Auditory-Visual Integration During the Attentional Blink

April Shi Min Ching, Jeesun Kim, Chris Davis

The MARCS Institute, Western Sydney University, Australia
a.ching@westernsydney.edu.au, j.kim@westernsydney.edu.au,
chris.davis@westernsydney.edu.au

Abstract

Prominent theories of consciousness such as Global Workspace Theory propose that consciousness is required for multimodal integration. We tested this proposal with a processing bottleneck known as the unmasked attention blink, which we used with synchronized auditory-visual stimulus streams to delay awareness of the onset of a stimulus in one modality but not the other. Event-related potentials (ERPs) were then used to examine auditory and visual integration processes in the context of the unmasked attentional blink. To index auditory-visual (AV) integration, we recorded ERPs following the presentation of auditory-visual (AV) and unimodal second targets (T2) in AV presentation streams, which were presented during or after the attentional blink period, 200-300 ms or 600-700ms after the onset of first targets (T1) respectively. The results showed that AV and unimodal ERP responses were more similar during the attentional blink than outside of it. This result suggests that AV integration was suppressed and visual and auditory information were processed independently during the attentional blink. AV integration occurred both before and during the time window of the P3 ERP component (300-500 ms), which is well-established as the earliest time window for attentional blink ERP effects. The attentional blink suppressed AV integration only at later stages of processing while that at earlier, pre-P3 latencies was relatively intact. We discuss the implications of this finding for theories linking consciousness and integration.

Index Terms: Multimodal integration, attentional blink, conscious perception

1. Introduction

That subjective experience is multisensory and coherent suggests that there may be an intimate link between multimodal integration and conscious awareness. This relationship has been formalised in several theories, the most prominent being Global Workspace Theory (GWT) [1], [2]. In brief, GWT posits a modular sensory processing system in which sensory inputs are processed in sequestered modality- or feature-specific networks. Processing is limited to these modules under nonconscious conditions; interaction between these modules is only possible via long-range connections (i.e., the global workspace), primarily in cortical regions, that when active underpin conscious perception. Subsequently, a key prediction of GWT is that consciousness is necessary for multimodal integration.

Recent studies have challenged this particular aspect of GWT by demonstrating examples of multisensory interactions under conditions of complete unawareness [3]–[5]. The results of these studies suggest that some degree of multisensory interaction is possible without global workspace access and as such indicate a relationship between consciousness and multisensory processing that is far more complex than is envisioned by a simple interpretation of the GWT. Indeed, in general, the findings of these studies provide the impetus for a more nuanced examination of multisensory phenomena in terms of the time-course of conscious perception.

In devising an alternative and subtler probe of GWT, we utilised a modified attentional blink (AB) paradigm [6], [7] and used event-related potentials (ERPs) to examine the connection between consciousness and integration. The aim of using a modified version of the AB paradigm was to briefly delay conscious access to stimulus information, rather than extinguish it entirely (as occurs in masking studies). This delay-based manipulation allowed us to investigate whether integration was affected by the relative timing of inputs into the putative global workspace - something that is not feasible in a masking paradigm, since a paradigm that inhibits conscious processes would be insensitive to the time-course of multimodal integration. If multisensory integration is inhibited by a delay in the conscious perception of a stimulus in one modality, it would suggest that in order to be integrated, AV stimuli must be perceived to occur at the same time. Conversely, if integration is unaffected by the perceived delay between the occurrence of AV stimuli, then it would appear that multisensory stimuli can be integrated as long as they both eventually enter the global workspace and become consciously reportable. The outcome here has the potential to add more to the understanding of the relationship between consciousness and multimodal integration, since, to our knowledge, GWT has no strong claims regarding the manner in which modality-specific information streams must access the global workspace for eventual integration.

1.1. The Unmasked Attentional Blink

The AB is a processing bottleneck that is typically elicited and studied using rapid stimulus sequences (e.g. rapid serial visual presentation; RSVP). In such sequences, report of a second target stimulus (T2) is usually poor if it appears within 200-500 ms after a first detected target (T1) within a rapid sequence of distractors; conversely, this performance deficit (i.e., the AB) is not observed if T2 is presented more than 500 ms after T1, or if there is no T1 detection task. This suggests that the use of limited cognitive resources by the processing of T1 inhibits the processing of closely following T2, causing subsequent report of T2 to suffer.

In the present study, we used a variant of AB referred to here as the unmasked AB - while the AB usually impairs report accuracy, this occurs only if T2 is immediately followed by an additional stimulus (i.e., is backward masked) and no performance deficit is observed when T2 is the last stimulus [8]–[11]. Since T2s during the unmasked AB are presumably subject to the same T1-associated processing bottleneck as in the masked AB, yet are not associated with poorer conscious perception, it suggests that AB-related interference is temporary and can be recovered from under certain situations (i.e., when no new perceptual information follows T2). This aspect of the unmasked AB allows us to systematically interfere with the timing of the later stages of stimulus processing without ultimately altering conscious perception.

The AB has been commonly interpreted as delayed consolidation of stimulus information [12]. These models of the AB generally postulate a two-stage processing pipeline in which stimuli first undergo perceptual processing to create high-level representations, then are consolidated and become consciously available in a serial fashion. This serial limitation means that the consolidation of T1 representations halts the same process for closely following T2 representations. If T2 is masked, the pre-conscious representation of T2 is overwritten by new incoming stimulus information (i.e., that of the mask) during the bottleneck, and so even though it has been processed it cannot be reported; if T2 is not followed by a mask, the T2 representation persists past the bottleneck and can become consciously available once resources are freed from the processing of T1. This interpretation of the unmasked AB is supported by what happens to the P3 ERP component [13] during AB. The P3 is a large positive deflection peaking at centro-parietal scalp electrodes which is associated with stimulus categorisation and working memory consolidation. When T2 is masked, the AB impairs detection and suppresses P3 amplitude [14]; when T2 is not masked, the behavioural deficit and P3 suppression are absent, but the peak of the P3 wave occurs relatively later in time [8], [9], [11].

Given the above background, the unmasked AB has interesting properties as a test of GWT. Delayed consolidation can be interpreted as delayed entry into the global workspace, and from this we can derive the clear and testable prediction that multimodal integration should be inhibited during the unmasked AB. We thus employed ERP measures to quantify and compare integration processes during and outside of the unmasked AB. ERP measures were especially critical as an index of multisensory processes, as we did not expect multisensory processing to be reflected in our behavioural data. Task performance was expected to be near ceiling in all conditions, since it would be unaffected by the unmasked AB and thus unlikely to reflect any processing changes arising from multimodal relative to unimodal presentation. In what follows, we describe how the ERP correlates of integration were determined in the present study.

1.2. Quantifying Multimodal Integration in an Attentional Blink Paradigm

An approach to quantifying integration processes using ERP measures was first developed by [15] and was subsequently adopted in other studies. Here, ERPs to auditory-visual (AV) items are compared to the summed ERPs of the constituent auditory and visual stimuli. If the response to AV items is equal to the summed responses of auditory and visual items presented separately, it is assumed that auditory and visual information conveyed by the AV items has been processed independently. On the other hand, if the AV response diverges (supra-additive or sub-additive) from the summed A and V responses, then this is taken to indicate an interaction between auditory and visual processing [15].

To implement our paradigm, we concurrently presented a rapid serial visual presentation (RSVP) stream (printed letters and digits) and a rapid serial auditory presentation (RSAP) stream (spoken letters and digits) to participants, who were instructed to monitor either the visual or auditory stream for targets while ignoring the other stream (i.e., attend-visual and attend-auditory conditions). A fairly fast presentation rate (10 items/s) was chosen to create demanding but manageable task conditions and discourage attending to the task-irrelevant stream. ERPs were recorded to conditions presenting identical item streams but with the following terminal items: auditory-visual (co-presented printed and spoken "J"; auditory (spoken "J" with blank screen), or visual (printed "J" with silence). In addition, a null terminal item condition (blank screen and silence) was also included. This design enabled the isolation and comparison of ERPs to T2s presented in auditory-visual and unimodal contexts, in an alternative but equivalent manner to that of [15] (see section 2.4).

Previous comparisons of multimodal and unimodal ERPs have usually been limited to specific ERP components, and tested for differences only within restricted, predetermined time windows. In the present study, we instead tested for differences across the entire ERP epoch by using a non-parametric statistical technique known as topological analysis of variance (TANOVA; see [16]). TANOVA compares global ERP activity for each time point in the epoch and identifies periods of divergence between auditory-visual and unimodal ERPs - i.e., ranges of time points corresponding to significant differences in activity between conditions. We expected to find more and wider time windows of integration in the AB absent conditions, where terminal items were presented outside the AB period and integration presumably proceeds as normal.

In brief, we examined whether integration processes are affected by the unmasked AB. To do this, we examined the pattern of divergence between auditory-visual and unimodal ERPs during AB and how it differs from that of the AB absent conditions. With regards to testing GWT, there are two hypotheses: if unimodal information must enter the global workspace in synchrony in order to be integrated, then the unmasked AB should prevent integration, and no time windows of divergence should be observed in the AB-present relative to AB-absent condition. Conversely, if unimodal information need only eventually enter the global workspace to be integrated, we might expect a delay of integration, which would be reflected as relatively later time windows of divergences in AB-present relative to AB-absent condition.

2. Method

2.1. Participants

25 students from Western Sydney University were recruited for the experiment after giving informed consent. One participant was excluded from analysis for poor task performance. All reported normal or corrected-to-normal vision, normal hearing, and no history of neurological/psychiatric disorders.

2.2. Stimuli and Procedure

There were 16 conditions in the experiment, created by a factorial combination of *attended modality* (visual or auditory), *intertarget lag* (short or long; i.e. temporal distance between T1 and T2), and *terminal item* (visual, auditory,

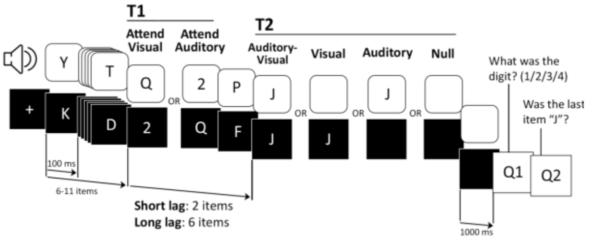

Figure 1: *Schematic depiction of a single experimental trial.*

visual-auditory, or null). The attended stream variable (i.e., attend visual; attend auditory) was blocked in two sessions that were presented on separate days; the order of these two blocks was counterbalanced across participants. The order of trials within each attended modality block (constituting 8 conditions) was fully and separately randomised for each participant. Each participant was presented with a total of 1152 trials (72 trials per condition; 576 trials per session). Stimuli presentation was controlled by the Psychophysics Toolbox Matlab package [17], [18].

All trials (see Figure 1) consisted of the simultaneous presentation of a visual and an auditory stimulus stream, each containing 8 to 18 printed and spoken digits/letters respectively. Each visual/auditory item was presented for 100 ms with no blank interstimulus interval. The onsets of items in both streams were fully synchronised. The attended stream contained one or two targets among randomly selected letter distractors; the unattended stream contained only letter distractors that were incongruent with the co-presented counterpart. Distractors were drawn from a set of 15 letters (B, C, D, E, F, K, L, M, N, P, R, S, T, Y, Z), with the constraint that letters did not repeat within the next 4 items in either stream. All trials contained a T1, which was a printed or spoken digit (1, 2, 3, or 4) presented randomly within the 5th through 12th positions. Each of the four digits appeared as T1 with equal frequency in each condition. Half of all trials contained a valid T2, which was the presence of a "J" as the last item of the stream (printed J in the attend visual block; spoken J in the attend auditory block). The content of the terminal item was assigned according to the terminal item modality condition - this was either a printed "J" with silence (visual); a spoken "J" with blank screen (auditory); co-presented printed and spoken "J" (visual-auditory); or blank screen with silence (null). Critically, T2-valid trials were in visual and visual-auditory terminal item conditions in the attend visual block, and in auditory and visual-auditory terminal item conditions in the attend auditory block. In the long intertarget lag conditions, T1s and the terminal item were separated by 5 items in both blocks. In the short intertarget lag conditions, T1s and the terminal item were separated by 1 item in the attend visual block, and 2 items in the attend auditory block.

Trials in each of the four terminal item conditions were created by generating a stimulus stream of T1 and distractors, then replacing the last visual-auditory item accordingly. This meant that for every visual-auditory item trial of a given attended stream type and intertarget lag combination, its counterparts in the remaining three terminal item conditions contained the exact same visual and auditory stimuli, with the exception of the terminal item.

At the beginning of each attended modality block, the participant was instructed to attend only to the assigned stream and ignore the other. Participants were asked to identify, amongst letter distractors, a digit within the attended stream, and whether the terminal item was the letter J. Each trial begun with the presentation of a white fixation cross in the middle of a black background for 1000 ms. This was followed by the start of the simultaneous visual and auditory serial presentation streams, and then a white '%' symbol and silence was presented for 500 ms. The participant was then presented with an unspeeded free-response prompt to identify the T1 digit. Upon entering their response, an unspeeded two-choice prompt for the detection of T2 (i.e. detected/not detected) was presented - specifically, participants were instructed to select 'not detected' if they did not actually perceive a T2 and refrain from guessing. Completing this prompt triggered the presentation of the next trial.

2.3. Electrophysiological Recording and Preprocessing

EEG signals were recorded using a 64-channel Biosemi ActiveTwo system and a sampling frequency of 512 Hz. EEG data were processed with EEGLab [19]. The scalp recordings were re-referenced against the average of the mastoid recordings and a 0.01 to 20 Hz bandpass filter was applied. Continuous data was visually inspected for movement and other artefacts. Infomax, an independent component analysis algorithm implemented in EEGLab, was applied to the remaining data and components reflecting typical artefacts (i.e., horizontal and vertical eye movements) were removed. Back-projected data was subsequently epoched using a 200 ms pre-stimulus baseline and 1000 ms following stimulus onset. The epoched data was base-line corrected and visually screened for residual artifacts. Each of the remaining 24 participants contributed an average of 61.6 trials to each condition (range: 40-72 trials) after the exclusion of artifactual and incorrect trials.

2.4. Global ERP Analyses

We first generated whole-scalp auditory-visual (AV) and unimodal (AO/VO) difference maps for each combination of attended modality and intertarget lag. AV maps were created by subtracting auditory/visual scalp ERPs from auditory-visual scalp ERPs for the attend visual/auditory conditions; AO/VO maps were created by subtracting null item (silence and black screen) scalp ERPs from visual/auditory scalp ERPs for the attend visual/auditory conditions. This subtraction procedure served to isolate the activity associated with T2s, presented in either AV or AO/VO contexts, from that of preceding stimuli.

For each combination of attended modality and intertarget lag, divergent activity between AV and AO/VO maps were defined as auditory-visual interactions. To detect these divergences, AV and AO/VO maps were submitted to global field power (GFP; [20]) and topographic ANOVA (TANOVA) analyses. The GFP analysis and TANOVA represent complementary analyses, aimed at detecting differences in field strength and topology respectively. Both analyses were carried out in RAGU, a randomisation statistics toolbox [16]. In the randomisation statistics approach, group and condition assignments for each participant are shuffled, and measurements based on multiple iterations of the randomised dataset are used to obtain an estimation of the null hypothesis and tested against the actual dataset.

GFP is analogous to the standard deviation of potentials across all scalp electrodes, and is a quantifier of map strength that is insensitive to topology. Comparisons of GFP values of scalp maps to auditory-visual and unimodal conditions were carried out via nonparametric randomisation tests at each time point of the recorded epoch. Time windows of divergent GFP were derived from significant differences of $p < .01$ in at least three consecutive time points in the waveform. This smaller alpha value was adopted to control for multiple comparisons – the joint probability of $p < .01$ in three consecutive frames $(0.01 \times 0.01 \times 0.01)$ is within the Bonferroni corrected threshold of $p < .05$ across 512 time points (0.05/512).

TANOVA involves a series of nonparametric randomisation tests carried out at each time point, applied to the GFP of difference maps between two conditions of interest (i.e., auditory-visual and unimodal). As the difference maps are taken from normalised data in which raw scalp maps have been scaled to unit variance, the GFP values in this particular analysis ("difference-GFP") do not reflect differences in field strength, but in topographic distribution. As with the GFP analysis, an adjusted alpha of $p < .01$ in at least three consecutive time points for derivation of time windows. Time windows of significant differences from the GFP tests and TANOVAs were then further inspected via scalp maps of mean (non-normalised) voltage and *t* values.

3. Results and Discussion

3.1. Auditory-Visual Integration Outside the Attentional Blink

Figure 2 depicts a summary of the global ERP analyses of T2-related activity for the long intertarget lag conditions. The GFP and TANOVA analyses indicated six and four time windows of significant differences in the attend visual and attend auditory blocks respectively, suggesting robust AV interactions at multiple stages of processing.

3.1.1. Early Latency Effects

At early-latency time windows #1 (24-47 ms) in the attend visual block and #1 (24-32 ms) in the attend auditory block, auditory-visual targets were generally associated with enhanced positivity at right-occipital sites and enhanced negativity at left fronto-temporal sites relative to unimodal (i.e. visual-only) targets. Although #1 in the attend auditory block did not reach the significance threshold, we included it for analysis due to its very similar effect latency and topography with #1 in the attend visual block. Their topographical and timing similarities seem to implicate the earliest component of the visual evoked potential known as the C1 [21] which reflects activity in the retinotopic area V1 in the calcarine cortex and is maximal at posterior scalp regions. While effects at such an early latency appear surprising, corroborating evidence exists from previous ERP-based comparisons of multimodal and unimodal processing [15], [22]–[24].

Inspection of difference waves (omitted for brevity) in the attend visual condition shows that the implicated time window aligns with the peak of the earliest deflection in both auditory-visual and unimodal waveforms, which supports the effect as a multimodal modulation of the visual processes underlying C1. How are auditory inputs able to influence visual processes at the V1 within a narrow time frame of 30 ms? One possibility is through neural pathways that connect primary auditory cortex and primary visual cortex including V1 either directly [25], [26] or through thalamic structures [27], which can mediate the transfer of low-level information between early-latency sensory processing regions. At this stage, simultaneously presented auditory and visual inputs are integrated in a bottom-up manner and, in this case, result in the increased recruitment of processes associated with the V1.

In the attend auditory condition however, the implicated time window was placed at the trough before the earliest deflection. The lack of a distinct C1 peak here is unsurprising since these difference waves represent responses to auditory items from which presumably all visual item-related processing has been subtracted. Here, the early latency effect suggests an automatic recruitment of V1-associated processes to sudden visual change (i.e., the presentation of a new item) that occurs in the auditory-visual condition but not unimodal, despite the irrelevance of visual stimuli.

3.1.2. P3-Related Effects

The majority of remaining time windows – #2, #3, #4 and #5 (193-213 ms, 256-270 ms, 290-315 ms, and 425-440 ms) in the attend visual block, and #2, #3 and #4 (320-365 ms, 387-413 ms and 450-457 ms) in the attend auditory block – all overlap the P3 waveform in the difference waves (omitted for brevity) and indicate an enhanced central or central-parietal positivity to auditory-visual targets compared to unimodal targets. These effects are best explained as modulations of the two major subcomponents of the P3 [13] – P3a, which has an earlier onset and fronto-central maximum, and P3b, which is relatively later and has a parietal maximum. The implication of frontal and temporal-parietal cortex activity in the generation of the P3 complex [13] suggests that AV processes at this latency are, in contrast to the stimulus-driven C1, mediated by higher brain regions linked to conscious processing.

In the attend visual conditions, the topographical effects suggest a faster and less variable onset of both P3a and P3b to auditory-visual targets compared to visual-only targets. These effects suggest a role for auditory information in eliciting an automatic orienting response to co-presented visual information (as indexed by P3a), resulting in more rapid categorisation and consolidation (as indexed by P3b).

Specific to the attend auditory condition were GFP effects indicating more intense fronto-centrally distributed positivity, or larger P3a elicitation to auditory-visual items than auditory-only items. Sub-threshold time window #4 also suggests larger P3b elicitation. The prevalence of GFP effects here suggests that visual inputs aid the detection of auditory targets more effectively than auditory inputs do for visual targets, probably due to the greater uncertainty inherent to the task.

3.1.3. Late Visual Effects

Finally, time window #6 (765-850 ms) in the attend visual block indicates a posterior negativity to unimodal targets compared to the more diffuse activity to auditory-visual targets. This negativity might be related to the visual offset negativity [28] which is associated with the offset of visual stimulation, which is more distinct in the unimodal condition due to the absence of overlapping auditory evoked effects.

3.2. Auditory-Visual Integration During the Attentional Blink

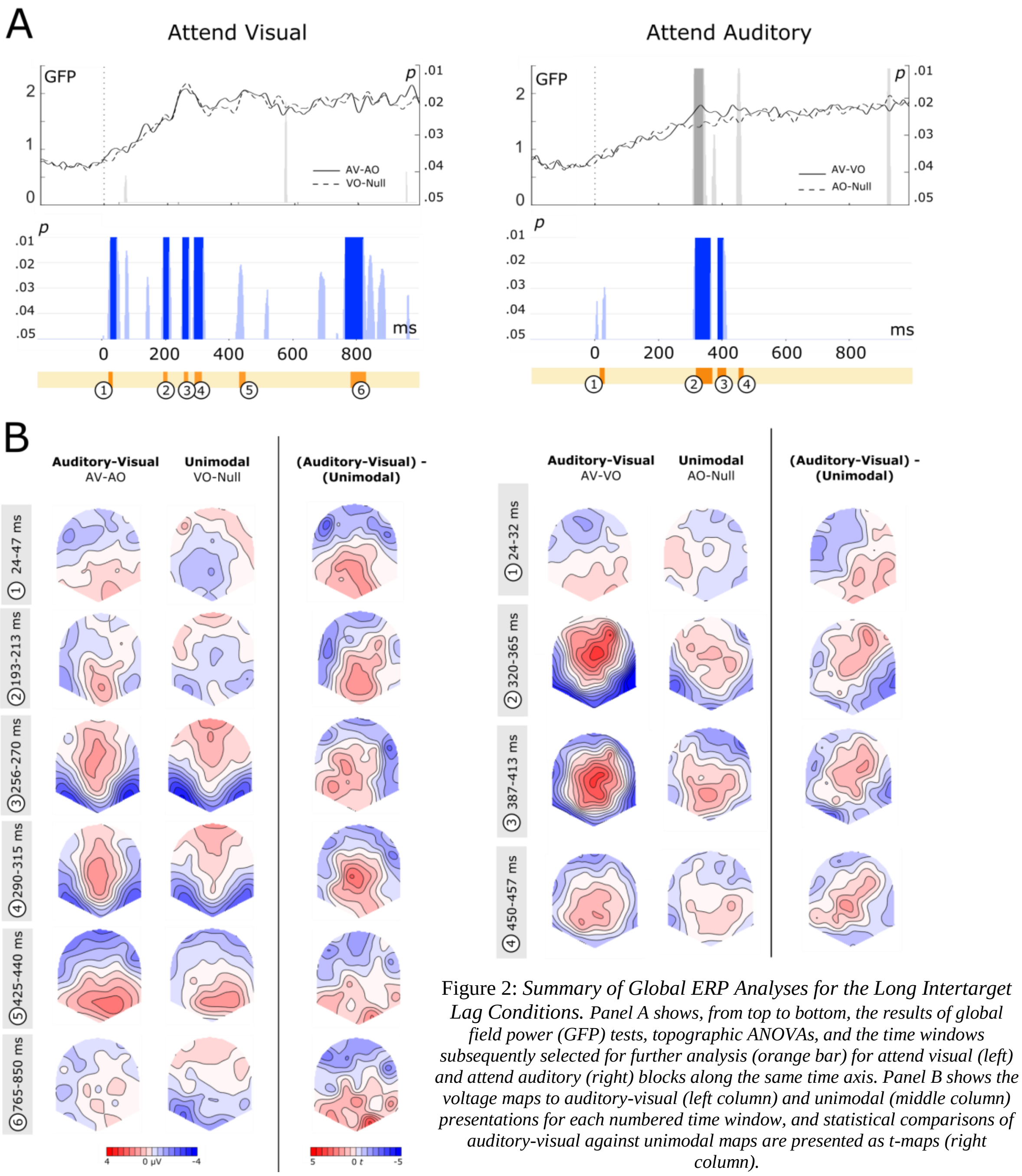

Figure 2: *Summary of Global ERP Analyses for the Long Intertarget Lag Conditions. Panel A shows, from top to bottom, the results of global field power (GFP) tests, topographic ANOVAs, and the time windows subsequently selected for further analysis (orange bar) for attend visual (left) and attend auditory (right) blocks along the same time axis. Panel B shows the voltage maps to auditory-visual (left column) and unimodal (middle column) presentations for each numbered time window, and statistical comparisons of auditory-visual against unimodal maps are presented as t-maps (right column).*

Figure 3 depicts a summary of the global ERP analyses of T2-related activity for the short intertarget lag conditions (attend auditory block omitted for brevity). The analyses indicated only one time window of significant difference in the attend visual block and none in the attend auditory block – time window #1 (53-61 ms) from the attend visual block seems to indicate effects similar to early-latency C1 effects discussed in the previous section.

Taken together, the unmasked AB appears to strongly suppress the majority of AV integration processes at the latency of the P3 and later, suggesting that attentional enhancement by the task-irrelevant modality is prevented by the unmasked AB despite normal task performance. This indicates that the relative entry of individual modalities into consciousness is critical to this specific integration process. Additionally, there is some evidence that AV processes at very early latencies remain relatively intact during the AB - this independence from consciousness agrees with the proposal of AV processes at C1 as a mainly stimulus-driven mechanism.

4. Conclusions

The present global ERP analyses indicate that delayed consciousness of stimuli during the unmasked AB can suppress multimodal enhancement around the latency of the P3 and beyond – this outcome is compatible with the GWT view that consciousness is necessary for multimodal integration, and adds to it the requirement for synchronised

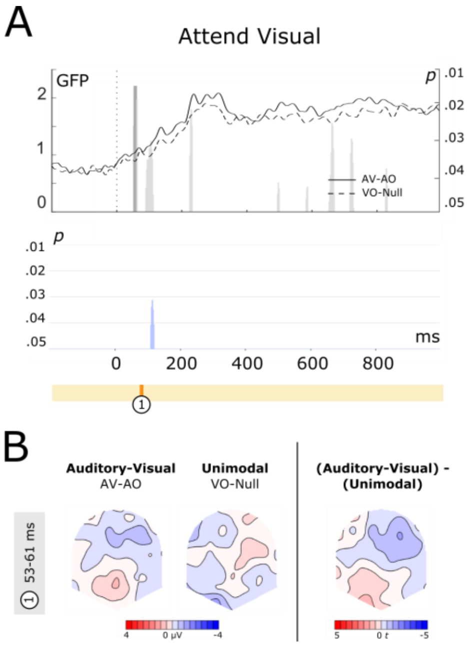

Figure 3: *Summary of Global ERP Analyses for the Short Intertarget Lag Condition in the Attend Visual Block.*

timings of entry for individual modalities. However, indications of early AV interactions that can occur regardless of the AB pose a challenge to GWT, which does not account for multisensory processes that are independent of consciousness. Further work will be required to distinguish between consciousness- dependent and -independent integration, to be able to reconcile GWT with a range of multisensory phenomena.

5. References

[1] B. J. Baars, ‘The global workspace theory of consciousness’, *Blackwell Companion Conscious.*, pp. 236–246, 2007.

[2] S. Dehaene and L. Naccache, ‘Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework’, *Cognition*, vol. 79, no. 1, pp. 1–37, 2001.

[3] N. Faivre, L. Mudrik, N. Schwartz, and C. Koch, ‘Multisensory integration in complete unawareness evidence from audiovisual congruency priming’, *Psychol. Sci.*, vol. 25, no. 11, pp. 2006–2016, 2014.

[4] R. B. Scott, J. Samaha, R. Chrisley, and Z. Dienes, ‘Prevailing theories of consciousness are challenged by novel cross-modal associations acquired between subliminal stimuli’, *Cognition*, vol. 175, pp. 169–185, Jun. 2018.

[5] A. S. M. Ching, J. Kim, and C. Davis, ‘Auditory–visual integration during nonconscious perception’, *Cortex*, vol. 117, pp. 1–15, Aug. 2019.

[6] D. E. Broadbent and M. H. Broadbent, ‘From detection to identification: Response to multiple targets in rapid serial visual presentation’, *Percept. Psychophys.*, vol. 42, no. 2, pp. 105–113, 1987.

[7] J. E. Raymond, K. L. Shapiro, and K. M. Arnell, ‘Temporary suppression of visual processing in an RSVP task: An attentional blink?’, *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 18, no. 3, p. 849, 1992.

[8] B. Brisson and M.-È. Bourassa, ‘Masking of a first target in the attentional blink attenuates the P 3 to the first target and delays the P 3 to the second target’, *Psychophysiology*, vol. 51, no. 7, pp. 611–619, 2014.

[9] R. Dell’Acqua *et al.*, ‘The Attentional Blink Impairs Detection and Delays Encoding of Visual Information: Evidence from Human Electrophysiology’, *J. Cogn. Neurosci.*, vol. 27, no. 4, pp. 720–735, Apr. 2015.

[10] B. Giesbrecht and V. D. Lollo, ‘Beyond the attentional blink: visual masking by object substitution.’, *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 24, no. 5, pp. 1454–1466, 1998.

[11] E. K. Vogel and S. J. Luck, ‘Delayed working memory consolidation during the attentional blink’, *Psychon. Bull. Rev.*, vol. 9, no. 4, pp. 739–743, Dec. 2002.

[12] P. E. Dux and R. Marois, ‘The attentional blink: A review of data and theory’, *Atten. Percept. Psychophys.*, vol. 71, no. 8, pp. 1683–1700, Nov. 2009.

[13] J. Polich, ‘Updating P300: an integrative theory of P3a and P3b’, *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2128–2148, 2007.

[14] E. K. Vogel, S. J. Luck, and K. L. Shapiro, ‘Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink.’, *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 24, no. 6, pp. 1656–1674, 1998.

[15] M. H. Giard and F. Peronnet, ‘Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study’, *J. Cogn. Neurosci.*, vol. 11, no. 5, pp. 473–490, Sep. 1999.

[16] T. Koenig, M. Kottlow, M. Stein, and L. Melie-García, ‘Ragu: A Free Tool for the Analysis of EEG and MEG Event-Related Scalp Field Data Using Global Randomization Statistics’, *Comput. Intell. Neurosci.*, vol. 2011, pp. 1–14, 2011.

[17] D. H. Brainard, ‘The psychophysics toolbox’, *Spat. Vis.*, vol. 10, pp. 433–436, 1997.

[18] M. Kleiner *et al.*, ‘What’s new in Psychtoolbox-3’, *Perception*, vol. 36, no. 14, p. 1, 2007.

[19] A. Delorme and S. Makeig, ‘EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis’, *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[20] D. Lehmann and W. Skrandies, ‘Spatial analysis of evoked potentials in man--a review.’, *Prog. Neurobiol.*, vol. 23, no. 3, pp. 227–250, 1984.

[21] K. Rauss, S. Schwartz, and G. Pourtois, ‘Top-down effects on early visual processing in humans: A predictive coding framework’, *Neurosci. Biobehav. Rev.*, vol. 35, no. 5, pp. 1237–1253, Apr. 2011.

[22] S. Molholm, W. Ritter, M. M. Murray, D. C. Javitt, C. E. Schroeder, and J. J. Foxe, ‘Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study’, *Cogn. Brain Res.*, vol. 14, no. 1, pp. 115–128, Jun. 2002.

[23] E. Van der Burg, D. Talsma, C. N. L. Olivers, C. Hickey, and J. Theeuwes, ‘Early multisensory interactions affect the competition among multiple visual objects’, *NeuroImage*, vol. 55, no. 3, pp. 1208–1218, Apr. 2011.

[24] D. Talsma, T. J. Doty, and M. G. Woldorff, ‘Selective Attention and Audiovisual Integration: Is Attending to Both Modalities a Prerequisite for Early Integration?’, *Cereb. Cortex*, vol. 17, no. 3, pp. 679–690, Mar. 2006.

[25] A. Falchier, S. Clavagnier, P. Barone, and H. Kennedy, ‘Anatomical Evidence of Multimodal Integration in Primate Striate Cortex’, *J. Neurosci.*, vol. 22, no. 13, pp. 5749–5759, Jul. 2002.

[26] K. S. Rockland and H. Ojima, ‘Multisensory convergence in calcarine visual areas in macaque monkey’, *Int. J. Psychophysiol.*, vol. 50, no. 1–2, pp. 19–26, Oct. 2003.

[27] R. L. van den Brink, M. X. Cohen, E. van der Burg, D. Talsma, M. E. Vissers, and H. A. Slagter, ‘Subcortical, Modality-Specific Pathways Contribute to Multisensory Processing in Humans’, *Cereb. Cortex*, vol. 24, no. 8, pp. 2169–2177, Aug. 2014.

[28] S. A. Huettel *et al.*, ‘Linking Hemodynamic and Electrophysiological Measures of Brain Activity: Evidence from Functional MRI and Intracranial Field Potentials’, *Cereb. Cortex*, vol. 14, no. 2, pp. 165–173, Feb. 2004.