CORTEX 194 (2026) 50–62

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Increasing visual uncertainty modulates multisensory decision-making

Xiangfu Yang [a], Weiping Yang [b,c], Yinghua Yu [a], Yoshimichi Ejima [a] and Jiajia Yang [a,*]

[a] Graduate School of Interdisciplinary Science and Engineering in Health Systems, Okayama University, Okayama, Japan

[b] Department of Psychology, Faculty of Education, Hubei University, Wuhan, China

[c] Brain and Cognition Research Center (BCRC), Faculty of Education, Hubei University, Wuhan, China

ARTICLE INFO

Article history:
Received 11 June 2025
Revised 5 November 2025
Accepted 5 November 2025
Action editor Vanja Kovic
Published online 20 November 2025

Keywords:
Multisensory decision-making
Visual uncertainty
Audiovisual integration
Event-related potential
Drift–diffusion model

ABSTRACT

The brain integrates and transforms information from multiple senses to make optimal decisions, a process that is critical for navigating complex environments with perceptual uncertainty. Despite a growing consensus that individuals adapt flexibly to uncertain sensory input, whether increasing visual uncertainty influences the decision process itself or other, non-decision sensory processes during multisensory decision-making are unclear. Here, an audiovisual categorization task was used to examine the responses of human participants (N = 30) to visual and audiovisual stimuli under low-, medium-, and high-uncertainty conditions. Modeling the behavioral data using a drift–diffusion model indicated that increased visual uncertainty in the audiovisual context decreased the evidence accumulation rate but had no effect on non-decision processes. Electrophysiological recordings confirmed and expanded upon these results: increased visual uncertainty in the audiovisual context reduced the amplitude during the late decision-making stage (300–380 msec) but had no effect on the amplitude during the early sensory encoding stage (140–220 msec). More importantly, electroencephalography analyses revealed that audiovisual integration in the early sensory encoding stage occurred robustly across all visual uncertainty conditions, whereas audiovisual integration in the late stage occurred only under medium and high visual uncertainty conditions. This study demonstrated that increased visual uncertainty modulates the decision process itself rather than early sensory encoding during multisensory decision-making. Moreover, multisensory integration strategies dynamically adapt to increasing visual uncertainty by engaging different mechanisms to maintain effective decision-making.

* *Corresponding author*. Graduate School of Interdisciplinary Science and Engineering in Health Systems, Okayama University, 3-1-1 Tsushima-Naka, Kita-ku, Okayama 700-8530, Japan.
E-mail address: yang@okayama-u.ac.jp (J. Yang).

https://doi.org/10.1016/j.cortex.2025.11.005

1. Introduction

When sensory input is processed in complex environments, the brain integrates information from multiple senses to make optimal multisensory decisions (Drugowitsch et al., 2014; Heekeren et al., 2008; Raposo et al., 2012; Stein & Stanford, 2008). Multisensory decision-making is affected by any potential mismatch in information content caused by sensory uncertainties (Alais & Burr, 2004; Ernst & Bülthoff, 2004; Otto & Mamassian, 2012; Rohe & Noppeney, 2015a). For example, as visual uncertainty increases, the reliability weight attributed to auditory information increases. However, to date, there is no neurophysiological evidence to clarify how the perceptual system integrates multiple types of sensory information effectively when visual uncertainty increases and when corresponding adjustments are made in multisensory decision-making.

Recent studies have demonstrated that multisensory integration can occur at different stages of perceptual decision-making. Although there is a growing consensus that multisensory information improves both speed and accuracy in perceptual tasks (Giard & Peronnet, 1999; Lippert et al., 2007), there is a long-standing debate on when and how multisensory information facilitates decisions, with two alternative hypotheses: multisensory integration begins during the sensory encoding process, contributing to the accumulation of sensory evidence, and subsequent decisions are based on an already multisensory representation (Cappe et al., 2010; Mercier et al., 2013; Molholm et al., 2002; Schroeder & Foxe, 2005; Zhao et al., 2018), or multisensory integration occurs at the same time as the decision formation stage, using unisensory evidence accumulated independently in the sensory encoding stage (Franzen et al., 2020; Kayser et al., 2017; Rohe & Noppeney, 2016). The temporal distinction between early integration during sensory encoding and late integration during decision formation highlights the functional differences between stages of multisensory processing. Early integration directly responds to sensory features, whereas late integration is based on the gradual reweighting of information according to multiple sensory reliability and decisional relevance (Bizley et al., 2016; Noppeney, 2021; Rohe & Noppeney, 2015b). In the audiovisual context, reliability weights are more strongly impacted by visual information than by auditory information, indicating a visual dominance effect (Colavita, 1974; Odegaard et al., 2015; Schmid et al., 2011; Sinnett et al., 2007). Visual information takes precedence over auditory information in terms of attention allocation and response selection when visual and auditory information are presented simultaneously. Consequently, visual uncertainty is more likely to alter the efficiency and outcome of audiovisual integration. Therefore, further exploration is needed to understand how increasing visual uncertainty affects the stages of audiovisual integration.

Multisensory information processing is believed to begin with unisensory representations in the primary sensory cortex, continue with sensory integration in parietal-temporal regions, and culminate as causal inferences in the frontal lobe (Cao et al., 2019). This distributed multilevel processing reflects the functional complexity of different brain regions in the multisensory integration process and the ability to flexibly integrate sensory inputs according to decision-making demands (Pasqualotto, 2016; Scheliga et al., 2023; Senkowski & Engel, 2024). Further research has shown that when uncertainty in visual and auditory information increases, audiovisual integration relies on input from the intraparietal sulcus, whereas clear audiovisual stimuli are processed through direct information exchange between sensory cortices (Regenbogen et al., 2018). However, how the whole-brain network readjusts its functional division of labor at various stages of audiovisual integration to adapt to decision-making demands when visual uncertainty increases still needs to be determined.

Here, we aimed to separate early sensory encoding from late decision formation in multisensory decision-making by combining event-related potentials (ERPs) with a drift–diffusion model (DDM) behavioral data. While DDM has been widely used to infer latent cognitive processes, recent research on perceptual decision-making suggests that conclusions based on DDM may be contradicted by complementary analyses of the neural signatures in some cases (McGovern et al., 2018; Purcell & Palmeri, 2017; Spieser et al., 2018), suggesting that it is critical to corroborate insights from behavioral modeling with neural evidence. To address the above questions, we employed a classic audiovisual categorization task under three levels of visual uncertainty (low, medium, and high), allowing us to systematically examine how uncertainty modulates different processing stages. Our hypothesis was that variations in information content and the weighting distribution induced by visual uncertainty notably affect the late decision formation stage; in contrast, visual uncertainty does not alter basic sensory features such as contrast or luminance and is therefore unlikely to influence the early sensory encoding stage.

2. Materials and methods

2.1. Participants

The required sample size was calculated using G*Power 3.1 software, with the medium effect size (f) set to .25, the type I error probability (α) set to .05, test power ($1 - \beta$ err prob) set to .95, number of groups set to 1, number of measurements set to 6, and correlation among repeated measures set to .5. Based on these parameters, the required sample size was determined to be 28 participants. The software does not directly implement a factorial 2 modalities (visual, audiovisual) $\times$ 3 uncertainties (high, medium, low) repeated-measures ANOVA, the power analysis was conducted as an approximation of this design. A total of 32 participants from Okayama University were recruited for the experiment. All participants had normal or corrected-to-normal vision, as verified by the Snellen eye test, and reported no history of hearing-related disorders. Two participants were excluded from the data analysis because of excessive artifacts in their electroencephalography (EEG) data, resulting in an insufficient number of remaining EEG segments. The final sample size was 30 participants (14 males; age: 25.3 $\pm$ 2.4 years). All the

participants were naive to the purpose of the study, provided written informed consent prior to participation, and received monetary compensation upon completion of the study. The Research Ethics Committee of Okayama University approved all the recruitment and experimental procedures.

2.2. Stimuli

The experimental stimuli were generated using MATLAB R2018b with routines from the Psychophysics and Video Toolbox (Brainard, 1997). The visual stimuli consisted of Gabor patches with two spatial frequencies: a high spatial frequency of 10 cycles per degree and a low spatial frequency of 5 cycles per degree. Each Gabor patch had a vertical orientation (90°), a size of 300 × 300 pixels, and a contrast level of 50%. Gaussian noise was then generated and superimposed on the Gabor patches; the level of visual evidence was controlled by adjusting the Gaussian noise intensity. Eleven visual stimuli with different signal-to-noise ratios (SNRs) were generated, ranging from completely noisy (0 % signal) to fully coherent (100 % signal) in 10 % increments. As the SNR increased, the visibility and clarity of the Gabor patch also increased. The parameters of the visual material were determined on the basis of previous research (Kupers et al., 2019; Murai & Yotsumoto, 2018; Parise & Ernst, 2017). The auditory stimuli consisted of two pure tones of different frequencies: a high-frequency pure tone at 1500 Hz and a low-frequency pure tone at 500 Hz. All auditory stimuli were normalized by the root mean square amplitude to ensure consistent loudness levels, with a standard intensity set at 65 dB(A) sound pressure level (SPL). Next, eleven different white noise loudness levels were generated—76, 71, 66, 61, 58, 55, 51, 47, 44, 41, and 37 dB(A) SPL—and superimposed on the pure tones to generate auditory stimuli with different SNRs. The SNR is expressed in dB SNR, where −11, −6, −1, −4, +7, +10, +14, +18, +21, +24, and +28; auditory evidence increases as the SNR increases (Binder et al., 2004). All the resulting audio files were normalized again to 65 dB(A) SPL to maintain consistent perceived loudness.

To systematically manipulate stimulus uncertainty, the different SNRs of visual and auditory stimuli were used to adjust the specific accuracy, which served as proxies for uncertainty. Three levels of uncertainty were defined on the basis of accuracy: high uncertainty corresponds to an accuracy of approximately 55 %, medium uncertainty is approximately 70 %, and low uncertainty is approximately 90 %. To achieve specific accuracy, a continuous adaptive staircase procedure based on the adaptive up-down method was used in each unisensory condition (Levitt, 1971). Each uncertainty condition corresponds to an independent staircase trajectory, which is used to dynamically adjust the SNR of the presented stimulus to achieve the desired accuracy. For example, in the medium uncertainty condition, a 2-down 1-up strategy was adopted: the SNR decreased after two consecutive correct responses and increased after each incorrect response. This method can make the accuracy converge steadily to approximately 70.7 %, thereby achieving systematic manipulation of different uncertainty conditions. Independent staircases were implemented for the visual (V) and auditory (A) conditions to adjust for stimulus uncertainty separately for each modality. For the audiovisual (AV) condition, the stimuli were constructed by pairing the current level (SNR) from the visual staircase and auditory staircase. The experiment involved three visual uncertainty conditions (high, medium, and low) and one medium auditory uncertainty condition. The mean SNRs across participants were as follows: HV = 14.9 % ± 2.3 %, MV = 35.3 % ± 4.2 %, LV = 77.1 % ± 3.6 %, and MA = +9.6 ± 2.1 dB (mean ± SD). The stimuli were presented using MATLAB 2018b with the Psychophysics Toolbox. Visual stimuli were displayed on a 22-inch 4K UHD monitor (Display++ UHD, Cambridge Research Systems), with a 3480 × 2160 IPS LCD panel, a 144 Hz refresh rate, a 2 msec pixel response time, and a wide color gamut of quantum dot LED backlights. Each visual stimulus subtended a visual angle of 8° × 8° and was presented against a mid-gray background (RGB: 128, 128, 128). Auditory stimuli were delivered via a Lenovo M220 speaker positioned directly behind the center of the screen. The sound intensity was calibrated to ~65 dB(A) using a sound level meter (AR814), with the microphone placed at the approximate location of the participant's head. The participants sat in a comfortable position approximately 65 cm from the screen.

2.3. Experimental design

Participants completed an audiovisual categorization task in which stimuli were randomly presented under one of the following seven conditions (Fig. 1A): three visual-only conditions with high, medium, or low uncertainty (HV, MV, or LV); one auditory-only condition with medium uncertainty (MA); and three audiovisual conditions combining visual and auditory inputs under different uncertainty conditions. Specifically, the audiovisual stimuli consisted of HV + MA (HAV), MV + MA (MAV), and LV + MA (LAV). In all audiovisual conditions, the spatial frequency of the Gabor patch was always congruent with the auditory tone: high-frequency Gabor patches were paired with high-frequency tones (1500 Hz), and low-frequency Gabor patches were paired with low-frequency tones (500 Hz). Participants were required to press the F key on the keyboard when they saw a high-spatial-frequency Gabor patch and heard a high-frequency tone and to press the J key on the keyboard when they saw a low-spatial-frequency Gabor patch and heard a low-frequency tone. The participants were instructed to respond as fast and as accurately as possible to each experience.

The experiment consisted of 840 trials that were evenly distributed across seven stimulus conditions: three visual uncertainty levels, one auditory condition, and three audiovisual uncertainty conditions. Each condition contained 120 trials. The stimulus was presented for 250 msec, followed by a response window of up to 1200 msec. The interstimulus interval (ISI) varied randomly between 1200 and 1500 msec to reduce anticipatory responses. The entire experiment lasted approximately 35 min. To minimize fatigue, participants were given a self-paced break after every 240 trials. The experiment was conducted in a quiet, dimly lit room to minimize external auditory and visual distractions. During EEG recording, participants were instructed to remain as still as possible and to minimize head and body movements to ensure high-quality signal acquisition.

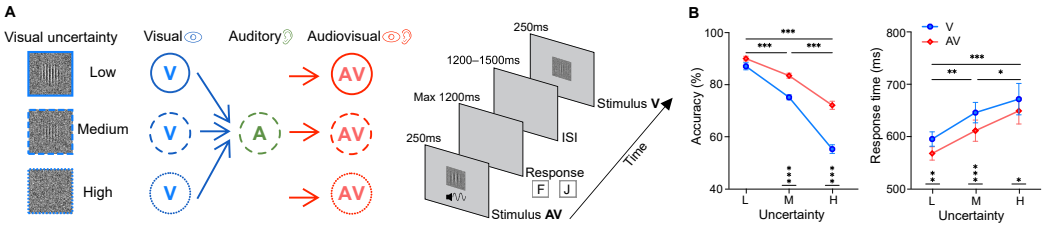

Fig. 1 – Experimental design and behavioral results. A. Participants performed an audiovisual categorization task under seven conditions: three visual (low, medium, and high uncertainty), one auditory (medium uncertainty), and three audiovisual (combinations of visual and auditory) tasks. Uncertainty levels were based on accuracy (~90 %, ~70 %, ~55 %). Each trial included a 250 msec stimulus, followed by a response window (max 1200 msec) and a jittered interstimulus interval (1200–1500 msec). Participants were required to press the F key for high-spatial-frequency Gabor patches or high-frequency sounds, and the J key for low-spatial-frequency Gabor patches or low-frequency sounds. B. Behavioral results. The error bars indicate the standard error of the mean (SEM). Significance levels: *$p < .05$, **$p < .01$, ***$p < .001$.

2.4. *Behavioral data analysis*

The behavioral data were first preprocessed to remove outlier responses. Trials with reaction times (RTs) less than 200 msec were excluded as anticipatory responses. Additionally, RTs falling more than 1.5 interquartile ranges below the first quartile or above the third quartile of each participant's RT distribution were excluded to reduce the influence of extreme values (Tukey, 1977; Voss et al., 2013). After exclusion, accuracy was defined as the proportion of correct responses relative to the total number of valid trials. The mean RTs were computed for each condition. To examine the effects of modality and uncertainty on accuracy and RT, a 2 (modality: visual, audiovisual) × 3 (uncertainty: high, medium, low) repeated-measures ANOVA was conducted using JASP version 0.9.2 (Love et al., 2019). *P* values of all reported post hoc comparisons were adjusted for multiple comparisons using the Holm–Bonferroni correction.

2.5. *Preprocessing of EEG recordings*

EEG data were recorded using a 64-channel MR-compatible system (BrainAmps MR-Plus, Brain Products GmbH, Germany). The setup included 62 Ag/AgCl passive scalp electrodes mounted on an elastic cap (BrainCap MR) according to the extended 10/10 system, along with two additional electrodes: one for horizontal electrooculography (HEOG) and one for vertical electrooculography (VEOG) eye movements. The FCz and AFz sites were used as reference and ground, respectively. Electrode impedance was kept below 10 kΩ using a conductive paste (ABRALYT HiCl 1000 gr, Germany). EEG signals were sampled at 1000 Hz and bandpass filtered online between .016 and 250 Hz.

EEG data were processed offline using Brain Vision Analyzer software (Brain Products GmbH, Germany). Raw signals were rereferenced to the averaged mastoid electrodes (TP9 and TP10) and bandpass filtered between .1 and 30 Hz with a 24 dB/oct slope. Powerline noise at 60 Hz (standard in Japan) was removed using a notch filter. Ocular artifacts, including eye movements and blinks, were corrected using independent component analysis (ICA). Artifactual components were identified and removed on the basis of their scalp topographies and temporal characteristics. Following ICA, automatic artifact rejection was applied to exclude trials with voltage steps exceeding 50 μV/ms or amplitude fluctuations greater than 100 μV. On average, 6.7 ± 2.1 ICs were removed per participant. Rejected trials for each condition were MA = 26.2 ± 19.8, HV = 25.7 ± 20.6, MV = 26.2 ± 21.8, LV = 28.63 ± 24.3, HAV = 28.4 ± 22.7, MAV = 28.7 ± 21.7, and LAV = 28.3 ± 24.1. The remaining trials were also visually inspected to ensure data quality. The cleaned EEG data were then segmented into epochs from −200 msec to 800 msec relative to stimulus onset and baseline corrected using the pre-stimulus interval (−200 to 0 msec).

2.6. *ERP analysis*

The aim of this study was to investigate the effects of modality and visual uncertainty on the early sensory encoding and late decision formation stages during multisensory decision-making. From the amplitude observations, the ERP amplitudes elicited by the visual and audiovisual stimuli began to separate after 140 msec, with further separation between uncertainty conditions emerging after 300 msec. On the basis of amplitude observations and related literature (Philiastides et al., 2006; Philiastides & Sajda, 2006), time windows of 140–220 msec and 300–380 msec, representing the early sensory encoding and late decision formation stages, respectively, were identified. Following evidence that posterior electrodes are sensitive to evidence accumulation and decision formation during perceptual decision-making (Kosciessa et al., 2021; O'Connell et al., 2012; Philiastides et al., 2014; Polania et al., 2014; van den Brink et al., 2021), the POz, PO3, and PO4 electrodes were selected as regions of interest (ROIs). For each condition, ERP amplitudes were first averaged within two time windows: 140–220 msec (early window) and 300–380 msec (late window). For each window, the mean amplitudes were extracted from the electrodes POz, PO3, and PO4 and then averaged across these three posterior sites to obtain a single amplitude measure per condition. To assess

the main effects and interactions between modality and visual uncertainty, 2 modalities (visual and audiovisual) × 3 uncertainty (high, medium, and low) repeated-measures ANOVAs were subsequently conducted on the averaged amplitudes in the early and late time windows.

Analyzing only the ERP amplitude elicited by visual and audiovisual stimuli risks overlooking the unique contribution of auditory input, limiting the interpretation of audiovisual integration. Therefore, the difference wave approach, defined as [Diff = AV − (A+ V)], was used to quantify the degree of audiovisual integration. Consistent with previous studies (Crosse et al., 2015; Giard & Peronnet, 1999; Li et al., 2025; Mercier et al., 2013; Molholm et al., 2002; Senkowski et al., 2007; Stevenson et al., 2014; Talsma et al., 2007; Yang et al., 2025; Zhao et al., 2018), audiovisual integration was assessed by comparing the ERPs for audiovisual stimuli (ERP AV) with the linear sum of the responses to the respective unisensory constituents (ERP A+ ERP V). This analytical approach relies on additive models to detect interactions of nonlinear neural responses. If there is a difference in the amplitude of the ERP AV and ERP A+ ERP V, it indicates that audiovisual integration has occurred. To comprehensively describe the spatiotemporal characteristics of audiovisual integration under the three visual uncertainty conditions, the time-varying difference waves at each electrode under the three conditions were calculated. Pointwise *t*-tests were then used to determine whether the difference wave [Diff = AV − (A+ V)] was significant. The statistical results are presented as intensity plots showing electrode location, time, and t values. Given the findings in the literature that multisensory integration often occurs in the parietal and frontal cortices (Bizley et al., 2016; Cao et al., 2019; Giard & Peronnet, 1999; Noppeney, 2021), two ROIs were selected: frontal electrodes (Fz, F1, F2) and parietal electrodes (Pz, P1, P2). Afterward, ERP difference waves were computed and averaged within two predefined time windows: 140–220 msec (early window) and 300–380 msec (late window). For each time window, mean amplitudes were obtained by averaging across electrodes from two ROIs. The average amplitudes of the early and late phases of each ROI were subsequently subjected to paired sample *t* tests to statistically evaluate whether audiovisual integration occurred.

2.7. Correlation analysis between behavioral models and neural signals

To characterize the underlying cognitive mechanisms of perceptual decision-making, we adopted the classic DDM, a well-established framework for modeling binary choice behavior (Ratcliff, 1978; Ratcliff et al., 2016). The DDM assumes that decision-making is made through a noisy accumulation process, in which sensory evidence accumulates over time from a starting point, and a response is triggered when the accumulated evidence reaches the correct response boundary or the incorrect response boundary. By simultaneously fitting the full reaction time distribution and response accuracy, the model estimates latent parameters that reflect distinct cognitive processes. The standard DDM contains four key parameters: the relative starting point (z), drift rate (v), decision boundary (a), and non-decision time (Ter) (Ratcliff & McKoon, 2008; Voss et al., 2013). The relative starting point reflects initial decision bias, the drift rate represents the speed and quality of evidence accumulation, the decision boundary denotes the amount of information required to make a decision (capturing the speed–accuracy tradeoff), and the non-decision time (Ter) accounts for processing stages unrelated to evidence accumulation, including early sensory encoding (Ter1) and motor execution (Ter2). In this study, since there was no initial preference for either option at the beginning of the task, the starting point (z) was fixed at .5 to reflect an unbiased starting point in the decision-making process. In addition, given that the response accuracy under each experimental condition was controlled through the experimental design and to simplify the model structure and focus on the modeling of perceptual processing and evidence accumulation, this study set the decision boundary parameter (a) to remain unchanged across different experimental conditions. Instead, the drift rate (v) and non-decision time (Ter) were modeled as parameters that varied with experimental conditions to capture the potential effects of different perceptual states on evidence accumulation and non-decision processes. To consider the impact of extreme values, this study set a 5 % outlier response probability. This modeling strategy is based on previous studies focusing on perception or multisensory decision-making tasks (Léon Franzen et al., 2020; van den Brink et al., 2021). Model fitting was performed using the Hierarchical Drift Diffusion Model (HDDM) toolbox (Wiecki et al., 2013), which applies hierarchical Bayesian inference via Markov Chain Monte Carlo (MCMC) sampling. For each model, 10,000 posterior samples were drawn, with the first 10 % discarded as burn-in and every second sample retained to reduce autocorrelation. Model convergence was assessed based on four MCMC chains using the R-hat statistic (Gelman & Rubin, 1992), with values close to 1 and below 1.02 indicating that the model provided reliable parameter estimates. Model fit was assessed using a posterior predictive test (PPC). Data were simulated for each parameter value by sampling 500 values from the posterior distribution. This included a graphical comparison of the observed and predicted RT distributions and an assessment of the 95 % confidence interval coverage (Pan et al., 2025).

To test for differences across conditions, the hypothesis was tested for Bayesian parameter comparison on the posterior distributions of the parameters obtained from the HDDM. Using a Bayesian framework provides flexible and robust estimates and credible intervals for the derived parameters, enabling direct hypothesis testing of the posterior distribution (Kruschke, 2021). Bayesian inference was used to perform grouped parameter comparisons, and the results for the hypotheses of interest were reported using posterior probabilities, with posterior probabilities $\geq$.95 considered statistically significant. Furthermore, to link the behavioral modeling with the neural data, the evaluation focused on whether the DDM parameters were correlated with the acquired ERP signal features. Specifically, the ERP amplitudes from the early (140–220 msec) and late (300–380 msec) time windows were included as second-level predictors in the diffusion model specification. The resulting regression coefficients were used to investigate whether there was a significant relationship between the ERP amplitude and diffusion model parameters.

3. Results

3.1. Behavioral measures

A continuous adaptive staircase procedure was implemented to determine the accuracy of participants under high-, medium- and low-uncertainty unisensory conditions (mean ± SD: HV = 55.4 % ± 9.1 %, MV = 75.2 % ± 5.3 %, LV = 87.1 % ± 7.5 %, MA = 72.1 % ± 5.2 %). To systematically examine the effects of modality and uncertainty, a 2 (modality: visual, audiovisual) × 3 (uncertainty: high, medium, low) repeated-measures ANOVA was conducted on both accuracy and response time (Fig. 1B). The results revealed that the accuracy was greater for audiovisual stimuli than for visual stimuli [$F(1, 29) = 82.33$, $p < .001$, $\eta^2_p = .740$; mean ± SD: HAV = 72.2 % ± 8.7 %, MAV = 83.5 % ± 5.4 %, LAV = 90.1 % ± 5.1 %]. Uncertainty level also played a significant role [$F(2, 58) = 183.13$, $p < .001$, $\eta^2_p = .863$], with accuracy increasing progressively from high-to low-uncertainty conditions [L > M: $t(29) = 7.04$, $p < .001$, $d = 1.31$; M > H: $t(29) = 11.89$, $p < .001$, $d = 2.21$; L > H: $t(29) = 18.93$, $p < .001$, $d = 3.53$]. The modality × uncertainty interaction effect was significant [$F(2, 58) = 42.07$, $p < .001$, $\eta^2_p = .592$]. Simple effect analysis revealed that the accuracy was greater for audiovisual stimuli than for visual stimuli under high- [$t(29) = 12.41$, $p < .001$, $d = 2.39$] and medium- [$t(29) = 6.11$, $p < .001$, $d = 1.18$] uncertainty conditions, but the difference was not significant under low-uncertainty conditions [$t(29) = 2.17$, $p = .100$, $d = .42$]. In addition, the results revealed that the response time was shorter for audiovisual stimuli than for visual stimuli [$F(1, 29) = 17.92$, $p < .001$, $\eta^2_p = .382$; mean ± SD: HV = 672 ± 142 msec, MV = 646 ± 108 msec, LV = 595 ± 75 msec, HAV = 649 ± 138 msec, MAV = 611 ± 111 msec, LAV = 568 ± 71 msec; *Ps* < .001] and increased progressively from high-to low-uncertainty conditions [$F(2, 58) = 20.95$, $p < .001$, $\eta^2_p = .419$; L < M: $t(29) = 3.84$, $p < .001$, $d = .42$; M < H: $t(29) = 2.59$, $p = .012$, $d = .29$; L < H: $t(29) = 6.43$, $p < .001$, $d = .71$]. The modality × uncertainty interaction was not significant [$F(2, 58) = 1.11$, $p = .327$, $\eta^2_p = .037$]. These findings revealed the advantages of multisensory stimuli in improving behavior performance, especially under conditions of high and medium visual uncertainty.

3.2. EEG recordings

To investigate the neural effects of visual uncertainty on multisensory decision-making, EEG recordings were collected and analyzed using event-related potentials (ERPs; Fig. 2A). ERP amplitudes were subjected to a 2 (modality: visual, audiovisual) × 3 (uncertainty: high, medium, low) repeated-measures ANOVA for two time windows representing distinct stages of processing (Fig. 2B). In the early sensory encoding window (140–220 msec), the ERP amplitude was greater for audiovisual stimuli than for visual stimuli [$F(1, 29) = 92.22$, $p < .001$, $\eta^2_p = .761$]. Neither a significant main effect of uncertainty [$F(2, 58) = 3.10$, $p = .067$, $\eta^2_p = .097$] nor an interaction effect between modality and uncertainty [$F(2, 58) = 1.63$, $p = .206$, $\eta^2_p = .053$] was found. In the late decision formation window (300–380 msec), the ERP amplitude was greater for audiovisual stimuli than for visual stimuli [$F(1, 29) = 24.26$, $p < .001$, $\eta^2_p = .456$] and increased progressively from high-uncertainty to low-uncertainty conditions [$F(1, 29) = 23.28$, $p < .001$, $\eta^2_p = .445$; L < M: $t(29) = 4.39$, $p < .001$, $d = .35$; M < H: $t(29) = 2.33$, $p = .023$, $d = .19$; L < H: $t(29) = 6.72$, $p < .001$, $d = .54$]. No significant interaction effect between modality and uncertainty was found [$F(2, 58) = .05$, $p = .948$, $\eta^2_p = .002$]. These results indicate that neural responses are driven primarily by modality with minimal influence from uncertainty during the early sensory encoding stage, whereas modality and uncertainty exert combined effects during the later decision formation stage.

3.3. Neural encoding of audiovisual integration

To further determine the temporal dynamics of audiovisual integration and its modulation by visual uncertainty, EEG data were analyzed by comparing the neural response to audiovisual stimuli with the summed response to the respective unisensory components. This difference wave [Diff = AV − (A+ V)] served as a quantitative index of multisensory integration. One-sample *t* tests were conducted to assess whether the difference significantly differed from zero under high-, medium-, and low-uncertainty conditions across two time windows and two regions of interest (ROIs; Fig. 2C–E). In the early time window (140–220 msec), significant differences were observed in the frontal region under high- [$t(29) = 6.08$, $p < .001$, $d = 1.11$] medium- [$t(29) = 4.86$, $p < .001$, $d = .89$] and low- [$t(29) = 4.04$, $p < .001$, $d = .74$] uncertainty conditions. Significant differences were also observed at the parietal electrodes under high- [$t(29) = 5.70$, $p < .001$, $d = 1.04$] medium- [$t(29) = 7.72$, $p < .001$, $d = 1.41$] and low- [$t(29) = 7.09$, $p < .001$, $d = 1.29$] uncertainty conditions. In contrast, in the late decision formation window (300–380 msec), significant differences under high- [$t(29) = 3.03$, $p = .005$, $d = .55$] and medium- [$t(29) = 2.38$, $p = .024$, $d = .43$] uncertainty conditions but not under low- [$t(29) = 1.47$, $p = .152$, $d = .27$] uncertainty conditions were observed at the frontal electrodes. At the parietal electrodes, no significant differences were found in the high- [$t(29) = 1.35$, $p = .187$, $d = .25$], medium- [$t(29) = 1.37$, $p = .180$, $d = .25$], or low- [$t(29) = .53$, $p = .603$, $d = .10$] uncertainty conditions. These results reveal a temporal and spatial dissociation: audiovisual integration during early sensory encoding is consistently present and widespread across uncertainty conditions, whereas audiovisual integration during late decision formation is more selective, occurring only under increased uncertainty and localized to the frontal electrodes. This pattern suggests that visual uncertainty predominantly modulates audiovisual integration at the later decision-making stage, likely reflecting adaptive adjustments in high-level processing rather than changes in early sensory encoding.

3.4. Relationship between the behavior model and the neural responses

A drift–diffusion model was used to gain a better understanding of the decision-making process (Fig. 3A). Posterior predictive checks indicated that the model adequately captured the distribution of the observed response times, as

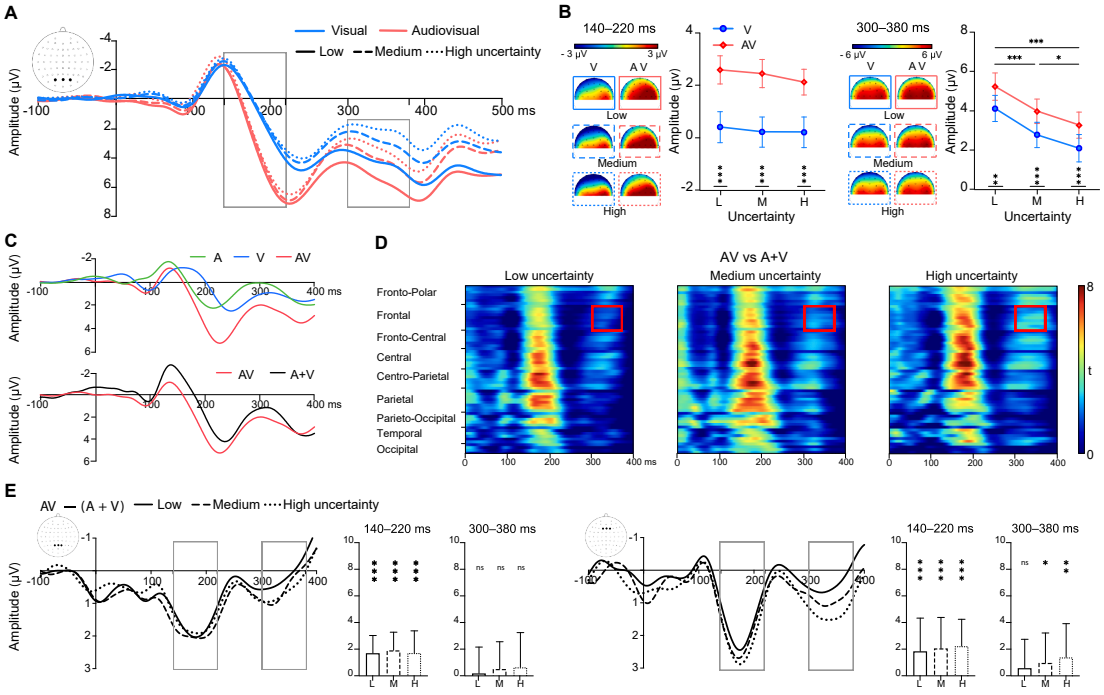

Fig. 2 – ERP waveforms and audiovisual integration under different visual uncertainty conditions. A. Grand-averaged ERP waveform for visual and audiovisual conditions at electrodes POz, PO3, and PO4. The shaded areas represent the time windows of the early sensory stage (140–220 msec) and late decision formation stage (300–380 msec). B. Topographic mapping and ERP amplitude data for each time window. C. Example ERP waveforms for audiovisual stimuli (ERP AV) and the linear sum of the responses to the respective unisensory constituents (ERP A+ ERP V) at the P4 electrode. Audiovisual integration is measured as the difference in amplitude between the ERP AV and ERP A+ ERP V. D. Spatiotemporal plot of t values for the difference waveform [Diff = AV − (A+ V)]. The electrodes within each section are arranged from the left-most lateral to the right-most lateral sites and from top to bottom. The red boxes indicate audiovisual integration at the frontal electrodes in the late decision-making stage. E. Average difference in waveforms at the Pz, P1, and P2 electrodes (left) and the Fz, F1, and F2 electrodes (right). ns: not significant.

evidenced by the close correspondence between the observed histograms and the posterior predictive distributions. The observed data fell within the predicted 95 % credible intervals (Fig. 3-B). Moreover, the observed data fell within the predicted 95 % credible intervals. The results revealed that non-decision time was not significantly different among the low-, medium- or high-uncertainty conditions for the visual and audiovisual conditions [$P_{posterior}$(L > M, V) = .733, $P_{posterior}$(M > H, V) = .590, $P_{posterior}$(L > H, V) = .810; $P_{posterior}$(L > M, AV) = .670, $P_{posterior}$(M > H, AV) = .419, $P_{posterior}$(L > H, AV) = .595]. However, non-decision time was greater for the visual condition than for the audiovisual condition in the low-, medium- and high-uncertainty conditions [$P_{posterior}$(V > AV, L) = .994, $P_{posterior}$(V > AV, M) = .992, $P_{posterior}$(V > AV, H) = .971]. In contrast, the drift rate increased progressively from the high-to low-uncertainty conditions for visual and audiovisual conditions [$P_{posterior}$(L > M, V) = 1, $P_{posterior}$(M > H, V) = 1, $P_{posterior}$(L > H, V) = 1; $P_{posterior}$(L > M, AV) = 1, $P_{posterior}$(M > H, AV) = 1, $P_{posterior}$(L > H, AV) = 1]. The drift rate was higher for the audiovisual condition than for the visual condition under medium- and high-uncertainty conditions [$P_{posterior}$(AV > V, M) = .966, $P_{posterior}$(AV > V, H) = 1], but not in the low-uncertainty condition [$P_{posterior}$(AV > V, L) = .815; Fig. 3C]. To examine the neural correlates of these DDM parameters, we calculated regression coefficients between ERP amplitudes and model estimates (Fig. 3D). In the early time window (140–220 msec), the ERP amplitude was significantly negatively correlated with the non-decision time ($\beta = -.013$, $P_{posterior} = 1$) and was significantly positively correlated with the drift rate ($\beta = .144$, $P_{posterior} = 1$). In the late decision formation window (300–380 msec), the ERP amplitude was not correlated with the non-decision time ($\beta = -.004$, $P_{posterior} = .895$) but was positively correlated with the drift rate ($\beta = .328$, $P_{posterior} = 1$). These results showed that visual uncertainty in the audiovisual context selectively modulates the drift rate without affecting non-decision time. Moreover, the results revealed a temporal dissociation. Early ERP activity was associated with both non-decision processes and evidence accumulation, whereas late ERP activity was associated with decision-related evidence accumulation.

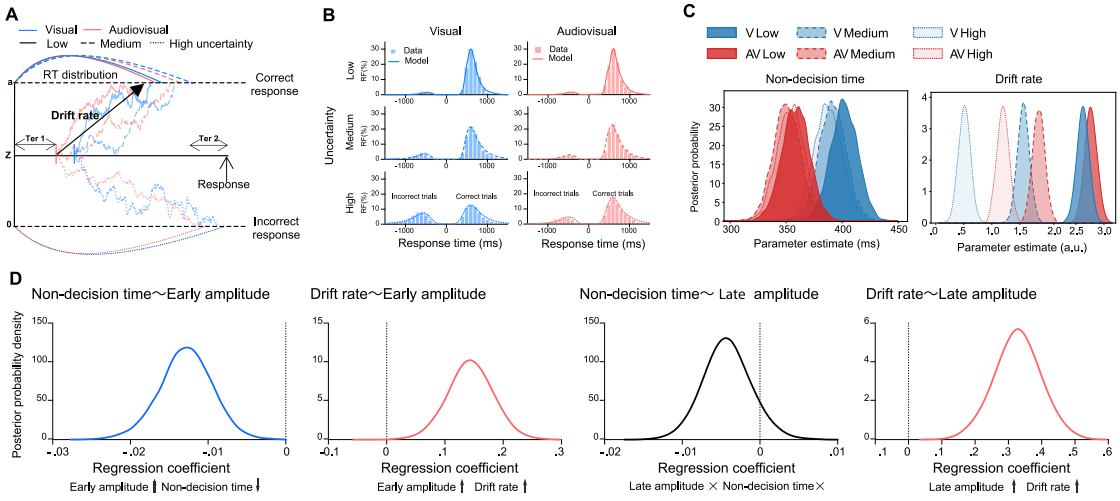

Fig. 3 – Behavioral modeling and neural–behavioral correlations. A. Schematic of the drift–diffusion model (DDM). Sensory evidence accumulates over time at a drift rate (v) toward a decision boundary (0 or a). Non-decision time (Ter) reflects processes outside decision formation, including early sensory encoding (Ter1) and motor execution (Ter2). The starting point (z) was fixed at a neutral value. B. HDDM fit (solid line) to the observed reaction time distributions (histograms) across participants. The figure includes incorrect (left) and correct (right) trials, with higher histogram values on the right indicating a greater proportion of correct choices. RF: response frequency. C. Posterior distributions of HDDM parameters. D. Bayesian posterior densities of non-decision time and drift rates estimated from the drift–diffusion model and how they varied as a function of early and late ERP amplitude. The peaks of the distributions reflect the most likely value of the parameter. Significance was assessed by at least 95 % of the distribution being to the left or right of zero.

4. Discussion

Multisensory input is critical for forming a unified and coherent perception of unfolding events, considerably improving both the speed and accuracy of behavioral performance. A key question in multisensory decision-making is how increasing visual uncertainty influences audiovisual integration. In our study, early audiovisual integration during the sensory encoding stage occurred robustly across all visual uncertainty conditions. In contrast, late audiovisual integration during the decision formation stage occurred under medium and high visual uncertainty conditions but not under low visual uncertainty conditions. By dissociating the temporal stages of integration, our study overcomes the limitations of traditional studies that focused primarily on early sensory integration and highlight the importance of late decision integration mechanisms. These results not only provide neurophysiological evidence for the current controversy on the stage and mechanism of multisensory integration but also expand our understanding of how multisensory integration adapts to dynamic changes in uncertain sensory information.

As observed in this study, visual uncertainty selectively affects audiovisual integration in the late stages but not in the early stages; this process may be interpreted using several theoretical frameworks. According to one, audiovisual integration involves stimulus-driven bottom-up and cognitive top-down processing in the context of multisensory decision-making (Keil & Senkowski, 2018; Macaluso et al., 2016; Senkowski & Engel, 2024; Talsma, 2015). Early-stage integration is generally stimulus driven and governed by bottom-up processes. Electrophysiological studies have shown that early audiovisual integration is influenced by basic sensory features approximately before 200 msec, such as visual contrast and auditory frequency, rather than sensory evidence or information content from the stimuli (Boyle et al., 2017; Senkowski et al., 2011; Yang et al., 2015). Alternatively, audiovisual integration in the early stages depends mainly on the temporal synchrony and spatial redundancy of multisensory input (Bertelson et al., 2000; Van der Burg et al., 2008; Vroomen et al., 2001). However, late-stage integration involves top-down mechanisms that assess the reliability of each sensory modality and dynamically allocate attentional resources (Donohue et al., 2015; Talsma et al., 2007; Tang et al., 2016). When unisensory uncertainty increases, top-down attentional control is recruited to reweight multisensory inputs on the basis of relative reliability (Van Ee et al., 2009). Under medium and high visual uncertainty conditions, the cognitive system compensates by increasing reliance on auditory information to make a decision (Alsius et al., 2005; Vercillo & Gori, 2015), thus promoting enhanced audiovisual integration during the decision-making stage. Such an allocation strategy reflects an adaptive response of the cognitive system to uncertainty in a variable multisensory environment.

Another alternative theoretical framework that may explain these findings is based on Bayesian causal inference and holds that the brain integrates or separates multisensory input on the basis of unisensory reliability and the probability of whether it comes from the same event source. This inference is dynamic and depends on the present sensations,

previous knowledge and causal structure inference (French & DeAngelis, 2020; Ma et al., 2023; Rohe et al., 2019). In the early stages, the brain tends to automatically integrate visual and auditory stimuli via temporal and spatial synchronization, a process referred to as forced fusion. This automatic integration occurs without causal structure judgments and is relatively insensitive to the amount of information in the visual stimuli, thereby explaining the robust early audiovisual integration observed across all visual uncertainty conditions. In contrast, the brain estimates the reliability of different unisensory and the probability of common causality to make optimal perceptual decisions in the late stages. Under low visual uncertainty, highly reliable visual input can almost independently drive perceptual decisions, leading the brain to treat auditory and visual signals as stemming from separate sources, thus reducing the need for integration (Odegaard et al., 2015; Shams & Beierholm, 2010). Under medium and high uncertainty, the reduced reliability of visual input prompts the brain to infer a common causal source and to integrate auditory information, minimizing the risk of disregarding the uncertainty of potentially relevant sensory evidence (Körding et al., 2007; Körding & Wolpert, 2004). The brain reassesses the causal structure, increasing the level of trust and willingness to integrate the auditory sense, thus activating late integration.

With respect to temporal dynamics, prior EEG studies have shown that the neural correlates of Bayesian causal inference emerge relatively late. EEG representational similarity analyses revealed that from approximately 200–400 msec post-stimulus, neural patterns reflected estimates of Bayesian causal inference estimates (Rohe et al., 2019). Furthermore, multivariate EEG decoding indicated that audiovisual integration was consistent with Bayesian causal inference at approximately 350 msec after the stimulus (Aller & Noppeney, 2019). These findings are consistent with the ERP time-domain analysis results of the present study, which revealed that increasing visual uncertainty affected audiovisual integration from 300 to 380 msec in the late decision stage. Notably, a recent study used supervised machine learning to analyze simultaneously recorded electroencephalograms (EEGs) and identified the neural signatures of two processing stages: sensory encoding and decision formation. Generalization analyses across experimental conditions and time further suggest an ongoing dynamic interplay between multisensory integration and decision-making processes (Mercier & Cappe, 2020). This means that even if no robust differences due to visual uncertainty were detected in the early window we defined, subtle fluctuations from the early period may still extend and affect subsequent decision-related processes. Together, these converging findings highlight that while visual uncertainty in audiovisual context predominantly modulates evidence accumulation and decision-related dynamics at later stages, early sensory influences cannot be entirely excluded and may interact with subsequent Bayesian weighting processes during multisensory decision formation.

The effect of visual uncertainty on audiovisual integration was concentrated in the frontal electrodes, which may reflect a driving role of the frontal cortex in general reasoning and adaptive behavior in the context of multisensory integration (Calvert, 2001; Noppeney et al., 2010; Rushworth et al., 2011). The frontal cortex has been shown to form beliefs about inferential states that are based on expectations and prior experience (Gau & Noppeney, 2016; Noppeney et al., 2010). Therefore, the role of the frontal cortex is not only to integrate sensory information but also to weigh competing strategies for forming the most appropriate sensory representation to guide behavior. In contrast, no audiovisual integration was observed at the parietal electrodes in the late stage, which is consistent with a general understanding of the functional differences between the parietal cortex and frontal cortex (Buschman & Miller, 2007; Erlich et al., 2015; Hanks et al., 2015). Although both cortices are involved in the accumulation of sensory information, the parietal cortex potentially encodes more of the sensory aspect, whereas the frontal cortex is responsible for the context-dependent transformation of this information to select the most appropriate response strategy. This functional differentiation provides support for understanding the stage-by-stage role of sensory integration in multisensory decision-making.

This study combined the analysis of behavioral models and neural indicators to reveal the correspondence between the DDM parameters and neural activity. Our behavioral model results revealed that audiovisual stimuli decreased the non-decision time, which is consistent with the findings of existing studies (Chau et al., 2021; Mercier & Cappe, 2020; Murray et al., 2020). Further analysis revealed that visual uncertainty selectively affects the drift rate but not the non-decision time, suggesting that visual uncertainty mainly affects the information accumulation process in the decision stage. In terms of neural data, late neural components (300–380 msec) were significantly correlated with the drift rate, which is consistent with the view that late components reflect post-perception or decision-related processing (Philiastides et al., 2006; Philiastides & Sajda, 2006; Ratcliff et al., 2009). In addition, early neural components (140–220 msec) were significantly correlated with the drift rate and non-decision time. Previous studies have shown that N1 is associated with the rate of information accumulation during decisional processing (Sui et al., 2023), and P200 and N200 explain single-trial evidence accumulation and preprocessing times (Nunez & Srinivasan, 2017). These findings suggest that early neural activity may play a dual role in perceptual decision-making: promoting faster stimulus encoding while also enhancing the efficiency of subsequent evidence accumulation. Nevertheless, some studies have reported insufficient consistency between DDM parameters and neural signals (McGovern et al., 2018; Purcell & Palmeri, 2017), which may be because neural decision-making processes are affected by the involvement of different brain regions, speed-accuracy trade-off strategies (Spieser et al., 2018), and multiple motion representations (Kohl et al., 2019). These factors may lead to neural activity that does not fully conform to the ideal assumptions of the DDM. Despite the above differences, our results still provide strong support for the establishment of a mapping between DDM parameters and ERP components in the context of multisensory decision-making, highlighting the theoretical value of combining DDM with neural

indicators under specific experimental conditions (O'Connell et al., 2018; Purcell & Palmeri, 2017).

Here, we used neural signals and behavior modeling measurements to systematically explore how visual uncertainty influences the specific processing stages involved in multisensory decision-making. The results revealed that visual uncertainty primarily affects the late decision formation stage and has less of an effect on the early sensory encoding stage. The synergy of multiple processing mechanisms, including bottom-up sensory input, top-down attentional regulation, and information weighting based on Bayesian causal inference, might be required when integrating audiovisual inputs under varying visual uncertainty conditions. Processing patterns involving multiple mechanisms reveal the flexibility and adaptability of cognitive systems when multisensory integration is performed in complex environments. Future studies should explore how visual uncertainty is represented in the neural networks of different brain regions, especially the connection patterns between the primary cortex and the parietal-frontal cortex.

CRediT authorship contribution statement

Xiangfu Yang: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Weiping Yang:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Conceptualization. **Yinghua Yu:** Writing – review & editing, Supervision, Investigation. **Yoshimichi Ejima:** Writing – review & editing, Supervision, Investigation. **Jiajia Yang:** Writing – review & editing, Validation, Resources, Project administration, Funding acquisition, Conceptualization.

Consent to participate

Informed consent was obtained from all the individual participants included in the study.

Consent to publish

The authors are responsible for the correctness of the statements provided in the manuscript. If accepted, it can be published in full.

Ethics approval

The questionnaire and methodology for this study were approved by the ethics committee of the university.

Data availability

All the data are true and have original data. The datasets and materials generated during the current study are available from the corresponding author upon reasonable request.

Funding

This work was supported by the JST FOREST Program (JPMJFR2041) KAKENHI (JP25K02549).

Declaration of competing interest

All the authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgement

We thank all the participants for their participation in the study.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2025.11.005.

Scientific transparency statement

DATA: All raw and processed data supporting this research are publicly available: https://osf.io/s3p6g/

CODE: All analysis code supporting this research is publicly available: https://osf.io/s3p6g/

MATERIALS: All study materials supporting this research are publicly available: https://osf.io/s3p6g/

DESIGN: This article reports, for all studies, how the author(s) determined all sample sizes, all data exclusions, all data inclusion and exclusion criteria, and whether inclusion and exclusion criteria were established prior to data analysis.

PRE-REGISTRATION: No part of the study procedures was pre-registered in a time-stamped, institutional registry prior to the research being conducted. No part of the analysis plans was pre-registered in a time-stamped, institutional registry prior to the research being conducted.

For full details, see the *Scientific Transparency Report* in the supplementary data to the online version of this article.

REFERENCES

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Aller, M., & Noppeney, U. (2019). To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference. *PLoS Biology, 17*(4), Article e3000210. https://doi.org/10.1371/journal.pbio.3000210

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology, 15*(9), 839–843. https://doi.org/10.1016/j.cub.2005.03.046

Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of

deliberate visual attention. *Perception and Psychophysics, 62*(2), 321–332. https://doi.org/10.3758/BF03205552
Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience, 7*(3), 295–301. https://doi.org/10.1038/nn1198
Bizley, J. K., Jones, G. P., & Town, S. M. (2016). Where are multisensory signals combined for perceptual decision-making? *Current Opinion in Neurobiology, 40*, 31–37. https://doi.org/10.1016/j.conb.2016.06.003
Boyle, S. C., Kayser, S. J., & Kayser, C. (2017). Neural correlates of multisensory reliability and perceptual weights emerge at early latencies during audio-visual integration. *European Journal of Neuroscience, 46*(10), 2565–2577. https://doi.org/10.1111/ejn.13724
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. https://www.ncbi.nlm.nih.gov/pubmed/9176952.
Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science, 315*(5820), 1860–1862. https://doi.org/10.1126/science.1138071
Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex, 11*(12), 1110–1123. https://doi.org/10.1093/cercor/11.12.1110
Cao, Y., Summerfield, C., Park, H., Giordano, B. L., & Kayser, C. (2019). Causal inference in the multisensory brain. *Neuron, 102*(5), 1076–1087. https://doi.org/10.1016/j.neuron.2019.03.043. e1078.
Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: Timing, topography, directionality, and sources. *Journal of Neuroscience, 30*(38), 12572–12580. https://doi.org/10.1523/JNEUROSCI.1099-10.2010
Chau, E., Murray, C. A., & Shams, L. (2021). Hierarchical drift diffusion modeling uncovers multisensory benefit in numerosity discrimination tasks. *PeerJ, 9*, Article e12273. https://doi.org/10.7717/peerj.12273
Colavita, F. B. (1974). Human sensory dominance. *Perception and Psychophysics, 16*(2), 409–412. https://doi.org/10.3758/BF03203962
Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. *Journal of Neuroscience, 35*(42), 14195–14204. https://doi.org/10.1523/JNEUROSCI.1829-15.2015
Donohue, S. E., Green, J. J., & Woldorff, M. G. (2015). The effects of attention on the temporal integration of multisensory stimuli. *Frontiers in Integrative Neuroscience, 9*, 32. https://doi.org/10.3389/fnint.2015.00032
Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E., & Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. *eLife, 3*, Article e03005. https://doi.org/10.7554/eLife.03005
Erlich, J. C., Brunton, B. W., Duan, C. A., Hanks, T. D., & Brody, C. D. (2015). Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. *eLife, 4*, Article e05457. https://doi.org/10.7554/eLife.05457
Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*(4), 162–169. https://doi.org/10.1016/j.tics.2004.02.002
Franzen, L., Delis, I., De Sousa, G., Kayser, C., & Philiastides, M. G. (2020). Auditory information enhances post-sensory visual evidence during rapid multisensory decision-making. *Nature Communications, 11*(1), 5440. https://doi.org/10.1038/s41467-020-19306-7
French, R. L., & DeAngelis, G. C. (2020). Multisensory neural processing: From cue integration to causal inference. *Current Opinion in Physiology, 16*, 8–13. https://doi.org/10.1016/j.cophys.2020.04.004
Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *Neuroimage, 124*, 876–886. https://doi.org/10.1016/j.neuroimage.2015.09.045
Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–472. https://doi.org/10.1214/ss/1177011136
Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473–490. https://doi.org/10.1162/089892999563544
Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., & Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature, 520*(7546), 220–223. https://doi.org/10.1038/nature14066
Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience, 9*(6), 467–479. https://doi.org/10.1038/nrn2374
Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One, 2*(9), Article e943. https://doi.org/10.1371/journal.pone.0000943
Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature, 427*(6971), 244–247. https://doi.org/10.1038/nature02169
Kayser, S. J., Philiastides, M. G., & Kayser, C. (2017). Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations. *Neuroimage, 148*, 31–41. https://doi.org/10.1016/j.neuroimage.2017.01.010
Keil, J., & Senkowski, D. (2018). Neural oscillations orchestrate multisensory processing. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry, 24*(6), 609–626. https://doi.org/10.1177/1073858418755352
Kohl, C., Spieser, L., Forster, B., Bestmann, S., & Yarrow, K. (2019). The neurodynamic decision variable in human multi-alternative perceptual choice. *Journal of Cognitive Neuroscience, 31*(2), 262–277. https://doi.org/10.1162/jocn_a_01347
Kosciessa, J. Q., Lindenberger, U., & Garrett, D. D. (2021). Thalamocortical excitability modulation guides human perception under uncertainty. *Nature Communications, 12*(1), 2430. https://doi.org/10.1038/s41467-021-22511-7
Kruschke, J. K. (2021). Bayesian analysis reporting guidelines. *Nature Human Behaviour, 5*(10), 1282–1291. https://doi.org/10.1038/s41562-021-01177-7
Kupers, E. R., Carrasco, M., & Winawer, J. (2019). Modeling visual performance differences "around" the visual field: A computational observer approach. *PLoS Computational Biology, 15*(5), Article e1007063. https://doi.org/10.1371/journal.pcbi.1007063
Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*(2B), 467–477. https://www.ncbi.nlm.nih.gov/pubmed/5541744.
Li, Z., Yang, W., Li, R., Luo, R., Yang, J., & Ren, Y. (2025). Beyond facilitating unisensory processing: Crossmodal associative memory training further modulates sensory integration. *Biological Psychology*. , Article 108995. https://doi.org/10.1016/j.biopsycho.2025.108995
Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research, 1173*, 102–109. https://doi.org/10.1016/j.brainres.2007.07.050
Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., & Epskamp, S. (2019). JASP: Graphical statistical software for common

statistical designs. *Journal of Statistical Software, 88*, 1–17. https://doi.org/10.18637/jss.v088.i02

Ma, W. J., Körding, K. P., & Goldreich, D. (2023). *Bayesian models of perception and action: An introduction*. MIT Press. ISBN: 9780262048378 (Hardcover), 9780262372824 (eBook).

Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J., & Adam, R. (2016). The curious incident of attention in multisensory integration: Bottom-up *vs* top-down. *Multisensory Research, 29*(6–7), 557–583. https://doi.org/10.1163/22134808-00002528

McGovern, D. P., Hayes, A., Kelly, S. P., & O'Connell, R. G. (2018). Reconciling age-related changes in behavioural and neural indices of human perceptual decision-making. *Nature Human Behaviour, 2*(12), 955–966. https://doi.org/10.1038/s41562-018-0465-6

Mercier, M. R., & Cappe, C. (2020). The interplay between multisensory integration and perceptual decision making. *Neuroimage, 222*, Article 116970. https://doi.org/10.1016/j.neuroimage.2020.116970

Mercier, M. R., Foxe, J. J., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H., & Molholm, S. (2013). Auditory-driven phase reset in visual cortex: Human electrocorticography reveals mechanisms of early multisensory integration. *Neuroimage, 79*, 19–29. https://doi.org/10.1016/j.neuroimage.2013.04.060

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. *Cognitive Brain Research, 14*(1), 115–128. https://doi.org/10.1016/S0926-6410(02)00066-6

Murai, Y., & Yotsumoto, Y. (2018). Optimal multisensory integration leads to optimal time estimation. *Scientific Reports, 8*(1), Article 13068. https://doi.org/10.1038/s41598-018-31468-5

Murray, C. A., de Larrea-Mancera, E. S. L., Glicksohn, A., Shams, L., & Seitz, A. R. (2020). Revealing multisensory benefit with diffusion modeling. *Journal of Mathematical Psychology, 99*, Article 102449. https://doi.org/10.1016/j.jmp.2020.102449

Noppeney, U. (2021). Perceptual inference, learning, and attention in a multisensory world. *Annual Review of Neuroscience, 44*, 449–473. https://doi.org/10.1146/annurev-neuro-100120-085519

Noppeney, U., Ostwald, D., & Werner, S. (2010). Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *Journal of Neuroscience, 30*(21), 7434–7446. https://doi.org/10.1523/JNEUROSCI.0455-10.2010

Nunez, M. D., Vandekerckhove, J., & Srinivasan, R. (2017). How attention influences perceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters. *Journal of Mathematical Psychology, 76*, 117–130. https://doi.org/10.1016/j.jmp.2016.03.003

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience, 15*(12), 1729–1735. https://doi.org/10.1038/nn.3248

O'Connell, R. G., Shadlen, M. N., Wong-Lin, K., & Kelly, S. P. (2018). Bridging neural and computational viewpoints on perceptual decision-making. *Trends in Neurosciences, 41*(11), 838–852. https://doi.org/10.1016/j.tins.2018.06.005

Odegaard, B., Wozny, D. R., & Shams, L. (2015). Biases in visual, auditory, and audiovisual perception of space. *PLoS Computational Biology, 11*(12), Article e1004649. https://doi.org/10.1371/journal.pcbi.1004649

Otto, T. U., & Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making. *Current Biology, 22*(15), 1391–1396. https://doi.org/10.1016/j.cub.2012.05.031

Pan, W. K., Geng, H. Y., Zhang, L., Fengler, A., Frank, M. J., Zhang, R. Y., & Hu, C. P. (2025). dockerHDDM: A user-friendly environment for Bayesian hierarchical drift-diffusion modeling. *Advances in Methods and Practices in Psychological Science, 8*(1), 26. https://doi.org/10.1177/25152459241298700

Parise, C. V., & Ernst, M. O. (2017). Noise, multisensory integration, and previous response in perceptual disambiguation. *PLoS Computational Biology, 13*(7), Article e1005546. https://doi.org/10.1371/journal.pcbi.1005546

Pasqualotto, A. (2016). Multisensory integration substantiates distributed and overlapping neural networks. *Behavioral and Brain Sciences, 39*, Article e127. https://doi.org/10.1017/S0140525X15001612

Philiastides, M. G., Heekeren, H. R., & Sajda, P. (2014). Human scalp potentials reflect a mixture of decision-related signals during perceptual choices. *Journal of Neuroscience, 34*(50), 16877–16889. https://doi.org/10.1523/JNEUROSCI.3012-14.2014

Philiastides, M. G., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: A timing diagram. *Journal of Neuroscience, 26*(35), 8965–8975. https://doi.org/10.1523/JNEUROSCI.1655-06.2006

Philiastides, M. G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex, 16*(4), 509–518. https://doi.org/10.1093/cercor/bhi130

Polanía, R., Krajbich, I., Grueschow, M., & Ruff, C. C. (2014). Neural oscillations and synchronization differentially support evidence accumulation in perceptual and value-based decision making. *Neuron, 82*(3), 709–720. https://doi.org/10.1016/j.neuron.2014.03.014

Purcell, B. A., & Palmeri, T. J. (2017). Relating accumulator model parameters and neural dynamics. *Journal of Mathematical Psychology, 76*, 156–171. https://doi.org/10.1016/j.jmp.2016.07.001

Raposo, D., Sheppard, J. P., Schrater, P. R., & Churchland, A. K. (2012). Multisensory decision-making in rats and humans. *Journal of Neuroscience, 32*(11), 3726–3735. https://doi.org/10.1523/JNEUROSCI.4998-11.2012

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences of the USA, 106*(16), 6539–6544. https://doi.org/10.1073/pnas.0812589106

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences, 20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Regenbogen, C., Seubert, J., Johansson, E., Finkelmeyer, A., Andersson, P., & Lundström, J. N. (2018). The intraparietal sulcus governs multisensory integration of audiovisual information based on task difficulty. *Human Brain Mapping, 39*(3), 1313–1326. https://doi.org/10.1002/hbm.23918

Rohe, T., Ehlis, A. C., & Noppeney, U. (2019). The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nature Communications, 10*(1), 1907. https://doi.org/10.1038/s41467-019-09664-2

Rohe, T., & Noppeney, U. (2015a). Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biology, 13*(2), Article e1002073. https://doi.org/10.1371/journal.pbio.1002073

Rohe, T., & Noppeney, U. (2015b). Sensory reliability shapes perceptual inference via two mechanisms. *Journal of Vision, 15*(5), 22. https://doi.org/10.1167/15.5.22

Rohe, T., & Noppeney, U. (2016). Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Current Biology, 26*(4), 509–514. https://doi.org/10.1016/j.cub.2015.12.056

Rushworth, M. F., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron, 70*(6), 1054–1069. https://doi.org/10.1016/j.neuron.2011.05.014

Scheliga, S., Kellermann, T., Lampert, A., Rolke, R., Spehr, M., & Habel, U. (2023). Neural correlates of multisensory integration in the human brain: An ALE meta-analysis. *Reviews in the Neurosciences, 34*(2), 223–245. https://doi.org/10.1515/revneuro-2022-0065

Schmid, C., Büchel, C., & Rose, M. (2011). The neural basis of visual dominance in the context of audio-visual object processing. *Neuroimage, 55*(1), 304–311. https://doi.org/10.1016/j.neuroimage.2010.11.051

Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level 'unisensory' processing. *Current Opinion in Neurobiology, 15*(4), 454–458. https://doi.org/10.1016/j.conb.2005.06.008

Senkowski, D., & Engel, A. K. (2024). Multi-timescale neural dynamics for multisensory integration. *Nature Reviews Neuroscience, 25*(9), 625–642. https://doi.org/10.1038/s41583-024-00845-7

Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *Neuroimage, 56*(4), 2200–2208. https://doi.org/10.1016/j.neuroimage.2011.03.075

Senkowski, D., Saint-Amour, D., Kelly, S. P., & Foxe, J. J. (2007). Multisensory processing of naturalistic objects in motion: A high-density electrical mapping and source estimation study. *Neuroimage, 36*(3), 877–888. https://doi.org/10.1016/j.neuroimage.2007.01.053

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences, 14*(9), 425–432. https://doi.org/10.1016/j.tics.2010.07.001

Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita effect revisited. *Perception & Psychophysics, 69*(5), 673–686. https://doi.org/10.3758/bf03193770

Spieser, L., Kohl, C., Forster, B., Bestmann, S., & Yarrow, K. (2018). Neurodynamic evidence supports a forced-excursion model of decision-making under speed/accuracy instructions. *eNeuro, 5*(3). https://doi.org/10.1523/ENEURO.0159-18.2018

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9*(4), 255–266. https://doi.org/10.1038/nrn2331

Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., Kurela, L. R., Siemann, J. K., James, T. W., & Wallace, M. T. (2014). Identifying and quantifying multisensory integration: A tutorial review. *Brain Topography, 27*(6), 707–730. https://doi.org/10.1007/s10548-014-0365-7

Sui, J., He, X., Golubickis, M., Svensson, S. L., & Macrae, C. N. (2023). Electrophysiological correlates of self-prioritization. *Consciousness and Cognition, 108*, Article 103475. https://doi.org/10.1016/j.concog.2023.103475

Talsma, D. (2015). Predictive coding and multisensory integration: An attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience, 9*, 19. https://doi.org/10.3389/fnint.2015.00019

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex, 17*(3), 679–690. https://doi.org/10.1093/cercor/bhk016

Tang, X., Wu, J., & Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neuroscience and Biobehavioral Reviews, 61*, 208–224. https://doi.org/10.1016/j.neubiorev.2015.11.002

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley. ISBN-13: 978-0201076165.

van den Brink, R. L., Murphy, P. R., Desender, K., de Ru, N., & Nieuwenhuis, S. (2021). Temporal expectation hastens decision onset but does not affect evidence quality. *Journal of Neuroscience, 41*(1), 130–143. https://doi.org/10.1523/JNEUROSCI.1103-20.2020

Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 34*(5), 1053–1065. https://doi.org/10.1037/0096-1523.34.5.1053

Van Ee, R., Van Boxtel, J. J., Parker, A. L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *Journal of Neuroscience, 29*(37), 11641–11649. https://doi.org/10.1523/JNEUROSCI.0873-09.2009

Vercillo, T., & Gori, M. (2015). Attention to sound improves auditory reliability in audio-tactile spatial optimal integration. *Frontiers in Integrative Neuroscience, 9*, 34. https://doi.org/10.3389/fnint.2015.00034

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology, 60*(6), 385–402. https://doi.org/10.1027/1618-3169/a000218

Vroomen, J., Bertelson, P., & De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics, 63*(4), 651–659. https://doi.org/10.3758/BF03194427

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics, 7*, 14. https://doi.org/10.3389/fninf.2013.00014

Yang, W., Yang, J., Gao, Y., Tang, X., Ren, Y., Takahashi, S., & Wu, J. (2015). Effects of sound frequency on audiovisual integration: An event-related potential study. *PLoS One, 10*(9), Article e0138296. https://doi.org/10.1371/journal.pone.0138296

Yang, X., Yang, W., Li, R., Lin, J., Yang, J., & Ren, Y. (2025). Audiovisual integration facilitates age-related perceptual decision making. *The Journals of Gerontology: Series B, 80*(6), gbaf037. https://doi.org/10.1093/geronb/gbaf037

Zhao, S., Wang, Y., Xu, H., Feng, C., & Feng, W. (2018). Early cross-modal interactions underlie the audiovisual bounce-inducing effect. *Neuroimage, 174*, 208–218. https://doi.org/10.1016/j.neuroimage.2018.03.036