, with no significant overall difference between the two tasks ($F<1$). The interaction between exposure lag and task was not significant ($F(4,72)=2.6$, $p=0.07$), although there was a tendency that, at an exposure lag of −200 ms, the effect of exposure was somewhat bigger in the simultaneous/successive task than in the sound-first/light-first task. Possibly, this may reflect that the simultaneous/successive task is more sensitive to shifts in criterion. When the effect of exposure lag was partitioned into linear, quadratic, cubic and higher order trends, there was a significant linear trend, $F(1,18)=57.98$, $p<0.001$, indicating that the PSS shifted, on average, about 6.7% in the direction of the lag. There was also a significant cubic trend $F(1,18)=7.94$, $p<0.01$, indicating that the effect levelled off when the exposure lag reached ~+200 or ~−200 ms.

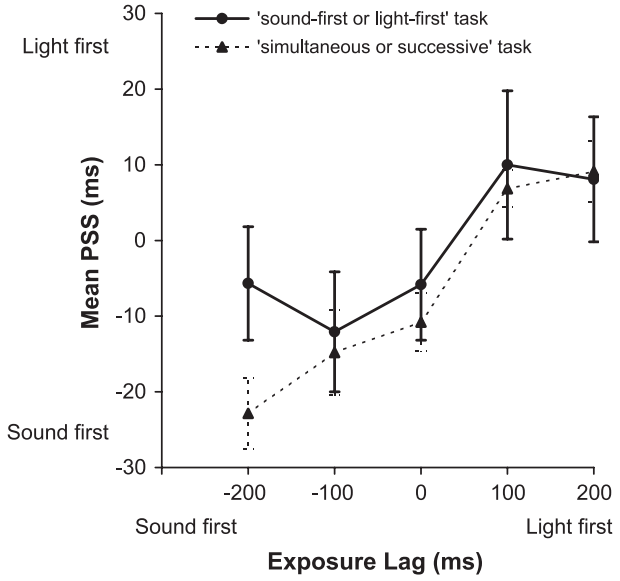

Fig. 1. The average point of subjective simultaneity (PSS) as a function of the audio-visual lag in the exposure phase. Error bars represent S.E.M. across participants. Participants either judged whether the sound or the light of the test stimulus was presented first (continuous line), or they judged whether the sound and light of the test stimulus were presented simultaneously or successively (dotted line). The PSS shifted, in both cases, towards the lag of the exposure phase.

4. Discussion

The present study showed for the first time that the point of subjective audio-visual simultaneity can be shifted towards a previously experienced temporal lag. This shift is interpreted as a manifestation of temporal recalibration: That is, when temporally offset but spatially co-localized audio-visual stimuli are paired, the criterion for simultaneity is shifted accordingly. The size of the shift is of the same order of magnitude as has been reported for recalibration in the spatial domain [7]. Moreover, the effect of exposure lag levelled off around ~±100 ms, which is also around the limit where a sound can capture the perceived onset of a light [5,11]. Beyond this limit, it becomes more likely that

data in the two modalities are not paired anymore, in which case there is no need for the perceptual system to recalibrate.[1]

Similar audio-visual induced aftereffects have now also been observed in the perception of space and speech [1,2,3]. The rule in all these cases is that whenever there is a moderate conflict between what is heard and what is seen, the brain takes advantage of the strength of each modality, such that the information that is most accurately coded in one modality changes the information encoded in the other, less accurate modality (see also Ref. [4]). In the present case, one might thus expect that since temporal information is more accurately coded in audition, vision has shifted towards audition. Admittedly, though, several other possibilities remain valid, as one might also argue that audition has been shifted towards vision, or it might even be the case that only the specific relation between the two modalities has changed. In future research, one may distinguish between these alternatives by checking whether adaptation to audio-visual temporal conflict generalizes to visual, auditory, or audio-visual tasks that measure the temporal occurrence of a stimulus.

One reasonable explanation how intersensory temporal recalibration might occur is that multi-modal neurons in the brain shift their temporal alignment preference during the exposure period to the discrepant stimuli. This implies that the intersensory representation of an object or event is dynamic, presumably to account for differences in the processing capacities of each sensory system. This may suggests that the way in which different sensory modalities are coordinated in time may not be because the brain employs a wide temporal window of integration, but because the window is actively changed, depending on previous experience.

[1] In a control experiment (N=10) in which participants judged which modality appeared first, we examined 'excessive' lags of the adaptors at SOAs of +350 ms (light-much-before-sound) and −350 ms (sound-much-before-light). In this case, there was no shift of the PSS (F<1) confirming that the audio-visual lag in the exposure phase has to be within limits for recalibration to occur.

References

[1] P. Bertelson, Ventriloquism: a case of crossmodal perceptual grouping, in: G. Aschersleben, T. Bachmann, J. Müsseler (Eds.), Cognitive Contributions to the Perception of Spatial and Temporal Events, Elsevier, North Holland, 1999, pp. 347–363.
[2] P. Bertelson, B. De Gelder, The psychology of multimodal perception, in: C. Spence, J.D. Driver (Eds.), Crossmodal Space and Crossmodal Attention, Oxford University Press, Oxford, 2004.
[3] P. Bertelson, J. Vroomen, B. De Gelder, Visual recalibration of auditory speech identification: a McGurk aftereffect, Psychol. Sci. 14 (2003) 592–597.
[4] M.O. Ernst, M.S. Banks, Humans integrate visual and haptic information in a statistically optimal fashion, Nature 415 (2002) 429–433.
[5] S. Morein-Zamir, S. Soto-Faraco, A. Kingstone, Auditory capture of vision: examining temporal ventriloquism, Cogn. Brain Res. 17 (2003) 154–163.
[6] M. Radeau, P. Bertelson, The after-effects of ventriloquism, Q. J. Exp. Psychol. 26 (1974) 63–71.
[7] M. Radeau, P. Bertelson, Auditory–visual interaction and the timing of inputs: Thomas (1941) revisited, Psychol. Res. 49 (1987) 17–22.
[8] C. Spence, D.I. Shore, R.M. Klein, Multisensory prior entry, J. Exp. Psychol. Gen. 130 (2001) 799–832.
[9] E.B. Titchener, Lectures on the Elementary Psychology of Feeling and Attention, Macmillan, New York, 1908.
[10] J. Vroomen, B. De Gelder, Perceptual effects of cross-modal stimulation: the cases of ventriloquism and the freezing phenomenon, in: G. Calvert, C. Spence, B.E. Stein (Eds.), Handbook of Multisensory Processes, MIT Press, Cambridge, 2004, pp. 141–150.
[11] J. Vroomen, B. De Gelder, Temporal ventriloquism: sound modulates the flash-lag effect, J. Exp. Psychol. Hum. Percept. Perform. 30 (2004) 513–518.
[12] R.B. Welch, Perceptual Modification: Adapting to Altered Sensory Environments, Academic Press, New York, 1978.