# Title Page

## Title: Prediction error processing speeds up in presence of multisensory contexts

**Authors:** Siddharth Talwar *, Priyanka Ghosh*, Arpan Banerjee#

*Authors having equal contributions, co first authors

**# Address for correspondence:** arpan@nbrc.ac.in

**Affiliation:** Cognitive Brain Dynamics Lab, National Brain Research Centre, NH-8, Manesar, Gurgaon, Haryana-122052, India

**Number of pages:** 32

**Number of figures:** 5

**Number of tables:** 4

**Number of words:**

**Abstract:** 234

**Significance statement:** 106

**Introduction:** 650

**Discussion:** 1432

## Conflict of interest

Authors declare no conflict of interest scenarios.

## Acknowledgements

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

## Abstract

Prediction errors in the brain are indexed by two event related potentials – MMN and P300 which are elicited whenever there is a violation of regularity in occurrence of a repetitive stimulus. While, MMN reflects the brain's ability to perform automatic comparisons between consecutive stimuli and provides an electrophysiological index of sensory error detection, P300 is associated with cognitive processes such as attention and memory. Till 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. However, the relationship between these two ERPs, in context of prediction error propagation is an open question. Secondly, whether these events represent modality specific or modality independent information processing needs to be tested in unisensory and multisensory environments. Our EEG study on 22 healthy human volunteers shows that multisensory prediction errors are temporally facilitated by the constituent unisensory modality that has shorter processing latencies at any stage of prediction error processing. Employing source localization using individual participant's MRI scan, our study reveals that the corresponding ERP generators underlying the temporal hierarchy of prediction mismatch signals in the brain transition from being modality sensitive to modality independent across both unisensory and multisensory contexts. Such knowledge can be of immense value in clinical research for determining the stages of various treatments in aging, schizophrenia and depression, and their efficacy on cognitive function.

**Abbreviations**

MMN – Mismatch negativity; ERP – Event related potential; EEG - Electroencephalography

## Significance statement

Failure of an expected event to occur needs to reinforce the internal model of environment in the brain thus, forming a key ingredient of any predictive coding framework. Just the way it does for human learning, the analysis of prediction errors can help guide machine learning in artificial intelligence by minimizing the error feedback. In order to facilitate such studies, it is important to first empirically establish how these prediction errors are processed in the human brain and how differently are they processed across various modalities because the environment around us is essentially multisensory in nature. Our electroencephalography study captures this precisely, both temporally and spatially.

## 1. Introduction

Brain attends to complex stimuli by minimizing its response to frequent events that do not require extensive processing thus, freeing up the cognitive resources to process unexpected events (Fontolan et al., 2014; Friston, 2005). To subserve this goal, a neural representation (internal model) is constructed based on the prior information from previous sensory inputs and any mismatch between the model and the subsequent sensory input gives rise to a prediction error (Kok et al., Winkler and Czigler, 2012). The prediction errors, at the cortical level, are known to generate two ERPs in the brain – the MMN (Mismatch Negativity) and the P300 (Calcus et al., 2015; Chennu et al., 2013). MMN appears as a negative deflection in the difference waveform between deviant and standard EEG signals around 100-250 ms post onset of a deviant stimulus with frontal/central/temporal generators in the brain (Näätänen et al., 2007). MMN can be elicited pre-attentively, during non-attentive states such as sleep (Molholm et al., 2005), or even in coma (Morlet and Fischer, 2014). P300, on the other hand, is a positive deflection in voltage with a latency between 250 to 500 ms distributed across fronto-central scalp locations (Polich, 2007a). P300 being a late component, is considered as a marker of attentive/evaluative processing of oddballs and has also been linked to decision making processes (Rohrbaugh et al., 1974).

Most prediction mismatches in our surroundings are perceived through not one but multiple sensory modalities (Luca et al., 2009), e.g. noticing the brake lights of a car in front while engaged listening to radio. From a predictive coding perspective, the internal mental model of sight, sound, smell, taste and touch are integrated with our existing cognitive schemata (Talsma, 2015). Processing of novel bottom-up stimuli in one modality can subsequently modify the neural representation and hence top-down prediction, of a stimulus in another sensory modality. If multisensory integration involves such complex interactions between top-down and bottom-up processes, then it should take place at multiple stages of processing, based on the complexity of the stimuli involved (Molholm et al., 2002).

The first distinct stage of multisensory integration was identified at about 100 ms of speech stimulus onset where the audio-visual N1 component peaked earlier than that of the auditory speech stimulus alone (Stekelenburg and Vroomen, 2007; Van Wassenhove et al., 2005). Interestingly, a few studies using oddball stimuli have also shown that multimodal stimulation (visual-audio-tactile and visual-audio) induces a significant early onset of latency (Marucci et al., 2021; Stefanics et al., 2005), and a significant increase in the amplitude of the P300 potentials (Fleming et al., 2020; Marucci et al., 2021) with respect to the corresponding unimodal stimulations. Feng et al.(2008), however did not observe such multisensory benefits for P300 latencies. In a recent study using the oddball paradigm (Shiramatsu et al., 2021), the latency of audio-visual MMN resembled closely with that of audio MMN as compared to the latency of visual MMN, indicating the advantage of audio MMN over visual MMN. However, a previous study (Sittiprapaporn, 2012) using a different oddball paradigm showed that visual and audio-visual MMN had shorter latencies that auditory MMN. Despite a huge body of research on MMN and P300, there is not one single comprehensive study that describes the relationship between the two ERPs, temporally and spatially, when there is interaction of inputs from various modalities.

Motivated by aforementioned lines of questioning, we hypothesized that (i) the speeded responses to multisensory stimuli are also seen in the middle/ late processing stages of prediction errors, i.e., for MMN/P300 and (ii) there exists sensory-cognitive dissociation in the source distribution of MMN and P300. Accordingly, we designed a multimodal oddball paradigm to systematically compare the markers of prediction errors between unisensory and multisensory stimuli using audio and visual modalities. The EEG data from human participants were analyzed in sensor and source space, the latter by co-registering the structural MRI scans from each subject with their EEG sensor locations.

## 2. Methods

### 2.1. Participants

22 healthy volunteers (9 males and 13 females) in the age group of 22 to 43 (mean=25.7, SD=±4.19) years participated in the study. All participants had no history of neurological, visual or audiological problems. All participants had normal or corrected-to-normal vision and were right handed. The study was carried out following the ethical guidelines and prior approval of Institutional Human Ethics Committee of National Brain Research Centre, India.

### 2.2. 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; third was ***bimodal,*** i.e., audio-visual; 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 of an audio standard and an audio oddball stimulus; the *visual only condition* comprised of a visual standard and a visual oddball stimulus; and the *audio-visual (AV) condition* comprised of an audio-visual standard and an audio-visual oddball stimulus. The remaining 2 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 stimuli. The contents of each condition are tabulated in **Table 1**. Each condition consisted of 400 trials, out of which oddballs constituted 14 percent 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 (5 conditions x 4 blocks) which were randomized and presented to the participants. The participants were prompted about the modality of the upcoming block before every run and in order to engage their attention throughout the entire block, they were asked to count the number of oddballs presented in each block and report at the end of the block.

The stimuli of visual only condition consisted of a standard blue square and a deviant red triangle. The auditory stimuli were inspired from musical notes, the standard as the C4 note and the deviant as the C5 note (higher octave), according to tuning of A440 pitch standard. 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 800-1200 ms (mean=1000 ms). AV and cross-modal conditions were constructed from combinations of the audio only and visual only stimuli (listed in Table 1).

## 2.3. 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 was acquired using the default reference near Cz, grounded to electrode AFz. Impedance of all electrodes were monitored and initially set below 5kΩ, and monitored in between blocks. Additionally, head digitization was performed using Polhemus Fastrak 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 3.0 T Philips Achieva MRI scanner (TR = 8.4 ms, FOV = 250 × 230 × 170, flip angle = 8°).

## 2.4. 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 was visually inspected further and 1 channel (F4 in 3 subjects) was interpolated to neighbouring electrodes. To identify and remove blink artifacts from the data, independent component analysis (ICA) was employed to each block. Only the blink component obtained as independent component (IC) from eye regions was visually identified and subsequently the IC was rejected. Each block was further 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. Subsequently, all trials were subjected to a low pass filter of <45 Hz, baseline correction was applied and the data was re-referenced to linked mastoids. Trials with signal amplitude greater than 100μV and lesser than -100μV were visually examined and at most 6% of all oddball trials were discarded per subject. To equate the number of standard and oddball trials from each participant, we chose the minimum number of trials that survived artifact rejection in any condition and those many trials were randomly sampled from each condition for each participant. 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).

## 2.5. Extracting MMN and P300 peaks

We defined MMN as the largest negative deflection (oddball-standard) at 120-250ms from stimulus onset (0 ms) and P300 was defined as the largest postive peak (oddball-standard) at 250-500ms for all conditions. To verify if the enhancement in magnitude of P300 and MMN in the oddball trials as compared to the standard trials were statistically significant across all conditions, we considered the grand-average ERPs from the top five sensors of each condition showing maximum difference (deviant-standard) in their voltage activities. The top 5 sensors showing maximum MMN activity were F1/z/2, FCz/2 in the a*udio only* condition; P1/3, PO3/5/7 in the *visual only* condition; P1/z/2/4, PO3 in the *AV* condition; POz/3/4/5, Oz in the *cross-audio* condition and CP1/3/5, P1/3 in the *cross-visual* condition. Similarly, electrodes F1/z/2, FC1/z showed maximum P300 activity in the a*udio only* condition; electrodes FC1/z, Cz, Cpz, Pz showed maximum activity in the *visual only* condition; electrodes Fz, FC1/z, Cz, Pz in the *AV* condition; electrodes F1/z/2, FC1/z in the *cross-audio* condition and electrodes F1/z, FC1/z/2 showed maximum P300 activity in the *cross-visual* condition. Since different sensors were under consideration for different modalities, we reduced the dimensionality across sensors by applying Principal component analysis (PCA) on the trial-level data (time-points X trials X sensors), separately for standard and deviant conditions. PCA computes eigenvectors of the covariance matrix ("principal axes") and sorts them by their eigenvalues (amount of explained variance). Upon applying PCA, we centered the data by subtracting the mean across the time-points and then this centered data was projected onto the principal axes to yield principal components ("scores"). We considered the projection of the first principal component from the top five sensors, in order to capture the maximum variance in the data. The original variables were further reconstructed from the principal component sub-space by multiplying the data with the projection matrix. To do this, we multiplied the transpose of the first eigenvector with the PCA scores projected along the first component and added back the mean. With this dimensionality reduced data (time-points X trials), we characterized the MMN and P300 latency windows for each condition for subsequent source localization. For both MMN and P300, we extracted 100ms (±50ms) windows centred around the MMN/P300 peaks at the trial averaged subtraction waveforms (Odd - Standard) for all the conditions.

## 2.6. Source Localization

Source localization was performed to obtain the MMN and P300 generators in the brain on the 100 ms latency windows extracted as described in the previous section. The EEG data was re-referenced with an average reference prior to source-localization. 19 participants' structural MRI data were re-sliced and segmented to identify the brain, skull and scalp tissues. 2 subjects' MRI scan 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 co-registration 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 in the segmented part of the headmodel 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. The 100 ms windows centered at P300 and MMN peaks (as obtained in section 2.6) were used to obtain the covariance matrix, separately for standard and oddball trials. Source localisation of each individual was performed using their respective headmodel, leadfield and electrode positions. 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 beneficial for exploratory source analysis, e g., where a prior hypotheses regarding approximate locations may not be available (Halder et al., 2019). Lambda of 0.1 was used as regularization parameter for localization of P3 and MMN. 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 more than 95 percent of the maximum value were thresholded. This was done seperately for each hemisphere.The clusters of P300 sources obtained using 100 ms windows were noisy (did not fall into any Brodmann area), because of which we recomputed the P300 sources using a longer time-window to obtain a better estimate of variance in the data. Thus, the P300 peaks were identified between 250-500 ms post-stimulus onset where maximum peak values of every trial were extracted along with their corresponding latencies. Based on these latencies, we calculated the mean and standard deviation for each modality and defined the windows as [mean-SD : mean+SD]. The covariance matrices were now obtained from these windows and the same steps were repeated to obtain the P300 sources.

## 3. Results

**Middle and late stages of prediction error processing conserved across unisensory and multisensory contexts**

For each condition, the time-series of the standard and the deviant categories were averaged across trials from all the participants and their differences (deviant-standard) were plotted to spot the MMN and P300 ERPs. Condition-wise topographies at the MMN and P300 peaks were visualized (**Figure1**) by taking the difference of deviant and standard conditions at the corresponding latencies. In the present study, to define the amplitudes and latencies of the prediction error markers, we have focussed mainly on the *audio only*, *visual only* and *AV* conditions due to unavailability of 'equivalent standard categories' to subtract from the other two *cross-modal* deviant conditions. However, the *cross-modal* conditions were used to investigate the robustness of the source distribution results across all five conditions which has been discussed in section 4. Student's t-test was performed on the dimensionality reduced data from the top 5 sensors of the 21 subjects which revealed that the magnitude of MMN in the oddball trials was significantly lower than that of the standard trials in *audio only* ($t(20)=5.85$, $SD=2.16$, $p<0.0001$), *visual only* ($t(20)=6.02$, $SD=1.33$, $p<0.0001$) and *audio-visual* ($t(20)=6.65$, $SD=1.82$, $p<0.0001$) conditions (**Figure2a-c**). Similarly, the P300 amplitudes were significantly higher for oddball trials in *audio only* ($t(20)=9.91$, $SD=2.49$, $p<0.0001$), *visual only* ($t(20)=6.18$, $SD=1.93$, $p<0.0001$) and *audio-visual* ($t(20)=6.26$, $SD=3.54$, $p<0.0001$) conditions (**Figure2d-f**).

**Multisensory context speeds the processing of prediction errors**

From the 100 ms MMN and P300 windows obtained (as described in section 2.6), the peak values (maximum negative amplitudes of subtraction waveform for MMN and maximum positive amplitude values from oddball conditions for P300) and their corresponding latencies were extracted for each

trial and subsequently, the means and standard deviations of latencies and amplitudes were computed for all conditions as listed in **Table 2** for P300 and in **Table 3** for MMN. A repeated-measures analyses of variances (rmANOVAs) on the subject-level data revealed a significant effect of modality on the amplitudes of both MMN ($F(2,40)=25.09$, $p<0.0001$, $\eta^2=0.56$) and P300 ($F(2,40)=3.65$, $p=0.03$, $\eta^2=0.15$). In case of MMN latency, Mauchly's test revealed a low p-value ($p = 0.04$) which indicated that the assumption of sphericity was violated. Hence, a Greenhouse-Geisser correction was applied ($\varepsilon =0.78$) and subsequently, a significant effect of modality was detected by rmANOVA on MMN latencies ($F(1.57,31.43)=383.51$, $p<0.0001$, $\eta^2=0.78$). The assumption of sphericity was not violated ($p>0.05$) for P300 latencies even though the p-value obtained through Mauchly's test was low($p=0.06$). rmANOVA showed a significant effect of modality on the P300 latencies ($F(2,40)=786.51$, $p<0.0001$, $\eta^2=0.97$).

Post-hoc comparisons with Bonferroni-adjusted p-values indicated significant differences in the P300 and MMN latencies and amplitudes across modalities (as shown in **Figure3**) reflecting a significant effect of conditions on the amplitudes and latencies of both the ERPs. While comparing the peak latencies of MMN, we observed that AV condition had the shortest MMN latency (AV < audio only/visual only at $p<0.0001$), followed by visual only condition (visual only < audio only at $p<0.0001$) while the slowest MMN latency was seen for the audio only condition. On the contrary, P300 latency was fastest for the audio only condition (audio only < visual only at $p<0.0001$; audio only < AV at $p=0.01$), followed by the AV condition (AV < visual only at $p<0.0001$). While comparing the MMN amplitudes, the MMN peak of the audio only condition showed the largest magnitude (audio only > visual only at $p<0.0001$; audio only > AV at $p=0.03$), followed by the AV condition (AV > visual only at $p=0.0002$). On the other hand, P300 amplitudes of visual only, audio only and AV conditions were not significantly different from each other.

**Spatial representations of unisensory and multisensory contexts in source space**

Using co-registration of individual participant MR data with their EEG sensor locations, we generated the source maps underlying MMN and P300 activity using eLORETA (details in Methods). The MMN sources were distributed throughout the brain and were different for different modalities (as illustrated in **Figure 4 a** and locations in **Table 4**). Parcels from Brainnetome atlas were interpolated to the same common structural template as used for the normalization of the individual sources. Only those parcels were chosen which included the sources or a part of them. The dorsolateral areas of the left and right superior frontal gyri were found to underlie the MMN response in the *audio only* condition; the caudal cuneus gyrus in medio-ventral occipital cortex on the left, the middle occipital gyrus in right lateral occipital cortex, the rostro-dorsal area of the right inferior parietal lobule and the dorso-lateral area of the right middle temporal gyrus elicited MMN in the *visual only* condition; and the *audio-visual* condition yielded sources that were located in the rostral areas of left superior parietal lobule and the caudal areas of multisensory regions STG (superior-temporal gyrus) and MTG (medial-temporal gyrus) on the right.

Interestingly, source analysis of P300 revealed many overlapping fronto-central areas for all the three sensory modality conditions (as illustrated in **Figure 4 b**). P300 was elicited at the dorso-lateral areas of left superior frontal gyri for *audio only, visual only* and *audio-visual* conditions; at right superior frontal gyri for the *audio only* and *audio-visual* conditions; at the trunk and caudal dorso-lateral areas in left precentral gyri for the *audio only* and *visual only* conditions, respectively; and additionally at the ventro-lateral area of right middle frontal gyrus for the *visual only* condition (details in **Table 4**).

## 4. Discussion

The first main outcome of this study is the demonstration that presence of multisensory contexts, i.e., simultaneous occurrence of audio and visual oddballs, result in faster processing speeds at various stages of prediction error processing in the brain. This follows previous studies which have shown that multisensory (audio + visual) contexts enhance the processing speed of both speech and non-speech stimuli (Begau et al., 2021; Diaconescu et al., 2011; Leone and Mccourt, 2015; van Wassenhove et al., 2005). The second major outcome was the comparison of modality specific versus modality agnostic relationship of early and late ERP components related to prediction error processing across unimodal,, bimodal and cross-modal contexts using combinations of auditory and visual stimuli in an oddball set-up. We found that the reorganization of brain regions along the temporal hierarchy of prediction error processing is reflected by an initial "modality-sensitive" phase, as indexed by MMN, that transitions to a "modality-independent" phase indexed by P300, as the mismatch information flows up the spatial hierarchy of the brain. In addition to MMN and P300, a further late ERP component is referred to index prediction errors in the brain. This component, known as the reorienting negativity or RON, appears as a negative deflection at 400-600ms in response to reorientation of attention to the predictive/standard stimuli after it has been switched (indexed by P300) towards the preceding deviant stimuli (Correa-Jaraba et al., 2016; Ungan et al., 2019). Even though our study could not cover this ERP since it did not come out as a prominent peak in the subtraction waveforms (**Figure 1**), especially in the audio only modality, we argue that RON too would, being a late component, have common brain generators like P300 across modalities and it is also shown to be centred upon fronto-central regions in extant literature (Correa-Jaraba et al., 2016; Justo-Guillén et al., 2019; Ungan et al., 2019). In the remaining text, we discuss the implications of the two key results.

As MMN and P300 are considered as the two main markers of predictive processing in this study, we investigated the latency differences between them for oddballs presented in unisensory and multisensory contexts. The MMN latencies obtained for our two stimulus oddball paradigm were in the following order: Audio-visual<Visual only<Audio only (**Table 3**). According to our results, the auditory MMN appears much later than the visual MMN which can be explained by the spatial hierarchy at which both the MMNs are generated in the predictive brain. Superior-frontal auditory MMN might be indicative of a stronger call for attention towards change in a single stimulus feature i.e., frequency change in tone, that requires an active comparison of the mismatch with the standard (for an average non-musician), as opposed to changes in color/shape in the visual oddball that easily get detected at the level of the visual areas in the occipital cortex. In the audio only condition, the travel time for further bottom up information flow was significantly reduced post MMN elicitation which resulted in fastest P300 elicitation (**Figure 3d**) at the superior-frontal region. On the contrary, visual mismatch information that was quickly resolved at the level of the occipital cortex during MMN, had to travel higher up until the superior-frontal area for its next level of mismatch processing, possibly resulting in slowest visual P300. Interestingly, fastest elicitation of audio-visual MMN (**Figure 3c**) may suggest that the faster visual component of the AV oddball temporally facilitated its slower auditory component which speeded up the process of change detection in an audio-visual context. Evidently, this facilitation for audio-visual deviants continued further up the deviant processing ladder and is also seen for AV P300 at the superior-frontal generators. The latency of AV oddball processing is however second to the audio only condition at the level of P300 but is significantly faster than the slowest visual P300, indicating that the audio-visual ERPs are always temporally facilitated by the fastest unisensory components at all stages of prediction error processing. The faster processing speed of AV deviant stimuli during P300 can perhaps be attributed to an association between the constituent unisensory audio and unisensory visual oddballs at the STG and MTG at an early pre-attentive stage during MMN. Although, unlike speech stimuli, in our paradigm the audio and visual stimuli together do not form any meaningful stimuli as such, the

simultaneous presentation of the audio and visual stimuli was sufficient to induce a multisensory context. It may also be noted that the two auditory stimuli presented, were pure tones with musical relation in terms of pitch (**Table 1**). Thus, this may indicate that the MMN and the P300 obtained for audio only condition represent the neural signatures of auditory pitch mismatch processing. Nonetheless, the results clearly reveal that presence of multisensory context shift the processing times towards the faster unisensory modality indicating a multisensory benefit for processing prediction errors.

Since, based on latencies one can postulate MMN as an early marker of predictive processing and P300 as a hierarchically superior marker, we sought to determine whether the underlying brain areas involved reflect any hierarchy of processing as well. For example, an earlier processing stage should reflect a modality specific component with involvement of unisensory processing regions to a greater degree, whereas a hierarchically later stage should reflect a domain general state of brain processing in the higher cortical regions of the predictive brain. Source localization results for MMN show scattered activations throughout the brain varying across all the sensory modalities (as shown in **Figure 5a**). For the *visual only* modality, we saw activations in the left medio-ventral occipital cortex, the right lateral occipital cortex, the right inferior parietal lobule and the right middle temporal gyrus; for *audio only* condition, we saw activations in the left and right superior frontal gyri; for the *audio-visual* modality, activations were seen in the left superior parietal lobule and the right superior-temporal and medial-temporal gyri which have been implicated as the areas of multisensory integration (Beauchamp, 2005; Callan et al., 2004); for *cross-visual* modality, we found activations in the left and the right superior parietal lobules; and that for *cross-audio* modality, in the left and the right lateral occipital cortex (**Figure 5a**). From these results, we can concur that different cortical areas are employed during the process of early prediction mismatch, based on the modality. In traditional literature, MMN was assumed to reflect the process of initiating a call for further processing of prediction mismatches potentially carrying important novel information (Escera et al.,

2000; Mäntysalo and Näätänen, 1987; Schröger, 1997). 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., 2004; David and Linden, 2005). Our source results revealed overlapping source activations for P300 in all the five modality conditions, localized around the fronto-central regions for *audio only, visual only* and *audio-visual* conditions (as shown in **Table 4**) and also at the right superior frontal gyrus, the left and the right precentral gyri for *cross-audio* condition and, at the left and the right superior frontal gyri for *cross-visual* condition (as shown in **Figure 5b**). Common brain generators for P300 across modalities suggest that this neural marker might particularly be responsible for 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.

Many recent and past studies have argued that the brain is not a passive input–output device but acts as a predictive system capable of anticipating the future (Bubic et al., 2010; Clark, 2013; Ficco et al., 2021; Friston, 2010). Thus, there arises a need to further understand the neural mechanisms by which the brain processes prediction errors in complex scenarios and multiple modality combinations. A detailed neuronal model of the auditory, visual and audio-visual cortices, based on the underlying processes of predictive mismatch that account for the critical features of MMN (Wacongne et al., 2012), P300 and RON could better explain the process of temporal facilitation and supra-additivity (if any) in the audio-visual modality. 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 (Boly et al., 2011), and attention (Polich, 2007b), 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.

## Author Contributions

ST, PG and AB conceived the study; ST and PG designed the stimulus; PG collected the data and wrote the initial draft; ST and PG analysed data; ST, PG and AB revised the draft; AB supervised the work.

## References


Beauchamp, M.S., 2005. See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. Curr. Opin. Neurobiol. 15, 145–153. https://doi.org/10.1016/J.CONB.2005.03.011

Begau, A., Klatt, L.I., Wascher, E., Schneider, D., Getzmann, S., 2021. Do congruent lip movements facilitate speech processing in a dynamic audiovisual multi-talker scenario? An ERP study with older and younger adults. Behav. Brain Res. 412, 113436. https://doi.org/10.1016/J.BBR.2021.113436

Bledowski, C., Prvulovic, D., Hoechstetter, K., Scherg, M., Wibral, M., Goebel, R., Linden, D.E.J., 2004. Localizing P300 Generators in Visual Target and Distractor Processing: A Combined Event-Related Potential and Functional Magnetic Resonance Imaging Study. J. Neurosci. 24, 9353. https://doi.org/10.1523/JNEUROSCI.1897-04.2004

Boly, M., Garrido, M.I., Gosseries, O., Bruno, M.A., 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 (80-. ). 332, 858–862. https://doi.org/10.1126/SCIENCE.1202043/SUPPL_FILE/1202043.BOLY.SOM.PDF

Bubic, A., Yves von Cramon, D., Schubotz, R.I., 2010. Prediction, cognition and the brain. Front. Hum. Neurosci. 4, 25. https://doi.org/10.3389/FNHUM.2010.00025/BIBTEX

Calcus, 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. https://doi.org/10.1016/J.CLINPH.2014.11.020

Callan, D.E., Jones, J.A., Munhall, K., Kroos, C., Callan, A.M., Vatikiotis-Bateson, E., 2004. Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. J. Cogn. Neurosci. 16, 805–816. https://doi.org/10.1162/089892904970771

Chennu, S., Noreika, V., Gueorguiev, D., Blenkmann, A., Kochen, S., Ibáñez, A., Owen, A.M., Bekinschtein, T.A., 2013. Expectation and Attention in Hierarchical Auditory Prediction. J. Neurosci. 33, 11194–11205. https://doi.org/10.1523/JNEUROSCI.0114-13.2013

Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 36, 181–204. https://doi.org/10.1017/S0140525X12000477

Correa-Jaraba, K.S., 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, 1–13. https://doi.org/10.3389/FNHUM.2016.00129

David, N., Linden, E.J., 2005. The P300: Where in the Brain Is It Produced and What Does It Tell Us? https://doi.org/10.1177/1073858405280524

Diaconescu, Andreea Oliviana, Alain, C., Mcintosh, A.R., Diaconescu, A O, 2011. The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. J Neurophysiol 106, 2896–2909. https://doi.org/10.1152/jn.00303.2011.-Perceptual

Escera, C., Alho, K., Schröger, E., Winkler, I., 2000. Involuntary Attention and Distractibility as Evaluated with Event-Related Brain Potentials. Audiol. Neurotol. 5, 151–166. https://doi.org/10.1159/000013877

Feng, T., Qiu, Y., Zhu, Y., Tong, S., 2008. Attention rivalry under irrelevant audiovisual stimulation.

Neurosci. Lett. 438, 6–9. https://doi.org/10.1016/J.NEULET.2008.04.049

Ficco, L., Mancuso, L., Manuello, J., Teneggi, A., Liloia, D., Duca, S., Costa, T., Kovacs, G.Z., Cauda, F., 2021. Disentangling predictive processing in the brain: a meta-analytic study in favour of a predictive network. Sci. Reports 2021 111 11, 1–14. https://doi.org/10.1038/s41598-021-95603-5

Fleming, J.T., Noyce, A.L., Shinn-Cunningham, B.G., 2020. Audio-visual spatial alignment improves integration in the presence of a competing audio-visual stimulus. Neuropsychologia 146, 107530. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2020.107530

Fontolan, L., Morillon, B., Liegeois-Chauvel, C., Giraud, A.L., 2014. The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nat. Commun. 2014 51 5, 1–10. https://doi.org/10.1038/ncomms5694

Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 2010 112 11, 127–138. https://doi.org/10.1038/nrn2787

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. B Biol. Sci. 360, 815–836. https://doi.org/10.1098/RSTB.2005.1622

Gramfort, A., Papadopoulo, T., Olivi, E., Clerc, M., 2010. OpenMEEG: Opensource software for quasistatic bioelectromagnetics. Biomed. Eng. Online 9, 1–20. https://doi.org/10.1186/1475-925X-9-45/TABLES/2

Halder, T., Talwar, S., Jaiswal, A.K., 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. https://doi.org/10.1523/eneuro.0170-19.2019

Justo-Guillén, E., Ricardo-Garcell, J., Rodríguez-Camacho, M., Rodríguez-Agudelo, Y., Lelo de Larrea-Mancera, E.S., Solís-Vivanco, R., 2019. Auditory mismatch detection, distraction, and

attentional reorientation (MMN-P3a-RON) in neurological and psychiatric disorders: A review. Int. J. Psychophysiol. 146, 85–100. https://doi.org/10.1016/J.IJPSYCHO.2019.09.010

Kok, P., Rahnev, D., Jehee, J.F.M., Lau, H.C., De Lange, F.P., 2012. Attention reverses the effect of prediction in silencing sensory signals. Cereb. Cortex 22, 2197–2206. https://doi.org/10.1093/CERCOR/BHR310

Leone, L.M., Mccourt, M.E., 2015. Dissociation of perception and action in audiovisual multisensory integration. Eur. J. Neurosci. 42, 2915. https://doi.org/10.1111/EJN.13087

Luca, M. Di, Machulla, T.-K., Ernst, M.O., 2009. Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. J. Vis. 9, 7–7. https://doi.org/10.1167/9.12.7

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. https://doi.org/10.1016/0301-0511(87)90001-9

Marucci, M., Di Flumeri, G., Borghini, G., Sciaraffa, N., Scandola, M., Pavone, E.F., 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. https://doi.org/10.1038/S41598-021-84196-8

Molholm, S., Martinez, A., Ritter, W., Javitt, D.C., Foxe, J.J., 2005. The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. Cereb. Cortex 15, 545–551. https://doi.org/10.1093/CERCOR/BHH155

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., Foxe, J.J., 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. https://doi.org/10.1016/S0926-6410(02)00066-6

Morlet, D., Fischer, C., 2014. MMN and Novelty P3 in Coma and Other Altered States of Consciousness: A Review. Brain Topogr. 27, 467. https://doi.org/10.1007/S10548-013-0335-5

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. https://doi.org/10.1016/J.CLINPH.2007.04.026

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011. https://doi.org/10.1155/2011/156869

Pascual-Marqui, R.D., 2007. Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization.

Polich, J., 2007a. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118, 2128–2148. https://doi.org/10.1016/J.CLINPH.2007.04.019

Polich, J., 2007b. Updating P300: An Integrative Theory of P3a and P3b. Clin. Neurophysiol. 118, 2128. https://doi.org/10.1016/J.CLINPH.2007.04.019

Rohrbaugh, J.W., Donchin, E., Eriksen, C.W., 1974. Decision making and the P300 component of the cortical evoked response. Percept. Psychophys. 1974 152 15, 368–374. https://doi.org/10.3758/BF03213960

Schröger, E., 1997. On the detection of auditory deviations: A pre-attentive activation model. Psychophysiology 34, 245–257. https://doi.org/10.1111/J.1469-8986.1997.TB02395.X

Shiramatsu, T.I., 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. https://doi.org/10.3389/FNHUM.2021.721476/BIBTEX

Sittiprapaporn, W., 2012. An effect of attention on mismatch negativity in audiovisual visual modalities. Pakistan J. Biol. Sci. 15, 542–546. https://doi.org/10.3923/pjbs.2012.542.546

Stefanics, G., Stavrinou, M., Sestieri, C., Ciancetta, L., Belardinelli, P., Cianflone, F., Bernáth, L., Hernádi, I., Pizzella, V., Romani, G.L., 2005. Cross-modal visual-auditory-somatosensory integration in a multimodal object recognition task in humans. Int. Congr. Ser. 1278, 163–166. https://doi.org/10.1016/J.ICS.2004.11.074

Stekelenburg, J.J., Vroomen, J., 2007. Neural Correlates of Multisensory Integration of Ecologically Valid Audiovisual Events. J. Cogn. Neurosci. 19, 1964–1973. https://doi.org/10.1162/JOCN.2007.19.12.1964

Talsma, D., 2015. Predictive coding and multisensory integration: an attentional account of the multisensory mind. Front. Integr. Neurosci. 9, 19. https://doi.org/10.3389/FNINT.2015.00019

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. https://doi.org/10.3389/FNHUM.2019.00058/BIBTEX

van Wassenhove, V., Grant, K.W., Poeppel, D., 2005. Visual speech speeds up the neural processing of auditory speech. PNAS 102, 1181–1186.

Van Wassenhove, V., Grant, K.W., Poeppel, D., 2005. Visual speech speeds up the neural processing of auditory speech. Proc. Natl. Acad. Sci. 102, 1181–1186. https://doi.org/10.1073/PNAS.0408949102

Wacongne, C., Changeux, J.P., Dehaene, S., 2012. A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. J. Neurosci. 32, 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Winkler, I., Czigler, I., 2012. Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. Int. J. Psychophysiol. 83, 132–143. https://doi.org/10.1016/J.IJPSYCHO.2011.10.001

## Legends, tables and figures

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

Table 2. Table shows the mean±SD of the peak P300 amplitudes and their corresponding latencies across all the trials for audio only, visual only and audio-visual conditions.

Table 3. Table shows the mean±SD of the peak MMN amplitudes and their corresponding latencies across all the trials for audio only, visual only and audio-visual conditions.

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

Figure 1. Grand averaged ERPs plotted across 21 subjects showing the differences between oddball and standard trials (oddball-standard) for (a) Audio, (b) Visual and (c) AV blocks. The corresponding topoplots indicate the distribution of ERPs across the brain at the identified MMN and P300 peaks. The topoplot on the top right displays the color code assigned to respective scalp channel locations used for plotting the ERPs in (a), (b) and (c).

Figure 2. Grand averaged ERPs computed from the temporal projections on the first principal component computed from the top 5 sensor time-series exhibiting highest MMN peak amplitude identified between 120-250 ms (left panels, a-c) and P300 peak amplitude identified between 250-500 ms (right panels, d-f). The shaded gray region (100 ms windows) represents a regime of significant difference (revealed by t-test) between standard and oddball conditions. The shadings with each of the oddball and standard trials reflect the standard error of mean.

Figure 3. The violin plots represent the MMN (a) peak amplitudes (maximum magnitude on the negative y axis of the subtraction waveform of oddball – standard trials of each subject) and their corresponding (c) latencies between 120-250 ms. The colored dots represent each subject's peak amplitude and peak latency value in μV and ms, respectively. The white dot at the centre of the gray box represents the median of the data and the gray box itself represents the inter-quartile range. Similarly, the P300 (b) peak amplitudes (between 250-500ms) and their corresponding (d) latencies are plotted for each subject.

Figure 4. eLORETA source localization results using time locked analysis (at threshold level 95%) representing the underlying (a) MMN and (b) P300 sources for Audio, Visual and Audio-Visual modalities.

Figure 5. eLORETA source localization results using time locked analysis (at threshold level 95%) representing the underlying (a) MMN and (b) P300 sources for Audio, Visual and Audio-Visual, Cross-visual and Cross-audio modalities.

**Table 1**

| Condition | **Standard / Frequent / Repetitive** | **Deviant / Oddball / Infrequent / Non-repetitive** |
|---|---|---|
| *Audio only* | 261.6 Hz | 523.3 Hz |
| *Visual only* | Blue Square | Red Triangle |
| *Audio-Visual* | 261.6 Hz and Blue Square | 523.3 Hz and Red Triangle |
| *Cross-Audio* | Blue Square | 523.3 Hz |
| *Cross-Visual* | 261.6 Hz | Red Triangle |

**Table 2**

| *Condition (P300)* | *Amplitude(μV)* | *Latency(ms)* |
|---|---|---|
| *Audio* | 16.49±10.42 | 327.41±30.99 |
| *Visual* | 14.57±8.18 | 399.93±32.04 |
| *AV* | 16.14±9.40 | 332.00±30.04 |

**Table 3**

| *Condition (MMN)* | *Amplitude(μV)* | *Latency(ms)* |
|---|---|---|
| *Audio* | 20.87±11.97 | 206.51±30.81 |
| *Visual* | 15.53±8.75 | 160.11±30.66 |
| *AV* | 18.73±10.41 | 150.13±30.69 |

**Table 4**

| Condition | MMN | P300 |
|---|---|---|
| **Audio** | SFG, Left Superior Frontal Gyrus A6dl, dorsolateral area 6<br>SFG, Right Superior Frontal Gyrus A8dl, dorsolateral area 8 | SFG, Left Superior Frontal Gyrus A6dl, dorsolateral area 6PrG<br>Left Precentral Gyrus A4t, area 4 (trunk region)<br>SFG, Right Superior Frontal Gyrus A6dl, dorsolateral area 6 |
| **Visual** | MVOcC, Left Medio-Ventral Occipital Cortex cCunG, caudal cuneus gyrus<br>LOcC, Right lateral Occipital Cortex mOccG, middle occipital gyrus<br>IPL, Right Inferior Parietal Lobule A39rd, rostro-dorsal area 39 (Hip3)<br>MTG, Right Middle Temporal Gyrus A37dl, dorsolateral area37 | SFG, Left Superior Frontal Gyrus A6dl, dorsolateral area 6<br>PrG, Left Precentral Gyrus A6cdl, caudal dorsolateral area 6<br>MFG, Right Middle Frontal Gyrus A8vl, ventrolateral area 8<br>MFG, Right Middle Frontal Gyrus A6vl, ventrolateral area 6 |
| **AV** | STG, Right Superior Temporal Gyrus A22c, caudal area 22<br>MTG, Right Middle Temporal Gyrus A21c, caudal area 21<br>SPL, Left Superior Parietal Lobule A7r, rostral area 7 | SFG, Left Superior Frontal Gyrus A6dl, dorsolateral area 6<br>SFG, Right Superior Frontal Gyrus A6dl, dorsolateral area 6 |

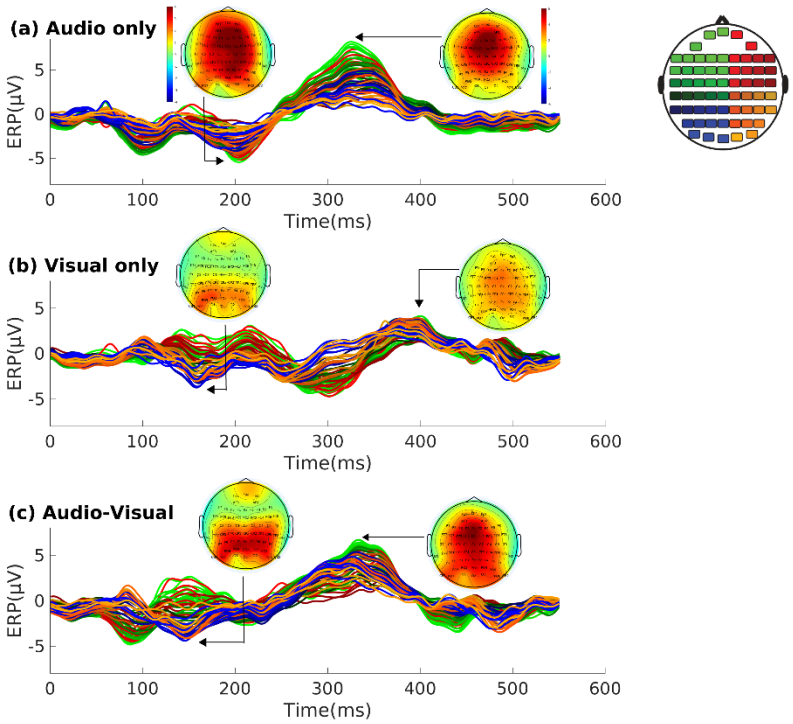


**Figure1**

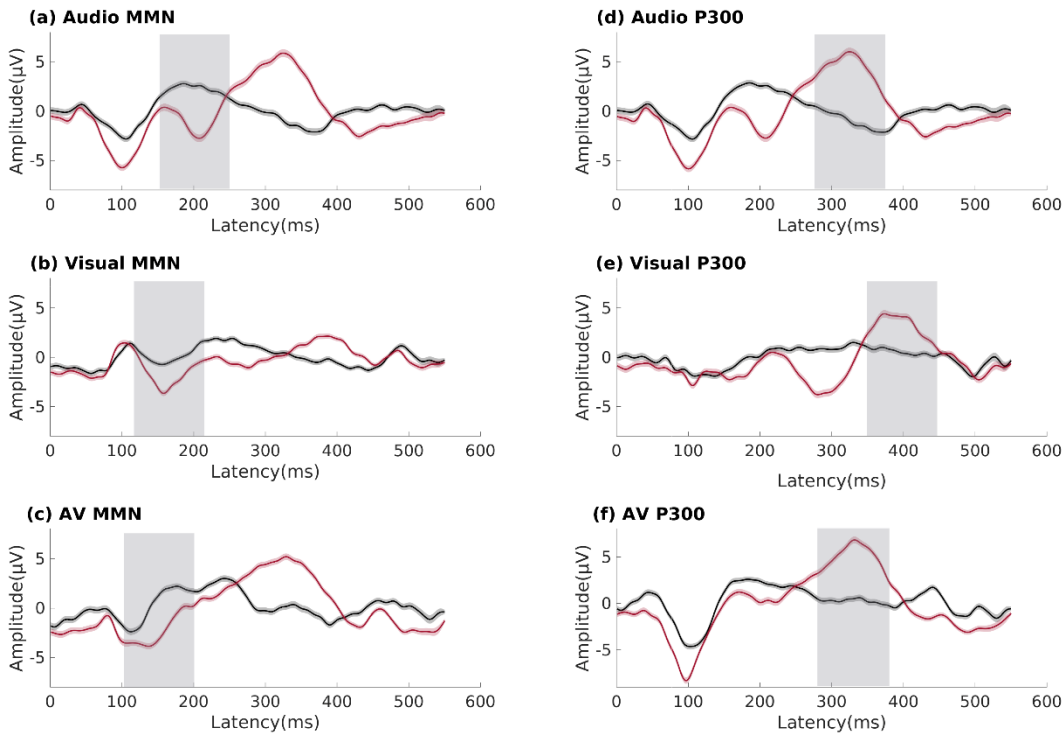


**Figure 2**

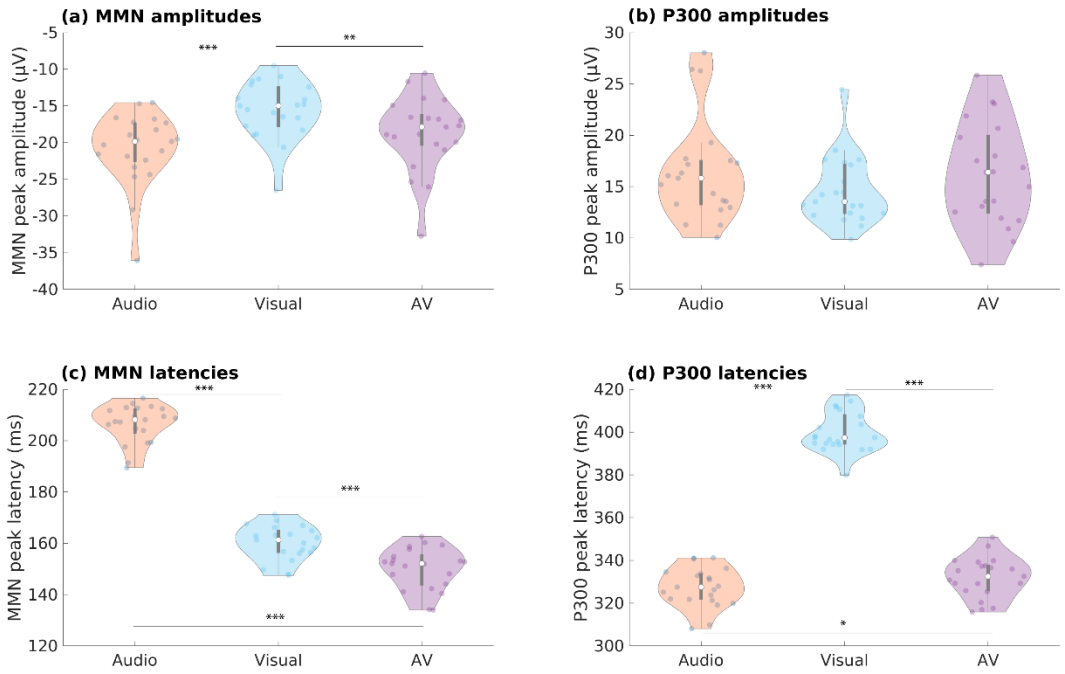


**Figure 3**

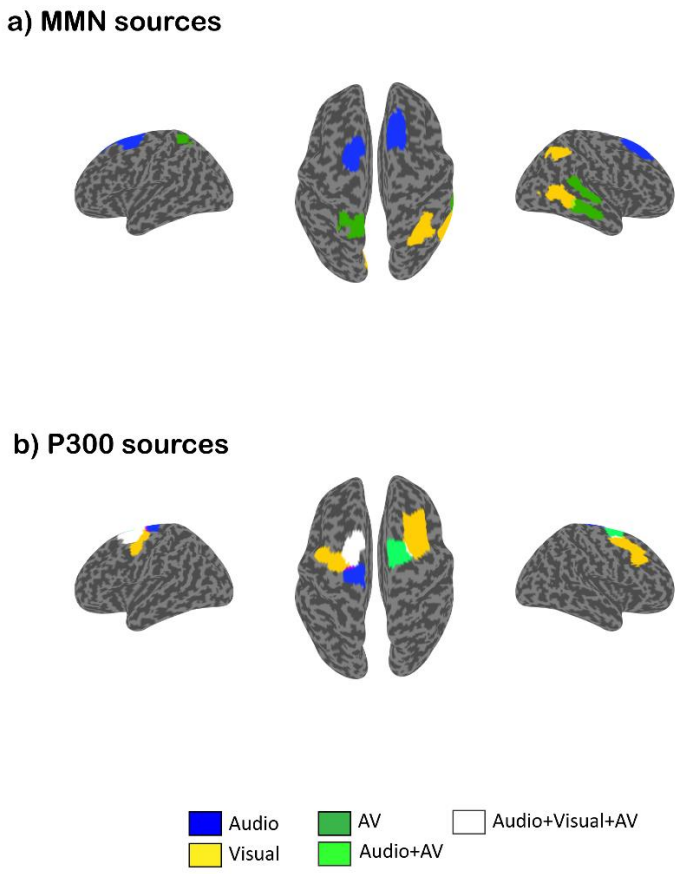


**Figure 4**

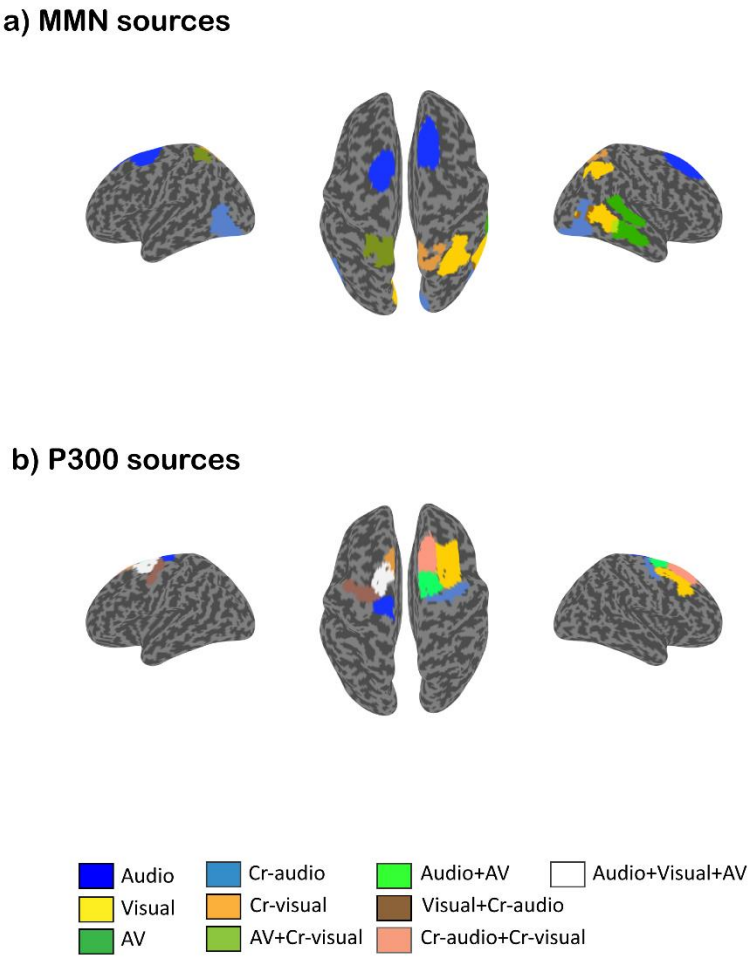
a) MMN sources
b) P300 sources
Audio
Visual
AV
Cr-audio
Cr-visual
AV+Cr-visual
Audio+AV
Visual+Cr-audio
Cr-audio+Cr-visual
Audio+Visual+AV

**Figure 5**