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.

Examining Multisensory Integration in Semantically Incongruent Contexts: Evidence from Cross-Modal Mismatch Negativity and N400 Responses

Xiaoxiao Chen
Syracuse University
Jarryd Osborne
Syracuse University
Elia F Soto
Louisiana State University
Huilin Sun
Lynn Lohnas
Syracuse University
Natalie Russo
nrusso@syr.edu
Syracuse University

Research Article

Keywords: audiovisual, multisensory integration, semantic congruency, mismatch responses, N400

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11261691/v1

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

Additional Declarations: No competing interests reported.

Examining Multisensory Integration in Semantically Incongruent Contexts: Evidence from Cross-Modal Mismatch Negativity and N400 Responses

Authors:

Xiaoxiao Chen[1]
xchen128@syr.edu
https://orcid.org/0000-0002-6806-1075

Jarryd Osborne[1]
josborne@syr.edu
https://orcid.org/0009-0009-9863-6288

Elia F Soto[2]
esoto@lsu.edu
https://orcid.org/0000-0002-6672-287X

Huilin Sun[3]
https://orcid.org/0000-0003-4132-0337

Lynn Lohnas[1]
ljlohnas@syr.edu
https://orcid.org/0000-0002-1204-8760

Natalie Russo[1], ✉ **Corresponding author**
nrusso@syr.edu
https://orcid.org/0000-0001-5229-3552

Affiliations: 1. Syracuse University
Department of Psychology, 765 Irving Ave, Suite 444,
Syracuse, NY 13244, USA

2. Louisiana State University
Department of Psychology, 236 Audubon Hall,
Baton Rouge, LA 70803, USA

3. Beijing, China

Acknowledgments:

The authors wish to thank all participants for their valuable contributions to this study, as well as the C.A.R.E. (Center for Autism Research and Electrophysiology) Lab graduate students and the undergraduate research assistants for their assistance with data collection.

Abstract

Multisensory integration (MSI), in which information from different sensory modalities is processed in an interdependent manner, is often considered to occur automatically and pre-attentively. The present study aimed to investigate how modality weighting shapes the temporal dynamics of audiovisual integration during the pre-attentive stage using EEG. Twenty-two adults completed a novel audiovisual oddball task. Naturalistic animal images and sounds were presented as a frequent audiovisual standard, with visual, auditory, or audiovisual incongruencies presented as deviants. Mean amplitude analyses showed that visual and audiovisual deviants elicited significant frontocentral mismatch responses, whereas the auditory deviant did not produce a canonical mismatch negativity. A significant main effect of visual input on early ERP amplitude indicated that visual deviance exerted a stronger influence on early neural processing than auditory deviance. In contrast, auditory deviance was associated with delayed processing, particularly when paired with the expected visual stimulus, suggesting a temporal cost for non-dominant auditory information. At later stages, N400 responses emerged for auditory and audiovisual deviants but not for the visual deviant, suggesting greater downstream processing of non-dominant auditory information while visual mismatches may be resolved earlier. Together, these findings indicate that visual dominance influences the temporal organization of multisensory processing. Such temporal flexibility may be one way to facilitate the efficient interpretation of complex sensory input.

Keywords: audiovisual, multisensory integration, semantic congruency, mismatch responses, N400

Statements and Declarations

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

1. Introduction

Multisensory integration (MSI) involves combining inputs from multiple sensory modalities—such as vision, hearing, and touch—to form a coherent perceptual experience. Classic models of MSI (Meredith 2002; Meredith and Stein 1986) propose that information processing unfolds hierarchically. Modality-specific sensory features are first registered and encoded before converging (Driver and Noesselt 2008) and being integrated across distributed neural pathways. Contemporary approaches reveal that multisensory convergence can occur early in cortical processing, including at stages traditionally considered unisensory (Lakatos et al. 2007; Schroeder and Foxe 2005). Understanding the neural mechanisms underlying MSI is essential for uncovering how perceptual experiences are constructed across the senses. Critically, this integration is not necessarily balanced across modalities; rather, modality-specific dominance constrains how strongly information from each sense is weighted during perception.

A well-documented example of such dominance is the *Colavita visual dominance effect* (Colavita 1974), which refers to the tendency for visual information to override or suppress concurrent auditory processing when both are presented simultaneously. In the original paradigm, participants were instructed to respond to either a tone or a light when presented individually, but to both when presented simultaneously. Participants were found to respond only to the light in most bimodal trials and frequently failed to report the concurrent auditory stimulus (Colavita 1974). This visual dominance has been replicated across task variations and stimulus types (Koppen et al. 2007, 2009; Sinnett et al. 2007), establishing it as a robust feature of audiovisual processing. The Colavita effect motivates the present study because it provides a behavioral benchmark against which neural indices of modality weighting can be evaluated: specifically, if visual input consistently dominates auditory input at the behavioral level, corresponding neural signatures of visual priority should be detectable in early electrophysiological responses during passive audiovisual processing.

Within multisensory contexts involving meaningful stimuli, semantic associations between modalities may also shape processing. Semantic congruency—the degree to which multisensory inputs refer to the same object or event—has been associated with facilitation effects, including

faster reaction times (RTs) and higher accuracy for congruent cross-modal pairings relative to unimodal conditions (Gottfried and Dolan 2003; Heikkilä et al. 2015; Laurienti et al. 2004; Molholm et al. 2002, 2004; Sinnett et al. 2008; Wang et al. 2025). Semantically incongruent information may still be integrated when presented synchronously (Calvert 2001; Martin 2007), yet such integration incurs processing costs: semantically incongruent pairings are processed more slowly and less accurately than congruent or unisensory stimuli (Laurienti et al. 2004; Molholm et al. 2004). Moreover, sensory modality dominance may persist regardless of semantic congruency (Koppen et al. 2008; Spence et al. 2012). Koppen et al. (2008) found that on correct bimodal trials in which participants pressed both auditory and visual response keys, separately recorded RTs were faster for semantically incongruent than congruent pairings, suggesting that mismatched pairings may have been processed with greater separability and reduced audiovisual integration. Together, these behavioral findings point to robust modality dominance—particularly visual dominance (Hirst et al. 2018)—across a range of multisensory contexts, while suggesting comparatively limited and conditional effects of semantic congruency.

These behavioral findings, however, do not specify the underlying neural dynamics or their temporal sequence. The gap between behavioral evidence of visual dominance and a mechanistic account of *how* that dominance constrains the neural encoding of incongruent meanings remains largely unaddressed. Prior work using behavioral measures cannot resolve this because reaction times collapse across all processing stages; a neural measure with millisecond-level resolution is necessary to trace the temporal unfolding of modality dominance and its downstream consequences for later analysis.

To address this gap, we examined the *mismatch negativity* (MMN; Näätänen and Alho 1997; Näätänen and Michie 1979), an early event-related potential (ERP), as an index of how changes in one or both modalities were detected within audiovisual pairings. The MMN is well established as a neural index of sensitivity to changes in sensory regularities, and differences in the timing and amplitude of multisensory MMNs may provide evidence for modality weighting in MSI. In a standard oddball paradigm (Kujala et al. 2007; Näätänen et al. 2007), a deviant stimulus is interspersed with a frequent standard. When measured with EEG, the difference wave between standard and deviant responses in the 100–250 ms window is referred to as the MMN. Historically,

the MMN has been studied as an auditory index of pre-attentive deviance detection. However, a growing body of research has established a visual analog—the visual MMN (vMMN)—which is elicited when a visual stimulus deviates from an established regularity (Czigler 2014; Kimura et al. 2011). Unlike the auditory MMN, which shows a frontocentral negativity, the classical vMMN responses typically show a parieto-occipital distribution with an occipital negativity around 150–300 ms after stimulus onset (Czigler 2014; Male 2025). To note, the present study designed a multisensory oddball paradigm using naturalistic animal sound–image pairings and uses the term 'MMN' in a modality-general sense to refer to the deviant-minus-standard difference wave in both auditory and visual channels, consistent with its use in recent audio-somatosensory mismatch paradigms (Butler et al. 2012).

The MMN—whether auditory or visual in origin—is a *pre-attentive* response that does not require participants to explicitly attend to or evaluate the deviant stimuli (Näätänen et al. 2007; Schröger 1997). This makes it particularly well-suited for isolating automatic deviance detection processes from voluntary, top-down attentional influences. By presenting multiple deviant types (visual, auditory, and audiovisual) within a single experimental block and examining the early ERP window (100–250 ms), this paradigm enables systematic investigation of how modality-specific deviance drives MSI in semantically incongruent context. Although effects of semantic congruency are inherent in modality-mismatched conditions and cannot be fully disentangled from them (Stekelenburg et al. 2004), mismatch responses observed within this early time window were interpreted primarily as reflecting the detection of audiovisual deviance rather than exclusively semantic incongruency. In addition, the term *modality weighting* is used throughout this paper to refer to the differential contribution of individual sensory channels to a neural mismatch response, operationalized as mean MMN amplitude and 50% fractional area latency as a function of deviant type (Butler et al. 2012). A modality is considered more heavily weighted to the extent that its deviance drives a larger and/or earlier mismatch response.

2. Current Study

In the present study, the audiovisual oddball paradigm includes images of a cat and a frog paired with either matching or mismatching animal sounds. This design yielded four audiovisual conditions: two semantically congruent combinations (cat–meow; frog–ribbit) and two semantically incongruent combinations (frog–meow; cat–ribbit). These stimuli were chosen from

familiar animal categories to minimize variability related to novelty or attentional capture. Previous work has shown that naturalistic sounds and images facilitate rapid access to associated semantic representations relative to lexical stimuli (Chen and Spence, 2010, 2013), making them well suited for probing multisensory processing. Within this paradigm, one bimodal stimulus (cat–meow) served as the standard, while three bimodal deviants were defined based on the sensory modality in which deviance occurred: a visual deviant (frog–meow), an auditory deviant (cat–ribbit), and an audiovisual deviant (frog–ribbit) differing from the standard in both modalities.

Butler et al. (2012) provided a key framework for using semantically laden bimodal mismatch responses to examine modality weighting. In their somatosensory–auditory oddball paradigm, MMN amplitudes were similar for auditory-only and auditory–somatosensory deviants, and the somatosensory-only mismatch response was attenuated relative to conditions that included an auditory mismatch. The present study extends this framework to a semantically meaningful audiovisual context. Two primary *hypotheses* were tested. First, drawing on the Colavita effect and prior electrophysiological evidence of visual dominance (Laurienti et al. 2004; Molholm et al. 2004), we predicted that visual deviance would contribute most strongly to early neural mismatch responses. Specifically, the visual MMN amplitude was predicted to be comparable to that of the audiovisual MMN, and both were predicted to be larger than any auditory mismatch response—reflecting neural-level visual dominance consistent with the Colavita effect. Second, we predicted differential temporal dynamics as a function of modality: because visual processing dominates the audiovisual representation, auditory deviance detection was predicted to incur a temporal cost, manifesting as delayed latency relative to visual and audiovisual deviants. Additionally, to holistically characterize the temporal progression of audiovisual mismatch processing and capture further integration of meaning across modalities, we observed a broader time window from 0-500 ms and examined the N400—a robust negativity that occurs around 400 ms after a stimulus violates its expected context (Kutas and Federmeier 2011; Kutas and Hillyard 1980). The N400 is sensitive not only to contextual expectancy but also to the semantic or conceptual meaning of incoming information (Kutas and Federmeier 2011). Because the audiovisual deviant (ribbit–frog) preserved a semantically congruent relationship between the auditory and visual information, it was expected to elicit a smaller N400 response during the later 300–500 ms time window than the semantically incongruent deviants (auditory: ribbit–cat; visual: meow–frog).

3. Methods

3.1 Participants

The study recruited 29 undergraduate students from Syracuse University's psychology research pool (SONA). Procedures were approved by the Institutional Review Board (IRB) at Syracuse University, and all participants provided informed consent. A priori power analysis was conducted using G*Power 3.1.9.7 (Faul et al. 2007) using parameters from comparable audiovisual ERP studies (Lauzon et al. 2022; Winkler et al. 2009): effect size = 0.3, power = 0.8, $\alpha = 0.05$, assumed correlation among repeated measures = 0.5, yielding a minimum recommended N of 17. All participants reported normal or corrected-to-normal vision and hearing and received course credit.

After EEG preprocessing (see Section 3.3), seven participants were excluded. Two were excluded due to an artifact rejection rate exceeding 70%, likely reflecting poor electrode contact unresolvable through interpolation that occurred at the time of data collection. Five participants were excluded due to the absence of reliable obligatory ERP components (P1 and/or N1) in the 50–150 ms window—standard indices used to verify that basic sensory processing occurred (Luck 2014). Specifically, three participants exhibited high-amplitude baseline fluctuations, and two showed flat or erratic waveforms lacking discernible peaks throughout the epoch. These exclusion criteria were established prior to data analysis. The final sample comprised 22 participants (13 female; *M* age = 19.78 years).

3.2 Stimuli and Task

Stimuli.

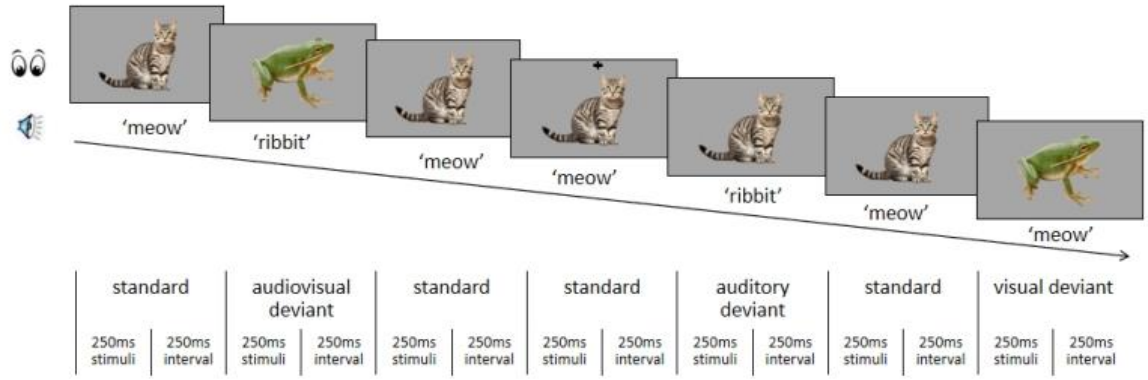
Sound files were obtained from audiomicro.com (03/14/2023) and processed in Praat 6.3.09 (Boersma and Weenink 2023). Both animal sounds were normalized to 75 dB SPL; Praat estimates of fundamental frequency were 363 Hz (ribbit) and 364 Hz (meow). Each stimulus included 25 ms linear onset and offset ramps to reduce transients. Prior psychophysical work (Keetels and Vroomen 2012) demonstrates that audiovisual simultaneity perception tolerates temporal offsets of approximately 25–50 ms, suggesting that the amplitude ramping is unlikely to disrupt multisensory temporal alignment. On each trial, auditory and visual stimuli were presented simultaneously for 250 ms. Images (5 cm × 5 cm) were displayed centrally on a gray background. Two speakers positioned to the left and right of the monitor (≈15° azimuth) delivered sounds at 60 dB SPL.

EEG Task.

The experiment was run in MATLAB R2022a (The MathWorks, Natick, MA). Standards occurred on 70% of trials; each deviant type occurred on 10% of trials (Lauzon et al. 2022). Stimuli were presented pseudorandomly with the constraint that no two deviants occurred consecutively. The experiment comprised 1,000 trials; each trial consisted of a 250 ms audiovisual stimulus followed by a 250 ms blank interval (see Fig.1).

Fig. 1

EEG task timeline



Behavioral Task.

To ensure participants were visually engaged, a black fixation cross appeared above the image on approximately 3–5% of trials. Participants were instructed to press the space bar whenever the fixation cross was visible. Trials containing fixation crosses and the immediately following trial were excluded from ERP analysis to minimize motor contamination. We acknowledge that this attention-check procedure directed participants' attention toward visual modality, which may have introduced a visual attentional bias (Fitzgerald and Todd 2020; Sinnett et al. 2007). Participants were encouraged to take breaks every 250 trials. Feedback on accuracy was provided to maintain engagement.

3.3 Electroencephalographic Recording (EEG)

An appropriately sized 128-channel high-density electrode cap (Electrical Geodesics Inc. Eugene, OR, USA) was fitted individually based on participants' head circumference. Continuous EEG was recorded via EGI NetStation (Electrical Geodesics 2003) at 250 Hz with the Cz electrode

as online reference. Impedances were kept below 50 kΩ. The hardware recording bandpass was 0.1–100 Hz (high-pass hardware filter: 0.1 Hz; low-pass hardware anti-aliasing filter: 100 Hz).

EEG Processing.

EEG was preprocessed offline in MATLAB R2022a using EEGLAB (Delorme and Makeig 2004). Post-recording digital filters were applied with a 2nd-order Butterworth bandpass filter (0.1–30 Hz low-cut and high-cut, respectively) and a 60 Hz notch filter to remove line noise. DC offsets were removed. Bad electrodes were identified via visual inspection of continuous EEG for later interpolation. Data were re-referenced to the average of the two mastoid electrodes (Cary et al. 2023; Näätänen et al. 2007). Segments containing large muscle artifacts, extreme voltage offsets, or break periods exceeding two seconds were removed (Kappenman et al. 2021). Bipolar HEOG and VEOG signals were constructed, and ICA was applied to a 64-channel scalp subset to identify and remove ocular artifact components; trained graduate students performed visual inspection of ICA components. Epochs from −100 to +500 ms relative to stimulus onset were extracted. Moving-window artifact rejection was applied, and epochs exceeding ±100 μV were rejected (Brandwein et al. 2010). Channels with more than 30% bad epochs were replaced via spherical spline interpolation.

Event-Related Potentials (ERPs) Analysis.

ERPs were computed by averaging EEG epochs within each condition. Based on the typical scalp distribution of the auditory MMN and visual mismatch responses elicited in multisensory contexts (Butler et al. 2012; Näätänen et al. 2007), the primary region of interest (ROI) comprised frontocentral electrodes E4, E16, E19, and Fz (Hedge et al. 2015; Sussman et al. 2003). Supplementary analyses were conducted at occipital electrodes (O1, Oz, O2) to characterize the visual response distribution (see Section 4.5). Mismatch difference waves were computed as deviant-minus-standard: visual MMN = meowfrog − meowcat; auditory MMN = ribbitcat − meowcat; audiovisual MMN = ribbitfrog − meowcat. Mean MMN amplitude was quantified within the 100–250 ms window. MMN latency was estimated using 50% fractional area latency (FAL), which is less sensitive to noise than peak latency (Luck, 2014). In addition, 50% fractional peak latency was computed to characterize onset and offset timing.

3.4 Data Analysis

Pairwise one-sample *t* tests were conducted on mean amplitudes in the 100–250 ms window to determine whether each deviant elicited a significant mismatch response relative to the standard. To test the primary hypotheses regarding modality weighting (Hypothesis 1: visual dominance in amplitude) and temporal dynamics (Hypothesis 2: delayed auditory latency), a 2 (Auditory input: meow vs. ribbit) × 2 (Visual input: cat vs. frog) repeated-measures ANOVA (rmANOVA) was conducted on mean ERP amplitude and 50% fractional area latency in the 100–250 ms window. The 2×2 factorial structure allows direct evaluation of main effects and interactions attributable to auditory and visual deviance independently and in combination, without requiring a canonical MMN for each condition. Significant main effects and interactions were followed up with pairwise *t*-tests with appropriate interpretation. False discovery rate (FDR) correction (Benjamini and Yekutieli 2001) was applied to time-resolved analyses to control for Type I error across time points. To examine later semantic processing, one-sample *t*-tests were conducted on mean difference-wave amplitudes in the 300–500 ms (N400) window for each deviant type. The N400 prediction that incongruent stimuli would elicit larger amplitudes than congruent stimuli presupposed a comparison between conditions that differed in semantic congruency. Because all deviants were also defined by sensory modality, the N400 results should be interpreted in the context of both sensory and semantic differences.

4. Results

Behavioral accuracy at detecting the fixation cross was high (M = 0.98, SD = 0.05), indicating good task engagement. The average artifact rejection rate after preprocessing was 13.3%, with the minimum number of accepted trials across participants being 680—sufficient to generate reliable ERP averages (Luck et al. 2021). Average accepted trials per condition were high: visual deviant M = 88.18 (SD = 9.18); auditory deviant M = 88.50 (SD = 8.93); audiovisual deviant M = 88.14 (SD = 9.94); standard M = 611.14 (SD = 62.27). Descriptive statistics for each stimulus are shown in Table 1. An overview of the data quality for ERPs is provided in Table 2 using the Root mean square of the standardized measurement error values (RMS($SM\hat{E}$)). The standardized measurement error ($SM\hat{E}$) for an ERP amplitude or latency is calculated using the average ERP waveform of a single participant and the root mean square (RMS) describes a set of $SM\hat{E}$ values in a group. *SD*s are approximately twice the RMS($SM\hat{E}$) values, consistent with recommended ratios (Luck et al. 2021), indicating that individual differences rather than measurement error drive the observed variance.

Table 1

Means (Standard Errors) of audiovisual pairs in 100–250 ms (N = 22)

	Standard Meow+Cat	**Visual Dev. Meow+Frog**	**Audio Dev. Ribbit+Cat**	**AV Dev. Ribbit+Frog**
Mean Amplitude (μV)	−0.73 (0.41)	−2.17 (0.54)	−0.52 (0.53)	−1.90 (0.60)
50% FAL (ms)	168.91 (5.78)	170.73 (3.96)	187.82 (4.88)	172.00 (4.02)

Table 2

RMS(SMÊ) by condition for mean ERP amplitude (μV)

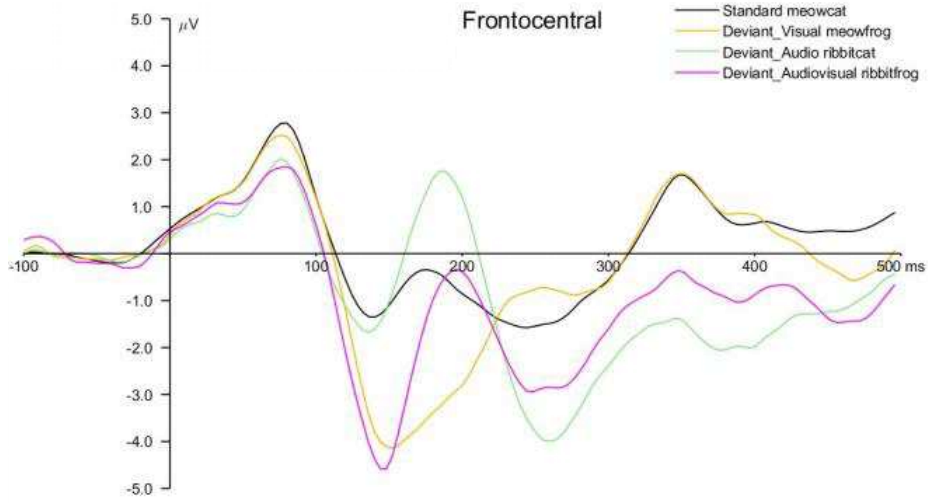
	Standard	**Visual Dev.**	**Audio Dev.**	**AV Dev.**
RMS(*SMÊ*)	0.38 / 0.45	1.01 / 1.14	1.01 / 1.17	0.94 / 1.10
SD	1.94	2.45	2.23	2.63

Note: Two RMS(*SMÊ*) values per cell correspond to 100–200 ms and 200–300 ms windows.

Grand-average ERPs for each condition across the frontocentral cluster are shown in Fig. 2. P1 morphology was comparable across conditions (peak ≈85 ms), indicating equivalent early sensory registration. Visual and audiovisual deviants showed greater negativity than the standard in the 100–250ms window. The auditory deviant showed a positive deflection during this interval followed by an anterior negativity in the 200–300 ms range. After 300 ms, the standard and visual deviant shifted positively, whereas auditory and audiovisual deviants remained negative-going.

Fig. 2

Grand-mean ERPs for all four conditions at the frontocentral cluster



Note. Standard (black), visual deviant (yellow), auditory deviant (green), audiovisual deviant (magenta).

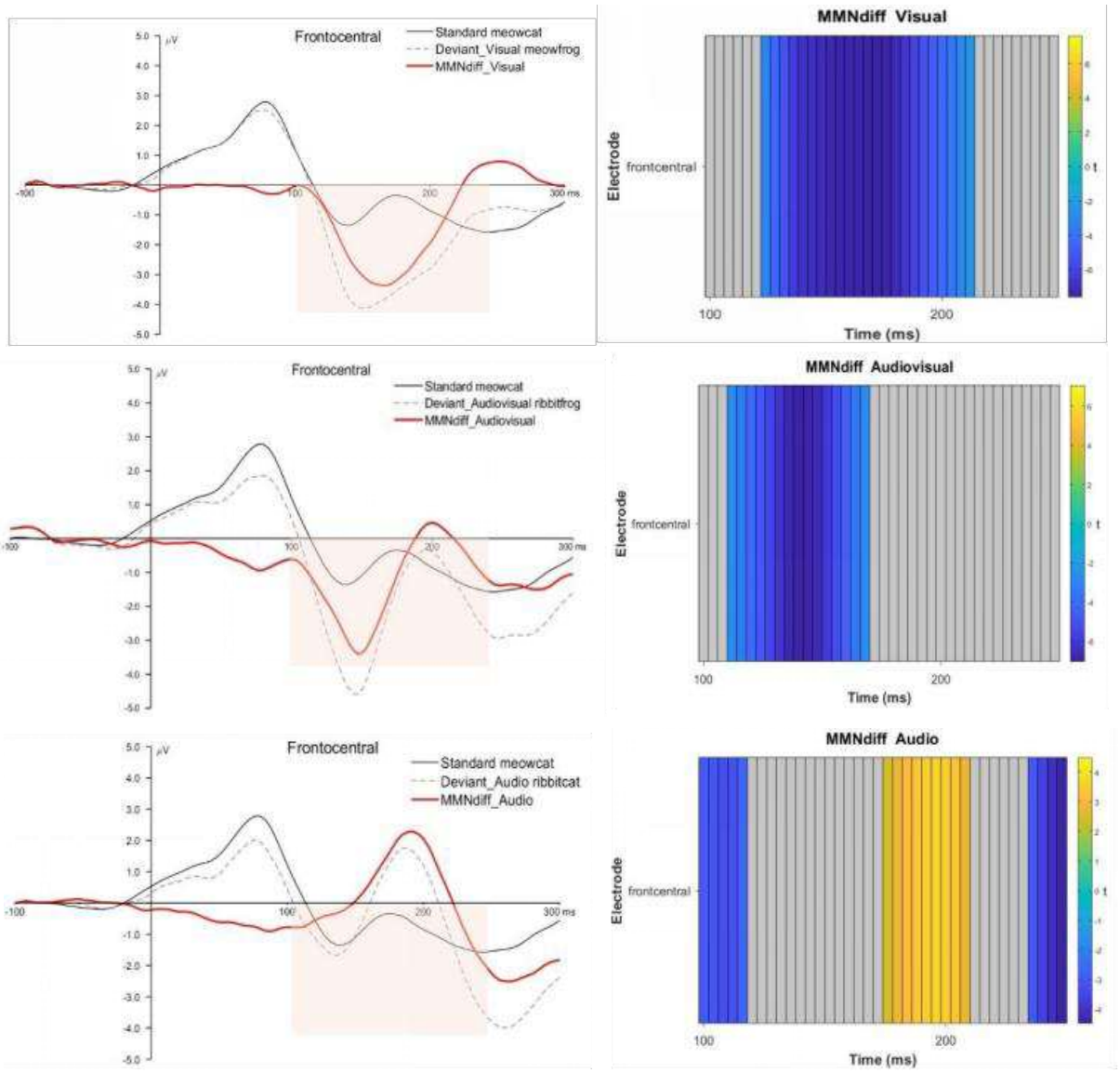
4.1 Mismatch Difference Waves

Pairwise *t*-tests in the 100–250 ms window confirmed a robust visual MMN ($t = -4.61$, $p < .001$, $d = -0.97$, 95% CI [−1.46, −0.46]) and audiovisual MMN ($t = -2.46$, $p = .023$, $d = -0.51$, 95% CI [−0.95, −0.07]) at frontocentral electrodes (see Fig. 3, left panel). No significant standard–deviant contrast was found for the auditory deviant ($p > .05$). These tests confirmed the presence of frontocentral mismatch responses for visual and audiovisual deviants, motivating subsequent analyses. FDR-corrected time-resolved analyses (Benjamini and Yekutieli, 2001) converged with these results (see Fig. 3, right panel). The visual MMN showed a sustained negative profile peaking at 168 ms. The audiovisual MMN peaked earlier, at 148 ms, with a narrower temporal profile. Pairwise *t*-tests on 50% fractional peak latencies revealed no significant onset difference between visual (132 ms) and audiovisual (118 ms) MMNs ($p > .05$). However, the audiovisual MMN offset (160 ms) was significantly earlier than the visual MMN offset (201 ms; $t = 3.40$, $p = .003$, $d = 0.71$, 95% CI [0.24, 1.17]).

For the auditory deviant, no canonical negative-going MMN was observed (See Fig. 3). Instead, the difference wave showed a positive deflection peaking around 200 ms, followed by an anterior negativity in the 200–300 ms range consistent with an N2 component. Both P2 and N2 components are associated with novelty detection and sensitivity to infrequent events (Folstein and Van Petten, 2008; Luck, 2014). The absence of a canonical auditory MMN is consistent with the hypothesis that visual continuity suppressed early auditory deviance detection—a pattern predicted by Colavita-like visual dominance and aligned with findings showing attenuated auditory sensitivity under concurrent visual processing (Laurienti et al. 2004; Molholm et al. 2004). This polarity reversal reduced the net mean amplitude in the 100–250 ms window, contributing to the absence of a significant standard–auditory deviant contrast.

Fig. 3

Difference waves and FDR-corrected time-resolved t-values



Note. In the left panel, difference waves are shown in red. In the right panel, significant *t*-values ($p < .05$) are highlighted, and the color bar indicates the magnitude and direction of the *t*-values.

4.2 Examination of Colavita Effects

Because the auditory deviant did not elicit a canonical MMN, direct MMN-amplitude comparisons across conditions were not appropriate. Instead, a 2 (Auditory: meow vs. ribbit) × 2 (Visual: cat vs. frog) rmANOVA was conducted on mean frontocentral ERP amplitude in the 100–250 ms window, treating all four stimulus conditions as a fully crossed factorial design. This approach tests modality contributions directly and does not presuppose canonical MMN morphology in any condition.

A significant main effect of visual input was observed, $F(1, 21) = 19.33$, $p < .001$, $\eta^2 = .48$, indicating that conditions containing the deviant image (frog) elicited significantly more negative responses than conditions with the standard image (cat), regardless of auditory input. This constitutes neural-level evidence of visual dominance consistent with the Colavita effect: visual deviance exerted a greater influence on early neural processing than auditory deviance. Neither the main effect of auditory input nor the Auditory × Visual interaction reached significance for mean amplitude (ps > .05).

4.3 Temporal Dynamics of MSI

The same 2×2 rmANOVA was conducted on 50% fractional area latency. A significant main effect of auditory input emerged, $F(1, 21) = 8.03$, $p = .010$, $\eta^2 = .28$, indicating that conditions containing the deviant sound (ribbit) showed delayed neural processing relative to conditions with the standard sound (meow). A significant Auditory × Visual interaction was also found, $F(1, 21) = 5.67$, $p = .027$, $\eta^2 = .21$. Post hoc comparisons revealed that auditory processing latencies did not differ as a function of sound when the visual stimulus was the deviant (frog). However, when the expected visual stimulus (cat) was paired with the auditory deviant (ribbit), processing was significantly slower than when cat was paired with the standard sound (meow). This suggests that delayed auditory deviance detection occurred specifically in the context of an expected visual stimulus—that is, visual context actively modulated auditory processing timing.

4.4 Investigation of Later Semantic Processing

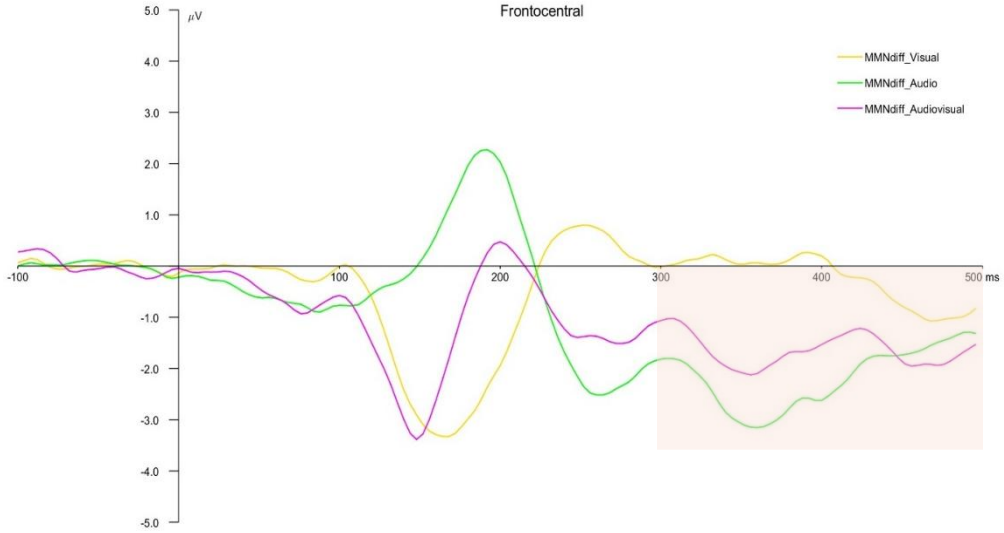
To examine later semantic binding, difference waves were computed for each deviant type relative to the standard over the 0–500 ms epoch (see Fig. 4). One-sample t-tests on mean amplitude in the 300–500 ms window revealed significant N400 effects for the auditory deviant ($t = -4.91$, $p < .001$, $d = -1.01$, 95% CI [−1.56, −0.52]) and the audiovisual deviant ($t = -3.13$, $p = .005$, $d = -0.67$, 95% CI [−1.12, −0.20]). The visual deviant did not produce a significant N400 ($p > .05$).

It should be noted that the predicted N400 pattern assumed that semantically incongruent pairings (auditory deviant: ribbit–cat; visual deviant: meow–frog) would show larger N400s than the semantically congruent audiovisual deviant (ribbit–frog). The observed pattern—with N400s for both the auditory and audiovisual deviants but not the visual deviant—is more complex than a simple congruency contrast. Conditions differ in both sensory modality and semantic congruency,

and these factors cannot be fully dissociated in the present design. The N400 results should therefore be interpreted as reflecting the interaction of sensory modality weighting and semantic mismatch, rather than as a pure index of semantic congruency.

Fig. 4

Grand-mean difference waves across 0–500 ms



Note. Auditory deviant (ribbit−cat minus meow−cat; green solid), visual deviant (meow−frog minus meow−cat; yellow solid), audiovisual deviant (ribbit−frog minus meow−cat; magenta solid).

4.5 Supplementary Analysis

We conducted additional time-resolved analyses over occipital electrodes (O1, Oz, O2) to determine whether visual and audiovisual deviants elicited classic posterior visual mismatch response in addition to the frontocentral effects. These results are presented as exploratory given that the occipital ROI was not part of the pre-specified analysis plan. Periods were considered significant when a series of *t* tests at locations of apparent differences identified at least five consecutive significant 4-ms epochs. We did not find any significance in occipital electrodes. The absence of a classical occipital vMMN suggests that the frontocentral negativity elicited by the visual deviant may not represent a purely visual mismatch response. Rather, this response may reflect visual modulation of auditory processing in an audiovisual context.

5. Discussion

This study used a modified audiovisual oddball paradigm to examine how modality weighting shapes early deviance detection during multisensory integration. ERP analyses in the 100–250 ms window characterized modality-dependent timing of mismatch processing, and N400 analyses (300–500 ms) assessed later semantic processing to relate early sensory weighting to subsequent semantic evaluation. Frontocentral mismatch responses were observed for visual and audiovisual deviants but not for the auditory deviant. Condition-level analyses revealed a dominant effect of visual deviance on early neural amplitude. Temporal analyses showed delayed processing for auditory deviants specifically when paired with the expected visual stimulus. At later stages, N400 effects were present for auditory and audiovisual deviants but absent for the visual deviant.

The pattern of early mismatch responses is consistent with neural-level visual dominance analogous to the Colavita effect (Colavita, 1974; Koppen et al. 2008). Visual deviants elicited frontocentral negativity of comparable amplitude to audiovisual deviants, whereas auditory deviants did not elicit a canonical negative-going MMN. Rather than indicating that auditory deviance went undetected, the auditory deviant condition produced a P2 followed by an anterior N2—components associated with novelty detection and the evaluation of infrequent stimuli (Folstein and Van Petten 2008; Luck 2014). These findings suggest that auditory deviance was detected, but at a later processing stage than visual or audiovisual mismatch, incurring a temporal cost consistent with the predicted delay.

We note that the frontocentral negativity observed for the visual deviant has a scalp distribution more consistent with auditory MMN or cross-modal modulation effects (Besle et al. 2005) than with the classical vMMN, which typically exhibits occipital polarity (Csikós et al. 2025; Czigler 2014). The frontocentral negativity may reflect cross-modal suppression—specifically, the fact that an unexpected visual stimulus disrupted the established audiovisual prediction (meow+cat), triggering a frontal mismatch signal regardless of modality (Stekelenburg et al. 2004). The supplementary occipital analysis provides additional evidence bearing on this question. Future studies using a controlled design with unimodal oddball blocks will be necessary to definitively assign the observed negativity to one of these mechanisms.

A clear dissociation emerged between early mismatch detection and later semantic processing. Whereas early responses (100–250 ms) were dominated by the visual modality, N400

effects (300–500 ms) were observed for auditory and audiovisual deviants but not for the visual deviant. The absence of an N400 for the visual deviant suggests that early visual dominance effectively resolved the sensory discrepancy within the early processing window, reducing the demand for further semantic evaluation. In contrast, the auditory deviant—whose change was suppressed at early stages—required extended semantic analysis, as evidenced by the robust N400. This dissociation is consistent with the idea that visual dominance modulates not only the amplitude and timing of early mismatch responses but also the allocation of later cognitive resources to semantic processing.

An important interpretive constraint is that the N400 pattern cannot be attributed purely to semantic congruency. The auditory deviant (ribbit–cat) is semantically incongruent and also lacks a canonical early mismatch response; the audiovisual deviant (ribbit–frog) is semantically congruent yet also shows an N400. The latter finding was unexpected and may reflect the persistence of cross-modal conflict with the standard in the audiovisual condition—even when sound and image are semantically congruent, the pairing deviates from the established standard (meow+cat) in both modalities, potentially sustaining a semantic discrepancy signal relative to the standard. Future designs should include a proper congruent versus incongruent contrast with identical sensory change to isolate semantic effects (Kaganovich et al. 2016).

Taken together, the present findings support competitive models of multisensory integration (*Biased Competition Hypothesis*; Desimone and Duncan 1995; Duncan 1996), in which stimuli across modalities compete for neural representation and the visual modality tends to dominate under many conditions (Colavita 1974). Our results extend this framework to a semantically rich, passive EEG context, suggesting that the brain dynamically reallocates processing resources such that early visual weighting shapes both the efficiency and the depth of subsequent semantic analysis.

6. Limitations and Future Directions

This study has several limitations that should inform its interpretation and motivate future work. Most critically, the absence of unimodal (auditory-only and visual-only) control blocks means that the design cannot fully dissociate violations of unimodal sensory regularities from violations of the learned audiovisual pairing or from semantic congruency effects (Schröger and Widmann 1998; Stekelenburg et al. 2004). Future studies should include a blocked design with

separate auditory-only and visual-only oddball sequences prior to or alongside the bimodal condition, enabling the unique contribution of audiovisual binding to be isolated. This approach would also make the present paradigm comparable to established multimodal MMN designs (Schröger and Widmann, 1998) and would allow polarity-reversal tests that we were unable to conduct here due to the mastoid reference scheme. Future studies should consider a linked-earlobe or nose reference to enable this diagnostic check.

Second, the attention-check task—requiring responses to a fixation cross—may have introduced a systematic visual attentional bias (Fitzgerald and Todd 2020; Sinnett et al. 2007). Although trials containing the fixation cross were excluded from ERP analysis, the overall task structure may have predisposed participants to orient to the visual modality across the session. Future work should consider alternative engagement monitoring strategies, such as eye-tracking or periodic verbal check-ins, that do not selectively cue visual attention.

Third, because all deviants carry both sensory and semantic differences relative to the standard, semantic congruency and sensory change are confounded. Future studies could incorporate semantically neutral audiovisual pairs (e.g., meow + image of a chair) to create an explicit congruency contrast independent of sensory novelty (Roberts et al. 2024).

Finally, the current findings may have implications for clinical populations that exhibit atypical multisensory temporal binding windows, including individuals with autism spectrum disorder (Foss-Feig et al. 2010; Kwakye et al. 2011; Taylor et al. 2010). Future research should explore whether the modality-weighting dynamics observed here generalize to or differ in these populations.

7. Conclusion

This study demonstrates that multisensory processing is shaped by modality weighting that influences both early deviance detection and later semantic processing stages. Visual input dominated early neural responses, suppressing canonical auditory mismatch morphology while leaving auditory deviance detectable at a later, post-MMN stage. The downstream consequence of this early visual dominance was a redistribution of semantic processing resources: auditory and audiovisual deviants—whose early processing was either delayed or suppressed—engaged late semantic analysis as indexed by the N400, whereas the visually resolved deviant did not. These findings underscore the importance of considering both modality-specific contributions and their temporal dynamics to understand how the brain constructs coherent meaning from competing

multisensory inputs. The design limitations identified—particularly the absence of unimodal control blocks—provide a clear roadmap for future work to further isolate the neural mechanisms of audiovisual semantic integration.

References

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188. https://doi.org/10.1214/aos/1013699998

Besle J, Fort A, Giard MH (2005) Is the auditory sensory memory sensitive to visual information? Exp Brain Res 166:337–344. https://doi.org/10.1007/s00221-005-2375-x

Boersma P, Weenink D (2023) Praat: doing phonetics by computer, version 6.3.09. https://praat.org. Accessed 14 Mar 2023

Brandwein AB, Foxe JJ, Russo NN, Altschuler TS, Gomes H, Molholm S (2010) The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. Cereb Cortex 21:1042–1055. https://doi.org/10.1093/cercor/bhq170

Butler JS, Foxe JJ, Fiebelkorn IC, Mercier MR, Molholm S (2012) Multisensory representation of frequency across audition and touch: high density electrical mapping reveals early sensory-perceptual coupling. J Neurosci 32:15338–15344. https://doi.org/10.1523/JNEUROSCI.1796-12.2012

Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110–1123. https://doi.org/10.1093/cercor/11.12.1110

Cary E, Pacheco D, Kaplan-Kahn E, McKernan E, Matsuba E, Prieve B, Russo N (2023) Brain signatures of early and late neural measures of auditory habituation and discrimination in autism and their relationship to autistic traits and sensory overresponsivity. J Autism Dev Disord 54:1344–1360. https://doi.org/10.1007/s10803-022-05866-8

Chen Y, Spence C (2010) When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures. Cognition 114:389–404. https://doi.org/10.1016/j.cognition.2009.10.012

Chen Y, Spence C (2013) The time-course of the cross-modal semantic modulation of visual picture processing by naturalistic sounds and spoken words. Multisens Res 26:371–386. https://doi.org/10.1163/22134808-00002420

Colavita FB (1974) Human sensory dominance. Percept Psychophys 16:409–412. https://doi.org/10.3758/BF03203962

Csikós N, Petro B, Gaál ZA, Czigler I (2025) Shaping the visual mismatch response: the influence of exogenous components and stimulus placement. Psychophysiology 62. https://doi.org/10.1111/psyp.70111

Czigler I (2014) Visual mismatch negativity and categorization. Brain Topogr 27:590–598. https://doi.org/10.1007/s10548-013-0316-8

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron 57:11–23. https://doi.org/10.1016/j.neuron.2007.12.013

Duncan J (1996) Cooperating brain systems in selective perception and action. Attention and Performance XVI, pp 549–578. https://doi.org/10.7551/mitpress/1479.003.0033

Duncan J (1996) Cooperating brain systems in selective perception and action. In: Inui T, McClelland JL (eds) Attention and performance XVI: Information integration in perception and communication. MIT Press, Cambridge, pp 381–402. DOI: 10.7551/mitpress/1479.003.0033

Faul F, Erdfelder E, Lang A, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39:175–191. https://doi.org/10.3758/BF03193146

Fitzgerald K, Todd J (2020) Making sense of mismatch negativity. Frontiers in Psychiatry 11:468. https://doi.org/10.3389/fpsyt.2020.00468

Folstein JR, Van Petten C (2008) Influence of cognitive control and mismatch on the N2 component of the ERP: a review. Psychophysiology 45:152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Foss-Feig JH, Kwakye LD, Cascio CJ, Burnette CP, Kadivar H, Stone WL, Wallace MT (2010) An extended multisensory temporal binding window in autism spectrum disorders. Exp Brain Res 203:381–389. https://doi.org/10.1007/s00221-010-2240-4

Gottfried JA, Dolan RJ (2003) The nose smells what the eye sees. Neuron 39:375–386. https://doi.org/10.1016/S0896-6273(03)00392-1

Hedge C, Stothart G, Todd Jones J, Rojas Frías P, Magee KL, Brooks JC (2015) A frontal attention mechanism in the visual mismatch negativity. Behav Brain Res 293:173–181. https://doi.org/10.1016/j.bbr.2015.07.022

Heikkilä J, Alho K, Hyvönen H, Tiippana K (2015) Audiovisual semantic congruency during encoding enhances memory performance. Exp Psychol 62:123–130. https://doi.org/10.1027/1618-3169/a000279

Hirst RJ, Cragg L, Allen HA (2018) Vision dominates audition in adults but not children: a meta-analysis of the Colavita effect. Neurosci Biobehav Rev 94:286–301. https://doi.org/10.1016/j.neubiorev.2018.07.012

Kaganovich N, Schumaker J, Rowland C (2016) Matching heard and seen speech: an ERP study of audiovisual word recognition. Brain Lang 157–158:14–24. https://doi.org/10.1016/j.bandl.2016.04.010

Kappenman ES, Farrens JL, Zhang W, Stewart AX, Luck SJ (2021) ERP CORE: an open resource for human event-related potential research. Neuroimage 225:117465. https://doi.org/10.1016/j.neuroimage.2020.117465

Keetels M, Vroomen J (2012) Perception of synchrony between the senses. In: Murray MM, Wallace MT (eds) The neural bases of multisensory processes. CRC Press/Taylor & Francis, pp 181–190. https://www.ncbi.nlm.nih.gov/books/NBK92837/

Kimura M, Schröger E, Czigler I (2011) Visual mismatch negativity and its importance in visual cognitive sciences. Neuroreport 22:669–673. https://doi.org/10.1097/WNR.0b013e32834973ba

Koppen C, Alsius A, Spence C (2007) Semantic congruency and the Colavita visual dominance effect. Exp Brain Res 184:533–546. https://doi.org/10.1007/s00221-007-1120-z

Koppen C, Levitan CA, Spence C (2009) A signal detection study of the Colavita visual dominance effect. Exp Brain Res 196:353–360. https://doi.org/10.1007/s00221-009-1853-y

Kujala T, Tervaniemi M, Schröger E (2007) The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. Biol Psychol 74:1–19. https://doi.org/10.1016/j.biopsycho.2006.06.001

Kutas M, Federmeier KD (2011) Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu Rev Psychol 62:621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207:203–205. https://doi.org/10.1126/science.7350657

Kwakye LD, Foss-Feig JH, Cascio CJ, Stone WL, Wallace MT (2011) Altered auditory and multisensory temporal processing in autism spectrum disorders. Front Integr Neurosci 4. https://doi.org/10.3389/fnint.2010.00129

Lakatos P, Chen C, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron 53:279–292. https://doi.org/10.1016/j.neuron.2006.12.011

Laurienti PJ, Kraft RA, Maldjian JA, Burdette JH, Wallace MT (2004) Semantic congruence is a critical factor in multisensory behavioral performance. Exp Brain Res 158:405–414. https://doi.org/10.1007/s00221-004-1913-2

Lauzon SA, Abraham AE, Curcin K, Butler BE, Stevenson RA (2022) The relationship between multisensory associative learning and multisensory integration. Neuropsychologia 174:108336. https://doi.org/10.1016/j.neuropsychologia.2022.108336

Luck SJ (2014) An introduction to the event-related potential technique, 2nd edn. MIT Press, Cambridge

Luck SJ, Stewart AX, Simmons AM, Rhemtulla M (2021) Standardized measurement error: a universal metric of data quality for averaged event-related potentials. Psychophysiology 58. https://doi.org/10.1111/psyp.13793

Male AG (2025) Predicting the unpredicted… brain response: a systematic review of the feature-related visual mismatch negativity (vMMN) and the experimental parameters that affect it. PLOS ONE 20. https://doi.org/10.1371/journal.pone.0314415

Martin A (2007) The representation of object concepts in the brain. Annu Rev Psychol 58:25–45. https://doi.org/10.1146/annurev.psych.57.102904.190143

Meredith M (2002) On the neuronal basis for multisensory convergence: a brief overview. Cogn Brain Res 14:31–40. https://doi.org/10.1016/S0926-6410(02)00059-9

Meredith MA, Stein BE (1986) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J Neurophysiol 56:640–662. https://doi.org/10.1152/jn.1986.56.3.640

Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual–auditory object recognition in humans: a high-density electrical mapping study. Cereb Cortex 14:452–465. https://doi.org/10.1093/cercor/bhh007

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cogn Brain Res 14:115–128. https://doi.org/10.1016/S0926-6410(02)00066-6

Näätänen R, Alho K (1997) Mismatch negativity—the measure for central sound representation accuracy. Audiol Neurootol 2:341–353. https://doi.org/10.1159/000259255

Näätänen R, Michie PT (1979) Early selective-attention effects on the evoked potential: a critical review and reinterpretation. Biol Psychol 8:81–136. https://doi.org/10.1016/0301-0511(79)90053-X

Näätänen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin Neurophysiol 118:2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Roberts K, Jentzsch I, Otto TU (2024) Semantic congruency modulates the speed-up of multisensory responses. Scientific Reports 14:567. https://doi.org/10.1038/s41598-023-50674-4

Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, ‘unisensory’ processing. Curr Opin Neurobiol 15:454–458. https://doi.org/10.1016/j.conb.2005.06.008

Schröger E (1997) On the detection of auditory deviations: a pre-attentive activation model. Psychophysiology 34:245–257. https://doi.org/10.1111/j.1469-8986.1997.tb02395.x

Schröger E, Widmann A (1998) Speeded responses to audiovisual signal changes result from bimodal integration. Psychophysiology 35:755–759. https://doi.org/10.1111/1469-8986.3560755

Sinnett S, Soto-Faraco S, Spence C (2008) The co-occurrence of multisensory competition and facilitation. Acta Psychol 128:153–161. https://doi.org/10.1016/j.actpsy.2007.12.002

Sinnett S, Spence C, Soto-Faraco S (2007) Visual dominance and attention: the Colavita effect revisited. Percept Psychophys 69:673–686. https://doi.org/10.3758/BF03193770

Stekelenburg JJ, Vroomen J, de Gelder B (2004) Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. Neurosci Lett 357:163–166. https://doi.org/10.1016/j.neulet.2003.12.085

Sussman E, Winkler I, Wang W (2003) MMN and attention: competition for deviance detection. Psychophysiology 40:430–435. https://doi.org/10.1111/1469-8986.00045

Taylor N, Isaac C, Milne E (2010) A comparison of the development of audiovisual integration in children with autism spectrum disorders and typically developing children. J Autism Dev Disord 40:1403–1411. https://doi.org/10.1007/s10803-010-1000-4

Wang J, Guo M, Zhang J, Bai Y, Ni G (2025) Early audiovisual integration in target processing under continuous noise: behavioral and EEG evidence. Neuropsychologia 211:109128. https://doi.org/10.1016/j.neuropsychologia.2025.109128

Winkler I, Horváth J, Weisz J, Trejo LJ (2009) Deviance detection in congruent audiovisual speech: evidence for implicit integrated audiovisual memory representations. Biol Psychol 82:281–292. https://doi.org/10.1016/j.biopsycho.2009.08.011