



Original article

Sensory modality affects emotional dimension perception and frontoparietal brain responses

Youngju Lee[†], Yongseong Lee[†], Yujeong Kim, Dongha Lee^{*}

Cognitive Science Research Group, Korea Brain Research Institute, Daegu, Republic of Korea

ARTICLE INFO

Keywords:

Emotion perception
Sensory modality
Emotional dimension
Frontoparietal network
Functional near-infrared spectroscopy

ABSTRACT

The ability to perceive other people's emotions is essential for social interaction and communication. Emotion perception involves complex cognitive processes triggered by visual, auditory, or both stimuli. However, it remains unclear how people use auditory, visual, and combined audiovisual cues to evaluate emotional valence and arousal. To address this, we conducted an emotion perception task using functional near-infrared spectroscopy to measure frontoparietal brain responses. Seven emotion categories (Angry, Calm, Disgust, Fear, Happy, Sad, and Surprise) portrayed by actors were presented under three modality conditions: audio-only (AO), audiovisual (AV), and video-only (VO). We observed significant differences between AO, AV, and VO in the behavioral evaluation of emotional valence and arousal. In addition, AV showed significant activations for emotion categories in the frontoparietal network, but not for AO or VO. Moreover, brain-behavior relationships showed that higher oxygenated hemoglobin (HbO) activities were associated with more positive valence ratings in the VO condition, but with stronger arousal ratings in the AV condition. These findings reveal that sensory modality affects brain activity (HbO) and behavior evaluation (valence and arousal ratings) in the perception of emotional dimensions.

Introduction

Humans need to understand other people's feelings appropriately to navigate social situations. Understanding how individuals perceive multimodal emotions under real-world conditions is important, as everyday social interactions rely on various emotional modalities. Individuals often analyze other people's emotions during face-to-face conversations, where visual (e.g., facial expressions) and auditory (e.g., tone of voice, pitch, and speed) information is synchronized. However, it is also common to encounter various unimodal emotional situations, such as eye contact (visual-only) and voicemails (audio-only). Earlier research on emotion perception has focused primarily on visual cues, particularly facial expressions, because they are essential in perceiving others' emotions and can be easily presented in experiments (Barrett et al., 2019; Gerdes et al., 2014). However, growing evidence indicates that auditory cues are equally important, whether presented alone or alongside visual information. Auditory emotion information is often processed automatically, facilitating attention and emotion perception (Rigoulot & Pell, 2012). However, it remains unclear how

sensory modalities influence emotion perception for valence and arousal.

Comparing unimodal (either visual or auditory) with multimodal approaches may clarify distinct roles of sensory modalities across emotion categories. Previous research shows that unimodal and multimodal emotional information processing can elicit different neural and behavioral responses (Husain et al., 2014; Klinge et al., 2010). Emotion perception may reflect sensory modality dominance, where one sensory channel contributes more strongly to interpretation. For instance, in emotion rating tasks, positive emotions tend to show greater similarity between audiovisual (AV) and video-only (VO) conditions, whereas negative emotions show greater similarity between AV and audio-only (AO) conditions (Barkhuysen et al., 2010). Similarly, the perception of happiness appears to rely more on visual cues, whereas that of anger depends more on auditory cues when individuals rate emotional intensity (Sou et al., 2022). Together, these findings highlight the importance of examining both unimodal and multimodal conditions to better understand how different sensory modalities contribute to emotion perception.

When auditory and visual emotional cues co-occur, individuals often

^{*} Corresponding author at: Cognitive Science Research Group, Korea Brain Research Institute, 61 Cheomdan-ro, Dong-gu, Daegu, Republic of Korea.

E-mail address: donghaee@kbri.re.kr (D. Lee).

[†] These authors contributed equally to this work.

perceive them as a unified emotional experience. This integration involves higher-order cognitive processing during multisensory emotion perception (Chen et al., 2010). A previous study reported that cognitive evaluation processes in frontal cortical areas are associated with multimodal emotion perception (Klasen et al., 2012). The combination of visual and auditory inputs within the same emotion category facilitates emotion identification (Gerdes et al., 2014; Klasen et al., 2012) and improves understanding of complex emotional states (Choo et al., 2025). Behaviorally, multisensory cues enhance both speed and accuracy, particularly when the signals are clear and congruent (Collignon et al., 2008; Paulmann & Pell, 2011). In contrast, incongruent auditory and visual emotional signals can produce dominance or interference effects, in which one modality influences attentional focus more strongly than the other (Peschard et al., 2018; Takagi et al., 2015).

Processing multimodal emotional information engages a distributed set of brain regions, among which the frontoparietal network, anchored by the dorsolateral prefrontal cortex (DLPFC) and the posterior parietal cortex (PPC), plays a particularly central role. This network constitutes a domain-general control system that flexibly couples with both sensory and limbic regions to support top-down attention as well as the evaluative integration of multisensory signals (Dixon et al., 2018; Dosenbach et al., 2008; Vincent et al., 2008). Within multimodal emotion perception, the DLPFC supports the appraisal of valence and arousal across modalities (Dixon et al., 2017; Kragel & LaBar, 2016), whereas the PPC and intraparietal sulcus underlie attention to competing auditory and visual emotional cues (Pourtois & Dhar, 2013; Regenbogen et al., 2018). Consistent evidence from fMRI and prior fNIRS work has implicated these frontoparietal nodes in audiovisual emotional integration (Choo et al., 2025; Klasen et al., 2012) as well as in the dimensional representation of valence and arousal (Lee et al., 2023). This network has primarily been studied in relation to emotion regulation, as well as attention adjustment to sensory cues (Sugiura et al., 2023; Vuilleumier & Huang, 2009). However, these areas are also involved at the perceptual level of emotions, highlighting that their emotion-regulating function is closely linked to emotion perception (Frajo-Apor et al., 2021; Whitehead et al., 2024). For further advances in emotion regulation applications, it is crucial to examine multimodal emotion perception involving the frontoparietal region.

Using functional near-infrared spectroscopy (fNIRS), the current study examined how multimodal and unimodal emotional stimuli are associated with the frontoparietal region. fNIRS was used because it is well suited for measuring neural activity in settings that provide a relatively naturalistic multimodal approximation (Pinti et al., 2018). Its tolerance to motion artifacts allows participants to process multimodal emotional stimuli in a more realistic manner compared to more restrictive imaging environments. In addition, fNIRS provides adequate spatial resolution at the cortical level, enabling the investigation of region-specific hemodynamic activity, although it is limited to superficial cortical areas. We hypothesized that the frontoparietal region would exhibit increased changes in activity under multimodal (AV) conditions compared to that under unimodal (AO or VO) conditions. We also examined whether frontoparietal activity varies across emotion categories, given the involvement of these regions in cognitive processes related to emotion regulation when perceiving emotional stimuli.

Materials and methods

Participants

Seventy healthy participants (35 male and 35 female, aged 20–34 years, mean = 25.3, standard deviation [SD] = 3.1) were recruited via student online community boards at local universities. A power analysis was conducted using G*Power, indicating that a minimum sample size of 43 participants would be required to achieve 95% power at an alpha level of 0.05, assuming a medium effect size ($f = 0.20$) across six within-subject conditions (actor gender $\times$ modality). Given the complexity of

the design, including multiple emotion categories and repeated-measures factors, a larger sample size was targeted to ensure adequate power for detecting both main and interaction effects. Head sizes were measured to adjust probe placement for different head circumferences (range 52.0–61.5 cm, mean = 57.1, SD = 2.0; after removal, range 52.0–61.0 cm, mean = 56.9, SD = 1.9). According to the fNIRS manual provided by Artinis, participants wore head caps of various sizes based on their head measurements, categorized as medium, large, and extra-large, corresponding to head circumferences of 52–57 cm, 57–59 cm, and 59–62 cm, respectively. All participants were right-handed, had normal or corrected-to-normal vision and hearing, and had no history of neurological or psychiatric disorders. This study was approved by the Institutional Review Board of the Korea Brain Research Institute (KBRI-202,111-HR-003-08), and all participants signed a consent form after being informed about the experimental process.

Experimental design and procedures

The 27-inch monitor used in the experiments had a resolution of 1920×1080 pixels, a refresh rate of 60 Hz, and a visual angle of 38.9×23.2 degrees. We used a chin rest to maintain a consistent distance of 57 cm from the eyes to the monitor. The auditory stimuli had a sound volume range of 50–79 dB.

All fNIRS data were collected using a Brite-MKII (Artinis Medical Systems, Elst, the Netherlands) and a Brite24 18-channel template; other parameters were set to default values or followed the Artinis provision (i.e., differential pathlength factor = 6, sampling rate: 25 Hz, wavelengths of 751 nm and 847 nm). Based on the 10–20 system, two five-channel clusters were placed symmetrically on both hemispheres of the dorsolateral prefrontal cortical areas and two four-channel clusters on the bilateral posterior parietal cortical areas (Fig. 1).

The stimulus set was obtained from The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) (Livingstone & Russo, 2018). The RAVDESS dataset comprises 12 male and 12 female actors who express eight assigned emotion categories (Angry, Calm, Disgust, Fear, Happy, Neutral, Sad, and Surprise) in either AO, AV, or VO versions, speaking an emotionally neutral sentence in English (i.e., kids are talking by the door, dogs are sitting by the door). Of those, 10 actors (five male and five female) were selected for the present study based on goodness rating scores (Livingstone & Russo, 2018). For the emotion categories, we included only strong-intensity stimuli to maximize emotional salience and ensure clear differentiation between categories, resulting in seven emotion categories. Because neutral stimuli are available only at a normal intensity level in the RAVDESS database, they were excluded.

We employed a block design in the experiment to acquire neural activity using fNIRS measurements, thereby assessing the immediate responses of the participants to each emotional stimulus.

Data acquisition and processing

Seven emotion categories (Angry, Calm, Disgust, Fear, Happy, Sad, and Surprise) portrayed by actors were presented under three modality conditions: audio-only (AO), audiovisual (AV), and video-only (VO). Each participant completed a total of 168 trials (28 trials $\times$ three modality conditions $\times$ two actor gender conditions) (Table 1). The task comprised six runs defined by actor gender: AO female, AO male, AV female, AV male, VO female, and VO male. The six runs were presented in a pseudo-randomized order. Because audiovisual stimuli could potentially influence the perception of subsequent unimodal stimuli, the order of the AV runs was counterbalanced across participants such that the AV runs appeared either first or last in the task sequence. Each run included 28 trials (seven emotions $\times$ two statements $\times$ two repetitions). Each trial began with a 4-s fixation, followed by a 4-s emotional stimulus (one of seven emotion categories), a 2-s resting phase, two rating phases (2 s each), and a final 4-s rest period. A total of 28 stimuli were presented

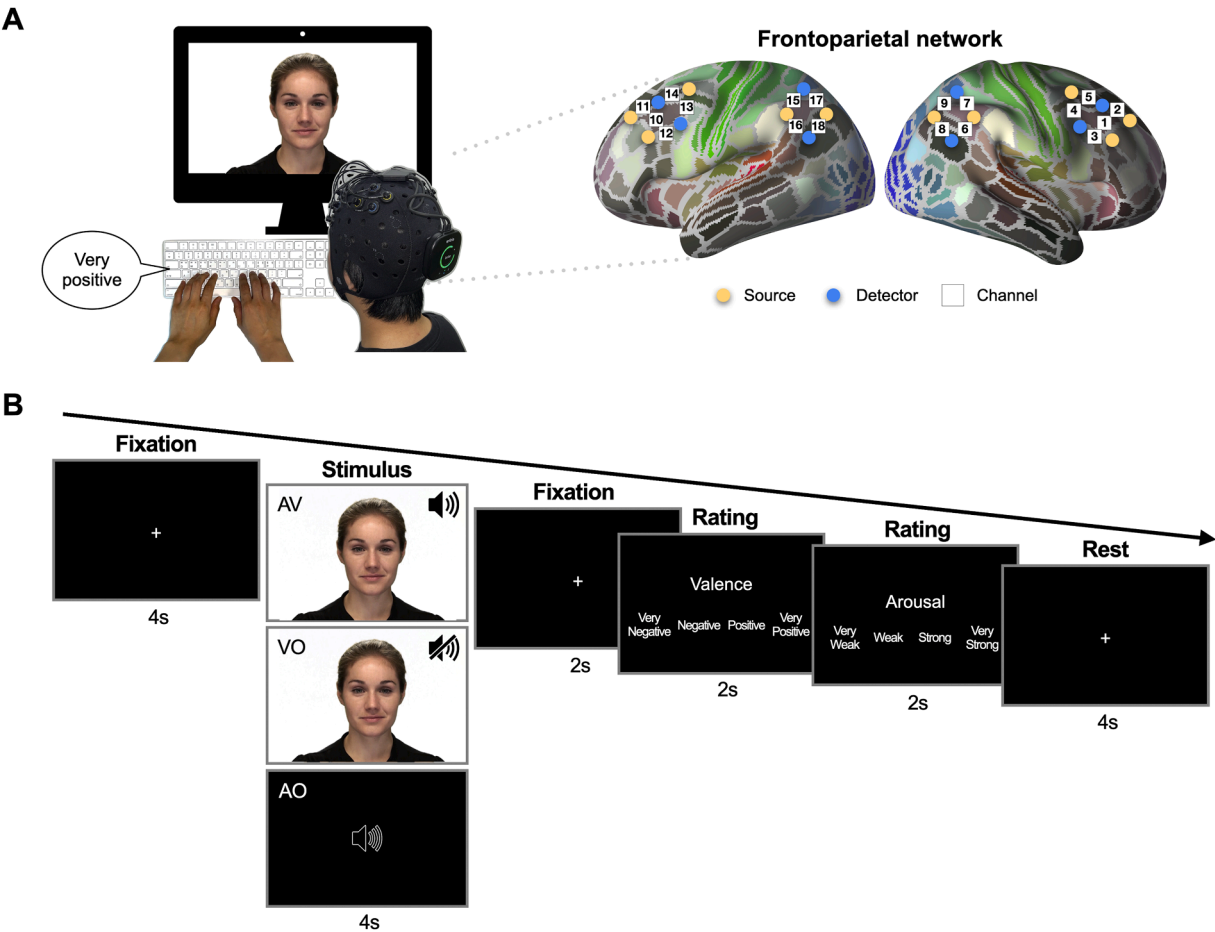


Fig. 1. Experimental paradigm and frontoparietal region of interest. (A) Experimental setup and fNIRS probe placement over the frontoparietal region of interest. In total, 18 channels were positioned over the bilateral dorsolateral prefrontal cortex (DLPFC) and posterior parietal cortex (PPC), with sources (yellow) and detectors (blue). Per hemisphere, the DLPFC channels comprised three sources and two detectors, and the PPC channels comprised two sources and two detectors. (B) Experimental paradigm. Stimuli were presented in one of three modality conditions: audio-only (AO), audiovisual (AV), or video-only (VO).

Table 1
Summary of the experimental design.

Factor	Levels	Trials per level	Notes
Modality	AO, AV, VO	56 trials per modality	Each modality included female and male actor runs
Actor Gender	Female, Male	84 trials per actor gender	Each actor gender appeared in all three modality conditions
Emotion Category	Angry, Calm, Disgust, Fear, Happy, Sad, Surprise	24 trials per emotion	Each emotion was repeated four times within each modality × actor gender run
Run	6 runs	28 trials per run	Each run defined by one modality × one actor gender condition
Repetition	4 repetition per emotion within each run	-	7 emotions × repetitions = 28 trials per run
Total trials	-	168 trials	28 trials × 6 runs

in each run (total duration = 456 s). The participants rated each emotional stimulus with emotional valence and arousal on a 4-point Likert scale for emotional valence (1 = “Very Negative,” 2 = “Negative,” 3 = “Positive,” 4 = “Very Positive”) and arousal (1 = “Very Weak,” 2 = “Weak,” 3 = “Strong,” 4 = “Very Strong”; Fig. 1).

In the experimental room, individuals sat in a comfortable position on a chair and received detailed task instructions. To help participants

understand the main experiment properly, we conducted practice trials where a male actor, who was not part of the main experiment, was presented in an AV condition. After the practice block, we ensured that the participant understood the task correctly and wore an fNIRS head cap to start the main task, with their head positioned on the chin rest. fNIRS and behavioral data from 63 participants (31 male and 32 female, aged 20–34 years, mean = 25.4, SD = 3.2) were analyzed. The fNIRS channel quality was monitored during data acquisition based on the signal quality indicators provided by the acquisition software, OxySoft (Artinis Medical Systems, Elst, the Netherlands). Before recording, hair directly beneath or near the optodes was moved away to improve optode-scalp contact and light transmission. Channels showing poor signal quality during the experiment were excluded from subsequent analyses. The overall channel removal rate was 1.94%, and excluded channels were not interpolated. As a result, participants who exhibited inconsistent behavioral responses or had low-quality fNIRS data were excluded (that is, either valence or arousal rating was missing in >20%; >90% of responses in either valence or arousal were the same; reaction time either in valence or arousal exceeds three SDs of the mean of all participants; or >25% of channels had low signal quality).

fNIRS data processing and analysis

The fNIRS data were preprocessed with Homer3 (MATLAB R2024b; MathWorks, Inc.). After converting the raw fNIRS light intensity into optical density (OD), motion artifacts were detected using a 0.5-second window, and the segment was extended by 1.0 s before and after the

detection window. Detection thresholds were set to 50 times the signal SD or an amplitude change of 5 OD units. Next, the data were band-pass-filtered, with 0.01 Hz for the high-pass and 0.1 Hz for the low-pass filter. OD data were converted to HbO (oxygenated hemoglobin) and HbR (deoxygenated hemoglobin) concentration changes using the modified Beer-Lambert law with a partial pathlength factor of 6.0 for both wavelengths. To estimate task-related hemodynamic response (HbO and HbR, respectively, in μmol), we applied a general linear model (GLM), $Y = X\beta + \epsilon$, to each run (that is, one of the three modality conditions). In this model, Y represents the observed hemoglobin concentration changes, X is the design matrix containing task regressors convolved with the HRF, β represents the estimated regression coefficients (beta values) reflecting the magnitude of task-related responses, and ϵ denotes the residual error. The parameters of GLMs were as follows: ordinary least squares were used to invert the design matrix; a gamma function modeled the HRF; a 1-second Gaussian filter was applied on the basis function; and three drift regressors were included. Beta values were obtained for each emotion condition relative to the resting phase within each run.

After conversion to HbO and HbR concentration changes, GLM analyses were conducted to estimate task-related beta values for each condition within a 0–4 s time window relative to stimulus onset. The primary analyses focused on HbO beta values because HbO responses are generally considered more sensitive to task-related changes in cortical activation and typically show a higher signal-to-noise ratio than HbR in fNIRS studies. HbR beta values were also extracted from the same GLM pipeline and analyzed as supplementary outcomes to evaluate whether they showed convergent or null patterns relative to HbO (Tables S4 and S5).

For the primary analyses, HbO beta values were averaged across all frontoparietal channels and treated as a single frontoparietal ROI. To explore possible hemispheric lateralization (Gainotti, 2019), supplementary analyses were additionally conducted by grouping channels into left and right frontoparietal ROIs, each including DLPFC and PPC channels from the corresponding hemisphere. These analyses were exploratory and were included to evaluate whether the primary findings were broadly comparable across hemispheres or showed evidence of lateralized patterns (Tables S5–S8, and Supplementary Fig. 3).

Statistical analysis of behavior ratings, brain activity, and brain-behavior relationships

To test whether the modality condition is emotion-specific for valence and arousal, we conducted a two-way 3 (Modality: AO, AV, VO) $\times$ 7 (Emotion: Angry, Calm, Disgust, Fear, Happy, Sad, Surprise) analysis of variance (ANOVA) for repeated measures.

For repeated-measures ANOVA, sphericity was assessed using Mauchly's test, and Greenhouse–Geisser corrections were applied when sphericity was violated. Partial eta squared (η^2) was reported for ANOVA results, and Cohen's d with 95% confidence intervals was reported for follow-up paired comparisons. As a post-hoc analysis, paired t -tests were conducted using Bonferroni correction that was applied by dividing α by the number of tests for multiple comparisons. The family of comparisons was defined separately for each simple effect: modality comparisons with each emotion category were corrected across three pairwise tests (adjusted p -value = 0.0167 ($\frac{0.05}{3}$)), and emotion-category comparisons within each modality condition were corrected across 21 pairwise tests (adjusted p -value = 0.0024 ($\frac{0.05}{21}$)).

To investigate the relationship of brain-behavior and its variation across modality conditions, participant-level Pearson's r correlations were computed between trial-by-trial HbO beta values and behavior ratings of emotional valence and arousal separately for each modality condition. Each correlation estimate was based on 56 trials per condition. The resulting r -values were Fisher z -transformed to approximate normality and stabilize variance prior to parametric analyses (Cohen

et al., 2003; Fisher, 1921). A one-way repeated-measures ANOVA was conducted, and as a post-hoc analysis, paired t -tests were used to test differences between modality conditions by applying Bonferroni's correction for multiple comparisons (adjusted p -value = 0.0167 ($\frac{0.05}{3}$)).

Results

Behavioral results across modality conditions and emotion categories

Fig. 2 displays behavioral ratings across emotion categories within each modality condition. We conducted a two-way 3 (Modality: AO, AV, VO) $\times$ 7 (Emotion: Angry, Calm, Disgust, Fear, Happy, Sad, Surprise) ANOVA for repeated measures. As the assumption of sphericity was violated, Greenhouse–Geisser-corrected degrees of freedom and p -values were reported. As a post-hoc analysis, paired t -tests were used to test differences between emotion categories by applying Bonferroni's correction.

For valence ratings, significant main effects were observed for Modality ($F(2, 124) = 6.15$, $p_{GG} = 0.004$, $\eta^2 = 0.042$) and Emotion ($F(6, 372) = 747.41$, $p_{GG} < 0.001$, $\eta^2 = 0.968$; Fig. 2A). The Modality $\times$ Emotion interaction was also significant ($F(12, 744) = 52.70$, $p_{GG} < 0.001$, $\eta^2 = 0.603$). In post hoc analyses, most pairwise comparisons between emotion categories were significant (Tables S1 and S2). For valence ratings, Fear and Sad did not differ significantly across any modality condition. In the AO condition, Calm and Surprise did not differ, whereas in the AV condition, Disgust and Fear showed no significant difference. All other emotion pairs were significantly different. Moreover, valence ratings differed by modality conditions in some emotion categories (Table 2, Supplementary Fig. 1A).

For arousal ratings, similar significant main effects were found for Modality ($F(2, 124) = 28.85$, $p_{GG} < 0.001$, $\eta^2 = 0.104$) and Emotion ($F(6, 372) = 290.83$, $p_{GG} < 0.001$, $\eta^2 = 0.866$; Fig. 2B) as well as the interaction was significant ($F(12, 744) = 18.45$, $p_{GG} < 0.001$, $\eta^2 = 0.167$). In post-hoc analysis, Sad and Surprise did not differ significantly across modality conditions. In the AO condition, Disgust did not differ from either Sad or Surprise. In the AV condition, Fear and Happy showed no significant difference. In the VO condition, significant differences were observed neither between Disgust and Fear nor between Fear and Happy. All remaining comparisons were significant. Arousal ratings also varied by modality conditions in some emotion categories (Table 3, Supplementary Fig. 1B).

Neural results across modality conditions and emotion categories

A two-way repeated-measures ANOVA (Modality $\times$ Emotion) was performed for the change in HbO beta values (Fig. 3). As the assumption of sphericity was violated, Greenhouse–Geisser corrected p -values were reported. As a post-hoc analysis, paired t -tests were used to test differences between emotion categories by applying Bonferroni's correction.

In the ANOVA analysis, significant main effects for Modality ($F(2, 124) = 5.18$, $p_{GG} = 0.007$, $\eta^2 = 0.021$) and Emotion ($F(6, 372) = 4.87$, $p_{GG} < 0.001$, $\eta^2 = 0.069$) were observed as well as significant interaction effects of Modality and Emotion ($F(12, 744) = 2.51$, $p_{GG} = 0.006$, $\eta^2 = 0.063$). In the post-hoc analyses, the HbO beta values in the AO and VO conditions had no significant differences across emotion categories (Table S3). However, the HbO beta values in the AV condition were significantly higher for Angry than for Calm (mean difference = 0.18, 95% CI [0.11, 0.25], $t(62) = 5.16$, $p < 0.001$, Cohen's $d = 0.65$), Disgust (mean difference = 0.14, 95% CI [0.07, 0.21], $t(62) = 3.97$, $p = 0.004$, Cohen's $d = 0.50$), Fear (mean difference = 0.15, 95% CI [0.08, 0.22], $t(62) = 4.36$, $p = 0.001$, Cohen's $d = 0.55$), Sad (mean difference = 0.13, 95% CI [0.07, 0.20], $t(62) = 4.06$, $p = 0.003$, Cohen's $d = 0.51$), and Surprise (mean difference = 0.15, 95% CI [0.08, 0.23], $t(62) = 4.05$, $p = 0.003$, Cohen's $d = 0.51$) (Table S3).

The post-hoc analysis showed that the modality main effect was only

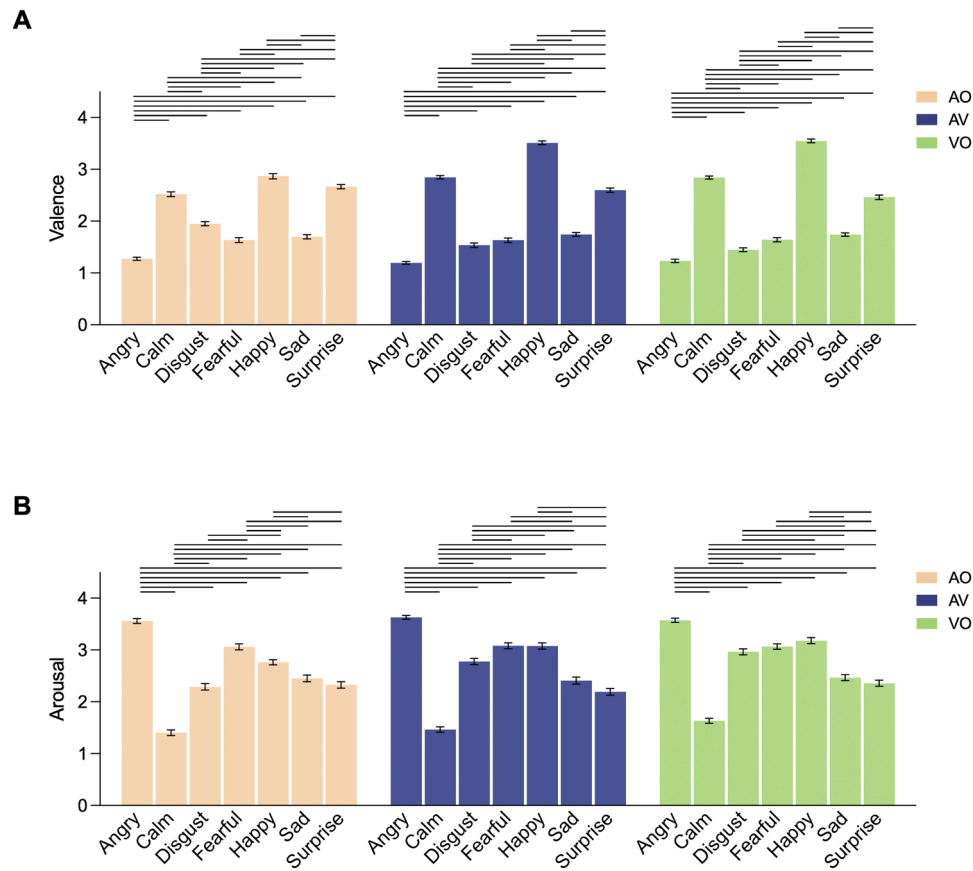


Fig. 2. Behavioral ratings across emotion categories within each modality condition (AO: audio-only; AV: audiovisual; VO: video-only). (A) Valence ratings. (B) Arousal ratings. Bars indicate group means and horizontal lines between bars denote significant pairwise differences between emotion categories based on two-way repeated-measures ANOVA (Emotion $\times$ Modality) with post hoc Bonferroni correction. Comparisons across modalities are not displayed for clarity.

Table 2

Valence ratings across modality and emotion categories. Post-hoc modality-pairwise comparisons following two-way repeated-measures ANOVA (Modality $\times$ Emotion) with Bonferroni correction ($m = 3$). SEM: standard error of the mean.

Emotion	Mean (SEM)			Bonferroni-corrected p (Cohen's d_z)		
	AO	AV	VO	AO vs. AV	AO vs. VO	AV vs. VO
Angry	1.27 (0.004)	1.19 (0.003)	1.23 (0.004)	0.054	0.820	0.587
Calm	2.52 (0.006)	2.85 (0.004)	2.84 (0.004)	3.36×10^{-8} (-0.83)	7.58×10^{-10} (-0.95)	1
Disgust	1.95 (0.005)	1.53 (0.005)	1.45 (0.005)	5.10×10^{-11} (1.03)	4.95×10^{-16} (1.41)	0.123
Fear	1.64 (0.006)	1.63 (0.006)	1.64 (0.005)	1	1	1
Happy	2.87 (0.006)	3.51 (0.005)	3.55 (0.005)	2.38×10^{-15} (-1.36)	9.67×10^{-18} (-1.54)	0.940
Sad	1.70 (0.005)	1.74 (0.005)	1.74 (0.004)	0.616	0.626	1
Surprise	2.67 (0.005)	2.60 (0.005)	2.46 (0.006)	0.480	2.29×10^{-4} (0.53)	1.81×10^{-5} (0.62)

Emotion $\times$ Modality interaction: $F(12, 744) = 52.70$, $pGG < 0.001$, $\eta^2 = 0.603$.
Modality: $F(2, 124) = 6.15$, $p = 0.003$, $\eta^2 = 0.042$.

Emotion: $F(6, 372) = 747.41$, $p < 0.001$, $\eta^2 = 0.968$.

Cohen's d_z is reported only for comparisons that reached statistical significance after Bonferroni correction.

Table 3

Arousal ratings across modality and emotion categories. Post-hoc modality-pairwise comparisons following two-way repeated-measures ANOVA (Modality $\times$ Emotion) with Bonferroni correction ($m = 3$). SEM: standard error of the mean.

Emotion	Mean (SEM)			Bonferroni-corrected p (Cohen's d_z)		
	AO	AV	VO	AO vs. AV	AO vs. VO	AV vs. VO
Angry	3.55 (0.006)	3.63 (0.005)	3.57 (0.005)	0.215	1.00	0.476
Calm	1.40 (0.007)	1.46 (0.007)	1.63 (0.006)	0.546	1.00×10^{-4} (-0.56)	8.00×10^{-4} (-0.49)
Disgust	2.28 (0.008)	2.78 (0.008)	2.96 (0.008)	8.21×10^{-11} (-1.02)	3.86×10^{-12} (-1.12)	0.003 (-0.43)
Fear	3.06 (0.008)	3.08 (0.007)	3.07 (0.007)	1	1	1
Happy	2.76 (0.006)	3.08 (0.008)	3.18 (0.007)	3.11×10^{-8} (-0.83)	6.08×10^{-10} (-0.96)	0.126
Sad	2.45 (0.008)	2.41 (0.009)	2.47 (0.007)	1	1	0.639
Surprise	2.32 (0.008)	2.19 (0.008)	2.36 (0.007)	0.092	1	0.009 (-0.39)

Emotion $\times$ Modality interaction: $F(12, 744) = 18.45$, $pGG < 0.001$, $\eta^2 = 0.167$.
Modality: $F(2, 124) = 28.85$, $p < 0.001$, $\eta^2 = 0.104$.

Emotion: $F(6, 372) = 290.83$, $p < 0.001$, $\eta^2 = 0.866$.

Cohen's d_z is reported only for comparisons that reached statistical significance after Bonferroni correction.

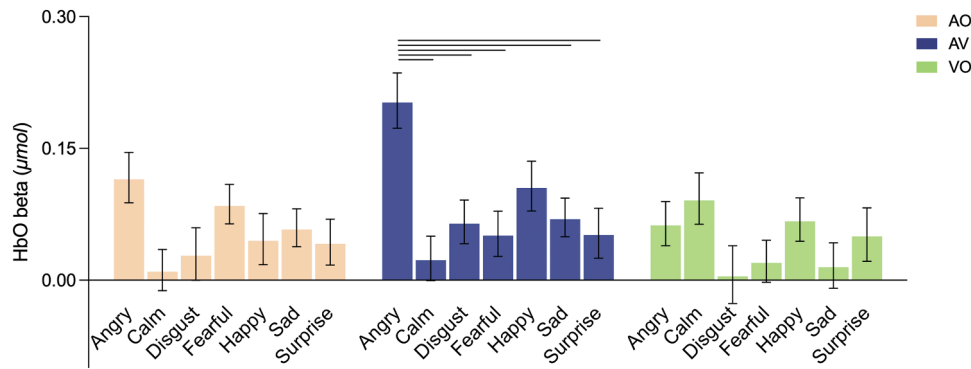


Fig. 3. Mean oxygenated hemoglobin concentration (HbO) beta values (μmol) across emotion categories within each modality condition (AO: audio-only; AV: audiovisual; VO: video-only). Bars represent group means. Horizontal lines between bars indicate significant pairwise differences between emotion categories identified by two-way repeated-measures ANOVA (Emotion $\times$ Modality) followed by post hoc Bonferroni correction. Comparisons across modalities are not displayed for clarity.

observed in the Angry conditions. The mean HbO beta value in the AV Angry condition was significantly higher than that in the AO angry condition (mean difference = 0.09, 95% CI $[-0.15, -0.02]$, $t(62) = -2.75$, $p = 0.023$, Cohen's $d_z = -0.35$) and the VO Angry condition (mean difference = 0.14, 95% CI $[0.08, 0.20]$, $t(62) = 4.58$, $p < 0.001$, Cohen's $d_z = 0.58$) (Table 4).

Correlation results between behavioral and neural responses across modality conditions

To test the modality effect on the coupling between brain and behavior, we conducted one-way repeated-measures ANOVA on correlations between HbO beta values and valence and arousal ratings.

The ANOVA showed significant modality differences in the correlations between HbO beta values and valence ratings ($F(2, 124) = 6.44$, $p = 0.002$, $\eta^2 = 0.199$). The post-hoc tests showed that the correlation between HbO beta values and valence score in the VO condition was significantly higher from those in the AO and AV conditions ($VO = 0.10 \pm 0.36$, $AO = -0.10 \pm 0.40$, VO vs. AO : $t(62) = -3.23$, $p = 0.006$, 95%

CI $[-0.37, -0.09]$, Cohen's $d_z = -0.41$; $VO = 0.10 \pm 0.36$, $AV = -0.11 \pm 0.38$, VO vs. AV : $t(62) = -3.34$, $p = 0.004$, 95% CI $[-0.39, -0.10]$, Cohen's $d_z = -0.42$). There were no significant modality differences between AO and AV conditions ($t(62) = 0.20$, $p = 0.841$, 95% CI $[-0.16, 0.19]$, Cohen's $d_z = 0.03$).

The ANOVA showed significant modality differences in the correlations between HbO beta values and arousal ratings ($F(2, 124) = 5.93$, $p = 0.003$, $\eta^2 = 0.120$). The post-hoc tests showed that the correlation between HbO beta values and arousal ratings in the VO condition was lower than that in the AV condition

($VO = -0.00 \pm 0.36$, $AV = 0.20 \pm 0.38$, VO vs. AV : $t(62) = 3.60$, $p = 0.002$, 95% CI $[0.10, 0.36]$, Cohen's $d_z = 0.45$). The modality difference of the correlation for the VO and AO conditions was not significant ($t(62) = 2.47$, $p = 0.049$, 95% CI $[0.03, 0.32]$, Cohen's $d_z = 0.31$). There were no significant modality differences between AO and AV conditions ($t(62) = -0.68$, $p = 0.497$, 95% CI $[-0.21, 0.10]$, Cohen's $d_z = -0.09$).

Discussion

Here, we report that sensory modality is not merely a delivery format of emotional stimuli but a core factor that shapes how emotions are perceived and processed in dissociable ways. To test this, we examined the perception of emotional dimensions at the neural and behavioral levels and their correlations.

Using AO, AV, and VO emotional stimuli, the current study examined the frontoparietal region as a region of interest during fNIRS measurements and emotion rating tasks assessing valence and arousal. Across behavioral, neural, and brain-behavior correlation analyses, our findings demonstrate that modality conditions play distinct roles in emotion perception. Consistent with our hypothesis, the multimodal (AV) condition was associated with greater neural activity in the frontoparietal region than the two unimodal conditions. Specifically, Angry in the AV condition showed the greatest change in the HbO beta value compared to other modality conditions and other emotions in each modality condition. Additionally, the correlation between HbO beta values and behavioral ratings—valence and arousal—varied across modality conditions, indicating similar processing mechanisms in AO and AV conditions.

Modality conditions in this study appear to play a role in representing ambiguous emotions along emotional dimensions. Behavioral results showed that sensory modality significantly influenced the perceived valence and arousal of Disgust and Surprise (Supplementary Fig. 1A, B). Valence ratings were more negative in the AV and VO conditions than those in the AO condition, whereas arousal ratings were higher in the VO conditions than in conditions that included auditory information. Notably, this pattern was specific to Disgust and Surprise and was not observed for other emotion categories. This is consistent

Table 4

HbO beta values (μmol) across modality and emotion categories. Post-hoc modality-pairwise comparisons following two-way repeated-measures ANOVA (Modality $\times$ Emotion) with Bonferroni correction ($m = 3$). SEM: standard error of the mean.

Emotion	Mean (SEM) (μmol)			Bonferroni-corrected p (Cohen's d_z)		
	AO	AV	VO	AO vs. AV	AO vs. VO	AV vs. VO
Angry	0.117 (0.029)	0.204 (0.032)	0.064 (0.025)	0.023 (-0.35)	0.288	1.00×10^{-4} (0.58)
Calm	0.012 (0.023)	0.025 (0.025)	0.093 (0.029)	1	0.086	0.141
Disgust	0.030 (0.030)	0.066 (0.025)	0.006 (0.033)	0.820	1	0.299
Fear	0.086 (0.022)	0.053 (0.026)	0.022 (0.024)	1	0.077	0.954
Happy	0.047 (0.029)	0.107 (0.028)	0.069 (0.025)	0.335	1	0.648
Sad	0.060 (0.021)	0.071 (0.022)	0.017 (0.026)	1	0.378	0.272
Surprise	0.043 (0.026)	0.053 (0.028)	0.052 (0.030)	1	1	1

Emotion $\times$ Modality interaction: $F(12, 744) = 2.51$, $p_{GG} = 0.003$, $\eta^2 = 0.063$. Modality: $F(2, 124) = 5.18$, $p = 0.007$, $\eta^2 = 0.021$.

Emotion: $F(6, 372) = 4.87$, $p < 0.001$, $\eta^2 = 0.069$.

Cohen's d_z is reported only for comparisons that reached statistical significance after Bonferroni correction.

with previous findings showing that auditory stimuli tend to elicit less negative evaluations than visual stimuli for Disgust (Croy et al., 2013) and Surprise (Paulmann et al., 2008). This may reflect the relatively ambiguous nature of these emotions, which leads individuals to rely more on the specific sensory cues available. Our results suggest that visual information may exert a stronger influence on emotion perception than auditory information. Thus, certain emotions may be more difficult to interpret when conveyed through a single sensory modality.

Moreover, negative emotions were perceived through multisensory processing to support higher-order cognitive functions. In the neural results, Angry was associated with higher HbO beta values in the frontoparietal region under the AV condition, compared with those levels for either other emotions or other modalities (Fig. 3, Supplementary Fig. 2). This finding is consistent with the results of a previous study showing that Angry is a target emotion that should be regulated through executive attentional and sensory integration in this brain region (Alia-Klein et al., 2018). More specifically, the increased activity appears to be related to the role of the dorsolateral prefrontal cortex, a part of the frontoparietal region that has been associated in previous research with the inhibition of adverse effects of negative emotions (Klimecki et al., 2018).

Interestingly, the relationship between the brain and behavior in perceiving emotions is varied according to the modality conditions. When comparing correlations between HbO beta values and behavioral ratings across modalities, AO and AV conditions showed, in contrast to the VO condition, similar patterns in both valence and arousal (Fig. 4). This suggests that the presence of auditory information may be associated with more consistent brain-behavior relationships in the frontoparietal region. For valence, significant negative correlations were observed between HbO beta values and valence ratings in the AO and AV conditions, whereas a positive correlation emerged in the VO condition. This pattern suggests that frontoparietal activity may be involved in evaluating emotional valence in the presence of auditory cues, regardless of the emotion category. This interpretation is consistent with previous findings indicating that auditory information can bias emotional evaluations toward more negative interpretations and may be associated with prefrontal modulation (Donhauser et al., 2014). In the present study, the negative correlations between HbO beta values and valence ratings in the AO and AV conditions may reflect this pattern. In contrast,

the positive correlations between HbO beta values and valence ratings observed in the VO condition reveal that the brain-behavior relationships may differ when emotional judgments rely primarily on visual information. Previous studies have shown that visual attention is associated with broader eye-movement patterns, particularly during the processing of positive emotional expressions (Calvo et al., 2018; Nusseck et al., 2008). Because of the involvement of frontoparietal regions in attentional adjustment, the positive association in the VO condition may reflect attentional processes engaged during visual valence judgments.

However, a different pattern was observed for arousal. Significant positive correlations between HbO beta values and arousal ratings were observed in the AO and AV conditions, whereas the VO condition showed no significant relationship and only a slight negative trend. This divergence indicates that auditory information may be associated more strongly with the coupling between perceived arousal and frontoparietal activity. One possible explanation is that auditory perceptual engagement is more strongly aligned with neural activity. In this context, auditory input may preferentially engage increased frontoparietal activity related to attentional allocation and evaluative processing. This result aligns with prior findings indicating that auditory signals play a dominant role in modulating arousal (Yang, 2025) by rapidly recruiting attentional systems when conveying emotional significance (Wickens, 2002) or by engaging the frontoparietal region in auditory processing (Seydell-Greenwald et al., 2014).

However, because the present analyses were correlational, these results do not establish a causal or functionally dominant role of visual or auditory information in valence and arousal perception. Taken together, the results suggest that frontoparietal activity may not directly represent valence or arousal *per se* but may instead be differentially coupled with behavioral judgments depending on the sensory modality available during emotion perception.

This study has several notable limitations. First, we examined only visual and auditory modalities, but emotions can be conveyed through additional sensory channels. For example, olfactory and bodily emotional stimuli can shape multimodal emotional experiences (Damon et al., 2021; Novak et al., 2015; Schirmer & Adolphs, 2017). Future studies should incorporate a broader range of sensory modalities to investigate emotion perception more comprehensively.

Second, all neural interpretations in the present study are restricted

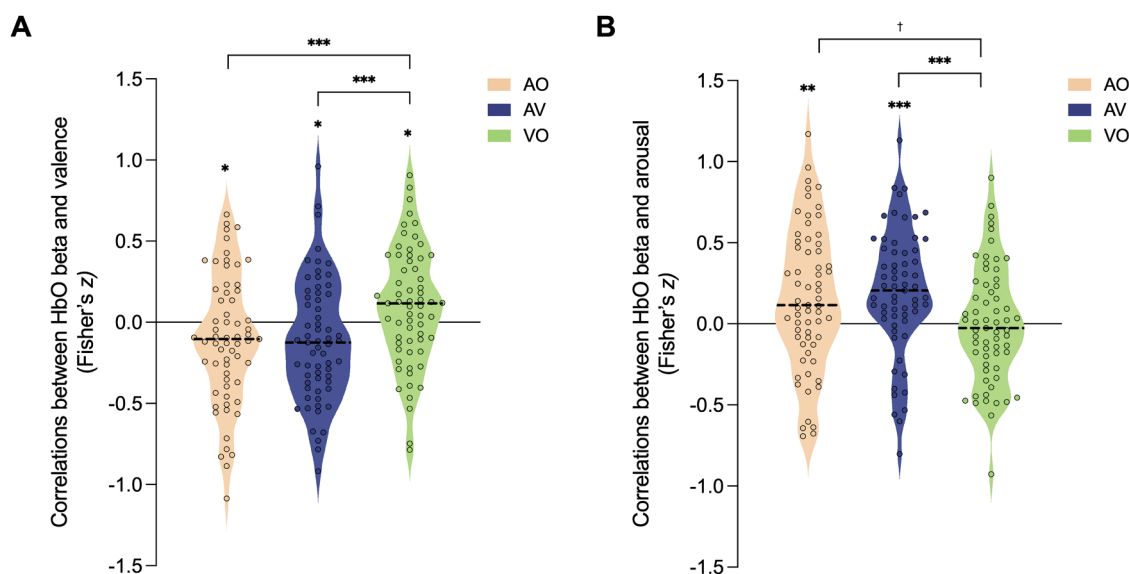


Fig. 4. Correlations between oxygenated hemoglobin concentration (HbO) beta values (μmol) and behavioral responses across modality conditions (AO: audio-only; AV: audiovisual; VO: video-only). (A) Correlations between HbO beta and valence. (B) Correlations between HbO beta and arousal. Dots represent individual correlation values, and violin plots denote distributions with medians indicated by horizontal lines. Asterisks on violin plots illustrate significant correlations from one-sample *t*-tests. Vertical lines between plots indicate significant differences across modalities based on repeated-measures ANOVA with post hoc Bonferroni correction. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, † $p < 0.057$.

to bilateral DLPFC and PPC. For instance, the study showed emotion-modality-specific activations, with the middle and lateral occipital cortices responding to visual emotional cues and the superior temporal gyrus responding to auditory emotional cues (Gao et al., 2024). As these regions are not measured, the present findings should not be generalized to whole-brain mechanisms.

Third, “Neutral” stimuli were excluded from the current design because the RAVDESS database does not include a neutral intensity level comparable to that of the other emotion categories, precluding a matched baseline condition. Although this decision was made to preserve stimulus consistency across conditions, it limits our ability to establish a baseline for interpreting the behavioral and neural responses observed in the present study. Without a neutral reference, it remains unclear whether the valence and arousal ratings, as well as the frontoparietal HbO beta values, associated with each emotion category reflect emotion-specific processing or more general responses to facial and vocal cues. Future studies should use stimulus sets in which valence and arousal are directly and continuously manipulated across a full range including neutral expressions, enabling a cleaner isolation of dimensional and modality-specific effects free from categorical confounds.

Another limitation is the hybrid nature of the experimental design. Although the study focused on valence and arousal ratings, the stimuli were selected from discrete emotion categories. Thus, modality effects may partly reflect category-specific properties instead of dimensional processing alone. This issue is particularly relevant for Angry stimuli, which may be more salient, intense, recognizable, or threat-relevant than other emotions. Normative ratings from the RAVDESS validation dataset showed that stimulus intensity varied across emotion and modality conditions, with Angry stimuli showing the highest intensity overall (refer to Table S9). Therefore, the stronger frontoparietal response to Angry stimuli in the AV condition may partly reflect stimulus-related properties. However, because this enhancement was not evident in the AV condition, intensity alone is unlikely to fully account for the pattern. Future studies should use stimulus sets that better control valence, arousal, intensity, and category membership, including intensity-matched neutral expressions.

However, the current study has several implications for better understanding how sensory modality affects behavioral ratings and frontoparietal brain responses during emotion perception. We revealed that modality-dependent differences varied across emotion categories. In particular, Angry stimuli in the audiovisual condition elicited higher brain activity in the frontoparietal network than in unimodal conditions, indicating that the integration of visual and auditory information may require greater evaluative processing for certain negative emotions. Moreover, the brain-behavior relationships showed that in the VO condition, higher HbO activities were associated with more positive valence ratings. In contrast, in the AV condition, higher HbO activities were associated with stronger arousal ratings.

This has practical applications for clinical populations, as previous studies have reported difficulties in integrating AV emotional cues (Gerdes et al., 2014; Maurage & Campanella, 2013; Stevenson et al., 2018). Specifically, these individuals may have difficulty perceiving multimodal emotions as less negative or less strong. Because such differences can be linked to experience-dependent neuronal reorganization (Husain et al., 2014; Klinge et al., 2010), future studies on multimodal emotion processing should investigate a broader range of individuals to better understand these variations.

Another important avenue for future research is the consideration of how individual differences may systematically modulate emotion perception. A growing body of research has shown that factors such as sex, trait anxiety, sensory processing sensitivity, alexithymia, and social cognitive ability can influence both behavioral and neural responses to emotional stimuli. For example, sex differences have been reported in the neural correlates of emotion processing (Whittle et al., 2011), and individuals with higher trait anxiety show enhanced sensitivity to threat-related expressions (Surcinelli et al., 2006). Sensory processing

sensitivity is associated with heightened neural responses to others' emotions (Acevedo et al., 2014), whereas alexithymia is linked to alterations in the early perceptual and attentional processing of emotional stimuli (Delle-Vigne et al., 2014). In addition, variability in social cognitive ability, including conditions such as autism-spectrum disorders and depression, influences the integration and interpretation of audiovisual emotional signals (Charbonneau et al., 2013; Li et al., 2022). Incorporating them as potential moderators in future designs will provide a more comprehensive and profound understanding of multimodal emotion perception. Taken together, we suggest that mapping emotional processing across modality conditions provides a more comprehensive understanding of how emotion is perceived by different individuals.

Data statement

The data and code in the present study are available upon request to the corresponding authors, subject to a data-sharing agreement with the Korea Brain Bank (<https://k-brainnet.kbri.re.kr/>).

Declaration of competing interest

The authors declare that there are no potential conflicts of interest.

Acknowledgements

This research was supported by KBRI basic research program through Korea Brain Research Institute funded by the Ministry of Science and ICT (26-BR-03-01, 26-BR-03-02, 26-BR-02-11) and the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2024-00401794).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.ijchp.2026.100699](https://doi.org/10.1016/j.ijchp.2026.100699).

References

- Acevedo, B. P., Aron, E. N., Aron, A., Sangster, M. D., Collins, N., & Brown, L. L. (2014). The highly sensitive brain: An fMRI study of sensory processing sensitivity and response to others' emotions. *Brain and Behavior*, 4(4), 580–594. <https://doi.org/10.1002/brb3.242>
- Alia-Klein, N., Preston-Campbell, R. N., Moeller, S. J., Parvaz, M. A., Bachi, K., Gan, G., & Goldstein, R. Z. (2018). Trait anger modulates neural activity in the fronto-parietal attention network. *PLOS ONE*, 13(4), Article e0194444. <https://doi.org/10.1371/journal.pone.0194444>
- Barkhuysen, P., Krahmer, E., & Swerts, M. (2010). Crossmodal and incremental perception of audiovisual cues to emotional speech. *Language and Speech*, 53(1), 3–30. <https://doi.org/10.1177/0023830909348993>
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, 20(1), 1–68. <https://doi.org/10.1177/1529100619832930>
- Calvo, M. G., Fernández-Martín, A., Gutiérrez-García, A., & Lundqvist, D. (2018). Selective eye fixations on diagnostic face regions of dynamic emotional expressions: KDEF-dyn database. *Scientific Reports*, 8(1), Article 17039. <https://doi.org/10.1038/s41598-018-35259-w>
- Charbonneau, G., Bertone, A., Lepore, F., Nassim, M., Lassonde, M., Mottron, L., & Collignon, O. (2013). Multilevel alterations in the processing of audio-visual emotion expressions in autism spectrum disorders. *Neuropsychologia*, 51(5), 1002–1010. <https://doi.org/10.1016/j.neuropsychologia.2013.02.009>
- Chen, Y.-H., Edgar, J. C., Holroyd, T., Dammers, J., Thönneßen, H., Roberts, T. P. L., & Mathiak, K. (2010). Neuromagnetic oscillations to emotional faces and prosody. *European Journal of Neuroscience*, 31(10), 1818–1827. <https://doi.org/10.1111/j.1460-9568.2010.07203.x>
- Choo, C. M., Bai, S., Privitera, A. J., & Chen, S.-H. A. (2025). Brain imaging studies of multisensory integration in emotion perception: A scoping review. *Neuroscience & Biobehavioral Reviews*, 172, Article 106118. <https://doi.org/10.1016/j.neubiorev.2025.106118>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (Producer). (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*, 3rd ed.

- Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research*, 1242, 126–135. <https://doi.org/10.1016/j.brainres.2008.04.023>
- Croy, I., Laqua, K., Suess, F., Joraschky, P., Ziemssen, T., & Hummel, T. (2013). The sensory channel of presentation alters subjective ratings and autonomic responses toward disgusting stimuli—Blood pressure, heart rate and skin conductance in response to visual, auditory, haptic and olfactory presented disgusting stimuli. [Original Research]. *Frontiers in Human Neuroscience*, 7, 510. <https://doi.org/10.3389/fnhum.2013.00510>
- Damon, F., Mezrai, N., Magnier, L., Leleu, A., Durand, K., & Schaal, B. (2021). Olfaction in the multisensory processing of faces: A narrative review of the influence of Human body odors. [Review]. *Frontiers in Psychology*, 12, Article 750944. <https://doi.org/10.3389/fpsyg.2021.750944>
- Delle-Vigne, D., Kornreich, C., Verbanck, P., & Campanella, S. (2014). Subclinical alexithymia modulates early audio-visual perceptual and attentional event-related potentials. [Original Research]. *Frontiers in Human Neuroscience*, 8, 106. <https://doi.org/10.3389/fnhum.2014.00106>
- Dixon, M. L., De La Vega, A., Mills, C., Andrews-Hanna, J., Spreng, R. N., Cole, M. W., & Christoff, K. (2018). Heterogeneity within the frontoparietal control network and its relationship to the default and dorsal attention networks. *Proceedings of the National Academy of Sciences*, 115(7), E1598–E1607. <https://doi.org/10.1073/pnas.1715766115>
- Dixon, M. L., Thiruchselvam, R., Todd, R., & Christoff, K. (2017). Emotion and the prefrontal cortex: An integrative review. *Psychological Bulletin*, 143(10), 1033–1081. <https://doi.org/10.1037/bul0000096>
- Donhauser, P. W., Belin, P., & Grosbras, M.-H. (2014). Biasing the perception of ambiguous vocal affect: A TMS study on frontal asymmetry. *Social Cognitive and Affective Neuroscience*, 9(7), 1046–1051. <https://doi.org/10.1093/scan/nst080>
- Dosenbach, N. U., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends Cogn Sci*, 12(3), 99–105. <https://doi.org/10.1016/j.tics.2008.01.001>
- Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3–32.
- Frajo-Apor, B., Pardeller, S., Kemmler, G., Mühlbacher, M., Welte, A.-S., Hörtnagl, C., & Hofer, A. (2021). The relationship between emotional intelligence and quality of life in schizophrenia and bipolar I disorder. *Quality of Life Research*, 30(9), 2475–2485. <https://doi.org/10.1007/s11136-021-02843-z>
- Gainotti, G. (2019). Emotions and the right hemisphere: Can new data clarify old models? *The Neuroscientist*, 25(3), 258–270. <https://doi.org/10.1177/1073858418785342>
- Gao, C., Oh, S., Yang, X., Stanley, J. M., & Shinkareva, S. V. (2024). Neural representations of emotions in visual, auditory, and modality-independent regions reflect idiosyncratic conceptual knowledge. *Human Brain Mapping*, 45(14), Article e70040. <https://doi.org/10.1002/hbm.70040>
- Gerdas, A. B. M., Wieser, M. J., & Alpers, G. W. (2014). Emotional pictures and sounds: A review of multimodal interactions of emotion cues in multiple domains. [Review]. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01351>
- Husain, F. T., Carpenter-Thompson, J. R., & Schmidt, S. A. (2014). The effect of mild-to-moderate hearing loss on auditory and emotion processing networks. [Original Research]. *Frontiers in Systems Neuroscience*, 8, 10. <https://doi.org/10.3389/fnsys.2014.00010>
- Klasen, M., Chen, Y.-H., & Mathiak, K. (2012). Multisensory emotions: Perception, combination and underlying neural processes. *Reviews in the Neurosciences*, 23(4), 381–392. <https://doi.org/10.1515/revneuro-2012-0040>
- Klimecki, O. M., Sander, D., & Vuilleumier, P. (2018). Distinct brain areas involved in anger versus punishment during social interactions. *Scientific Reports*, 8(1), Article 10556. <https://doi.org/10.1038/s41598-018-28863-3>
- Klinge, C., Röder, B., & Büchel, C. (2010). Increased amygdala activation to emotional auditory stimuli in the blind. *Brain: A Journal of Neurology*, 133(6), 1729–1736. <https://doi.org/10.1093/brain/awq102>
- Kragel, P. A., & LaBar, K. S. (2016). Decoding the nature of emotion in the brain. *Trends in Cognitive Sciences*, 20(6), 444–455. <https://doi.org/10.1016/j.tics.2016.03.011>
- Lee, Y., Seo, Y., Lee, Y., & Lee, D. (2023). Dimensional emotions are represented by distinct topographical brain networks. *International Journal of Clinical and Health Psychology: IJCHP*, 23(4), Article 100408. <https://doi.org/10.1016/j.ijchp.2023.100408>
- Li, L., Li, R., Shen, F., Wang, X., Zou, T., Deng, C., & Chen, H. (2022). Negative bias effects during audiovisual emotional processing in major depression disorder. *Human Brain Mapping*, 43(4), 1449–1462. <https://doi.org/10.1002/hbm.25735>
- Livingstone, S. R., & Russo, F. A. (2018). The Ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLOS ONE*, 13(5), Article e0196391. <https://doi.org/10.1371/journal.pone.0196391>
- Maurage, P., & Campanella, S. (2013). Experimental and clinical usefulness of crossmodal paradigms in psychiatry: An illustration from emotional processing in alcohol-dependence. [Review]. *Frontiers in Human Neuroscience*, 7, 394. <https://doi.org/10.3389/fnhum.2013.00394>
- Novak, L. R., Gitelman, D. R., Schuyler, B., & Li, W. (2015). Olfactory-visual integration facilitates perception of subthreshold negative emotion. *Neuropsychologia*, 77, 288–297. <https://doi.org/10.1016/j.neuropsychologia.2015.09.005>
- Nussek, M., Cunningham, D. W., Wallraven, C., & Bühlhoff, H. H. (2008). The contribution of different facial regions to the recognition of conversational expressions (Vol. 8, pp. 1–23). US: Assn for Research in Vision & Ophthalmology (ARVO).
- Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion*, 35(2), 192–201. <https://doi.org/10.1007/s11031-011-9206-0>
- Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). How aging affects the recognition of emotional speech. *Brain and Language*, 104(3), 262–269. <https://doi.org/10.1016/j.bandl.2007.03.002>
- Peschard, V., Philippot, P., & Gilboa-Schechtman, E. (2018). Involuntary processing of social dominance cues from bimodal face-voice displays. *Cognition and Emotion*, 32(1), 13–23. <https://doi.org/10.1080/02699931.2016.1266304>
- Pinti, P., Aichelburg, C., Gilbert, S., Hamilton, A., Hirsch, J., Burgess, P., & Tachtsidis, I. (2018). A review on the use of wearable functional near-infrared spectroscopy in naturalistic environments. *Japanese Psychological Research*, 60(4), 347–373. <https://doi.org/10.1111/jpr.12206>
- Pourtois, G., & Dhar, M. (2013). Integration of face and voice during emotion perception: Is there anything gained for the perceptual system beyond stimulus modality redundancy? In P. Belin, S. Campanella, & T. Ethofer (Eds.), *Integrating face and voice in person perception* (pp. 181–206). New York, NY: Springer New York.
- 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>
- Rigoulot, S., & Pell, M. D. (2012). Seeing emotion with your ears: Emotional prosody implicitly guides visual attention to faces. *PLOS ONE*, 7(1), Article e30740. <https://doi.org/10.1371/journal.pone.0030740>
- Schirmer, A., & Adolphs, R. (2017). Emotion perception from face, voice, and touch: Comparisons and convergence. *Trends in Cognitive Sciences*, 21(3), 216–228. <https://doi.org/10.1016/j.tics.2017.01.001>
- Seydell-Greenwald, A., Greenberg, A. S., & Rauschecker, J. P. (2014). Are you listening? Brain activation associated with sustained nonspatial auditory attention in the presence and absence of stimulation. *Human Brain Mapping*, 35(5), 2233–2252. <https://doi.org/10.1002/hbm.22323>
- Sou, K. L., Say, A., & Xu, H. (2022). Unity assumption in audiovisual emotion perception. [Original Research]. *Frontiers in Neuroscience*, 16, Article 782318. <https://doi.org/10.3389/fnins.2022.782318>
- Stevenson, R. A., Segers, M., Ncube, B. L., Black, K. R., Bebko, J. M., Ferber, S., & Barense, M. D. (2018). The cascading influence of multisensory processing on speech perception in autism. *Autism: The International Journal of Research and Practice*, 22(5), 609–624. <https://doi.org/10.1177/1362361317704413>
- Sugiura, M., Katayori, Y., Muratsubaki, T., Shiratori, M., Hanawa, S., Nejad, K. K., & Fukudo, S. (2023). Automatic adaptive emotion regulation is associated with lower emotion-related activation in the frontoparietal cortex and other cortical regions with multi-componential organization. [Original Research]. *Frontiers in Behavioral Neuroscience*, 17, Article 1059158. <https://doi.org/10.3389/fnbeh.2023.1059158>
- Surcinelli, P., Codispoti, M., Montebarducci, O., Rossi, N., & Baldaro, B. (2006). Facial emotion recognition in trait anxiety. *Journal of Anxiety Disorders*, 20(1), 110–117. <https://doi.org/10.1016/j.janxdis.2004.11.010>
- Takagi, S., Hiramatsu, S., Tabei, K.-i., & Tanaka, A. (2015). Multisensory perception of the six basic emotions is modulated by attentional instruction and unattended modality. [Original Research]. *Frontiers in Integrative Neuroscience*, 9, 1. <https://doi.org/10.3389/fnint.2015.00001>
- Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *Journal of Neurophysiology*, 100(6), 3328–3342. <https://doi.org/10.1152/jn.90355.2008>
- Vuilleumier, P., & Huang, Y.-M. (2009). Emotional attention: Uncovering the mechanisms of affective biases in perception. *Current Directions in Psychological Science*, 18(3), 148–152. <https://doi.org/10.1111/j.1467-8721.2009.01626.x>
- Whitehead, J. C., Spiouas, I., & Armony, J. L. (2024). Individual differences in the evaluation of ambiguous visual and auditory threat-related expressions. *European Journal of Neuroscience*, 59(3), 370–393. <https://doi.org/10.1111/ejn.16220>
- Whittle, S., Yücel, M., Yap, M. B. H., & Allen, N. B. (2011). Sex differences in the neural correlates of emotion: Evidence from neuroimaging. *Biological Psychology*, 87(3), 319–333. <https://doi.org/10.1016/j.biopsycho.2011.05.003>
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 159–177. <https://doi.org/10.1080/14639220210123806>
- Yang, M. (2025). How auditory sensations affect human emotional responses to acoustic environments. *Journal of Environmental Management*, 379, Article 124742. <https://doi.org/10.1016/j.jenvman.2025.124742>