

Unsupervised Characterization of Prediction Error Markers in Unisensory and Multisensory Streams Reveal the Spatiotemporal Hierarchy of Cortical Information Processing

 Priyanka Ghosh,^{*} Siddharth Talwar,^{*} and  Arpan Banerjee

Cognitive Brain Dynamics Lab, National Brain Research Centre, Gurgaon 122052, India

Abstract

Elicited upon violation of regularity in stimulus presentation, mismatch negativity (MMN) reflects the brain's ability to perform automatic comparisons between consecutive stimuli and provides an electrophysiological index of sensory error detection whereas P300 is associated with cognitive processes such as updating of the working memory. To date, there has been extensive research on the roles of MMN and P300 individually, because of their potential to be used as clinical markers of consciousness and attention, respectively. Here, we intend to explore with an unsupervised and rigorous source estimation approach, the underlying cortical generators of MMN and P300, in the context of prediction error propagation along the hierarchies of brain information processing in healthy human participants. The existing methods of characterizing the two ERPs involve only approximate estimations of their amplitudes and latencies based on specific sensors of interest. Our objective is twofold: first, we introduce a novel data-driven unsupervised approach to compute latencies and amplitude of ERP components accurately on an individual-subject basis and reconfirm earlier findings. Second, we demonstrate that in multisensory environments, MMN generators seem to reflect a significant overlap of “modality-specific” and “modality-independent” information processing while P300 generators mark a shift toward completely “modality-independent” processing. Advancing earlier understanding that multisensory contexts speed up early sensory processing, our study reveals that temporal facilitation extends to even the later components of prediction error processing, using EEG experiments. Such knowledge can be of value to clinical research for characterizing the key developmental stages of lifespan aging, schizophrenia, and depression.

Key words: MMN; multisensory facilitation; P300; prediction error; source localization

Significance Statement

Widely studied ERPs—MMN and P300—have been considered in the literature to be neural markers of prediction error propagation. Here, we use a data-driven unsupervised method to identify the amplitude and latencies of MMN and P300 from EEG-ERP signals and accurately estimate their underlying cortical sources, in auditory, visual, and audiovisual contexts. Subsequently, we could demonstrate that the multisensory/audiovisual contexts speed up the processing times of prediction mismatch compared with unisensory scenarios, and modality-specific processing at the MMN stage gives way to modality-independent processing at the P300 stage, thus capturing the hierarchy of prediction error processing. The results confirm previously speculated understanding of the cortical processing hierarchies often found in the literature of perceptual and cognitive processes.

Received July 17, 2023; revised March 19, 2024; accepted March 20, 2024.

The authors declare no competing financial interests.

Author contributions: P.G., S.T., and A.B. designed research; P.G. and A.B. performed research; P.G. and S.T. contributed unpublished reagents/analytic tools; P.G. and S.T. analyzed data; P.G. and A.B. wrote the paper.

We acknowledge the support of NBRC core funds and the computational support from the Computational Facility, NBRC. This project is funded by the NBRC Flagship program BT/MEDIII/NBRC/Flagship/Program/2019: Comparative mapping of common mental disorders (CMD) over lifespan.

Continued on next page.

Introduction

Prediction error propagation along sensory streams is conceptualized to be organized hierarchically (Friston, 2005), consequently updating the internal model of the world at each hierarchical stage (Wacongne et al., 2011). Prediction mismatches are perceived through multiple sensory modalities (Di Luca et al., 2009); however, the internal mental models of sight, sound, smell, taste, and touch are hypothesized to be integrated with our existing cognitive schemata (Talsma, 2015). On the other hand, studies have provided evidence that multisensory integration can also happen in parallel and, at times speed up the speed of processing (Molholm et al., 2002; Van Wassenhove et al., 2005; Dhamala et al., 2007). This raises the question—what are the time scales corresponding to modality-specific and modality-independent processing during prediction mismatch? The answer can unravel the fundamental neural operations driving the processing of prediction errors regardless of a specific sensory stream. The first stage of multisensory integration was identified ~100 ms of speech stimulus onset, from the appearance of a faster audiovisual N100—a robust ERP component, compared to the unisensory auditory analog (Van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2007). Such evidences of multisensory facilitation are, however, not established for the later stages of information processing. The overarching goal of this article is to use the ERPs from human electroencephalogram (EEG) recordings in tandem with unsupervised signal processing approaches to index the temporal markers of prediction errors and their underlying cortical sources across different sensory modalities.

Prediction errors are indexed by the mismatch negativity (MMN) and the P300 components in the literature (Chennu et al., 2013; Stefanics et al., 2014; Calcus et al., 2015; Aitchison and Lengyel, 2017; Banellis et al., 2020). MMN is a negative deflection ~100–250 ms poststimulus onset, in the difference waveform between deviant and standard EEG signals (Näätänen et al., 1978, 2007; Näätänen and Michie, 1979; Sams et al., 1985; Garrido et al., 2009). P300 is a positive deflection with a latency between 250 and 500 ms (Sutton et al., 1967; Giard et al., 1990; Polich, 2007). The key difference is that MMN is known to index any deviance in the absence of consciousness (Atienza et al., 2001; Koelsch et al., 2006) whereas P300 indexes possible attention switches and, hence, a conscious perception of stimulus change (Sutton et al., 1967; Giard et al., 1990; Polich, 2007). The latency of audiovisual MMN (avMMN) resembled closely to the latency of audio MMN (and not visual MMN; Shiramatsu et al., 2021), but a previous study (Sittiprapaporn, 2012) reported shorter visual and audiovisual MMN latencies than auditory MMN. Similarly, others have reported a significant early onset of P300 latency (Stefanics et al., 2005; Marucci et al., 2021) as well as increased amplitude (Fleming et al., 2020; Marucci et al., 2021) during multimodal stimulation (visual-audio-tactile and visual-audio) as compared to corresponding unimodal stimulations, but another study by Feng et al. (2008) did not observe such multisensory benefits for P300 latencies. Thus, latency and amplitude of MMN/P300 vary with experimental parameters, for example, stimulus duration, occurrence probability, and discrimination difficulty (Magliero et al., 1984; Patel and Azzam, 2005; Polich, 2007) as well as physiological variables, for example, age and attention (Polich et al., 1990; Blackwood, 2000; Dinteren et al., 2014; Erickson et al., 2016) leading to such disparate findings.

In this article, we hypothesize (1) the faster responses to multisensory stimuli can be seen in the middle/late processing stages of prediction errors, that is, for MMN/P300, and (2) there exists a modality-specific sensory-cognitive dissociation in the source distribution of MMN and P300. Previous MMN and P300 studies have been limited in their use of averaging methods across electrodes or of choosing just the conventional central electrodes (e.g., Fz, Cz, and Pz; Gonsalvez and Polich, 2002; Gonsalvez et al., 2007), which may not effectively capture the true MMN and P300 latencies in the first place. Additionally, cross-sensory modality, where the modality of the frequent and the deviant stimuli are different, may recruit different brain regions impacting the processing speeds of deviant stimuli. To resolve these issues, we employed an unsupervised dimensionality reduction technique across the sensor space to create comparable filters on which sensor data from different conditions, unisensory, multisensory, and cross-modal, can be projected for characterization of the ERP latencies/amplitudes across different sensory modalities. This novel approach to determining latencies provides an accurate estimation of modality-specific changes in the prediction error markers. Furthermore, we perform rigorous source analysis using coregistration with individual subject's MRI data to reveal

*P.G. and S.T. contributed equally to this work.

Correspondence should be addressed to Arpan Banerjee at arpan@nbc.ac.in.

Copyright © 2024 Ghosh et al.
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

the overlapping cortical networks and specialized brain regions specific to unimodal, multimodal, and cross-modal MMN and P300.

Materials and Methods

Ethics statement. The study was carried out following the ethical guidelines and prior approval of the Institutional Human Ethics Committee (IHEC) of the National Brain Research Centre, India. Written informed consent was obtained from all participants before the commencement of the experiment, and they were remunerated for the time of their participation.

Participants. Twenty-two healthy volunteers (nine males and 13 females) in the age group of 22–43 (mean = 25.7, SD = ± 4.19) years participated in the study. They had normal or corrected-to-normal vision and were right-handed. All participants had university degrees or higher and reported no history of neurological or audiological problems. They were requested to avoid the intake of any medication/stimulant (e.g., sedative and coffee) before the experiment.

Stimuli. The experiment consisted of five different conditions and each condition consisted of two categories of stimuli, i.e., repetitive/frequent/standard and non-repetitive/oddball/deviant. Two of the five conditions presented were unimodal, i.e., audio only and visual only; the third was bimodal, i.e., audiovisual; and the remaining two were cross-modal in nature. In the first three conditions, the standard and the oddball stimuli were of the same sensory modality, which means that the audio-only condition comprised an audio standard and an audio oddball stimulus; the visual-only condition comprised a visual standard and a visual oddball stimulus; and the audiovisual (AV) condition comprised of an audiovisual standard and an audiovisual oddball stimulus. The remaining two conditions, namely, the cross-audio consisted of an audio deviant and a visual standard; and the cross-visual consisted of a visual oddball and an audio repetitive stimulus. The contents of each condition are tabulated in Table 1. Each condition consisted of 400 trials, out of which oddballs constituted 14% of the total trials. Each condition was presented in the form of a block of 100 stimuli, where the standard and oddball stimuli of a particular condition were presented in random order and the number of oddball stimuli varied across each block. There were 20 such blocks (five conditions $\times$ 4 blocks) that were randomized and presented to the participants. The participants were asked to count the number of oddballs presented in each block and report at the end of the block to ensure they attended to the whole stimulus (the results of counting accuracy for each participant are presented as Extended Data Table 1-1).

The stimuli of the visual-only condition consisted of a standard blue square and a deviant red triangle. The auditory stimuli were inspired by musical notes, the standard as the C4 note and the deviant as the C5 note (higher octave), according to the tuning of the A440 pitch standard. All stimuli were presented on a white background on a 21" LED screen (1,280 $\times$ 1,024 pixels). The participants were asked to keep their eyes open and fixate on a central cross on the screen during the presentation of all auditory stimuli. The inter-stimulus interval also consisted of the same cross-fixation. The length of each oddball and standard stimulus was 350 ms, and the inter-trial interval ranged between 800 and 1,200 ms (mean = 1,000 ms). AV and cross-modal conditions were constructed from combinations of the audio-only and the visual-only stimuli (listed in Table 1).

Data acquisition. EEG was recorded with 64 Ag/AgCl electrodes, using a Neuroscan System (Compumedics NeuroScan, SynAmps2). The electrodes were attached to an elastic cap in a 10–20 international system montage. The data were acquired at a sampling rate of 1,000 Hz with the default reference near Cz, grounded to electrode AFz. During the experiment, the participants were seated comfortably at a distance of 60–70 cm from the monitor in a sound-attenuated and dimly lit room. The participants were requested to make minimal body movements and blink normally during the experiment. The impedance of all electrodes was initially set below 5 k Ω and was also monitored in between blocks. Additionally, head digitization was performed using Polhemus Fastrak (Polhemus) to mark the position of the electrodes and the fiducials based on the placement of the cap on individual participants at the end of the entire EEG session. Individual T1-weighted structural images (MPRAGE) were also obtained using a 3.0T Philips Achieva MRI scanner (TR = 8.4 ms, FOV = 250 $\times$ 230 $\times$ 170, flip angle = 8°).

Table 1. The table lists the standard and the deviant stimuli used in our oddball paradigm for all the five sensory modality conditions

Condition	Standard/frequent/repetitive	Deviant/oddball/infrequent/nonrepetitive
Unimodal audio (A)	261.6 Hz	523.3 Hz
Unimodal visual (V)	Blue square	Red triangle
Bimodal audiovisual (AV)	261.6 Hz and blue square	523.3 Hz and red triangle
Cross-modal (audio deviant CrA)	Blue square	523.3 Hz
Cross-modal (visual deviant CrV)	261.6 Hz	Red triangle

Extended Data Table 1-1 has details on the accuracy of participants.

Preprocessing. EEG was acquired from 22 participants out of which data from 1 participant was discarded due to noisy recordings. The raw data of the remaining 21 participants were imported, and each block was epoched. A high-pass filter of 0.1 Hz was applied to the data to remove slow drifts in the signal. The data were visually inspected further, and one channel (F4 in three subjects) was interpolated to neighboring electrodes. To identify and remove blink artifacts from the data, independent component analysis (ICA) was employed for each block. Only the blink component obtained as the independent component (IC) from eye regions was visually identified and subsequently, that IC was rejected. Furthermore, all blocks were subjected to a low-pass filter of <45 Hz and then, each block was epoched to trials of [-500 550] ms where 0 ms marked the onset of the stimulus. The trials were further divided into standard and oddball categories, baseline correction was applied using the prestimulus activity and the data was rereferenced to linked mastoids. Trials with signal amplitude >100 μ V and lesser than -100 μ V were removed and at most 6% of all oddball trials were discarded per subject. Then we equalized the number of oddball and standard trials within each separate condition for each participant by randomly selecting standard trials. The following are the total number of oddballs or standards in each condition with mean and standard deviation across participants—audio-only condition was 1,416 (mean = 67.4286, SD = 5.835), visual-only was 1,397 (mean = 66.5238, SD = 8.2257), AV was 1,450 (mean = 69.0476, SD = 2.8544), cross-audio was 1,468 (mean = 69.9048, SD = 1.5781), and cross-visual was 1,461 (mean = 69.5714, SD = 1.1650). All analyses were done using the MATLAB-based FieldTrip toolbox developed at Donders Institute for Brain, Cognition and Behaviour in Nijmegen, Netherlands (Oostenveld et al., 2011), and custom-written scripts in MATLAB (www.mathworks.com).

Unsupervised MMN and P300 peak extraction. MMN and P300 ERPs were observed in all five conditions: audio only (A), visual only (V), audiovisual (AV), cross-audio (CrA), and cross-visual (CrV) as seen in Figure 1, obtained by subtracting the grand average of standard waveforms from their respective deviant waveforms (waveforms of standard and deviant conditions in midline sensors are presented as Extended Data Figure 1-1). As evident from the peaks in the plots, precise estimation of MMN and P300 peak latencies/amplitudes when considering the activities of all sensors can be a challenging task as peaks from all sensors make the overall peak diffused. Averaging the EEG activity across sensors or selecting only specific central sensors for characterizing MMN and P300 peaks warrant only approximate estimations of their amplitudes and latencies. To eliminate all such researchers' degrees of freedom, we identify the peaks in our study using a novel unsupervised data-driven method. Unlike the averaging method where a few peaks get lost in the process of averaging due to a set of sensors forming dipoles, our method effectively captures maximum variance in the waveforms across the entire sensor space. Due to the multivariate nature of this method, it is more reliable than the peak detection algorithms that essentially capture the local extrema for a single sensor time series (Dinh et al., 2021). For our analyses, we first pooled in trials (550 ms post-stimulus onset) from all the five conditions, i.e., A, V, AV, CrA, and CrV of the standard and oddball categories, and generated a common spatial filter using principal component analysis (PCA) on the difference (oddball-standard) waveforms. PCA computes the eigenvectors of the covariance matrix ("principal axes") and sorts them by their eigenvalues (amount of explained variance). The loadings or "coefficients" obtained as the output of the MATLAB function *pca* were extracted to create the common spatial filter. To capture maximum variance in the data, we used the first four principal components as they together explained >80% of the data. In the ERP data (before dimensionality reduction), the signal-to-noise ratio expressed as $100 \times [(\text{peak to peak amp-std dev}) / \text{peak to peak amplitude}]$ was found to be ~74–79% across all participants and conditions (presented as Extended Data Table 1-2). Thus, choosing a variance threshold of 80% retains all signal components that are relevant and ignores mostly the noise components.

Thus, choosing a variance threshold of 80% takes care of all signals that are relevant and ignores mostly the noise components. The common spatial filter obtained using PCA was then projected to the individual standard and oddball waveforms of each condition. The dimensionality of the data was hence reduced from timepoints $\times$ sensors $\times$ trials to timepoints $\times$ trials. A common spatial filter was used to optimize the differences between standard and deviant categories across all conditions. Figure 2A illustrates the "scores" of the first four principal components as topoplots along with their corresponding eigenvalues (λ) in the principal component subspace. After projecting the common spatial filter (first four PCs) on the standard and deviant waveforms of each condition, we computed the average of the four projected activities and subject-wise, inspected the presence of MMN and P300 ERPs in audio (Fig. 2B), visual (Fig. 2C), audiovisual (Fig. 2D), cross-audio (Fig. 2E), and cross-visual conditions (Fig. 2F). The difference wave (oddball – standard) for each condition projected on common spatial filter (Extended Data Figure 2-1A) didn't show much difference from projections on condition-specific filters (Extended Data Figure 2-1B) computed using PCA.

Source localization. Source localization was performed to obtain the MMN and P300 generators in the brain across various modalities. The EEG data across all sensors were re-referenced with an average reference prior to source localization. Nineteen participants' structural MRI data were resliced and segmented to identify the brain, skull, and scalp tissues. Two subjects' MRI scans could not be obtained because of their incompatibility with the fMRI scanner. The origin (0,0,0) of all the T1 images was set to the anterior commissure. Participant-specific headmodel was computed using the OpenMEEG toolbox (Gramfort et al., 2010), using realistic conductivity values. The Polhemus data was imported to place the sensor locations on the head model of each participant. To obtain high accuracy of electrode positions, individual coregistration was employed by firstly visually marking the fiducials (nasion, left preauricular, and right preauricular) in the MRI volume and finally matching the marked points with the fiducial locations as per the Polhemus data. Next, the sources were placed

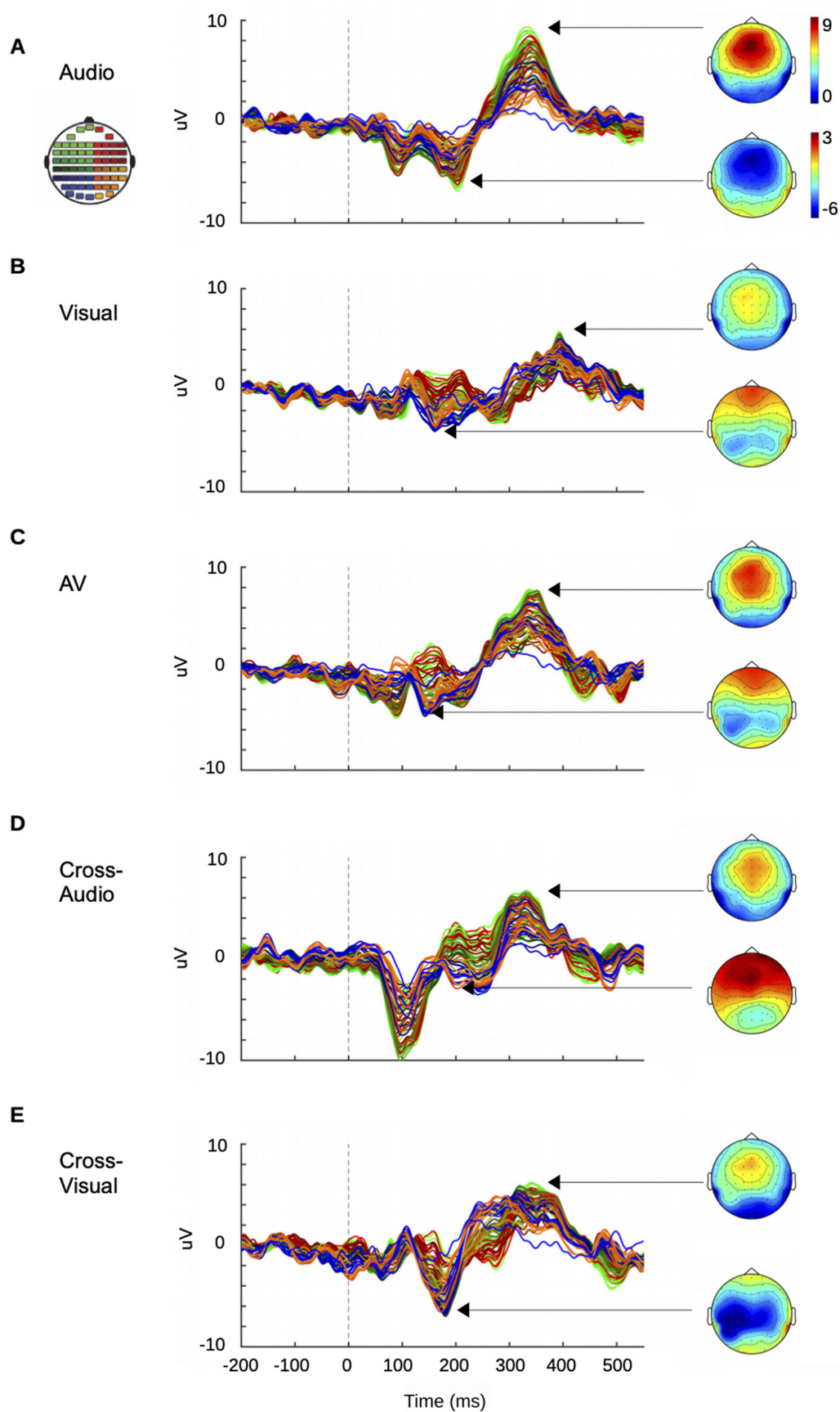


Figure 1. Continued.

in the segmented part of the head model at a distance of 5 mm from each other and the leadfield matrix was computed i.e., a transfer function between each source point and each electrode. Source localization of each individual was performed using their respective headmodel, leadfield, and electrode positions. For the inverse solution, covariance across all EEG sensors was computed for each condition. eLORETA, belonging to a class of current density measures to calculate the distribution of neuronal sources, was used to solve the inverse problem (Pascual-Marqui, 2007). eLORETA also generates the least amount of false-positives; hence, it is beneficial for exploratory source analysis, for example, where prior hypotheses regarding approximate locations may not be available (Halder et al., 2019). Lambda of 0.1 was used as the regularization parameter for the localization of P300 and MMN ERPs. After localization, each individual's source intensities were interpolated to their respective MRI volume. Further, to calculate the grand average of the source values, the interpolated images were normalized to a common structural template. Finally, we subtracted the voxel intensities of oddball and standard categories and the voxels having intensities >99th percentile were thresholded. This was done separately for each condition and hemisphere.

Code and data accessibility. Codes used to analyze the data and generate the figures can be obtained from the bitbucket repository using the following link: https://bitbucket.org/cbdl/pe_erpsandsourceanalysis/src/master/. Pre-processed and deidentifiable EEG data can be downloaded from the OSF repository: <https://osf.io/v9nu5/>. Copyright of the data and codes are held by the National Brain Research Centre, an autonomous institution of the Government of India.

Results

Temporal characterization of prediction error markers across modalities

The P300 peaks were recognized in the difference waveforms as the most positive deflections and the MMN peaks as the most negative deflections (except for Cross-Audio (CrA)) where it was second most negative) across all modalities/conditions. We performed Student's *t* test on the dimensionality-reduced subject-level data to identify the temporal windows where MMN and P300 peaks in the deviant waveform showed significant differences in activities from their corresponding standard waveform in each condition. A timepoint-wise *t* test was conducted to identify the significant segments where, based on our data, significant activity ($p < 0.001$) corresponding to consecutive timepoints ≥ 10 ms were deemed significant. To maintain uniformity in the variances across conditions, the window length was restricted to 100 ms, centered at the peak. This window length was sufficient to include the descending and ascending limbs of MMN peaks and the ascending and descending limbs of the P300 peaks. The significant windows for MMN were found between 136 and 235 ms for audio only (Fig. 2G), 137–236 ms for visual only (Fig. 2H), 98–197 ms for audiovisual (Fig. 2I), 118–217 ms for cross-audio (Fig. 2J), and 121–220 ms for cross-visual (Fig. 2K) conditions. Similarly, the significant windows for P300 were found between 296 and 395 ms for audio-only (Fig. 2G), 346–445 ms for visual-only (Fig. 2H), 287–386 ms for audiovisual (Fig. 2I), 278–377 ms for cross-audio (Fig. 2J), and 264–363 ms for cross-visual conditions (Fig. 2K). For amplitude and latency characterization of MMN and P300, the minimum and maximum amplitudes, respectively, from the difference (deviant-standard) waveforms along with their corresponding latencies were extracted from the 100 ms windows for each subject across the audio only (A), visual only (V), audiovisual (AV), cross-audio (CrA), and cross-visual (CrV) conditions. Subsequently, the group mean and standard deviation of latencies and amplitudes were obtained for all the five conditions as listed in Table 2 for MMN and Table 3 for P300. We statistically investigated if the change of modality/condition indeed had an effect on the latencies and amplitudes of MMN/P300 peaks by conducting a repeated-measures analysis of variances (rmANOVA) on the subject-level data. The results revealed a significant effect of modality on the latencies of both MMN ($F_{(4,80)} = 15.487$, $p < 0.0001$, $\eta^2 = 0.436$) and P300 ($F_{(4,80)} = 33.390$, $p < 0.0001$, $\eta^2 = 0.625$), as well as on the amplitudes of MMN ($F_{(4,80)} = 3.231$, $p = 0.016$, $\eta^2 = 0.1391$) and P300 ($F_{(4,80)} = 6.8$, $p = 0.0001$, $\eta^2 = 0.254$). rmANOVA in this case is the best statistical test to check the interaction of modality with latencies/amplitudes because a simple ANOVA would otherwise inflate Type I error (false-positives) in between-subject effects and Type II error (false-negatives) in within-subject effects. An important point to keep in mind while using univariate rmANOVA is that the variances of the differences between all combinations of measurements should be equal. This is known as the sphericity (or circular) assumption, which is strongly assumed for within-subject rmANOVA statistics. If the sphericity assumption is violated, within-subject rmANOVA statistics are meaningless. We used Mauchly's test to verify that the assumption of sphericity was not violated ($p > 0.05$) in our comparisons. Furthermore, to employ a multiple comparisons correction that provides a good statistical power, post hoc analyses were performed using Bonferroni's test to compare the peak MMN and P300 latencies/amplitudes between different conditions. As seen in Figure 3, the peak latency of audiovisual MMN (avMMN) was significantly faster than both audio MMN ($p = 0.012$) and visual MMN ($p < 0.0001$). Similarly, the

Figure 1. Grand average of difference ERPs (deviant-standard) plotted across 21 subjects for conditions (A) audio-only, (B) visual-only, (C) audiovisual, (D) cross-audio, and (E) cross-visual conditions. Stimulus onset is at 0 ms and the activity before that [–200 to 0 ms] corresponds to the prestimulus baseline. The topoplots indicate the distribution of the respective MMN and P300 peaks in the sensor space. The topoplot on the top-left displays the color code assigned to respective scalp channel locations used for plotting the ERPs in A–E. For a curious reader, ERP from each condition, deviant, and standard are plotted at midline sensors in Extended Data Figure 1–1, along with SNR calculations in Extended Data Table 1–2.

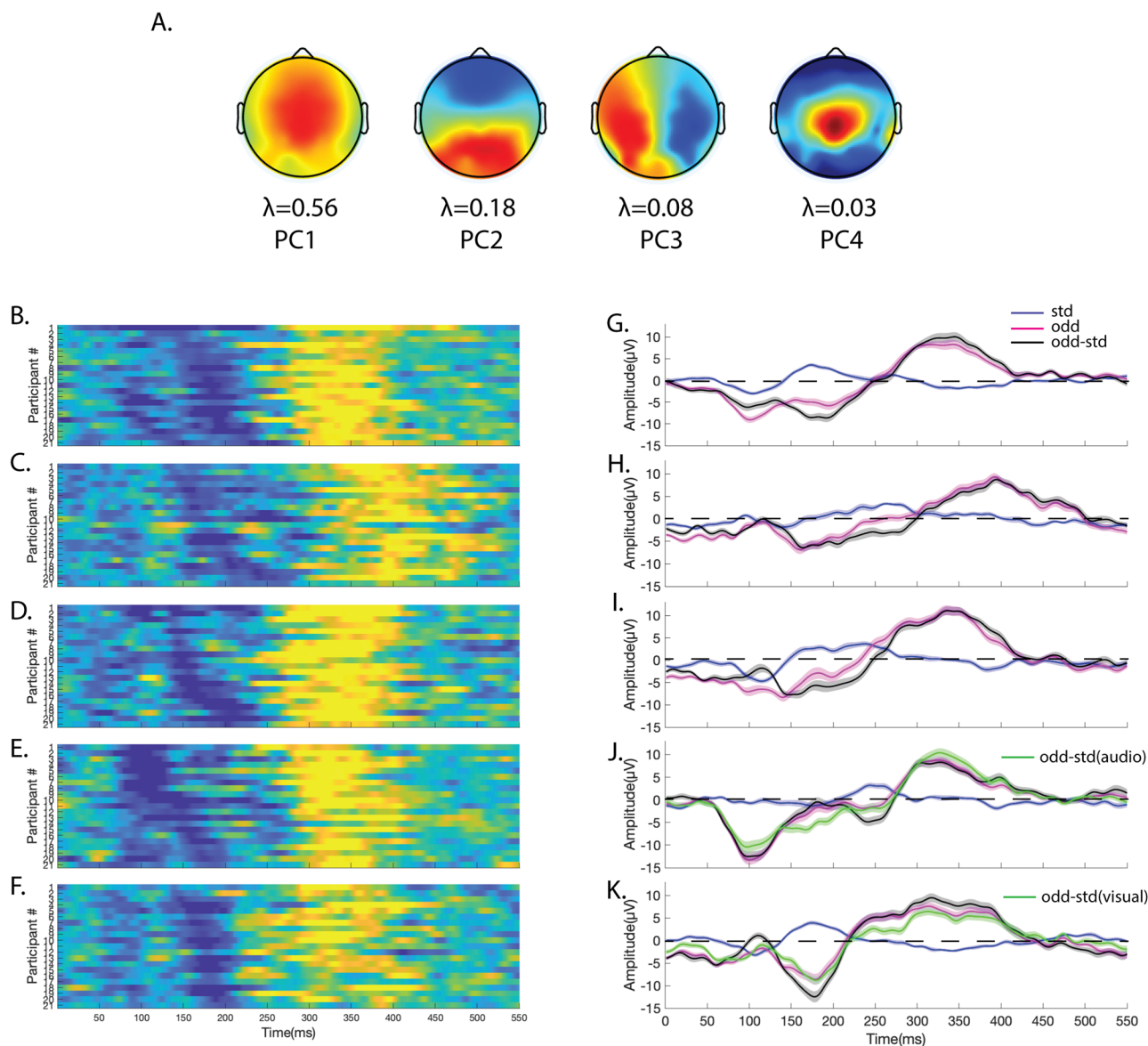


Figure 2. **A**, The topoplots correspond to the first four principal components (left to right) of the common spatial filter (Extended Data Figure 2-1) computed through PCA on the difference waveform of deviant minus standard trials together from all conditions. The λ values represent the fraction of variance explained by each principal component. The heatmaps on the left depict the time course of difference waveforms of individual participants after projection of the common spatial filter on the standard and deviant data of **(B)** audio-only, **(C)** visual-only, **(D)** audiovisual, **(E)** cross-audio, and **(F)** cross-visual conditions. On the right are the grand-averaged plots of the standard (blue), deviant (magenta), and difference (black) waveforms along with their shaded SEMs for **(G)** audio-only, **(H)** visual-only, **(I)** audiovisual, **(J)** cross-audio, and **(K)** cross-visual conditions.

Table 2. Table shows the mean $\pm$ SD of the peak MMN amplitudes and their corresponding latencies across all the subjects for audio-only, visual-only, audiovisual (AV), cross-audio (CrA), and cross-visual (CrV) conditions

MMN	Latency (ms)	Amplitude (μ V)
Audio	180.81 $\pm$ 22.72	-10.91 $\pm$ 3.33
Visual	187.76 $\pm$ 31.38	-8.74 $\pm$ 3.06
AV	154.28 $\pm$ 28.88	-11.18 $\pm$ 3.19
CrA	142.90 $\pm$ 22.92	-11.37 $\pm$ 3.24
CrV	171.81 $\pm$ 15.19	-10.11 $\pm$ 3.49

peak latency of cross-audio MMN (craMMN) was significantly faster than both audio MMN ($p = 0.0004$) and visual MMN ($p < 0.0001$). Interestingly, cross-audio MMN was also significantly faster than cross-visual MMN ($p = 0.001$). Again for P300, the latencies of all non unisensory modalities—avP300, craP300, and crvP300—were significantly faster than the latency

Table 3. The table shows the mean $\pm$ SD of the peak P300 amplitudes and their corresponding latencies across all the subjects for audio-only, visual-only, audiovisual (AV), cross-audio (CrA), and cross-visual (CrV) conditions

P300	Latency (ms)	Amplitude (μ V)
Audio	337.34 $\pm$ 24.45	12.13 $\pm$ 4.61
Visual	390.62 $\pm$ 24.53	10.87 $\pm$ 3.34
AV	330.38 $\pm$ 25.76	13.29 $\pm$ 3.87
CrA	328.81 $\pm$ 23.61	12.59 $\pm$ 4.12
CrV	312.24 $\pm$ 29.92	9.23 $\pm$ 3.50

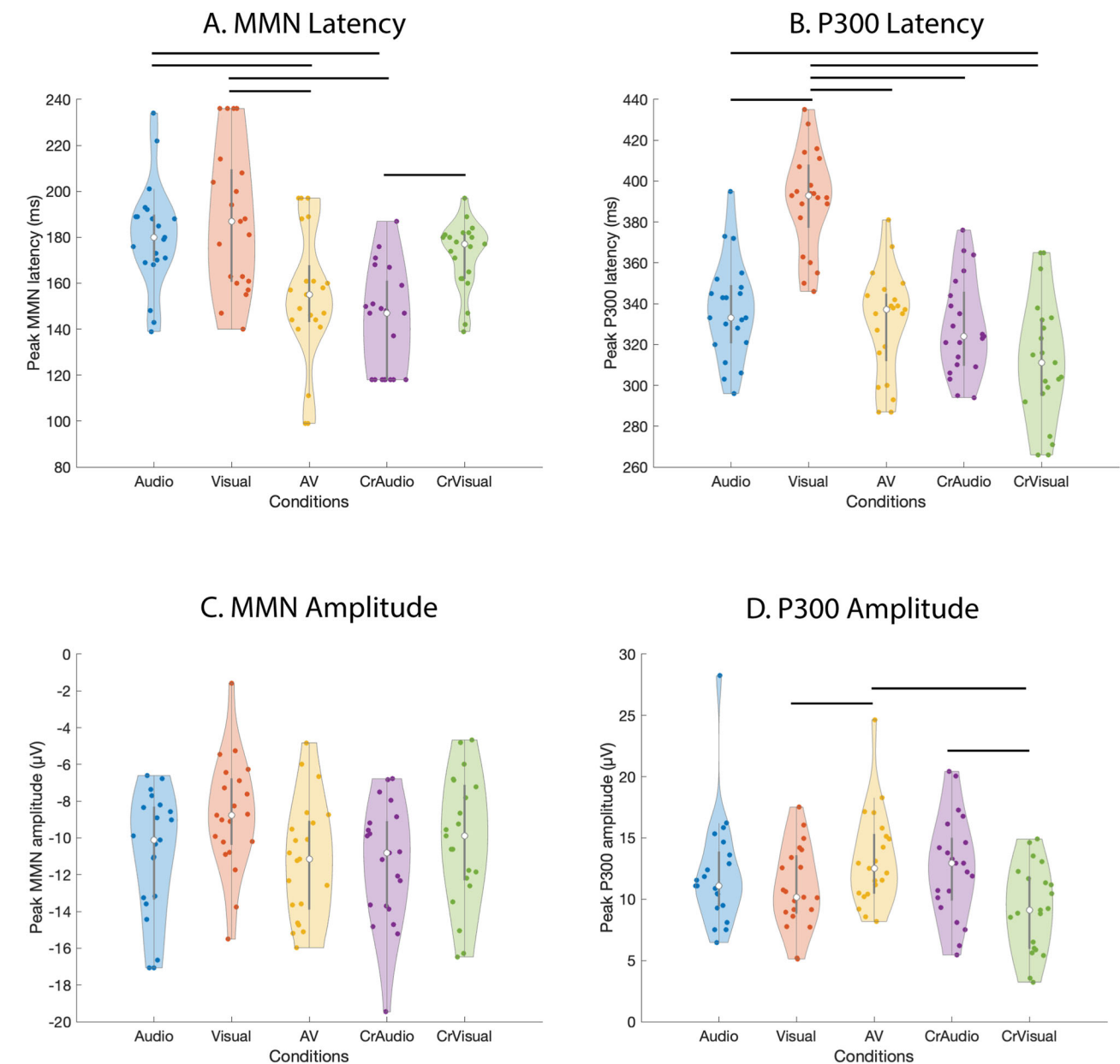


Figure 3. The violin plots represent the (A) MMN peak latencies corresponding to the (C) MMN peak amplitudes (maximum magnitude of the difference waveforms on the negative y-axis) of each participant. Similarly, the (B) P300 peak latencies corresponding to the (D) P300s peak amplitudes are plotted for each participant. The colored dots represent each participant's peak amplitude and peak latency values in μ V and ms, respectively. The white dot at the center of the gray box in each violin represents the median of the data, and the gray box itself represents the inter-quartile range. The horizontal lines represent a significant difference ($p < 0.05$) between any two conditions obtained through Bonferroni-corrected post hoc analysis.

of the unisensory visual modality ($p < 0.0001$). The peak latency of crvP300 was also significantly faster than the peak latency of aP300 ($p = 0.01$). Between the unisensory modalities, the peak latency of vP300 was significantly slower than the peak latency of aP300 ($p < 0.0001$). The peak P300 amplitude of audiovisual condition was significantly higher than the peak P300 amplitudes of visual ($p = 0.01$) and cross-visual ($p < 0.0001$) conditions. Additionally, the peak P300 amplitude of cross-audio condition was significantly higher than the peak P300 amplitudes of cross-visual condition ($p = 0.03$). No significant differences were seen between the peak MMN amplitudes across all conditions.

Spatial characterization of prediction error markers across modalities

The 100 ms windows of each modality were used to obtain the covariance matrices, separately for standard and oddball trials of MMN and P300. Using coregistration of individual participant MR data with their EEG sensor locations, we generated the source maps underlying MMN and P300 activity using eLORETA (details in Materials and Methods). The Brainnetome atlas was interpolated to the same common structural template as used for the normalization of the individual sources to locate the sources. As revealed in Figure 4, the MMN sources were distributed throughout the brain and were different for different modalities, while P300 displayed overlapping maps (locations in Table 4).

The left superior frontal, paracentral, right superior, and middle temporal gyri, bilateral precentral gyrus, and bilateral superior parietal lobule were found to underlie the MMN response in the audio-only condition; the left superior frontal gyrus and right posterior superior temporal sulcus and inferior parietal lobule elicited the MMN in the visual-only condition; and the audiovisual condition yielded sources that were located in the left middle frontal gyrus, fusiform area, parahippocampal gyrus, medioventral and lateral occipital cortex, and right middle temporal gyrus, posterior superior temporal gyrus, and inferior parietal lobule. In the cross-modal conditions, left superior and middle frontal gyri, right superior and middle temporal gyri, and inferior parietal lobule were found in cross-modal audio condition, while the cross-modal visual condition was elicited from the bilateral middle and superior frontal gyri (details in Table 4). In summary, all conditions had distributed source maps, however, some commonalities were present, such as all conditions were elicited from the frontal regions and all conditions except cross-visual included the parietal regions and importantly, different clusters in the temporal area.

In contrast, source analysis of P300 revealed overlapping frontal areas for all the five conditions (as illustrated in Fig. 4B). P300 was elicited at the bilateral frontal gyrus in all five conditions, the right middle frontal gyrus in audio-only and cross-audio task, and the bilateral middle frontal and cingulate cortex in the AV (details in Table 4).

Discussion

We employed two-stimulus oddball tasks using audio-alone (A), visual-alone (V), synchronous audiovisual (AV), cross-modal audio (CrA), and cross-modal visual (CrV) conditions to evoke MMN and P300 components—ERP-level markers of prediction error processing (Chennu et al., 2013; Stefanics et al., 2014; Calcus et al., 2015; Aitchison and Lengyel, 2017; Banellis et al., 2020). Addressing our first hypothesis, we elucidate how the presence of multimodal and cross-modal audiovisual contexts speed up the processing speeds of prediction errors at the middle (MMN) and later (P300) stages of mismatch information processing. Our study provides a novel data-driven method to accurately confirm the understanding of cortical information processing and, hence, can serve as a powerful tool to formulate “internal models” trying to understand prediction error processing at different stages/hierarchies in the brain (Wacongne et al., 2012). In the statistical learning literature, PCA is recognized as an unsupervised approach that involves the diagonalization of the covariance matrix computed from two random variables without explicit assumption of a specific relationship (model) between them, for example, in the case of regression (which assumes a linear model; Jolliffe, 2002). No single study before ours has demonstrated this in healthy humans so explicitly, involving both the prediction error markers across different modality conditions and pinpointing the exact temporal scale when modality-specific processing is completed in the cortex. While addressing our second hypothesis, we found that the modality-specific and modality-independent nature of MMN and P300 is complex. As we move from MMN to P300, we see an overall loss in the modality specificity in terms of cortical sources. While MMN sources had a significant overlap over frontal cortical locations, unlike P300 sources, there were also areas such as in temporal and parietal regions specific to a particular modality. An important caveat to note is that while PCA allows the unsupervised comparison, it is also possible if the threshold for variance is not empirically based, noise in one condition leaks into the common spatial filter, thus creating an unnecessary bias for latency computation. So, it is advisable that the threshold eigenvalues for signal rank determination are empirically motivated.

Multisensory facilitation of MMN and P300

Previous studies have evidenced auditory cortical activity to operate at a faster timescale than visual cortical activity (Williams et al., 2004; Kumar et al., 2020). As expected, in our study, both auditory mismatch negativity (aMMN) and aP300 ERPs were faster than visual MMN (vMMN) and vP300 ERPs, respectively. But what happens when these two modalities come together in different combinations still needs to be deciphered. When audio and visual oddballs are presented synchronously against a background of synchronous audiovisual standards, avMMN is significantly faster than both aMMN and vMMN. As the mismatch signal further propagates onto the P300 level, aP300 and avP300 attain comparable latencies, avP300 still being significantly faster than vP300. This suggests the presence of multisensory facilitation

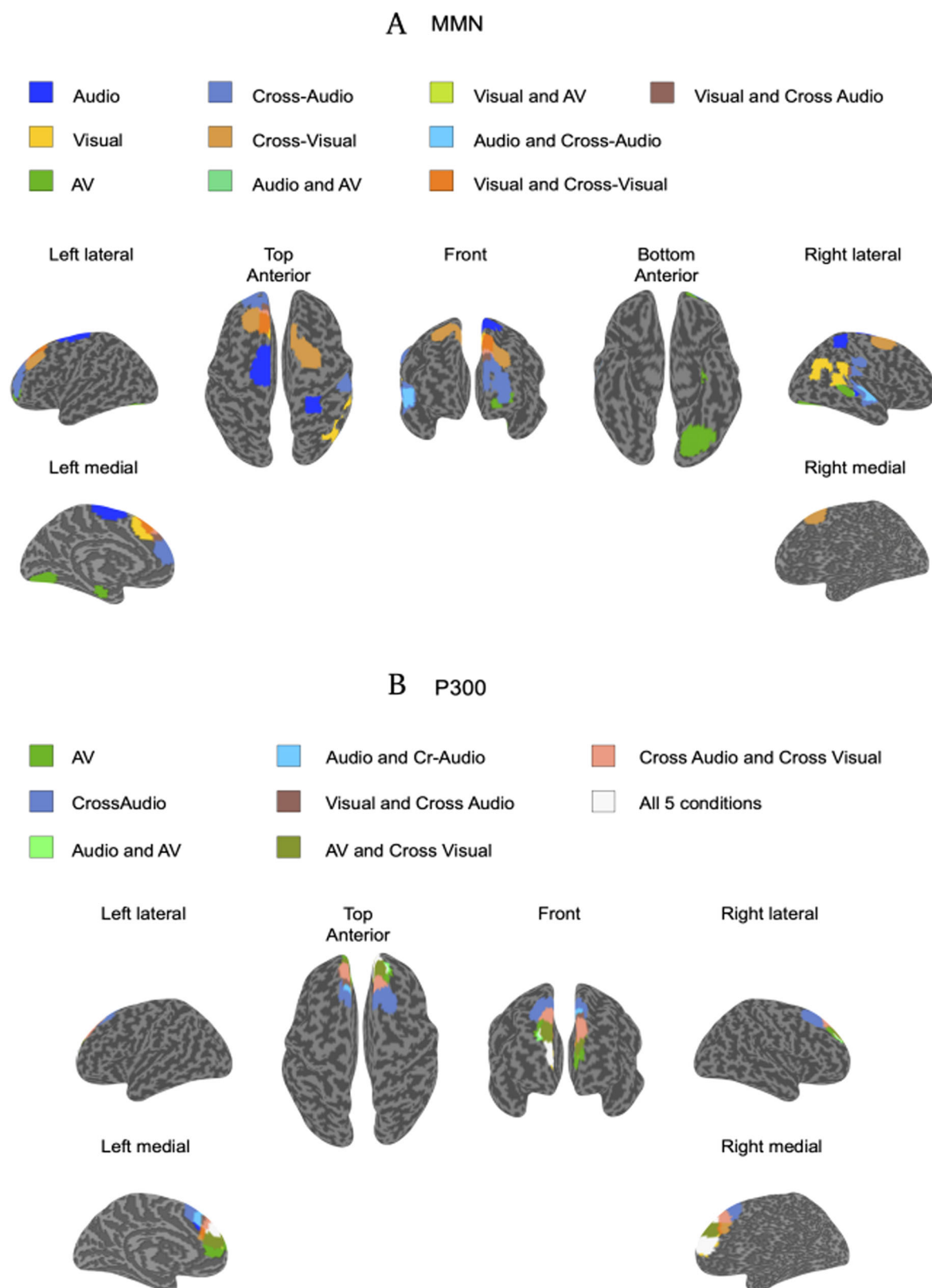


Figure 4. eLORETA source localization results using time-locked analysis (99th percentile and above) representing the underlying **(A)** MMN and **(B)** P300 sources for audio, visual, audiovisual, cross-audio, and cross-visual modalities.

Table 4. The table lists the brain areas underlying the peak MMN and P300 activations across the audio-only, visual-only, and audiovisual conditions

Condition	MMN	P300
Audio-only	SFG, left superior frontal gyrus PrG, left precentral gyrus PrG, right precentral gyrus PCL, left paracentral lobule STG, right superior temporal gyrus MTG, right middle temporal gyrus aSTS, anterior superior temporal sulcus SPL, right superior parietal lobule	SFG, right superior frontal gyrus SFG, left superior frontal gyrus MFG, right middle frontal gyrus
Visual only	SFG, left superior frontal gyrus pSTS, right posterior superior temporal sulcus IPL, right inferior parietal lobule	SFG, right superior frontal gyrus SFG, left superior frontal gyrus
Audiovisual	MFG, left middle frontal gyrus MTG, right middle temporal gyrus Left fusiform gyrus Left parahippocampal gyrus pSTS, right posterior superior temporal sulcus IPL, right inferior parietal lobule MVOcC, left medioventral occipital cortex LOcC, left lateral occipital cortex	SFG, right superior frontal gyrus SFG, left superior frontal gyrus MFG, right middle frontal gyrus MFG, left middle frontal gyrus Left cingulate gyrus
Cross-audio	SFG, left superior frontal gyrus MFG, left middle frontal gyrus STG, right superior temporal gyrus MTG, right middle temporal gyrus IPL, right inferior parietal lobule	SFG, right superior frontal gyrus SFG, left superior frontal gyrus MFG, right middle frontal gyrus
Cross-visual	SFG, right superior frontal gyrus SFG, left superior frontal gyrus MFG, left middle frontal gyrus MFG, right middle frontal gyrus	SFG, right superior frontal gyrus SFG, left superior frontal gyrus

where the faster auditory component in the AV oddball temporally facilitated its slower visual component, which speeded up the process of change detection in the audiovisual condition. On the other hand, despite having the same audio deviant in cross-audio condition as used in the audio-only condition (and the same visual deviant in cross-visual condition as presented in the visual-only condition), the latency of crMMN was significantly shorter than aMMN while that of crvP300 was significantly shorter than vP300, indicating that the processing speed of oddballs is not only determined by the modality of the oddball itself but by that of the background standard as well.

The speeded appearance of the prediction error markers in the audiovisual condition can possibly be attributed to an early coupling in the information processing hierarchy between the constituent unisensory audio and unisensory visual deviant stimuli during an early sensory processing stage (e.g., at N100; [Van Wassenhove et al., 2005](#); [Stekelenburg and Vroomen, 2007](#)). An alternate account could be that the unisensory systems operate largely independently to extract feature-level information, before multisensory convergence and integration marked by avMMN at the superior temporal gyrus and middle temporal gyrus, the higher-order association areas ([Callan et al., 2004](#); [Beauchamp, 2005](#)). If there were no intersensory coupling at the sensory-perceptual level, the MMN for combined visual and auditory deviants should reflect a linear summation of the individual auditory and visual MMNs ([Molholm et al., 2002](#), for auditory and somatosensory deviants). In our results, however, we saw a clear nonlinearity in the multisensory response where the amplitudes of the avMMN and avP300 were smaller than the sum of amplitudes of aMMN and vMMN as well as that of aP300 and vP300, respectively ([Fig. 3C,D](#)). This suggests that the aMMN and aP300 responses were minimized when paired with the concurrent visual stimulation, which can be attributed to the bimodal nature of our audiovisual stimuli as the presence of two stimuli gives the brain more flexibility to choose between deviant features to process the AV oddballs efficiently and, hence, conserve its neural resources during change detection in accordance to the free energy principle ([Friston, 2010](#)). To minimize refractoriness in the sensory cortical feature-specific neurons ([Jacobsen and Schröger, 2001](#)), there may also be a switch between which features the brain attends to, during the course of the oddball experiment. In the future, comparing audio and visual deviant stimuli against a matched audiovisual standard can further explain if the deviant stimuli alone can bring forth any change in the processing speed of prediction errors and also mitigate differences arising due to different baseline noise in standards. To better understand the nature and extent of audiovisual interactions, it will be essential to design advanced paradigms that form meaningful stimuli in both unisensory and multisensory constructs.

Loss of specificity from MMN to P300

Source localization results for MMN showed heterogeneous activations in the brain across different sensory modalities (as shown in Fig. 4A). From these results, we concur that mostly the primary and secondary sensory areas distributed throughout the brain along with other higher cortical areas are employed during the process of early prediction mismatch, based on the modality. In traditional literature, MMN is widely considered a “perceptual” prediction error signal carrying important novel information that initiates a call for further processing of prediction mismatches (Mäntysalo and Näätänen, 1987; Schröger, 1997; Escera et al., 2000). Such novel information should necessarily feed into the higher cognitive processing areas for further evaluation, as is reflected by the occurrence of the late event-related potential, P300 (Bledowski et al., 2004b; David and Linden, 2005). Our source results revealed common source activations for P300 localized around the fronto-central regions for audio-only, visual-only, audiovisual, cross-audio, and cross-visual conditions (as shown in Fig. 4B). Common brain generators for P300 across modalities suggest that this neural marker might particularly be responsible for domain-general higher evaluative processes like keeping a count of the number of oddballs as in our paradigm, after a change detection signal has been relayed by MMN generators of corresponding modalities in the brain. A caveat to note here is that the MMN stage does not necessarily reflect complete modality dependence, and some amount of overlapping cortical sources is probably responsible for its generation. Nonetheless, our results highlight that the existence of modality-dependent cortical modules almost disappears only at P300 stage suggesting a reorganization of brain regions along the temporal hierarchy of prediction error processing (MMN followed by P300) as the mismatch information flows up the spatial hierarchy of the brain (sensory to higher-order regions). Though such a transition may appear unsurprising in terms of neurocognitive processing, we questioned at what stage in the information processing hierarchy is the modality specificity to the sensory stream lost? We found that the processing of prediction error becomes completely amodal at the P300 level itself and is not something that is attained at further later stages during the reorienting negativity (RON) that is elicited at ~400–600 ms (Correa-Jaraba et al., 2016; Ugan et al., 2019).

An important aspect of our present study is that the hierarchical stages of prediction error processing can be explained by multiple underlying internal models of error propagation (Aitchison and Lengyel, 2017). A detailed neuronal model of the auditory, visual, and audiovisual cortices, based on the underlying processes of predictive mismatch that account for the critical features of MMN, P300, and RON could better explain the process of temporal facilitation and supra-additivity in the audiovisual modality (Molholm et al., 2002; Pattamadilok and Sato, 2022). Our empirical investigations in the context of hierarchical processing of prediction errors would have implications beyond the theoretical domain as well. Based on the models of MMN and P300 responses from patients with disorders of consciousness like in vegetative and minimally conscious states (Daltrozzo et al., 2007; Boly et al., 2011), attention (Polich, 2007; Szurmi et al., 2011), and schizophrenia (Blackwood, 2000; Erickson et al., 2016), researchers can isolate the deficits in predictive information flow that might underlie these states of profound cognitive and neurological dysfunction. Such foundational advances can be of extreme value to clinical neuroscience researchers. From a neural networks’ perspective, two well-known attentional networks, the dorsal attention network (DAN) and the ventral attention network (VAN), have been largely reported in oddball studies (Clark et al., 2000; Stevens et al., 2000; Bledowski et al., 2004a). The VAN in particular has been exclusively involved in the detection of deviant stimuli (Palaniyappan and Liddle, 2012; Kim, 2014) and is activated at both MMN and P300 stages (Justen and Herbert, 2018). Further connectivity analysis between the regions underlying the VAN during MMN and P300 latencies across various modalities can further our understanding of the function of this important attentional network beyond recent evidence in the frequency domain (Ghosh et al., 2021). Although a few studies have also attempted to draw a relationship between prestimulus brain states and P300 (Reinhart et al., 2011; Karch et al., 2016), a detailed source connectivity analysis will be informative about the mechanisms of P300 and MMN generation, thus providing an important direction to future research.

References

- Aitchison L, Lengyel M (2017) With or without you: predictive coding and Bayesian inference in the brain. *Curr Opin Neurobiol* 46:219–227.
- Atienza M, Cantero JL, Escera C (2001) Auditory information processing during human sleep as revealed by event-related brain potentials. *Clin Neurophysiol* 112:2031–2045.
- Banellis L, Sokoliuk R, Wild CJ, Bowman H, Cruse D (2020) Event-related potentials reflect prediction errors and pop-out during comprehension of degraded speech. *Neurosci Conscious* 2020:niaa022.
- Beauchamp MS (2005) See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr Opin Neurobiol* 15:145–153.
- Blackwood D (2000) P300, a state and a trait marker in schizophrenia. *Lancet* 355:771–772.
- Bledowski C, Prvulovic D, Goebel R, Zanella FE, Linden DEJ (2004a) Attentional systems in target and distractor processing: a combined ERP and fMRI study. *NeuroImage* 22:530–540.
- Bledowski C, Prvulovic D, Hoehstetter K, Scherg M, Wibral M, Goebel R, Linden DEJ (2004b) Localizing P300 generators in visual target and distractor processing: a combined event-related potential and functional magnetic resonance imaging study. *J Neurosci* 24:9353.
- Boly M, Garrido MI, Gosseries O, Bruno MA, Boveroux P, Schnakers C, Massimini M, Litvak V, Laureys S, Friston K (2011) Preserved feedforward but impaired top-down processes in the vegetative state. *Science* 332:858–862.
- Calculus A, Deltenre P, Hoonhorst I, Collet G, Markessis E, Colin C (2015) MMN and P300 are both modulated by the featured/featureless nature of deviant stimuli. *Clin Neurophysiol* 126:1727–1734.

- Callan DE, Jones JA, Munhall K, Kroos C, Callan AM, Vatikiotis-Bateson E (2004) Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *J Cogn Neurosci* 16:805–816.
- Chennu S, Noreika V, Gueorguiev D, Blenkmann A, Kochen S, Ibáñez A, Owen AM, Bekinschtein TA (2013) Expectation and attention in hierarchical auditory prediction. *J Neurosci* 33:11194–11205.
- Clark VP, Fannon S, Lai S, Benson R, Bauer L (2000) Responses to rare visual target and distractor stimuli using event-related fMRI. *J Neurophysiol* 83:3133–3139.
- Correa-Jaraba KS, Cid-Fernández S, Lindín M, Díaz F (2016) Involuntary capture and voluntary reorienting of attention decline in middle-aged and old participants. *Front Hum Neurosci* 10:129.
- Daltrozzo J, Wioland N, Mutschler V, Kotchoubey B (2007) Predicting coma and other low responsive patients outcome using event-related brain potentials: a meta-analysis. *Clin Neurophysiol* 118:606–614.
- David N, Linden EJ (2005) The P300: where in the brain is it produced and what does it tell us? *Neuroscientist* 11:563–576.
- Dhamala M, Assisi CG, Jirsa VK, Steinberg FL, Kelso JAS (2007) Multisensory integration for timing engages different brain networks. *Neuroimage* 34:764–773.
- Di Luca M, Machulla TK, Ernst MO (2009) Recalibration of multisensory simultaneity: cross-modal transfer coincides with a change in perceptual latency. *J Vis* 9:7.
- Dinh TH, Singh AK, Linh Trung N, Nguyen DN, Lin C-T (2021) EEG peak detection in cognitive conflict processing using summit navigator and clustering-based ranking. *IEEE Trans Neural Syst Rehabil Eng* 30:1548–1556.
- Dinteren R, Arns M, Jongsma MLA, Kessels RPC (2014) P300 development across the lifespan: a systematic review and meta-analysis. *PLoS One* 9:e87347.
- Erickson MA, Ruffle A, Gold JM (2016) A meta-analysis of mismatch negativity in schizophrenia: from clinical risk to disease specificity and progression. *Biol Psychiatry* 79:980.
- Escera C, Alho K, Schröger E, Winkler I (2000) Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiol Neurootol* 5:151–166.
- Feng T, Qiu Y, Zhu Y, Tong S (2008) Attention rivalry under irrelevant audiovisual stimulation. *Neurosci Lett* 438:6–9.
- Fleming JT, Noyce AL, Shinn-Cunningham BG (2020) Audio-visual spatial alignment improves integration in the presence of a competing audio-visual stimulus. *Neuropsychologia* 146:107530.
- Friston KJ (2005) A theory of cortical responses. *Phil Trans R Soc B* 360:815–836.
- Friston K (2010) The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11:127–138.
- Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009) The mismatch negativity: a review of underlying mechanisms. *Clin Neurophysiol* 120:453.
- Ghosh P, Roy D, Banerjee A (2021) Organization of directed functional connectivity among nodes of ventral attention network reveals the common network mechanisms underlying saliency processing across distinct spatial and spatio-temporal scales. *NeuroImage* 231:117869.
- Giard MH, Perrin F, Pernier J, Bouchet P (1990) Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. *Psychophysiology* 27:627–640.
- Gonsalvez CJ, Barry RJ, Rushby JA, Polich J (2007) Target-to-target interval, intensity, and P300 from an auditory single-stimulus task. *Psychophysiology* 44:245–250.
- Gonsalvez CJ, Polich J (2002) P300 amplitude is determined by target-to-target interval. *Psychophysiology* 39:388–396.
- Gramfort A, Papadopoulos T, Olivi E, Clerc M (2010) OpenMEEG: open-source software for quasistatic bioelectromagnetics. *Biomed Eng Online* 9:1–20.
- Halder T, Talwar S, Jaiswal AK, Banerjee A (2019) Quantitative evaluation in estimating sources underlying brain oscillations using current source density methods and beamformer approaches. *eNeuro* 6:ENEURO.0170-19.2019.
- Jacobsen T, Schröger E (2001) Is there pre-attentive memory-based comparison of pitch? *Psychophysiology* 38:723–727.
- Jolliffe IT (2002) *Principal component analysis*. New York: Springer-Verlag.
- Justen C, Herbert C (2018) The spatio-temporal dynamics of deviance and target detection in the passive and active auditory oddball paradigm: a sLORETA study. *BMC Neurosci* 19:25.
- Karch S, Loy F, Krause D, Schwarz S, Kiesewetter J, Segmiller F, Chrobok AI, Keiser D, Pogarell O (2016) Increased event-related potentials and alpha-, beta-, and gamma-activity associated with intentional actions. *Front Psychol* 7:7.
- Kim H (2014) Involvement of the dorsal and ventral attention networks in oddball stimulus processing: a meta-analysis. *Hum Brain Mapp* 35:2265–2284.
- Koelsch S, Heinke W, Sammler D, Olthoff D (2006) Auditory processing during deep propofol sedation and recovery from unconsciousness. *Clin Neurophysiol* 117:1746–1759.
- Kumar VG, Dutta S, Talwar S, Roy D, Banerjee A (2020) Biophysical mechanisms governing large-scale brain network dynamics underlying individual-specific variability of perception. *Eur J Neurosci* 52:3746–3762.
- Magliero A, Bashore TR, Coles MGH, Donchin E (1984) On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology* 21:171–186.
- Mäntysalo S, Näätänen R (1987) The duration of a neuronal trace of an auditory stimulus as indicated by event-related potentials. *Biol Psychol* 24:183–195.
- Marucci M, Di Flumeri G, Borghini G, Sciaraffa N, Scandola M, Pavone EF, Babiloni F, Betti V, Aricò P (2021) The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Sci Rep* 11:4831.
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* 14:115–128.
- Näätänen R, Gaillard AWK, Mäntysalo S (1978) Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica* 42:313–329.
- Näätänen R, Michie PT (1979) Early selective-attention effects on the evoked potential: a critical review and reinterpretation. *Biol Psychol* 8:81–136.
- Näätänen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol* 118:2544–2590.
- Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011:156869.
- Palaniyappan L, Liddle PF (2012) Does the salience network play a cardinal role in psychosis? An emerging hypothesis of insular dysfunction. *J Psychiatry Neurosci* 37:17–27.
- Pascual-Marqui, RD (2007) Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization. Available at: <http://arxiv.org/abs/0710.3341>
- Patel SH, Azzam PN (2005) Characterization of N200 and P300: selected studies of the event-related potential. *Int J Med Sci* 2:147.
- Pattamadilok K, Sato M (2022) How are visemes and graphemes integrated with speech sounds during spoken word recognition? ERP evidence for supra-additive responses during audiovisual compared with auditory speech processing. *Brain Lang* 225:105058.
- Polich J (2007) Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol* 118:2128–2148.
- Polich J, Ladish C, Burns T (1990) Normal variation of P300 in children: age, memory span, and head size. *Int J Psychophysiol* 9:237–248.
- Reinhart RMG, Mathalon DH, Roach BJ, Ford JM (2011) Relationships between pre-stimulus gamma power and subsequent P300 and reaction time breakdown in schizophrenia. *Int J Psychophysiol* 79:16–24.

- Sams M, Paavilainen P, Alho K, Näätänen R (1985) Auditory frequency discrimination and event-related potentials. *Electroencephalogr Clin Neurophysiol* 62:437–448.
- Schröger E (1997) On the detection of auditory deviations: a pre-attentive activation model. *Psychophysiology* 34:245–257.
- Shiramatsu TI, Mori K, Ishizu K, Takahashi H (2021) Auditory, visual, and cross-modal mismatch negativities in the rat auditory and visual cortices. *Front Hum Neurosci* 15:536.
- Sittiprapaporn W (2012) An effect of attention on mismatch negativity in audiovisual visual modalities. *Pak J Biol Sci* 15:542–546.
- Stefanics G, Stavrinou M, Sestieri C, Ciancetta L, Belardinelli P, Cianflone F, Bernáth L, Hernádi I, Pizzella V, Romani GL (2005) Cross-modal visual-auditory-somatosensory integration in a multi-modal object recognition task in humans. *Int Congr Ser* 1278:163–166.
- Stefanics G, Stefanics G, Kremláček J, Czigler I (2014) Visual mismatch negativity: a predictive coding view. *Front Hum Neurosci* 8:666.
- Stekelenburg JJ, Vroomen J (2007) Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci* 19:1964–1973.
- Stevens AA, Skudlarski P, Gatenby JC, Gore JC (2000) Event-related fMRI of auditory and visual oddball tasks. *Magn Reson Imaging* 18:495–502.
- Sutton S, Tueting P, Zubin J, John ER (1967) Information delivery and the sensory evoked potential. *Science* 155:1436–1439.
- Szuromi B, Czobor P, Komlósi S, Bitter I (2011) P300 deficits in adults with attention deficit hyperactivity disorder: a meta-analysis. *Psychol Med* 41:1529–1538.
- Talsma D (2015) Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Front Integr Neurosci* 9:19.
- Ungan P, Karsilar H, Yagcioglu S (2019) Pre-attentive mismatch response and involuntary attention switching to a deviance in an earlier-than-usual auditory stimulus: an ERP study. *Front Hum Neurosci* 13:58.
- Van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A* 102:1181–1186.
- Wacongne C, Changeux JP, Dehaene S (2012) A neuronal model of predictive coding accounting for the mismatch negativity. *J Neurosci* 32:3665–3678.
- Wacongne C, Labyt E, Van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S (2011) Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc Natl Acad Sci U S A* 108:20754–20759.
- Williams PE, Mechler F, Gordon J, Shapley R, Hawken MJ (2004) Entrainment to video displays in primary visual cortex of macaque and humans. *J Neurosci* 24:8278–8288.