BRILL

Multisensory Research 26 (2013) 561–568

brill.com/msr

Short Note

The Influence of Previous Environmental History on Audio-Visual Binding Occurs during Visual-Weighted but not Auditory-Weighted Environments

Jonathan M. P. Wilbiks * **and Benjamin J. Dyson**

Department of Psychology, Ryerson University, Toronto, ON, M5B2K3, Canada

Received 2 August 2013; accepted 19 November 2013

Abstract

Although there is substantial evidence for the adjustment of audio-visual binding as a function of the distribution of audio-visual lag, it is not currently clear whether adjustment can take place as a function of task demands. To address this, participants took part in competitive binding paradigms whereby a temporally roving auditory stimulus was assigned to one of two visual anchors (visual-weighted; VAV), or, a temporally roving visual stimulus was assigned to one of two auditory anchors (auditory-weighted; AVA). Using a blocked design it was possible to assess the malleability of audio-visual binding as a function of both the repetition and change of paradigm. VAV performance showed sensitivity to preceding contexts, echoing previous 'repulsive' effects shown in recalibration literature. AVA performance showed no sensitivity to preceding contexts. Despite the use of identical equi-probable temporal distributions in both paradigms, data support the contention that visual contexts may be more sensitive than auditory contexts in being influenced by previous environmental history of temporal events.

Keywords

Multisensory processing, auditory perception, visual perception, perceptual recalibration

1. Introduction

Auditory and visual sensory systems show remarkable flexibility in updating multi-sensory processes on the basis of previous discrepancies between times of arrival for sound and vision. First, there are natural constraints associated with audio-visual processing due to differences in the transmission speeds of sound and light in air. Even if sound and light are released simulta-

* To whom correspondence should be addressed. E-mail: jwilbiks@psych.ryerson.ca

© Koninklijke Brill NV, Leiden, 2013

DOI:10.1163/22134808-00002432

neously from the same external source, the auditory portion of the signal will typically arrive later than the visual portion. Within the approximate limit of 10–15 meters (Sugita and Suzuki, 2003), we make allowances for differential transmission speeds such that perceived audio-visual simultaneity is maximal when audio arrives at the ear slightly later than vision arrives at the eye (*auditory lag*; Vroomen and Keetels, 2010; although see Heron *et al.*, 2007, and Lewald and Guski, 2004). Thus, thunder is often said to be 'caused by' lightning (even though both are technically caused by the same electrostatic discharge) and furthermore causation would be more likely to be attributed to flashes that precede rather than follow the rumble. Second, studies have also shown how indices of audio-visual integration such as the current point of subjective simultaneity and temporal order judgments can be recalibrated by the prior distribution of audio-visual lag during an experimental session (e.g., Fujisaki *et al.*, 2004; Heron *et al.*, 2012; Vroomen *et al.*, 2004). These findings indicate that when presented with auditory and visual stimuli with a systematic manipulation of temporal asynchrony, participants show movement toward the temporal manipulation during a block, and away from it after the block has ended. This movement away from previously calibrated responding has come to be called a *repulsive aftereffect*, and has been demonstrated many times in the literature with regard to temporal asynchronies (cf. Fujisaki *et al.*, 2004; Heron *et al.*, 2012; Vroomen *et al.*, 2004).

To extend our understanding of the malleability of audio-visual integration, we examined whether the previous stimulus context within which audio-visual integration took place impacted on current binding decisions. Importantly, we also wanted to explore this independently of differences in temporal distribution. To this end, we created two different kinds of *competitive binding* environments wherein participants assigned a temporally roving to-be-bound stimulus in one modality to either a primary or secondary temporally static anchor in a different modality (see Fig. 1). Each of the three stimuli could vary according to magnitude (size in the case of vision, intensity in the case of audition), while the to-be-bound stimulus also varied in its time of presentation relative to the anchors. In one of these environments the anchors were visual and the roving stimulus was auditory (hereafter, VAV; Experiment 1) thus making the context 'visual-weighted'. In the other environment the anchors were auditory and the roving stimulus was visual (hereafter, AVA; Experiment 2) thus making the context 'auditory-weighted'. Despite holding the temporal factors constant between VAV and AVA contexts, we predicted that there would be an overall difference between Experiments 1 and 2 in terms of roving stimulus assignment. Given the previously observed preference for auditory lag conditions during audio-visual binding (Soto-Faraco and Alsisus, 2009; van Wassenhove *et al.*, 2007), we predicted primary anchor selection would be more likely during VAV contexts than AVA contexts. That is, par-

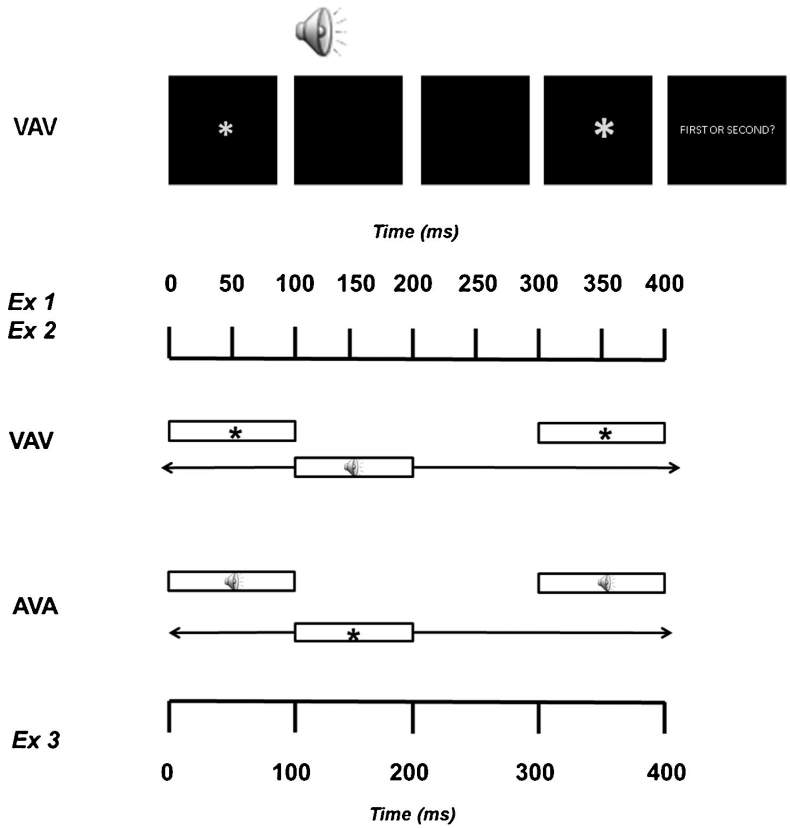

Figure 1. Schematic of procedure in Experiments 1, 2 and 3. Each trial consisted of two anchors of the same modality (V–V in the VAV paradigm, A–A in the AVA paradigm), which were presented for 100 ms each with a 200 ms ISI between them. The roving stimulus (A in the VAV paradigm, V in the AVA paradigm) was also presented for 100 ms, and its onset could be at one of nine time points in Experiments 1 and 2, and one of five time points in Experiment 3. The insert at the top shows the actual stimulus presentation associated with the example of the VAV trial depicted in the schematic.

ticipants would be generally more likely to bind the roving auditory stimulus to the first visual anchor in VAV but more likely to bind the roving visual stimulus to the second auditory anchor in AVA.

In Experiment 3, participants completed two blocks of one context followed by two blocks of the other context (i.e., VAV–VAV–AVA–AVA, or, AVA–AVA–VAV–VAV). This data should reveal the extent to which context constrains the multisensory binding process, both as a function of contextual repetition and of change. Firstly, we predicted that initial performance in VAV and AVA contexts would reproduce the hypothesised difference between Experiment 1 and 2, with AVA contexts leading to more secondary anchor binding. Following performance in this initial block, we predicted repeated exposure to the same context in a second block of trials should reveal response shifts based on prior paradigm sampling. Taking our cue from the recalibration literature, if the influence of previous contexts is repulsive on subsequent performance (e.g., Heron *et al.*, 2012), then a second VAV block should weaken primary anchor selection and a second AVA block should weaken secondary anchor selection. Following a similar logic as a result of task switch in the third block, AVA judgments preceded by VAV should be shifted towards the secondary anchor whereas VAV judgments preceded by AVA should be shifted towards the primary anchor. Maintenance of the alternate condition in a fourth block should show shifts similar to those in the second block.

2. Method

Twenty-seven different participants were analysed in each of Experiments 1 (mean age 20.8 (sd = 5.4) years, 26 females, 27 right-handed individuals), 2 (mean age 19.7 (sd = 3.8) years, 21 females, 23 right-handed individuals) and 3 (mean age 20.2 (sd = 2.0) years, 23 females, 27 right-handed individuals). All participants self-reported normal or corrected-to-normal vision and hearing, and the experimental procedure was approved by the Research Ethics Board at Ryerson University. Due to additional experimental manipulations reported elsewhere (Wilbiks and Dyson, 2013), visual stimuli (asterisks) differed in size (48 or 96 point Chicago font, Experiments 1 and 2; 24 or 96 point Chicago font, Experiment 3) and auditory stimuli (1 kHz tones) differed in intensity (66 or 71 dB, Experiments 1 and 2; 56 or 71 dB, Experiment 3). Stimulus presentation and responding was controlled by PsyScope (Cohen *et al.*, 1993). Visual stimuli were presented centrally on a computer monitor located 57 cm away from the participant. In order to promote audiovisual binding conditions (Calvert *et al.*, 2004), all sounds were presented binaurally from free-field speakers, which were positioned immediately on either side of the computer monitor. In VAV conditions, asterisks served as the anchors (fixed in time) and a tone served as the to-be-bound stimulus (roving in time; see Fig. 1 for the temporal arrangement of stimuli). In AVA conditions, tones served as the anchors and an asterisk served as the to-be-bound stimulus. The two anchors (of the same modality) varied independently, such that there were eight independent combinations of stimulus factors in each modality context (e.g., in VAV: V1 [small, large] × A [quiet, loud] × V2 [small, large]). All stimuli were presented for 100 ms and an equal number of all combinations of size and intensity were used within each experiment. Following a variable delay at the start of the trial, primary anchor onset to secondary anchor offset covered a 400 ms period (see Fig. 1). In Experiments 1 and 2, the timing of the roving stimulus varied in 50 ms intervals across the epoch (0, 50, 100, 150, 200, 250, 300, 350, 400 ms) while in Experiment 3, the timing of the roving stimulus varied in 100 ms intervals across the epoch (0, 100, 200, 300, 400 ms). All possible roving timing was equiprobable, thereby making the chosen temporal positions equally probable. In all blocks, trial order was randomised with respect to temporal positioning of the roving stimulus, the sizes of the visual stimuli, and the intensities of the auditory stimuli. Participants were asked to determine the causality of the roving stimulus by pressing one button on a PsyScope button box for the first anchor and a second button for the second anchor. In Experiments 1 and 2, participants completed only the VAV or AVA task, thereby establishing baseline measures of performance in each paradigm. As a result of the use of a between-participant design in Experiments 1 and 2 and within-participant design in Experiment 3, the total number of trials completed per condition was larger than that of Experiment 3. In Experiment 3, participants completed two blocks of 240 trials of VAV and AVA in counterbalanced order. Consequently, to maintain parity with Experiment 3, baseline calculations were based on the first 240 trials of either VAV in Experiment 1 or AVA in Experiment 2.

3. Results and Discussion

Between Experiments 1 and 2 ($t[52] = 3.256$, $p = 0.002$) and within Experiment 3 ($t[26] = 2.634$, $p = 0.014$), VAV judgments led to more primary anchor responses than AVA judgments. Neither average VAV nor AVA responding in Experiment 3 differed with the comparable values in Experiments 1 and 2 (0.417 *versus* 0.436; $t[52] = 0.839$, $p = 0.405$, and, 0.486 *versus* 0.497; $t[52] = 0.599$, $p = 0.552$, respectively). This allays concerns of response bias as a result of the repeated measures design in the latter experiment. Using a mixed block (one, two) × order (first, second) ANOVA in Experiment 3, VAV responding showed an effect for block ($F[1, 25] = 5.442$, $\text{MSE} = 0.003$, $p = 0.028$, $\eta_p^2 = 0.179$) and a trend for order ($F[1, 25] = 3.850$, $\text{MSE} = 0.017$, $p = 0.061$, $\eta_p^2 = 0.133$; block × order interaction: $F < 1$). In contrast, AVA responding showed no effects either as a function of block ($F < 1$) or order ($F[1, 25] = 1.315$, $\text{MSE} = 0.006$, $p = 0.262$, $\eta_p^2 = 0.050$; order × block interaction: $F < 1$ — see Fig. 2). An examination of individual differences (see Fig. 3) reveals variation in the probability of secondary anchor responding both as a function of condition (VAV, AVA) and block (1, 2). The inclusion of one potential outlier in VAV serves to weaken rather than strengthen the reported modulation of VAV responding.

The comparison between Experiments 1 and 2, and within Experiment 3, revealed increased binding to the primary anchor in VAV relative to AVA (Experiment 3 data is also reported in Wilbiks and Dyson, 2013). This is consistent with previous observations for auditory lag tolerance during audio-

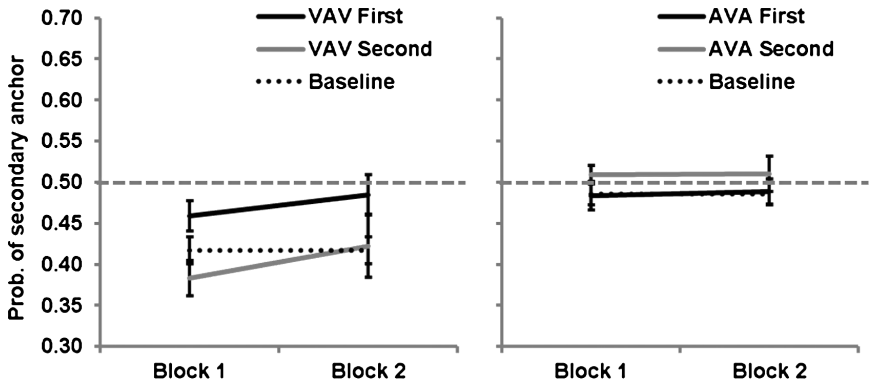

Figure 2. Signal-to-source realignment as a function of task (VAV, AVA), block (one, two) and order of task completion (first [black lines], second [grey lines]). The grey dotted line represents equivalent attribution to first and second anchor. Black dotted lines denote baseline responding, based on the first 240 trials in Experiments 1 and 2; they have been plotted twice for the sake of comparison. Solid lines show data based on repetition of blocks in Experiment 3. VAV judgments (left graph) showed sensitivity to prior context as a function of both block and order. AVA judgments (right graph) were not sensitive to prior contexts, showing no difference between order of completion or block number. Error bars represent standard error.

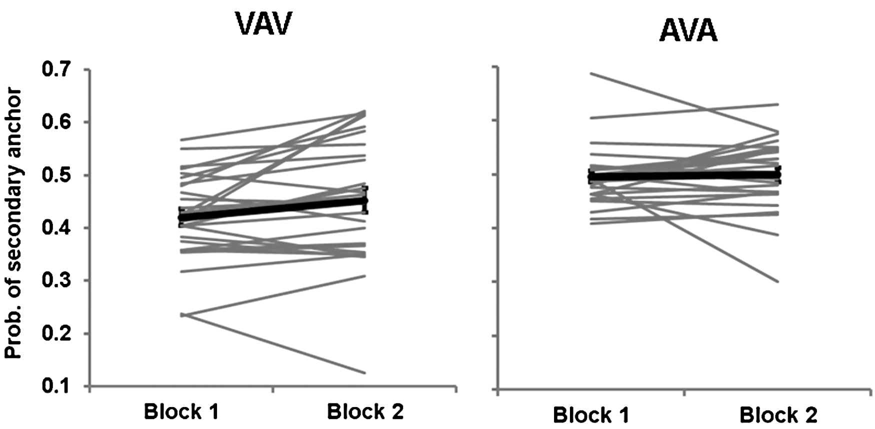

Figure 3. Individual realignment as a function of task (VAV, AVA) and block (one, two). Grey lines depict individual performance; black lines depict group average performance (after van Ee *et al.*, 2009). Error bars represent standard error.

visual binding, and the observation that auditory stimuli tend to be assigned to visual stimuli that precede rather than follow them (e.g., Dixon and Spitz, 1980; Lewald and Guski, 2003; Soto-Faraco and Alsisus, 2009; Spence and Squire, 2003). Reductions in primary anchor assignment for the second block of VAV and a trend to increase in primary anchor assignment when participants switched into VAV responding from AVA responding are consistent with the observation of repulsive recalibration effects (Heron *et al.*, 2012). We are however reluctant to label the findings in the current study 'recalibration effects', since such effects are usually the result of explicitly manipulating the temporal distribution of the environment (e.g., Fujisaki *et al.*, 2004; Heron *et al.*, 2012; Vroomen *et al.*, 2004). In the current study, temporal distributions were constant across VAV and AVA paradigms and, critically, were equiprobable in nature. Therefore, the effect can only be attributed to global context effects associated with the repetition or change of visual-weighted (VAV) or auditory-weighted (AVA) environments. On the basis of the observation that only the block main effect reached statistical significance, it appears that such effects are stronger when the context is maintained (eg., VAV–*VAV*) relative to changed (eg., AVA–*VAV*). Nevertheless, the direction of the effects during VAV performance remains consistent with repulsive recalibration effects, and the observation that audio-visual binding is antagonistic to the structure of previous environmental exposure joins a larger population of similarly 'repulsive' sequential effects found within the literature such as motion after-effect (e.g., Mather *et al.*, 2007) and negative priming (e.g., Tipper, 2001).

The data appear to reflect a greater willingness to allow previous performance to influence current behaviour during vision-weighted contexts. Evidence for the malleability of temporal assignment during VAV contexts but not during AVA contexts is consistent with the idea that audition and temporal processing are intimately linked (Burr *et al.*, 2009; although see Vroomen and Keetels, 2010). In an auditory-weighted paradigm (i.e., AVA) we see no significant impact of prior context, presumably since the temporal information is more reliable than in a visual-weighted environment (i.e., VAV). It seems the more auditory information we are provided with compared to visual information, the more confident we can be about the current temporal arrangement of the stimuli, and the less we are influenced by previous responding. In this experimental series, this concept is manifest in audio-visual binding malleability during VAV blocks, with no such variation in AVA blocks. Thus, with respect to making decisions regarding the assignment of signal-to-source, visual contexts appear less robust than auditory contexts. In terms of the relative weight placed on different sensory systems during temporal decision making, it seems we are willing to believe our ears much more than we believe our eyes.

Acknowledgement

B.J.D. is supported by an Early Researcher Award granted by the Ontario Ministry of Research and Innovation.

References

Burr, D., Banks, M. S. and Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration, *Exp. Brain Res.* **198**, 49–57.

Calvert, G. A., Spence, C. and Stein, B. E. (2004). *The Handbook of Multisensory Processing*. MIT Press, Cambridge, MA, USA.

Cohen, J., MacWhinney, B., Flatt, M. and Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers, *Behav. Res. Methods Instrum. Comput.* **25**, 257–271.

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

Fujisaki, W., Shimojo, S., Kashino, S. and Nishida, S. (2004). Recalibration of audio-visual simultaneity, *Nat. Neurosci.* **7**, 773–778.

Heron, J., Roach, N. W., Hanson, J. V., McGraw, P. V. and Whitaker, D. (2012). Audiovisual time perception is spatially specific, *Exp. Brain Res.* **218**, 477–485.

Heron, J., Whitaker, D., McGraw, P. V. and Horoshenkov, K. V. (2007). Adaptation minimizes distance-related audiovisual delays, *J. Vis.* **7**, 1–8.

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

Lewald, J. and Guski, R. (2004). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli, *Neurosci. Lett.* **357**, 119–122.

Mather, G., Pavan, A., Campana, G. and Casco, C. (2007). The motion aftereffect reloaded, *Tr. Cogn. Sci.* **12**, 481–487.

Soto-Faraco, S. and Alsisus, A. (2009). Deconstructing the McGurk–MacDonald illusion, *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 580–587.

Spence, C. and Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony, *Curr. Biol.* **13**, R519–R521.

Sugita, Y. and Suzuki, Y. (2003). Audiovisual perception: Implicit estimation of sound-arrival time, *Nature* **421**, 911.

Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views, *Q. J. Exp. Psychol.* **54A**, 321–343.

van Ee, R., van Boxtel, J. J. A., Parker, A. I. and Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection, *J. Neurosci.* **29**, 11641–11649.

van Wassenhove, V., Grant, K. W. and Poeppel, D. (2007). Temporal window of integration in auditory–visual speech perception, *Neuropsychologica* **45**, 598–607.

Vroomen, J. and Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review, *Atten. Percept. Psychophys.* **72**, 871–884.

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

Wilbiks, J. M. P. and Dyson, B. J. (2013). Effects of temporal asynchrony and magnitude congruency on competitive audio-visual binding, *Atten. Percept. Psychophys.* **75**, 1883–1891.