Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Sensory Noise Traits Predict Illusion Susceptibility and Multisensory Mismatch Negativity

Hexin Xu
The University of Tokyo
Amit Yaron
International Research Center for Neurointelligence (WPI-IRCN), UTIAS, The University of Tokyo
Yiyuan Teresa Huang
International Research Center for Neurointelligence (WPI-IRCN), UTIAS, The University of Tokyo
Zenas C. Chao
International Research Center for Neurointelligence (WPI-IRCN), UTIAS, The University of Tokyo
Tomoyo Isoguchi Shiramatsu
The University of Tokyo
Hirokazu Takahashi
takahashi@i.u-tokyo.ac.jp

The University of Tokyo

Article

Keywords: Multisensory integration, Audiovisual illusions, Bayesian causal inference, Mismatch negativity, Predictive coding, Individual differences

Posted Date: July 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10121253/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Individual differences in audiovisual illusion susceptibility are well documented, yet their computational origins remain unclear. Standard audiovisual illusion paradigms typically measure susceptibility in a single direction, confounding sensory precision and causal inference within one behavioural index. To dissociate these components, we developed a bidirectional audiovisual rabbit task in which auditory and visual signals could capture each other's percept across multiple sound levels. This design enabled fitting a Bayesian causal inference (BCI) model to individual behaviour, yielding separable latent traits: auditory noise, visual noise, sound-level sensitivity of auditory noise, and common-cause prior. Behavioral analyses showed that sensory-noise traits and common-cause prior were associated with illusion measures, supporting the validity of the computational decomposition. We then tested whether these traits predicted EEG markers of multisensory mismatch processing in a passive audiovisual oddball paradigm. Higher auditory noise was associated with less negative occipital mismatch negativity in the congruent audiovisual condition, indicating weaker multisensory mismatch responses. This association was specific to the occipital audiovisual component and was not observed for unimodal MMN responses. Together, these findings show that behavior-derived sensory-noise traits capture meaningful individual differences in multisensory perception and predict neural signatures of multisensory mismatch processing.

1. Introduction

Human perception emerges through the integration of signals from multiple sensory channels, allowing the brain to form coherent and reliable representations of the environment. Such multisensory integration enhances perceptual accuracy, accelerates learning, and provides robustness when one sensory modality is noisy or unavailable [1–4]. A large body of psychophysical work has shown that multisensory combination often approaches statistical optimality, with each cue weighted in proportion to its moment-to-moment reliability [5–7]. This reliability-based weighting enables the perceptual system to re-balance sensory inputs according to their signal-to-noise ratios [8].

Beyond reliability weighting, **Bayesian causal-inference (BCI) frameworks** propose that the brain infers whether signals from different modalities arise from a common or distinct cause and integrates them only when a common cause is likely [9–13]. These probabilistic accounts explain why perception sometimes fuses incongruent signals and sometimes segregates them. At the neural-computational level, **predictive-coding theories** provide a complementary account in which perception depends on the minimisation of prediction errors—differences between expected and actual sensory input—through hierarchical message passing [14, 15]. Recent work further suggests that crossmodal predictions can themselves be organized hierarchically across distributed brain networks, providing a concrete account of how unimodal predictions are integrated into crossmodal knowledge [16]. Within this predictive architecture, the *precision* (inverse variance) assigned to prediction errors governs the relative influence of each modality. Thus, sensory dominance—the tendency for one sense to outweigh another—can partly

be viewed as the behavioural manifestation of differences in precision weighting across sensory hierarchies.

Within a BCI framework, perceptual reports can be decomposed into separable computational components, including modality-specific sensory noise and an observer's prior tendency to treat audiovisual signals as arising from a common cause. This is important for individual-differences research because similar overt illusion rates may arise from different latent combinations of sensory precision and binding tendency. Bayesian causal-inference models therefore provide a principled way to move from descriptive illusion susceptibility to interpretable computational traits that can be related to neural measures.

Audiovisual illusions as behavioural windows on multisensory inference

Audiovisual (AV) illusions have proven particularly incisive for revealing the dynamics of multisensory weighting. In the **sound-induced flash illusion (SIFI)**, two brief beeps accompanying a single flash often elicit the perception of two flashes [17, 18], demonstrating auditory capture of visual numerosity. In contrast, the **McGurk effect** shows that incongruent lip movements can alter the perceived speech sound [19], revealing visual dominance in speech perception. These and related phenomena highlight that dominance is context-dependent: the same observer may exhibit auditory bias in temporal tasks but visual bias in spatial ones [5, 6, 20, 21].

More recently, [22] introduced the *audiovisual rabbit illusions*, in which sequences of beeps and flashes create illusory intermediate flashes or suppress real ones. These postdictive illusions demonstrate that multisensory integration operates dynamically and retrospectively, and recent causal-inference accounts suggest that such rabbit-like phenomena reflect flexible multisensory binding rather than simple obligatory fusion [23]. Crucially, not all observers experience the illusion to the same degree, implying substantial **inter-individual variability** in multisensory inference.

Unlike many audiovisual illusion paradigms that primarily probe dominance in one direction, the bidirectional rabbit paradigm allows auditory and visual capture to be assessed within a common task structure. In the present study, we manipulated auditory level to alter the reliability of the auditory stream, extending earlier behavioural work on this bidirectional audiovisual rabbit variant [24](Fig. 1A), and allowing us to test whether changes in auditory precision shift the strength or direction of the illusion across individuals. This makes the bidirectional audiovisual rabbit paradigm especially suitable for asking whether inter-individual variability is better explained by latent sensory noise and binding tendency than by descriptive illusion rates alone.

Individual differences in multisensory binding

Recent research has established that susceptibility to AV illusions is not merely random noise but can reflect stable, trait-like features of sensory processing. Individuals with a wider temporal binding window are more susceptible to both SIFI and McGurk effects [25, 26], and differences in reliability reweighting

have been linked to age, cognitive capacity, and clinical profiles [27, 28]. These findings suggest that individual variations in sensory dominance arise from enduring differences in neural precision, temporal alignment, or prior expectations. A BCI model can quantify these differences as participant-specific latent traits, such as auditory noise, visual noise, and common-cause prior, providing a mechanistic account of why perceptual behaviour varies across individuals and a computational bridge to neural measures.

Mismatch negativity as a neural index of prediction error

A promising neural correlate of these individual differences is the **mismatch negativity (MMN)** component of event-related potentials. MMN—first described by [29]—is reliably evoked by violations of regularity and is interpreted as a precision-weighted prediction-error signal [30–34]. More broadly, this view is consistent with evidence that auditory cortical neurons are sensitive to complex statistical regularities over extended timescales, supporting the idea that deviance responses are grounded in learned regularity structure rather than in simple local change alone [35]. While originally described in audition, homologous **visual MMN (vMMN)** has been documented and interpreted within the same predictive-coding architecture [36]. Larger MMN responses are commonly taken to index stronger neural sensitivity to violations of regularity, consistent with more precise internal predictions. If sensory noise shapes the precision of prediction-error signals, then BCI-derived sensory-noise parameters may predict individual variability in MMN amplitudes, particularly under multisensory conditions.

Our approach builds on evidence that audiovisual cortical responses exhibit early nonlinear interactions rather than simple linear summation [37–39], and that these interactions depend on signal strength according to inverse effectiveness (multisensory gain is greatest when each unimodal stimulus is weak on its own) [40–42]. Against this background, the present study asked whether individual differences in the bidirectional audiovisual rabbit illusion can be decomposed into latent computational traits and whether those traits predict neural markers of multisensory prediction-error processing. To address this question, we fit a BCI model to each participant's perceptual reports to estimate auditory noise, visual noise, sound-level sensitivity of auditory noise, and common-cause prior. We then tested whether these traits tracked individual differences in illusion susceptibility and predicted MMN amplitudes in independent passive unimodal and audiovisual oddball conditions.

2. Materials and Methods

2.1. Participants

Prior to participation, all individuals received detailed explanations and instructions regarding the behavioural and EEG experiments. Written informed consent was obtained from all participants, who were assured that their personal information would be handled confidentially. Participants were informed of their right to withdraw at any time without penalty. The experimental protocols were approved by the Ethics Committee at the Graduate School of Information Science and Technology, the

University of Tokyo (Permit Number: UT-IST-RE-230111, UT-IST-RE-220520-2). All experiments were performed in accordance with relevant institutional guidelines and regulations and with the Declaration of Helsinki.

The behavioural audiovisual rabbit experiment was completed by 28 healthy adults (20 males and 8 females). Of these, 15 participants (12 males and 3 females) also completed the passive EEG oddball experiment. For the behavioural BCI analyses, 26 participants (18 males and 8 females) were retained. Two participants were excluded from the BCI-based behavioural analyses: one because sensitivity analyses identified an atypical response profile that exerted a disproportionately strong influence on the model fit, and one because only a partial behavioural dataset was available. All participants had normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. Participant counts are therefore reported separately for raw behavioural summaries (N = 28), behavioural BCI correlations (N = 26), and EEG/MMN analyses (N = 15).

During the experiment, participants were seated 60 cm from a monitor. All experiments were conducted in a dimly lit soundproof room. Participants were encouraged to rest as needed during the experiment.

2.2. Bi-directional Audiovisual Rabbit Illusion Experiment

Visual and auditory stimuli were presented using MATLAB (The MathWorks, Natick, MA, USA) and Psychtoolbox-3 [43, 44]. Stimulus timing followed the original audiovisual rabbit illusion paradigm [22]. Auditory stimuli consisted of sequences of two or three pure tones (or beeps; 800 Hz, 10 ms duration) presented with headphones. In three-beep sequences, the onset interval between the beeps was 68 ms, but in two-beep sequences, the middle beep was omitted, resulting in a 136 ms onset interval. The amplitude of the beeps was individualized, starting from each participant's beep hearing level (BHL, level 0), which was determined in a pilot hearing test, and increasing in 5-dB steps up to 40 dB above BHL (level 40). In half of the trials, white noise was presented from 1000 ms before the onset of the first beep until 1000 ms after the final beep. The noise amplitude was also individualized based on the hearing test. This white noise was introduced to reduce beep detectability and increase stimulus ambiguity.

Visual stimuli were displayed on a monitor with a 60 Hz refresh rate. The stimuli consisted of rectangular flashes presented 10° of visual angle below a central fixation cross. Each flash subtended 0.28° × 1.2° and was presented for 15 ms on a dark gray background. The background luminance was set to 30% of the monitor's maximum brightness, and the flashes were presented at 80% of the maximum brightness. For two-flash sequences, flashes appeared at the leftmost and the rightmost positions, separated by 2.84° and by an onset interval of 134 ms. For three-flash sequences, flashes appeared at the left, middle, and right positions, separated by 1.42° and by an onset interval of 67 ms. A 23-ms delay between flashes and beeps followed the timing used in the original audiovisual rabbit study, in which no major perceptual difference attributable to this delay was reported (Fig. 1B, 1C).

The task comprised eight conditions crossing beep number (2 or 3), flash number (2 or 3), and noise presence (absent or present): 2b2f0n (two beeps two flashes no noise), 2b2f1n, 3b3f0n, 3b3f1n, 2b3f0n,

2b3f1n, 3b2f0n, and 3b2f1n. Congruent conditions (2b2f0n, 2b2f1n, 3b3f0n and 3b3f1n) included five trials per volume level, whereas incongruent conditions (2b3f0n, 2b3f1n, 3b2f0n and 3b2f1n) included ten trials per volume level.

Participants were instructed to maintain fixation on the central cross throughout the task. On each trial, one of the eight conditions was randomly presented, and participants reported the number of beeps and flashes they perceived, either 2 or 3, using the number pad on a keyboard. Two rest breaks of approximately 10 min were offered during the experiment.

2.3. EEG Experimental Design

Visual and auditory stimuli were presented using MATLAB and Psychtoolbox-3, with stimulus presentation synchronized to EEG recordings via a parallel port. Visual stimuli were presented on a monitor with a 60-Hz refresh rate. They were identical in size and brightness to the flashes used in the audiovisual rabbit illusion task but differed in orientation: flashes were tilted ± 18° from vertical, measured 0.28° in width and 1.2° in height, and were presented on a dark-gray background at 80% of maximum brightness. The background luminance was set to 30% of maximum brightness (Fig. 1D). Each flash lasted 100 ms and was presented every 500 ms, leaving a 400-ms interval between stimulus offset and the next one. A light-blue fixation dot was displayed at the center of the screen, and participants were instructed to maintain gaze on the dot throughout the experiment.

Auditory stimuli consisted of pure tones of 800 Hz or 1600 Hz, presented through two speakers positioned 70 cm apart in the soundproof room. Each tone lasted 100 ms and was presented every 500 ms, aligned to the visual stimulus timing.

The EEG session included several paradigms; the present study analyzed the passive oddball paradigms relevant to the current question. The main analyses focused on three conditions: unimodal auditory oddball, unimodal visual oddball, and congruent audiovisual oddball (Fig. 1E). Each oddball condition comprised 400 trials. Each trial contained either a standard or a deviant event. Additional paradigms, including other oddball variants acquired for separate objectives, were not analyzed here and will be reported elsewhere.

Unimodal visual oddball: ±18° flashes served as standard or deviant stimuli, with two counterbalanced assignments: in one, + 18° flashes were standards and − 18° flashes were deviant; in the other, this assignment was reversed (Fig. 1E (i)).

Unimodal auditory oddball: 800-Hz and 1600-Hz tones served as standard or deviant stimuli, with two counterbalanced assignments: in one, 800-Hz tones were standards and 1600-Hz tones were deviants; in the other, this assignment was reversed (Fig. 1E (ii)).

Congruent audiovisual oddball

auditory and visual stimuli were simultaneously standard or deviant. For example, in one assignment, 800-Hz tones paired with + 18° flashes served as standards, and 1600-Hz tones paired with − 18° flashes served as deviants; in the other, this assignment was reversed (Fig. 1E (iii)).

In all analyzed oddball conditions, deviants accounted for 10% of trials and standards for 90%.

2.4. EEG Data Acquisition

EEG data were recorded using a 64-channel cap arranged according to the international 10–20 system (Standard 64Ch-actiCAP-Slim, Easycap GmbH, Germany) with Ag/AgCl actiCAP-Slim electrodes and BrainAmp EEG amplifier. Electrode impedances were kept below 5 kΩ. Signals were recorded at 1000 Hz with an online bandpass filter of 0.1–100 Hz. FCz was used as the online reference, and the ground electrode was placed on the forehead.

2.5. Bayesian Causal Inference Model

Perceptual reports from the bidirectional audiovisual rabbit task were modeled within a Bayesian causal inference (BCI) framework. We compared four conceptual model families: **segregation**, **forced fusion**, **classical BCI**, and **postdictive BCI**. In all families, participants were assumed to form latent auditory and visual count estimates under sensory uncertainty and then generate overt beep and flash reports from those latent estimates. The segregation model treated auditory and visual signals as separate, the forced-fusion model assumed obligatory integration, the classical BCI model assumed a stable prior probability that the two modalities arose from a common cause, and the postdictive BCI model allowed this common-cause probability to vary with stimulus context. The postdictive BCI model therefore extended the classical BCI model by allowing causal inference to depend on the stimulus sequence, consistent with the postdictive nature of the audiovisual rabbit illusion. For the two BCI families, overt reports were generated using a probability-matching readout, in which the probability of reporting each count was proportional to its posterior probability [45].

The participant-level parameters of primary interest were auditory noise (σ_a), visual noise (σ_v), sound-level sensitivity of auditory noise ($\sigma_{a,level}$), and the prior probability of a common cause (p_{common}). Auditory and visual noise captured each participant's baseline sensory uncertainty. To account for the sound-level manipulation, auditory noise was allowed to vary systematically with sound level, such that larger values of $\sigma_{a,level}$ indicated a stronger volume-dependent reduction in auditory noise. Thus, $\sigma_{a,level}$ did not represent higher auditory noise per se, but the degree to which auditory noise decreased as sound level increased. In the BCI families, p_{common} captured each participant's baseline tendency to treat auditory and visual signals as originating from a common source.

Models were fit hierarchically to aggregated perceptual report counts using custom Python code implemented in PyMC [46]. Data were aggregated at the participant $\times$ sound level $\times$ stimulus-condition level. Because reports were binary for each modality, either 2 or 3, separate binomial outcomes were used for beep and flash reports. For each row, the model predicted the probability of reporting three beeps and the probability of reporting three flashes, and these predictions were linked to the observed

counts. Global criterion and lapse parameters were included to capture report bias and occasional random responses. Model comparison was based on Pareto-smoothed importance-sampling leave-one-out cross-validation (PSIS-LOO [47] and stacking weights [48].

Posterior means of the participant-level parameters were extracted for the individual-differences analyses. For correlation analyses, σ_a and σ_v were analyzed on log transformation and z-scoring, whereas p_{common} was analyzed after logit transformation and z-scoring. Sound-level sensitivity of auditory noise was analyzed using its z-scored latent scale parameter ($\sigma_{a,level,lat}$), with higher values indicating stronger sound-level-dependent decreases in auditory noise. Posterior predictive checks were used to assess whether the selected model reproduced the main level-dependent structure of the behavioural data.

2.6. Data Analysis

Audiovisual rabbit illusion test results. Mean reported beep and flash numbers were calculated for each sound level and stimulus condition. Illusion intensities were calculated from the two noisy incongruent conditions, 2b3f1n and 3b2f1n. Auditory illusion intensity quantified the extent to which beep reports were biased toward the visually presented flash number. Specifically, in the 2b3f1n condition, the auditory error term was defined as the mean reported beep number minus 2, whereas in the 3b2f1n condition, it was defined as 3 minus the mean reported beep number. Auditory illusion intensity was defined as the mean of these two auditory error terms. Visual illusion intensity was defined analogously: in the 3b2f1n condition the visual error term was the mean reported flash number minus 2, whereas in the 2b3f1n condition, it was defined as 3 minus the mean reported flash number. Total illusion intensity was defined as the sum of auditory and visual illusion intensities.

EEG data. Continuous EEG data were bandpass filtered from 0.1 to 30 Hz and downsampled to 250 Hz. For artifact correction, ICA weights were estimated separately for each participant from data pooled across available conditions. For ICA estimation only, the pooled data were high-pass filtered at 1 Hz. ICA was performed using the Picard algorithm, and components classified by ICLabel as eye or muscle artifacts with probabilities ≥ 0.8 were removed. Epochs from − 100 to 350 ms relative to stimulus onset were extracted, and epochs exceeding ± 150 μV at any channel were rejected. Standard and deviant ERPs were averaged separately, transformed to current source density (CSD), baseline-corrected from − 100 to 0 ms, and subtracted to obtain MMN difference waves.

Current source density (CSD). To eliminate the influence of the EEG reference electrode and improve spatial specificity, ERPs were transformed to CSD using a spherical spline algorithm [49] as implemented in MNE-Python (regularization parameter $\lambda^2 = 10^{-5}$, spline stiffness m = 4, head radius = 10 cm)[50]. CSD values are reported in $\mu V/cm^2$.

MMN computation. For each participant and condition, the ERPs were computed by averaging epochs separately for standard and deviant trials. CSD transformation was applied to the averaged ERP, followed by baseline correction from − 100 to 0 ms. MMN difference waves were computed as deviant ERPs

minus standard ERPs. The resulting ROI-based MMN measures are referred to as frontocentral auditory MMN (A_{FC}), occipital visual MMN (V_{OC}), frontocentral audiovisual MMN (AV_{FC}), and occipital audiovisual MMN (AV_{OC}). To determine whether a reliable MMN response was present at the group level, mean MMN amplitudes within each predefined ROI and time window were tested against zero using one-sample t-tests. Significant negative deviations from zero were interpreted as evidence of a canonical MMN response.

Regions of interest (ROI) and time windows. Based on prior MMN literature [30, 36], two ROIs were defined: a frontocentral ROI (Fz, FCz, Cz, F3, F4, FC1, FC2) for auditory MMN responses, and an occipital ROI (Oz, O1, O2, POz, PO3, PO4, PO7, PO8) for visual MMN responses. Time windows were 100–250 ms for auditory conditions and 150–300 ms for visual conditions. For the congruent audiovisual condition, the frontocentral ROI was analyzed using the auditory time window whereas the occipital ROI was analyzed using the visual time window. For each participant, MMN amplitude was defined as the mean CSD amplitude across channels and time points within the corresponding ROI and window.

Individual-differences analysis. For behavioural correlations between BCI-derived traits and illusion measures in the larger behavioural BCI sample (N = 26), Pearson correlations were used as the primary analysis because the variables of interest were continuous and the main question concerned linear associations between latent traits and behavioural measures. Spearman rank correlations were additionally computed as a secondary robustness check.

To address multiple comparisons in the primary behavioural analysis, BH-FDR correction was applied within a pre-specified family of four tests: σ_a × auditory illusion intensity, σ_v × visual illusion intensity, $\sigma_{a,level}$ × auditory illusion intensity, and p_{common} × total illusion intensity.

For EEG/MMN correlations in the EEG subsample (N = 15), Spearman rank correlations were used as the primary test because of the modest sample size, with Pearson correlations additionally reported as secondary support for the principal σ_a – AV_{OC} association. Correlations were computed between transformed BCI-derived parameters and MMN amplitudes across participants. To assess reference-scheme robustness, EEG/MMN correlations were computed under four reference schemes: original FCz reference, average reference, mastoid reference (TP9/TP10 average), and CSD. Statistical significance was assessed at $\alpha = 0.05$ (two-tailed). To address multiple comparisons in the EEG/MMN analyses, BH-FDR correction was applied within two pre-specified families: (i) AV MMN correlations (4 tests: σ_a and σ_v × AV_{OC} and AV_{FC}) and (ii) unimodal MMN correlations (4 tests: σ_a and σ_v × A_{FC} and V_{OC}, Supplementary Fig. S1).

Nonparametric permutation tests (10,000 iterations) were used to assess the robustness of the primary σ_a–AV_{OC} finding. For descriptive visualization of the primary σ_a–MMN association, participants in the EEG sample were divided into high- and low- σ_a groups using a median split, and grand-average standard, deviant, and difference waveforms were plotted for the AV_{OC} and AV_{FC} ROIs in supplementary. This grouped analysis was used for visualization only and was not treated as a primary inferential test.

3. Results

3.1. Perceptual Report on the Audiovisual Rabbit Illusion

The AV rabbit effect was examined using congruent (2b2f, 3b3f) and incongruent (2b3f, 3b2f) beep–flash combinations (Fig. 2A). In the congruent no-noise conditions, reports closely matched the physical stimulus counts: participants reported 2.01 beeps and 2.04 flashes in the 2b2f0n condition (Fig. 2A(i)) and 2.98 beeps and 2.95 flashes in 3b3f0n (Fig. 2A(iv)). In the incongruent no-noise conditions, flash reports were shifted toward the beep count: In 2b3f0n (Fig. 2A(ii)), beep reports remained close to the physical count (2.03), whereas flash reports shifted from three toward two (mean = 2.37); Conversely, in 3b2f0n (Fig. 2A(iii)), beep reports were close to the physical count (2.96), whereas flash reports shifted from two toward three (mean = 2.57).

Wilcoxon signed-rank tests confirmed that flash reports in the incongruent conditions were significantly biased toward the beep count, and all effects remained significant after BH-FDR correction. Under noisy low-volume conditions (0–10 dB above BHL), auditory reports became less veridical, consistent with reduced auditory detectability: mean beep reports were 2.34 in 2b3f1n and 2.66 in 3b2f1n (Table 1). Flash reports also remained shifted away from the physical flash count in the incongruent conditions, indicating that cross-modal ambiguity persisted under noise.

Table 1
Mean numbers for low volume level (0–10 dB above BHL) of AV rabbit test

Conditions	Beep	Flash	Conditions	Beep	Flash
2b3f0n	2.04	2.38	3b2f0n	2.95	2.56
2b3f1n	2.34	2.56	3b2f1n	2.66	2.35

Using the illusion intensity metrics, we further characterized individual sensory-dominance patterns. Illusion strength showed substantial inter-individual variability, with auditory illusion intensity ranging from 0.04 to 0.72, visual illusion intensity ranging from 0.01 to 0.80, and total illusion intensity ranging from 0.19 to 1.01 (Fig. 2B). These ranges indicate marked heterogeneity in the relative balance of auditory- versus visual-dominant capture across participants.

3.2. BCI Model Parameters

We next asked which computational family best accounted for the behavioural reports. Four conceptual families were compared: segregation, forced fusion, classical Bayesian causal inference (BCI), and postdictive BCI (Fig. 3A; Supplementary Table S1). Segregation assumes that auditory and visual counts are estimated independently, forced fusion assumes obligatory integration, classical BCI estimates a stable participant-level common-cause prior, and postdictive BCI allows this common-cause probability to vary with mismatch direction.

Based on PSIS-LOO model comparison in the larger behavioural BCI sample (N = 26), postdictive BCI provided the strongest predictive account of the behavioural data (stacking weight = 0.808, ELPD-LOO = -4201.18; Fig. 3C). It outperformed classical BCI (ΔELPD = 26.07) and greatly exceeded segregation and forced fusion (ΔELPD = 1447.99 and 2092.38, respectively).

Because postdictive BCI showed the best out-of-sample predictive performance, its participant-level posterior means were taken forward as computational traits for the individual-differences analyses. These traits comprised auditory noise (σ_a), visual noise (σ_v), sound-level sensitivity of auditory noise (σ_a,level), and common-cause prior (p_{common}). Posterior predictive checks showed that the selected model reproduced the main level-dependent structure of both beep and flash reports across conditions (Fig. 3D, E; Supplementary Fig. S2), supporting its use as the computational basis for the subsequent behavioural and EEG correlations. Full sampling and fit diagnostics are provided in Supplementary Table S2.

3.3. Behavioural Correlates of Model-Derived Traits

To examine how individually estimated BCI parameters related to illusion behaviour, we correlated model parameters with illusion measures across the larger behavioural BCI sample (N = 26), using Pearson correlations as the primary analysis. Auditory noise (σ_a) was positively associated with auditory illusion intensity ($r = +0.780$, $p < 0.0001$, BH-FDR $q < 0.0001$; Fig. 4A), and visual noise (σ_v) was positively associated with visual illusion intensity ($r = +0.621$, $p = 0.0007$, $q = 0.0009$; Fig. 4B). In addition, sound-level sensitivity of auditory noise ($\sigma_{a,level}$) was negatively associated with auditory illusion intensity ($r = -0.594$, $p = 0.0014$, $q = 0.0014$; Fig. 4C), and a stronger common-cause prior (p_{common}) was associated with greater total illusion intensity ($r = +0.885$, $p < 0.0001$, $q < 0.0001$; Fig. 4D).

Spearman rank correlations, computed as a secondary robustness check, preserved the same overall directional pattern. The associations involving σ_a, σ_v, and p_{common} remained significant, whereas the $\sigma_{a,level}$ association was attenuated and no longer significant. Together, these results indicate that the behavioural expression of the AV rabbit illusion was robustly related to modality-specific sensory uncertainty and common-cause tendency, whereas the sound-level sensitivity parameter appeared to capture a weaker or less robust dimension of auditory reliability.

3.4. Group-Level MMN Responses

Grand-average CSD-referenced MMN responses are shown in Fig. 5. Temporally, the aMMN showed a fronto-central negativity between 100 and 250 ms, peaking at 164 ms (− 0.035 μV/cm², maximal at F4, Fig. 5A). The vMMN exhibited an occipital negativity between 150 and 300 ms, peaking at 228 ms (− 0.033 μV/cm², maximal at PO7, Fig. 5B). The avMMN included both a fronto-central component (152 ms, − 0.045 μV/cm², Fig. 5C) and an occipital component (248 ms, − 0.026 μV/cm², Fig. 5D), with the fronto-central audiovisual response numerically exceeding the unimodal auditory MMN, consistent with a possible cross-modal enhancement of prediction-error responses.

Across conditions, MMN topographies were modality-specific: the auditory MMN (aMMN) was maximal over fronto-central electrodes, whereas the visual MMN (vMMN) was maximal over occipital–parietal sites. The congruent audiovisual condition (avMMN) exhibited negativities in frontal regions, consistent with combined modality contributions (Fig. 5E).

One-sample t-tests confirmed significant group-level negativity for the aMMN at fronto-central electrodes (mean = − 0.014 μV/cm², t(14) = − 2.54, p = 0.024) and for the avMMN fronto-central component (mean = − 0.020 μV/cm², t(14) = − 3.37, p = 0.005). The vMMN at occipital sites was marginally significant (mean = − 0.019 μV/cm², t(14) = − 2.08, p = 0.057), whereas the avMMN occipital component was not significant (mean = − 0.010 μV/cm², t(14) = − 0.91, p = 0.378), consistent with substantial inter-individual variability (see Section 3.5).

3.5. Auditory Noise Predicts Occipital MMN in the Congruent Audiovisual Condition

The primary analysis tested whether BCI model parameters predicted CSD MMN amplitudes across participants in the EEG subsample (N = 15). The primary finding was a significant positive correlation between auditory sensory noise (σ_a, log z-scored) and occipital MMN amplitude in the congruent audiovisual condition (AV_{OC}; Spearman ρ = +0.707, p = 0.0043, BH-FDR q = 0.0173; Fig. 6). Because MMN amplitudes are negative-going responses, this positive correlation indicates that participants with higher auditory noise showed less negative, and therefore weaker, occipital MMN responses in the congruent audiovisual condition. This pattern suggests that auditory sensory noise is associated with individual variability in occipital responses to audiovisual deviance.

The corresponding Pearson correlation was r = + 0.773 (p = 0.0007), and a nonparametric permutation test (10,000 iterations, two-tailed) yielded p = 0.0038. Bootstrap resampling further yielded a 95% confidence interval on Spearman's ρ of [0.258, 0.903], supporting the robustness of the association. Moreover, the correlation between σ_a and AV_{OC} MMN amplitude was temporally specific to the visual MMN window (150–300 ms), as confirmed by a point-wise Spearman ρ(t) analysis across the full epoch (Supplementary Fig. S3). The same positive association was also observed under alternative EEG reference schemes (original ρ = 0.664, p = 0.0086;average ρ = 0.739, p = 0.0024；mastoid ρ = 0.839, p = 0.0001), indicating that the effect was not specific to the CSD transformation. By contrast, σ_a did not significantly predict the frontocentral congruent audiovisual MMN (AV_{FC}: ρ = −0.154, p = 0.5844), unimodal auditory MMN in the frontocentral ROI (A_{FC}: ρ = −0.236, p = 0.3966), or unimodal visual MMN in the occipital ROI (V_{OC}: ρ = +0.400, p = 0.1408).

Thus, the σ_a–MMN association was selectively expressed in the occipital component of the audiovisual deviance response, rather than reflecting a general relationship between auditory noise and MMN amplitude. A median-split visualization of σ_a is shown in Supplementary Fig. S4. Consistent with the primary correlation result, high-σ_a participants showed a weaker AV_{OC} difference wave than low-σ_a participants, whereas AV_{FC} waveforms showed little corresponding separation. This visualization is

descriptive and does not replace the continuous-trait analysis. No comparable associations were observed for visual sensory noise (σ_v), suggesting that the observed brain–behaviour link was specific to auditory noise rather than reflecting a general relationship between sensory uncertainty and MMN amplitude.

3.6. Multiple Comparisons

BH-FDR correction was applied within the pre-specified AV MMN, unimodal MMN, and behavioural correlation families. The $\sigma_a \times AV_{OC}$ association was the only AV MMN association that remained significant after correction (BH-adjusted q = 0.0173), whereas no unimodal MMN association survived correction. All four primary associations remained significant after BH-FDR correction ($\sigma_a \times$ auditory illusion, $\sigma_v \times$ visual illusion, $\sigma_{a,level} \times$ auditory illusion, and $p_{common} \times$ total illusion), Spearman correlations for the same behavioural pairs are reported in Supplementary.

4. Discussion

This study asked whether individual differences in audiovisual illusion susceptibility can be decomposed into computationally interpretable traits, and whether these traits predict independent neural markers of multisensory mismatch processing. Using a bidirectional audiovisual rabbit task and a Bayesian causal-inference model, we found that illusion behaviour was associated with separable latent components, including auditory noise, visual noise, sound-level sensitivity of auditory noise, and common-cause prior. Critically, auditory noise estimated from behavioural reports also predicted occipital MMN amplitude in a separate passive audiovisual oddball paradigm: participants with higher auditory noise showed less negative occipital mismatch responses in the congruent audiovisual condition. This cross-paradigm association suggests that behaviour-derived sensory-noise traits capture individual differences that extend beyond the task in which they were estimated.

The $\sigma_a \times AV_{OC}$ association provides a possible link between auditory precision and multisensory prediction-error processing. Within predictive-coding accounts, perceptual inference depends not only on the magnitude of sensory mismatch but also on the precision assigned to prediction errors [14, 15, 32, 36]. MMN has been widely discussed as a neural response to violations of sensory regularity, and both auditory and visual MMN have been interpreted within predictive-processing frameworks in which the precision assigned to sensory evidence shapes the impact of deviance signals [10, 30–32, 36]. Recent neuronal evidence that auditory cortical neurons can encode negative prediction errors to omitted sounds in predictable sequences further supports the link between sequence regularity and mismatch-related responses [51]. In the present study, higher auditory noise can be interpreted as lower auditory precision in the BCI model. If auditory input contributes to establishing or stabilizing predictions about audiovisual events, then reduced auditory precision may weaken the cross-modal predictive signal available to visual cortical processing. This does not prove that the AV_{OC} response is a prediction-error signal in a strict mechanistic sense, but it supports the more cautious interpretation that auditory precision is linked to occipital mismatch-related activity under multisensory conditions.

The selectivity of this association further strengthens this interpretation. Auditory noise did not significantly predict frontocentral audiovisual MMN, unimodal auditory MMN, or unimodal visual MMN, and visual noise did not show a comparable association with audiovisual MMN. Thus, the finding cannot be reduced to a general relationship between sensory uncertainty and ERP amplitude, nor to globally weaker neural responsiveness in high-noise participants. Instead, the effect was expressed specifically in the occipital component of the congruent audiovisual deviance response, where auditory and visual features jointly defined the expected event structure. One possible interpretation is that the AV_{OC} response is not simply a visual MMN elicited alongside sound, but a multisensory mismatch response whose magnitude depends partly on the reliability of the auditory channel contributing to the audiovisual event model. This interpretation is broadly consistent with evidence that audiovisual cortical responses involve nonlinear cross-modal interactions rather than simple linear summation, and that multisensory response strength depends on stimulus reliability and effectiveness [37, 38, 40, 41, 52].

The behavioural findings further support the value of decomposing illusion susceptibility into latent computational traits rather than treating it as a single descriptive score. In the larger behavioural sample, auditory noise was associated with auditory illusion intensity, visual noise with visual illusion intensity, and common-cause prior with total illusion intensity. These findings are consistent with cue-reliability and causal-inference accounts, in which multisensory percepts depend on modality-specific sensory uncertainty and the observer's tendency to infer a shared audiovisual source [5, 6, 9, 53, 54]. The sound-level sensitivity parameter showed a weaker pattern: it was significant in the primary Pearson analysis but not robust in the Spearman analysis. This suggests that context-dependent changes in auditory reliability may contribute to illusion susceptibility, but this parameter should be interpreted cautiously until replicated in larger samples with denser sampling of sound levels. This distinction is important because multisensory weighting depends not only on average cue reliability, but also on how reliability changes across stimulus contexts [5, 6, 9].

More broadly, the BCI model served as more than a descriptive fit to behavioural reports. In causal-inference accounts of multisensory perception, perceptual reports reflect a combination of modality-specific sensory uncertainty and the inferred probability that signals arise from a common source [9, 53, 54]. This decomposition is especially useful for audiovisual rabbit illusions, whose postdictive structure makes overt illusion rates likely to reflect both sensory precision and causal-inference processes [22, 23, 55]. The fact that auditory noise generalized from an active illusion task to a passive oddball paradigm suggests that the model captured a meaningful individual-difference dimension rather than only task-specific response bias. At the same time, the overall pattern was not redundant: common-cause prior was most strongly expressed in total illusion intensity, whereas auditory noise was the parameter that predicted audiovisual occipital mismatch responses. This dissociation suggests that different latent traits may be preferentially expressed in different behavioural and neural readouts.

Several limitations should be considered. First, the EEG sample was modest (N = 15), which limits the precision of effect-size estimates and the ability to detect weaker brain–behaviour associations. Although the σ_a–AV_{OC} association survived correction within the pre-specified audiovisual MMN family

and was supported by permutation testing and reference-scheme robustness checks, independent replication in a larger sample is essential. Second, the model comparison should be interpreted relative to the tested model families rather than as proof that the selected implementation is uniquely correct. Third, the individual-differences analyses used posterior mean parameter estimates, rather than fully propagating posterior uncertainty into the EEG correlation stage. Finally, although consistency across reference schemes argues against a major reference-related confound, CSD-based occipital measures may still be affected by individual differences in signal quality.

These findings may also have broader implications for individual-difference and translational research, although such applications require direct testing. Because MMN is widely discussed as a translational biomarker, linking computationally defined sensory-noise traits to MMN responses may provide a framework for future studies of atypical multisensory processing in clinical populations [56–59]. In applied contexts such as virtual reality, where presence and usability depend on effective multisensory synchronization, user-specific sensory-noise estimates may eventually inform more individualized multisensory rendering strategies [60, 61]. The present results do not establish clinical or applied utility, but they illustrate a general strategy for moving from descriptive illusion susceptibility to computationally decomposed individual differences.

In conclusion, sensory-noise traits estimated from a Bayesian causal-inference model captured meaningful individual differences in audiovisual illusion susceptibility and related them to neural mismatch responses. In particular, auditory noise was linked to weaker occipital audiovisual MMN responses, suggesting that auditory precision contributes to individual variability in multisensory mismatch processing. These findings support a computational approach in which behavioural susceptibility to audiovisual illusions is decomposed into separable latent traits, some of which can be related to neural measures across tasks. Future work with larger samples, repeated sessions, and circuit-level methods such as source localization or dynamic causal modelling will be important for testing the stability and neural mechanisms of these traits.

Declarations

Amit Yaron

Conceptualization, Methodology, Formal Analysis, Visualization, Writing – Review & Editing.

Yi-yuan Teresa Huang

Methodology, Formal Analysis, Supervision, Writing – Review & Editing.

Zenas C. Chao

Conceptualization, Methodology, Funding Acquisition, Resources, Supervision, Writing – Review & Editing.

Tomoyo Isoguchi Shiramatsu

Conceptualization, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing.

Hirokazu Takahashi

Conceptualization, Methodology, Funding Acquisition, Project Administration, Resources, Supervision, Writing – Review & Editing.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Fundings

This work was partly supported by JSPS KAKENHI (23H03023, 25H02600, 25K22825, 26H02517), AMED (24wm0625401h0001), JST (JPMJPR22S8), the Asahi Glass Foundation, the Secom Science and Technology Foundation. The World Premier International Research Center Initiative (WPI), MEXT, Japan.

Author Contribution

H. X.: Conceptualization, Methodology, Investigation, Formal Analysis, Visualization, Writing – Original Draft, Writing – Review & Editing.A. Y.: Conceptualization, Methodology, Formal Analysis, Visualization, Writing – Review & Editing.Y. T. H.: Methodology, Formal Analysis, Supervision, Writing – Review & Editing.Z. C. C.: Conceptualization, Methodology, Funding Acquisition, Resources, Supervision, Writing – Review & Editing.T. I. S.: Conceptualization, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing.H. T.: Conceptualization, Methodology, Funding Acquisition, Project Administration, Resources, Supervision, Writing – Review & Editing.

Data Availability

The processed data and analysis codes supporting the results of this study are available in Zenodo at 10.5281/zenodo.20797916

References

1. Oakland, T., Black, J. L., Stanford, G., Nussbaum, N. L. & Balise, R. R. An evaluation of the dyslexia training program: a multisensory method for promoting reading in students with reading disabilities. *J. Learn. Disabil.* **31**, 140–147 (1998).

2. Shams, L. & Seitz, A. R. Benefits of multisensory learning. *Trends Cogn. Sci.* **12**, 411–417 (2008).
3. Denervaud, S., Gentaz, E., Matusz, P. J. & Murray, M. M. Multisensory Gains in Simple Detection Predict Global Cognition in Schoolchildren. *Sci. Rep.* **10**, 1394 (2020).
4. Mathias, B., Andrä, C., Schwager, A., Macedonia, M. & von Kriegstein, K. Twelve- and Fourteen-Year-Old School Children Differentially Benefit from Sensorimotor- and Multisensory-Enriched Vocabulary Training. *Educ. Psychol. Rev.* **34**, 1739–1770 (2022).
5. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
6. Alais, D. & Burr, D. The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* **14**, 257–262 (2004).
7. Deneve, S. & Pouget, A. Bayesian multisensory integration and cross-modal spatial links. *J. Physiol. Paris*. **98**, 249–258 (2004).
8. Haroush, K., Hochstein, S. & Deouell, L. Y. Momentary fluctuations in allocation of attention: cross-modal effects of visual task load on auditory discrimination. *J. Cogn. Neurosci.* **22**, 1440–1451 (2010).
9. Körding, K. P. et al. Causal Inference in Multisensory Perception. *PLOS ONE*. **2**, e943 (2007).
10. Arnal, L. H. & Giraud, A. L. Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* **16**, 390–398 (2012).
11. Rohe, T. & Noppeney, U. Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLoS Biol.* **13**, e1002073 (2015).
12. Noppeney, U., Jones, S. A., Rohe, T. & Ferrari, A. See what you hear – How the brain forms representations across the senses. *Neuroforum* **24**, A169–A181 (2018).
13. Rohe, T., Ehlis, A. C. & Noppeney, U. The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nat. Commun.* **10**, 1907 (2019).
14. Friston, K. A theory of cortical responses. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **360**, 815–836 (2005).
15. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).
16. Huang, Y. T. et al. Crossmodal hierarchical predictive coding for audiovisual sequences in the human brain. *Commun. Biol.* **7**, 965 (2024).
17. Shams, L., Kamitani, Y. & Shimojo, S. What you see is what you hear. *Nature* **408**, 788–788 (2000).
18. Shams, L., Kamitani, Y. & Shimojo, S. Visual illusion induced by sound. *Brain Res. Cogn. Brain Res.* **14**, 147–152 (2002).
19. Mcgurk, H. & Macdonald, J. Hearing lips and seeing voices. *Nature* **264**, 746–748 (1976).
20. Bonath, B. et al. Neural basis of the ventriloquist illusion. *Curr. Biol.* **17**, 1697–1703 (2007).
21. Stekelenburg, J. J., Keetels, M. & Vroomen, J. Multisensory integration of speech sounds with letters vs. visual speech: only visual speech induces the mismatch negativity. *Eur. J. Neurosci.* **47**, 1135–

1145 (2018).

22. Stiles, N. R. B., Li, M., Levitan, C. A., Kamitani, Y. & Shimojo, S. What you saw is what you will hear: Two new illusions with audiovisual postdictive effects. *PLOS ONE*. **13**, e0204217 (2018).
23. Günaydın, G., Moran, J. K., Rohe, T. & Senkowski, D. Causal inference shapes crossmodal postdiction in multisensory integration. *Sci. Rep.* **16**, 7490 (2026).
24. Xu, H., Yaron, A., Shiramatsu, T. I. & Takahashi, H. Common Mechanism Underlying Multimodal Integration. in *15th Biomedical Engineering International Conference (BMEiCON)* 1–5 (2023). 1–5 (2023). (2023). 10.1109/BMEiCON60347.2023.10322007
25. Keil, J., Müller, N., Ihssen, N. & Weisz, N. On the variability of the McGurk effect: audiovisual integration depends on prestimulus brain states. *Cereb. Cortex*. **22**, 221–231 (2012).
26. Stevenson, R. A., Zemtsov, R. K. & Wallace, M. T. Individual Differences in the Multisensory Temporal Binding Window Predict Susceptibility to Audiovisual Illusions. *J. Exp. Psychol. Hum. Percept. Perform.* **38**, 1517–1529 (2012).
27. Wallace, M. T. & Stevenson, R. A. The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia* **64**, 105–123 (2014).
28. Noel, J. P., De Niear, M. A., Stevenson, R., Alais, D. & Wallace, M. T. Atypical rapid audio-visual temporal recalibration in autism spectrum disorders. *Autism Res.* **10**, 121–129 (2017).
29. Näätänen, R., Gaillard, A. W. & Mäntysalo, S. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol. (Amst)*. **42**, 313–329 (1978).
30. Näätänen, R., Paavilainen, P., Rinne, T. & Alho, K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* **118**, 2544–2590 (2007).
31. Näätänen, R. et al. The mismatch negativity (MMN)--a unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clin. Neurophysiol.* **123**, 424–458 (2012).
32. Garrido, M. I., Kilner, J. M., Stephan, K. E. & Friston, K. J. The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* **120**, 453–463 (2009).
33. Wacongne, C., Changeux, J. P. & Dehaene, S. A. Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *J. Neurosci.* **32**, 3665–3678 (2012).
34. Fitzgerald, K. & Todd, J. Making Sense of Mismatch Negativity. *Front Psychiatry* **11**, (2020).
35. Yaron, A., Hershenhoren, I. & Nelken, I. Sensitivity to Complex Statistical Regularities in Rat Auditory Cortex. *Neuron* **76**, 603–615 (2012).
36. Stefanics, G., Kremláček, J. & Czigler, I. Visual mismatch negativity: a predictive coding view. *Front Hum. Neurosci* **8**, (2014).
37. Giard, M. H. & Peronnet, F. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J. Cogn. Neurosci.* **11**, 473–490 (1999).
38. Senkowski, D., Saint-Amour, D., Höfle, M. & Foxe, J. J. Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *Neuroimage* **56**, 2200–2208 (2011).

39. Shiramatsu, T. I., Mori, K., Ishizu, K., Takahashi, H. & Auditory Visual, and Cross-Modal Mismatch Negativities in the Rat Auditory and Visual Cortices. *Front Hum. Neurosci* **15**, (2021).
40. Meredith, M. A. & Stein, B. E. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.* **56**, 640–662 (1986).
41. Fort, A., Delpuech, C., Pernier, J. & Giard, M. H. Early auditory-visual interactions in human cortex during nonredundant target identification. *Brain Res. Cogn. Brain Res.* **14**, 20–30 (2002).
42. Stekelenburg, J. J. & Vroomen, J. Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cogn. Neurosci.* **19**, 1964–1973 (2007).
43. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
44. Pelli, D. G. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
45. Wozny, D. R., Beierholm, U. R. & Shams, L. Probability Matching as a Computational Strategy Used in Perception. *PLoS Comput. Biol.* **6**, e1000871 (2010).
46. Abril-Pla, O. et al. PyMC: a modern, and comprehensive probabilistic programming framework in Python. *PeerJ Comput. Sci.* **9**, e1516 (2023).
47. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
48. Yao, Y., Vehtari, A., Simpson, D. & Gelman, A. Using Stacking to Average Bayesian Predictive Distributions (with Discussion). *Bayesian Anal.* **13**, 917–1007 (2018).
49. Kayser, J. & Tenke, C. E. Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. *Clin. Neurophysiol.* **117**, 348–368 (2006).
50. Gramfort, A. et al. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, 267 (2013).
51. Yaron, A. et al. Auditory cortex neurons that encode negative prediction errors respond to omissions of sounds in a predictable sequence. *PLoS Biol.* **23**, e3003242 (2025).
52. Ullsperger, P., Erdmann, U., Freude, G. & Dehoff, W. When sound and picture do not fit: Mismatch negativity and sensory interaction. *Int. J. Psychophysiol.* **59**, 3–7 (2006).
53. Shams, L. & Beierholm, U. R. Causal inference in perception. *Trends Cogn. Sci.* **14**, 425–432 (2010).
54. Rohe, T. & Noppeney, U. Sensory reliability shapes perceptual inference via two mechanisms. *J. Vis.* **15**, 22 (2015).
55. Odegaard, B., Wozny, D. R. & Shams, L. Biases in Visual, Auditory, and Audiovisual Perception of Space. *PLoS Comput. Biol.* **11**, e1004649 (2015).
56. Light, G. A. & Näätänen, R. Mismatch negativity is a breakthrough biomarker for understanding and treating psychotic disorders. *Proc. Natl. Acad. Sci. U S A*. **110**, 15175–15176 (2013).
57. Perez, V. B. et al. Mismatch Negativity is a Sensitive and Predictive Biomarker of Perceptual Learning During Auditory Cognitive Training in Schizophrenia. *Neuropsychopharmacology* **42**, 2206–2213 (2017).

58. Van de Cruys, S. et al. Precise minds in uncertain worlds: predictive coding in autism. *Psychol. Rev.* **121**, 649–675 (2014).
59. Lawson, R. P., Mathys, C. & Rees, G. Adults with autism overestimate the volatility of the sensory environment. *Nat. Neurosci.* **20**, 1293–1299 (2017).
60. Ibrahim, A. et al. ARbis Pictus: A Study of Vocabulary Learning with Augmented Reality. *IEEE Trans. Vis. Comput. Graph*. **24**, 2867–2874 (2018).
61. Glenn, C. P. & Coxon, M. Individual differences in processing multisensory information predict presence in different virtual reality environments. *Virtual Real.* **29**, 9 (2024).

Figures

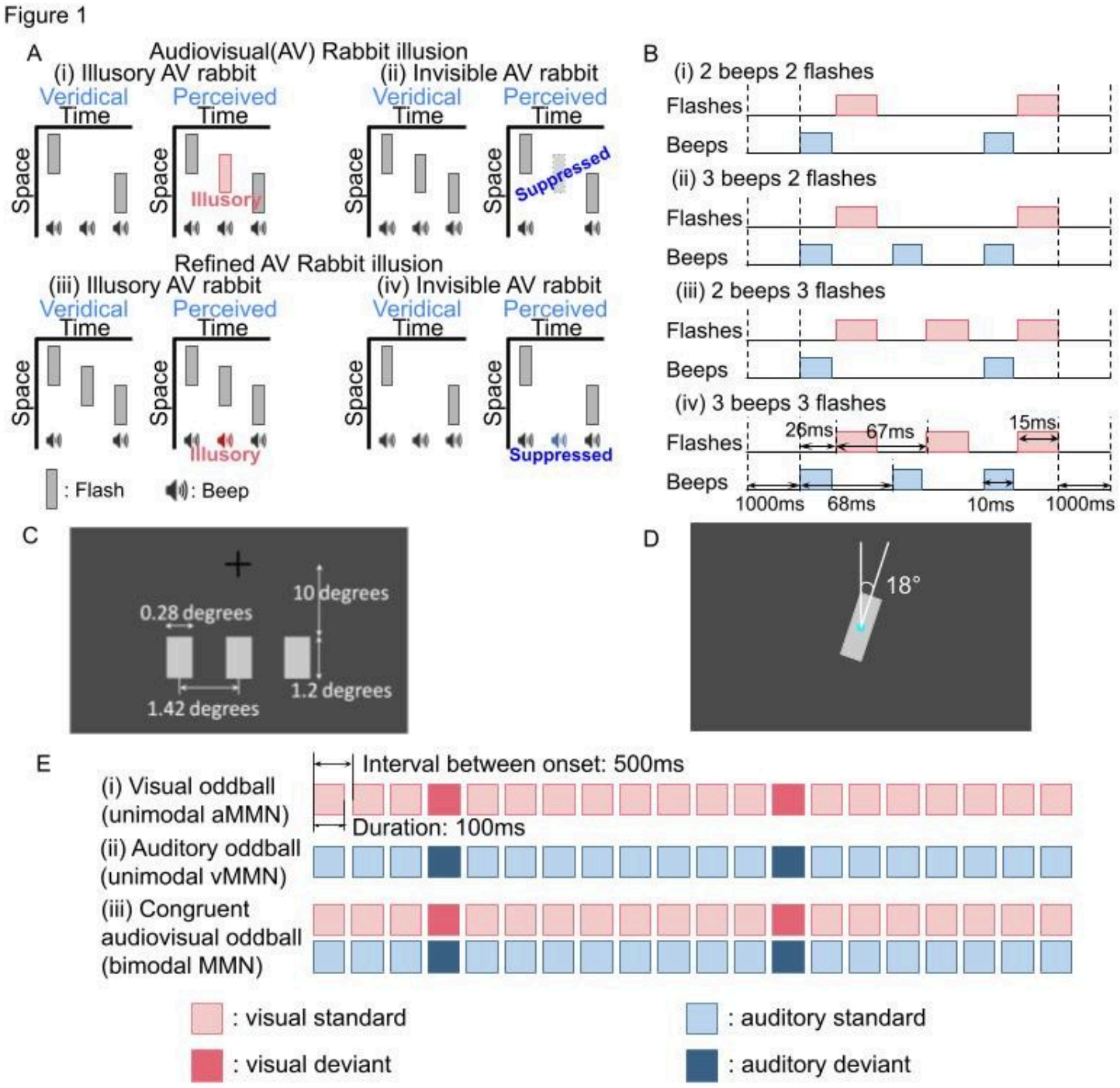


Figure 1

Experimental design of the bidirectional audiovisual rabbit illusion and MMN paradigms.

(A) Schematic illustration of the audiovisual (AV) rabbit illusion. In the classical illusion, auditory input biases visual perception: (i) in the 3-beeps/2-flashes condition, an additional illusory flash is perceived, whereas (ii) in the 2-beeps/3-flashes condition, one flash may be perceptually suppressed. In the refined bidirectional paradigm, visual input can also bias auditory perception: (iii) in the 2-beeps/3-flashes condition, an illusory beep may be perceived, whereas (iv) in the 3-beeps/2-flashes condition, one beep

may be perceptually suppressed. These conditions demonstrate bidirectional audiovisual interactions. **(B)** Temporal structure of audiovisual rabbit stimuli. Sequences of two or three beeps and flashes were presented with fixed onset intervals, with relative timing preserved across conditions. **(C)**Spatial configuration of visual stimuli in the audiovisual rabbit task. Rectangular flashes (0.28° × 1.2°) were presented 10° below fixation, with spatial separations of 1.42° in the three-flash condition. **(D)** Visual stimuli used in the oddball paradigm. Tilted bars (±18° from vertical), matched in size to the rabbit stimuli, served as standard and deviant stimuli. **(E)**Oddball paradigms used for mismatch negativity (MMN) recording. (i) Visual oddball (vMMN), (ii) auditory oddball (aMMN), and (iii) congruent audiovisual oddball (avMMN). Stimuli were presented for 100 ms with a 500 ms inter-onset interval. Deviant stimuli occurred with a probability of 10% among standards (90%). MMN responses were computed as the difference between deviant and standard ERPs for each condition.

Figure 2

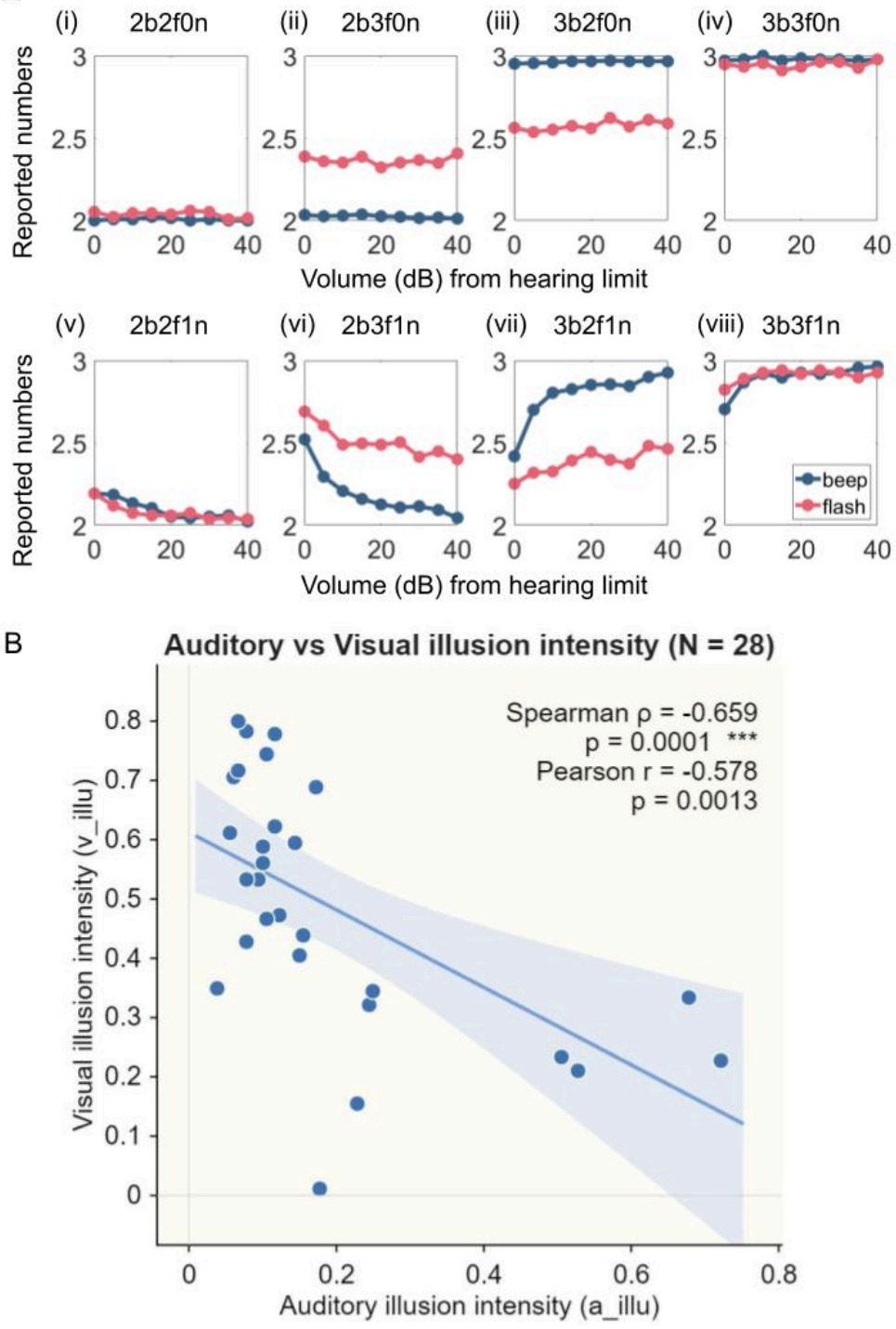


Figure 2

Overall behavioral results, illusion intensities and its relation to MMN integration.

(A) Average reported number of beeps and flashes across all participants in no noise conditions(i)2 beeps 2 flashes (2b2f0n) (ii)2 beeps 3 flashes (2b3f0n) (iii) 3 beeps 2 flashes (3b2f0n) (iv) 3 beeps 3 flashes (3b3f0n), and noisy conditions (v) 2 beeps 2 flashes (2b2f1n) (vi) 2 beeps 3 flashes (2b3f1n) (vii)

3 beeps 2 flashes (3b2f1n) (viii) 3 beeps 3 flashes (3b3f1n). (B) Auditory and visual illusion intensities of all participants. There were auditory dominant participants had stronger flash illusion and weaker beep illusion (upper-left dots) and visual dominant participants had weaker flash illusion and stronger beep illusion (bottom-right dots).

Figure 3

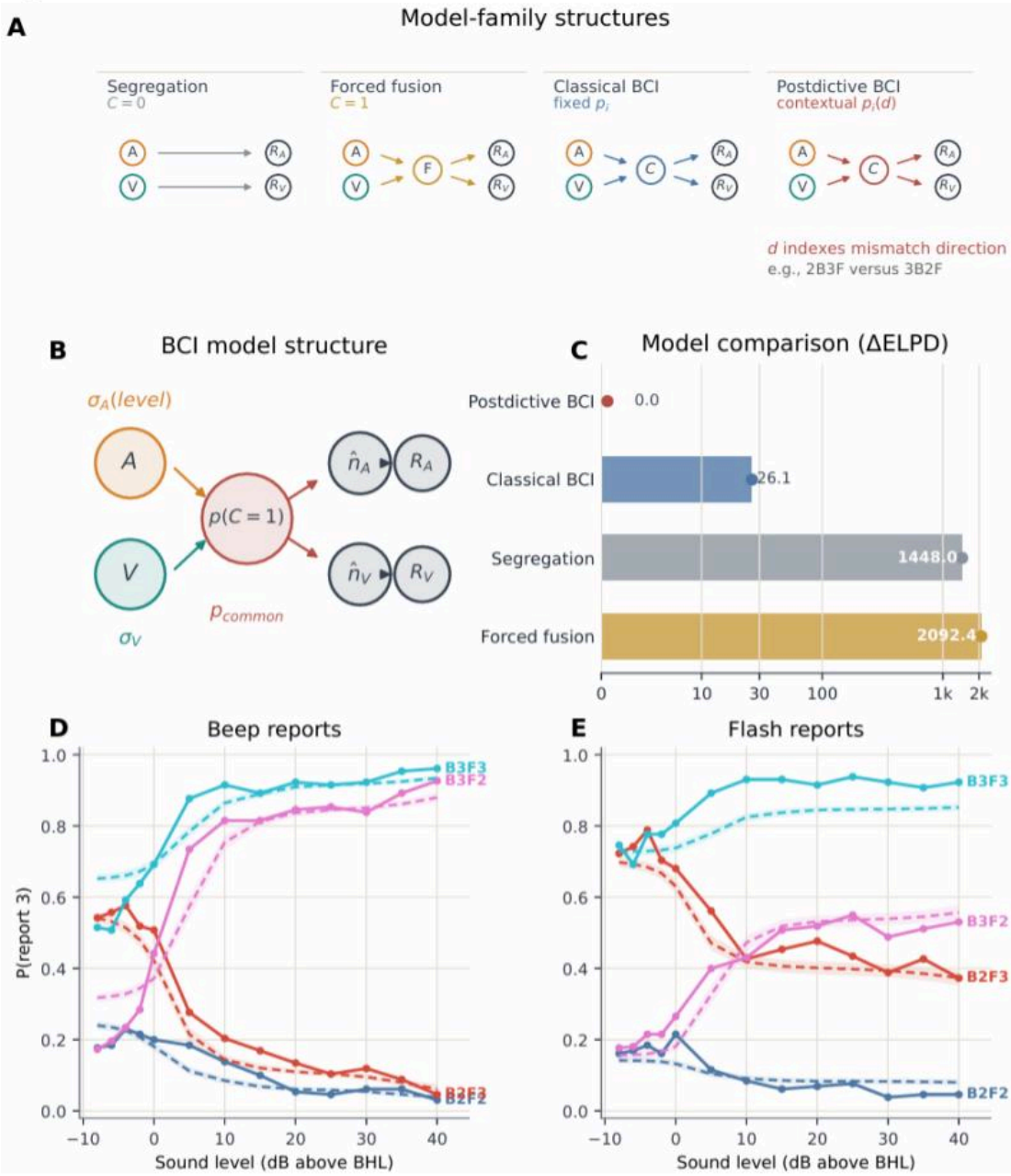


Figure 3

Bayesian causal-inference model families, selected model, and posterior predictive fit.
(A) Schematic comparison of the four model families tested: segregation, forced fusion, classical BCI, and postdictive BCI. In the postdictive BCI family, d indexes mismatch direction, allowing the common-cause probability to vary across contexts such as 2B3F versus 3B2F. (B) Structure of the selected BCI model used to estimate participant-level traits, including auditory noise, visual noise, sound-level-dependent auditory noise, and common-cause prior. (C) PSIS-LOO model comparison, expressed as ΔELPD from the best model. Lower values indicate better out-of-sample predictive performance. (D,E) Posterior predictive checks for beep reports and flash reports. Solid lines show observed report probabilities, dashed lines show posterior predictive means, and shaded ribbons show 95% posterior intervals. Condition labels indicate physical beep and flash counts.

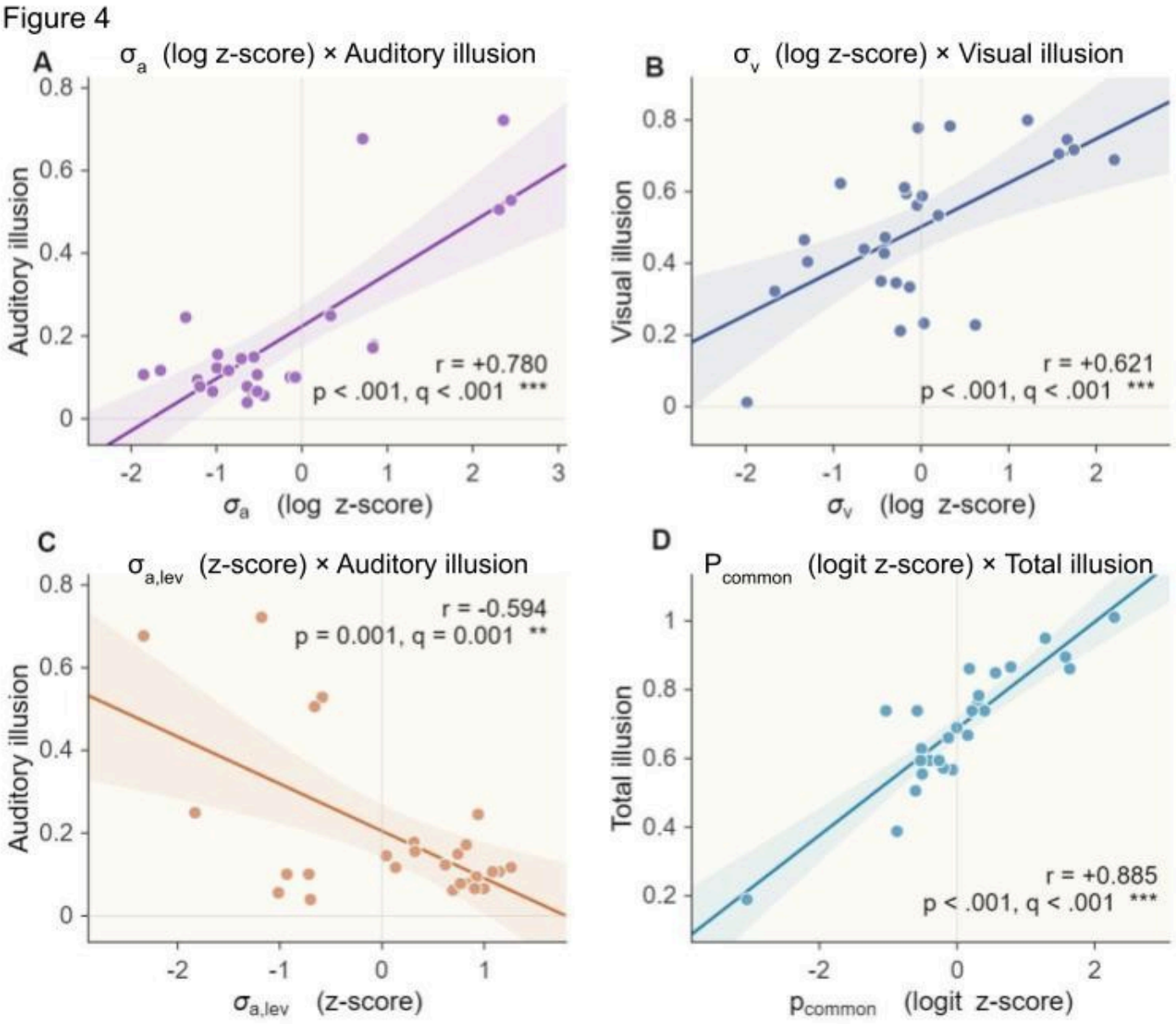


Figure 4

Behavioural correlates of BCI-derived latent traits.

Scatter plots showing relationships between individual BCI model parameters and illusion intensity measures.

(A) Auditory noise (σ𝒶, log z-scored) positively correlates with auditory illusion intensity.

(B) Visual noise (σ𝓋, log z-scored) positively correlates with visual illusion intensity.

(C) Sound-level sensitivity of auditory noise ($\sigma_{a,level}$, z-scored) negatively correlates with auditory illusion intensity, indicating that participants whose auditory representations improved more strongly with sound level showed reduced auditory illusion intensity.

(D) Common-cause prior (p_{common}, logit z-scored) strongly correlates with total illusion intensity.

Solid lines indicate linear fits from Pearson correlation analyses, with shaded regions representing 95% confidence intervals. Correlations were computed in the larger behavioural BCI sample (N = 26), reported p-values are two-tailed and corrected using the Benjamini–Hochberg false discovery rate (BH-FDR) within the pre-specified behavioural family of tests.

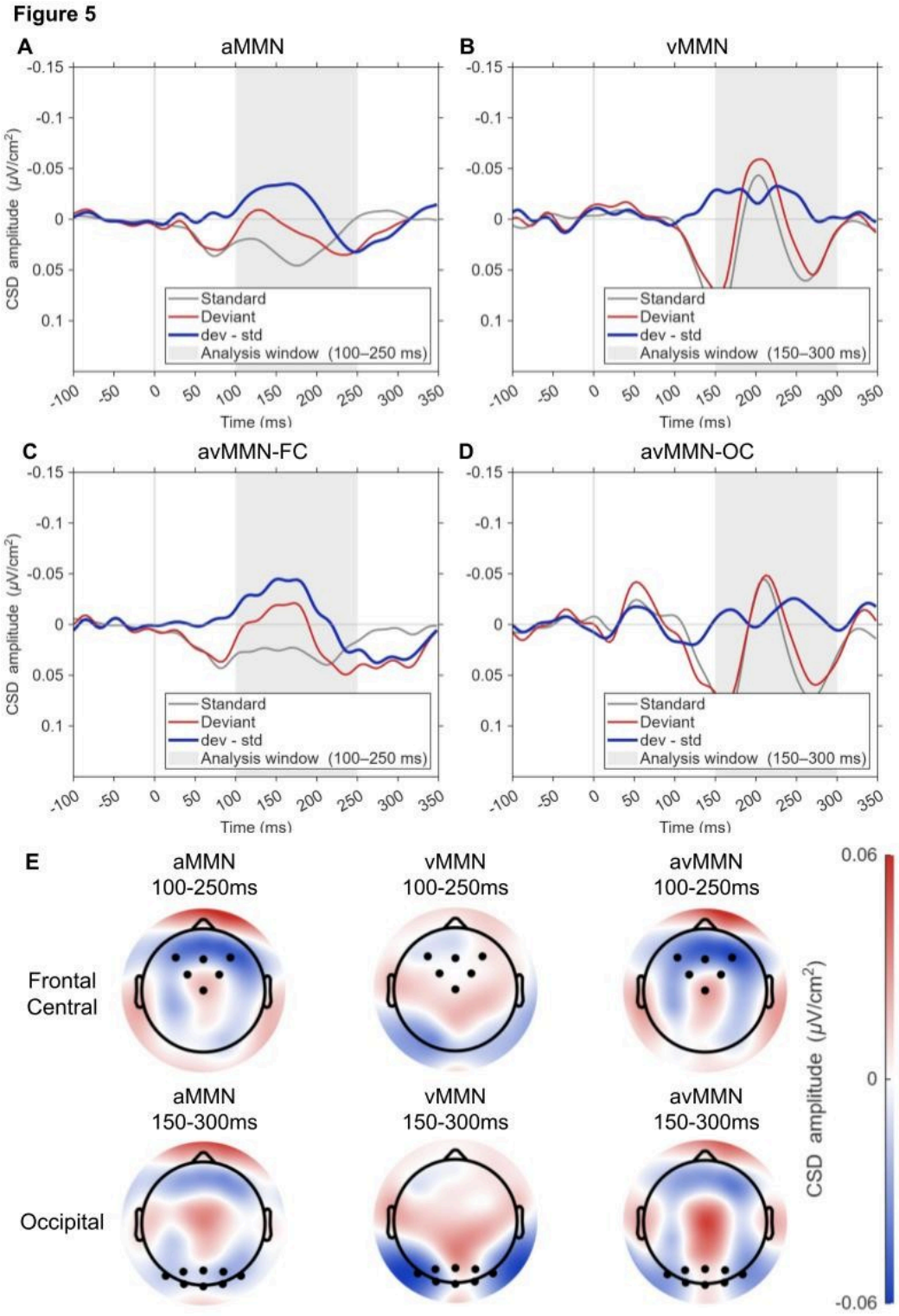


Figure 5

Group-level MMN responses across unimodal and audiovisual conditions (CSD, N = 15).
Grand-average event-related potentials (ERPs) are shown for standard (black), deviant (red), and MMN difference waves (deviant − standard, blue) following current source density (CSD) transformation. Shaded regions indicate the predefined analysis windows.

(A) Auditory MMN (aMMN) at fronto-central electrodes (Fz, FCz, Cz, F3, F4, FC1, FC2), showing a negative deflection between 100 and 250 ms.
(B) Visual MMN (vMMN) at occipital electrodes (Oz, O1, O2, POz, PO3, PO4, PO7, PO8), showing a negative deflection between 150 and 300 ms.
(C) Audiovisual MMN (avMMN) at fronto-central electrodes, showing a negativity within the auditory time window.
(D) Audiovisual MMN at occipital electrodes, showing a later negative component within the visual time window.

(E) Scalp topographies of mean CSD-referenced MMN amplitudes averaged within the predefined analysis windows. For the auditory MMN (aMMN), topographies were averaged over 100–250 ms; for the visual MMN (vMMN), over 150–300 ms. For the congruent audiovisual condition (avMMN), topographies are shown for both the fronto-central auditory time window (100–250 ms) and the occipital visual time window (150–300 ms). The aMMN showed a fronto-central distribution, the vMMN showed an occipital–parietal distribution, and the avMMN exhibited both patterns.

One-sample t-tests against zero confirmed significant MMN responses for the aMMN fronto-central component and the audiovisual fronto-central component, whereas the occipital visual MMN was marginal and the occipital audiovisual MMN was not significant. These results indicate modality-specific spatial distributions of MMN responses and greater inter-individual variability in occipital components.

Figure 6

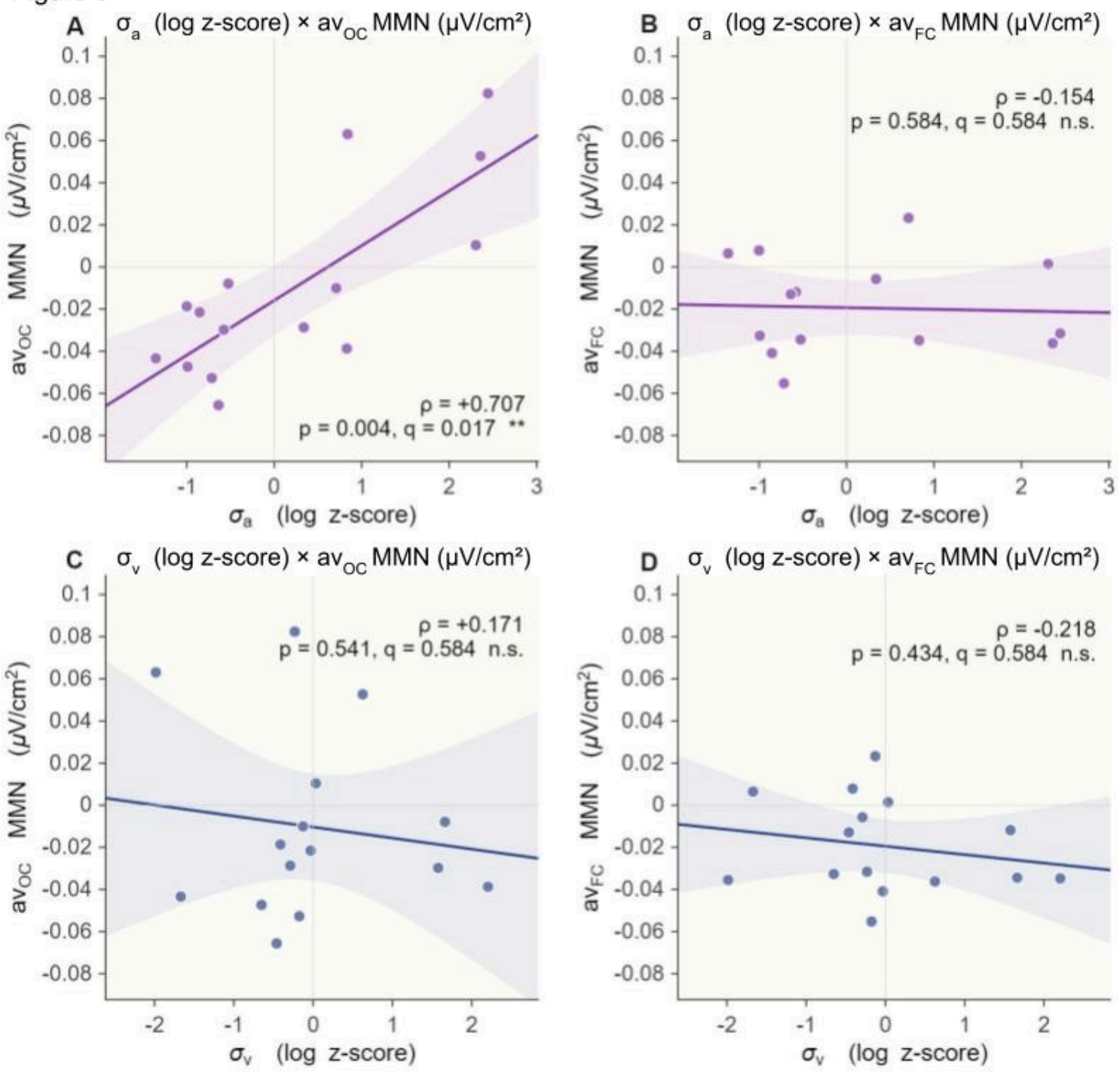


Figure 6

Specific association between auditory sensory noise and occipital audiovisual MMN (CSD reference, N = 15).

Scatter plots showing relationships between BCI-derived sensory noise parameters and MMN amplitudes in predefined regions of interest. **(A)** Auditory sensory noise (σ𝒶, log z-scored) positively correlates with occipital audiovisual MMN (AV_{OC}), such that higher σ𝒶 is associated with less negative (weaker) MMN amplitudes (Spearman ρ = +0.707, p = 0.0043, BH-FDR q = 0.017). **(B)** Auditory sensory

noise shows no significant association with frontocentral audiovisual MMN (AV_{FC}). **(C)** Visual sensory noise (σ_V, log z-scored) shows no significant association with occipital audiovisual MMN (AV_{OC}). **(D)** Visual sensory noise shows no significant association with frontocentral audiovisual MMN (AV_{FC}). Solid lines indicate linear fits, and shaded regions represent 95% confidence intervals. These results demonstrate that the observed brain–behaviour relationship is specific to auditory sensory noise and to the occipital component of audiovisual deviance processing.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.docx