Studia Psychologica, Vol. 68, No. 3, 2026, 280-298
https://doi.org/10.31577/sp.2026.03.949

Development and Psychometric Evaluation of a Multimodal Emotion Recognition Task Battery: Evidence from an Indian Sample

Meenakshi Shukla[1], Rakesh Pandey[2]
[1] Department of Psychology, University of Allahabad, Prayagraj, India-211002
[2] Department of Psychology, Banaras Hindu University, Varanasi, India-221005

Emotion recognition is fundamental to social functioning, yet most existing measures rely on explicit processing within a single modality, despite everyday emotion perception occurring implicitly and across multiple channels. This study reports the development and psychometric evaluation of a multimodal emotion recognition task battery assessing implicit and explicit processing across visual, auditory, and cross-modal modalities. Data were obtained from Indian adults (*N* = 175). The psychometric properties of the tasks were evaluated using item-level indices and internal consistency estimates, and inter-task associations were examined using correlations. An exploratory factor analysis was conducted to examine shared performance variance across tasks. The tasks demonstrated acceptable internal consistency across modalities and processing levels. Inter-task correlations indicated moderate convergence. Factor analysis revealed a dominant shared component accounting for approximately 69% of the variance, alongside residual variance suggestive of task-related differences. These findings indicate that the battery provides an initial set of reliable behavioural measures of emotion recognition. The results support substantial shared variance across tasks while allowing for additional task-specific influences. This work contributes a psychometrically informed tool for investigating emotion recognition in a non-Western sample.

Key words: emotion recognition, implicit processing, explicit processing, visual modality, auditory modality

Emotions form a crucial component of personal and social life. The abilities of perception, expression, understanding, and application of emotions play a central role in maintaining interpersonal functioning, psychological well-being, and adaptive social behaviour (Denham et al., 2003; Eisenberg et al., 2000). Emotion recognition is particularly important because emotional expressions provide signals that facilitate adaptive responses to significant events and interpersonal situations (Ekman, 1992). Given its significance, researchers have developed numerous tests to assess emotion recognition. However, these

Correspondence concerning this article should be addressed to Prof. Rakesh Pandey, Department of Psychology, Banaras Hindu University, India-221005. E-mail: pandeyr@bhu.ac.in, rpan_in@yahoo.com

Received December 31, 2025

measures vary considerably in the way they operationalise the construct, often focusing on specific tasks or modalities and relying predominantly on explicit emotion labelling.

An important theoretical question concerns whether emotion recognition represents a single overarching ability or consists of multiple distinct competencies tied to specific modalities and contexts. Some models propose that emotion recognition reflects a unified construct, whereas others suggest that it comprises separate modality-specific or task-specific skills. Bänziger, Grandjean, and Scherer (2009) raised this issue while developing the Multimodal Emotion Recognition Test (MERT), which assessed emotion recognition across four modalities: video, audio, audio–video, and still images. Their findings provided initial support for a general underlying factor, while also indicating modality-related influences, leading them to propose separate auditory and visual components rather than a strictly unidimensional structure.

Subsequent studies have examined this issue using a range of methodological approaches. Schlegel, Grandjean, and Scherer (2012) used confirmatory factor analysis and found little evidence for modality specificity, supporting the existence of a supramodal emotion recognition ability. Connolly et al. (2020) extended this work through latent variable modelling in large samples, demonstrating that individuals who perform well in one modality, such as facial expressions, also tend to perform well in others, such as vocal expressions. A meta-analysis by Schlegel et al. (2017) further reported a moderate but consistent correlation across emotion recognition tasks, suggesting the presence of shared variance alongside task-specific components. More recently, Mayer et al. (2024) situated emotion recognition within the broader framework of emotional intelligence, arguing that while a general factor is evident, distinct variance components associated with modality and task demands should also be acknowledged. Together, these findings indicate substantial shared variance across tasks, alongside possible task-specific influences.

Despite the variability in behavioural findings, converging evidence from neuroscience suggests that emotion recognition engages shared neural systems across sensory modalities. Neuroimaging studies indicate that regions such as the medial prefrontal cortex and superior temporal sulcus respond to emotional information conveyed through facial, vocal, and bodily cues (Peelen et al., 2010), while the posterior superior temporal gyrus is involved in processing emotional expressions across both auditory and visual inputs (Sievers et al., 2018). The amygdala, a key structure in emotional processing, receives input from multiple sensory channels (Amaral, 2003; Nishijo et al., 1988) and contributes to the processing of emotional signals across modalities, including visual (Phan et al., 2002) and auditory (Klinge et al., 2010) cues. While such findings support the presence of shared neural mechanisms, they do not preclude behavioural variability arising from task demands or processing requirements.

Building on this body of work, the present study focuses on the development and psychometric evaluation of a multimodal emotion recognition task battery designed to assess emotion recognition across different sensory modalities and levels of processing. Unlike many existing measures that emphasise explicit recognition within a single modality, the present battery comprises tasks that assess both implicit and explicit emotion recognition across visual, auditory, and cross-modal conditions. Prior emotion recognition tasks have varied widely in stimulus format and semantic constraint, particularly in the auditory domain, motivating the present task design choices discussed below. By examining performance across six distinct tasks, the study aims to evaluate the extent of shared performance variance in emotion recognition tasks while also considering the presence of task-linked variability.

In addition to addressing theoretical questions regarding the structure of emotion recognition, the study makes a methodological contribution by providing a psychometric evaluation of a multimodal emotion recognition battery in an Indian sample. Most existing emotion recognition tasks have been developed and validated primarily in Western populations, and relatively little is known about their performance characteristics in non-Western contexts. Although the facial stimuli used in the present study were drawn from established international datasets, the contribution of the study lies in examining task functioning, reliability, and performance associations within an Indian population rather than in the cultural specificity of the stimuli themselves.

Existing emotion recognition measures have differed not only in modality but also in the nature of the stimuli and the degree of semantic constraint. In the auditory domain, several widely used tasks rely on musical excerpts or affective vocalisations, such as the Musical Emotional Bursts (Paquette et al., 2013) and the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS; Livingstone & Russo, 2018), whereas other corpora use emotionally intoned words or sentences that may carry non-neutral semantic content (FAU Aibo Emotion Corpus; Burkhardt et al., 2005). These variations are important because semantic content and stimulus format can shape task demands and performance. In addition, theoretical accounts suggest that implicit and explicit emotion recognition may engage partly different processing requirements (Ethofer et al., 2009), and that facial and vocal cues can differ in the type of emotional information they convey, with facial cues often providing stronger information about valence and vocal cues about arousal (Mauss & Robinson, 2009). Together, this literature supports the need for task batteries that vary systematically across modality and processing level while maintaining careful control over stimulus characteristics.

Accordingly, the primary objectives of the present study were: 1) to develop and psychometrically evaluate a multimodal emotion recognition task battery assessing implicit and explicit processing across visual, auditory, and cross-modal tasks, and 2) to examine the dimensionality of the construct of emotion recognition based on shared variance across task performances using exploratory factor analytic methods. By addressing these objectives, the study seeks to contribute to ongoing debates on the dimensionality of emotion recognition while providing a validated behavioural tool assessing emotion recognition performance in a non-Western sample. The multimodal emotion recognition task battery developed in the present study is hereafter referred to as the *Multimodal Emotion Recognition Implicit–Explicit Tasks (MERIT)* battery.

Method

Participants

A purposive sampling procedure was followed to recruit participants. All participants recruited in the present study were Asian Indian in ethnicity and belonged to a middle socio-economic status. The socio-economic status of the participants was defined in terms of their monthly per capita income, according to the criteria set out in the BG Prasad socio-economic classification (Khairnar et al., 2016). Individuals who regularly smoked or consumed alcohol, had a history of mental illness, uncorrected vision or hearing problems, were undergoing treatment or medication for any physical or mental health condition, or were pregnant or lactating, were excluded from the study. Additional medical conditions, such as thyroid disorders, renal disease, or cardiac conditions, were also part of the exclusion criteria. These criteria were applied to minimise potential confounding influences on emotion recognition performance, including alterations in emotional processing, cognitive functioning, or sensory perception that may arise from clinical conditions, medical treatments, or physiological changes. The Ethics Committee, Institute of Science, Banaras Hindu University, India, provided ethics approval for the present study (Ref. No.: I.Sc./ECM IX/2016–17/02), and written informed consent was obtained from each participant.

The first part of the study focused on the development of emotion recognition tasks to be included in the multimodal emotion recognition task battery, and the second part focused on the validation of the newly developed multimodal task battery, including its factor structure. Since the two parts of the study involved samples recruited for different purposes, with partial overlap in participants, the method, results, and discussion of each part have been presented separately, followed by a General Discussion.

Part 1: Development and preliminary psychometric evaluation of a multimodal emotion recognition task battery

Method

Participants

The study consisted of two phases: 1) task-wise psychometric screening to ensure adequate item functioning and internal consistency, and 2) an exploratory analysis of performance patterns across the full task set. Participants for Part 1 were recruited from undergraduate and postgraduate colleges, as well as from community samples in the local region, using convenience sampling via advertisements on social media platforms (Facebook and WhatsApp). The participants recruited for the initial task-wise screening differed across the six tasks. Initially, 111 participants were approached. The six tasks were developed sequentially, and data were collected to evaluate item-level performance characteristics (e.g., item difficulty and response distribution) and basic psychometric properties at the item and scale levels, including internal consistency, as each task was finalised. As a result, participant recruitment occurred in phases, leading to varying sample sizes across tasks (initial N = 111). Not all participants completed every task due to phased task completion and time constraints. Accordingly, the task-wise sample sizes reported in Table 1 reflect these partially overlapping samples rather than independent participant groups. Given that the aim of Part 1 was preliminary item-level psychometric screening rather than hypothesis testing or precise parameter estimation, the task-wise sample sizes were deemed appropriate for evaluating basic reliability indices and identifying poorly functioning tasks prior to further validation. Accordingly, a formal power analysis was not conducted, as the primary aim was item screening across various tasks and the evaluation of item functioning. Once all tasks were finalised, a larger sample (N = 175) was recruited to examine performance relationships across the complete task set and evaluate its underlying performance structure. Demographic information for participants who completed the initial task-wise screening phase is presented in Table 1.

A total of 175 participants (84 men, 91 women) were recruited to collect data to examine the performance structure of emotion

Table 1 *Demographic information of the participants recruited for the task-wise psychometric screening*

Emotion recognition tasks	Sample size	Gender distribution	Age range (years)	Average Age (years)	Average Education (years)
Visual Implicit	101	80 women, 21 men	18-46	25.50 ± 5.80	15.99 ± 2.35
Auditory Implicit	75	54 women, 21 men	18-41	24.19 ± 3.79	16.05 ± 2.38
Visual target-Auditory response options	111	82 women, 29 men	18-41	23.63 ± 3.63	15.87 ± 2.24
Auditory target-Visual response options	100	70 women, 30 men	18-41	23.65 ± 3.61	15.93 ± 2.25
Visual Explicit	103	82 women, 21 men	18-53	25.93 ± 7.04	16.04 ± 2.66
Auditory Explicit	99	73 women, 26 men	18-52	24.02 ± 4.64	15.94 ± 2.34

Note. Sample sizes vary across tasks because Part 1 used staged task development and task-wise screening. Participants were recruited in phases as each task was finalised; therefore, the task-wise samples are *partially overlapping* (some participants completed more than one task) but *not fully identical* across all tasks.

recognition across tasks. This included some participants who had provided their data for the earlier task-wise screening analyses. These participants ranged in age from 19 to 62 years (Mean ± standard deviation (*SD*) = 34.73 ± 14.09). The mean number of years of formal education obtained by these participants was 14.73 ± 3.68 years.

Measures

The emotion recognition test battery we previously developed was used in this study. It comprises six computerised emotion recognition tasks designed to vary systematically by modality (visual, auditory, cross-modal) and processing level (implicit vs. explicit). Four tasks required implicit emotion processing (emotion matching), and two required explicit emotion processing (emotion labelling). Two tasks each assessed emotion processing in the visual, auditory, and audio-visual modalities. Each of the six tasks comprised 24 trials (four trials for each of the six basic emotions: happiness, sadness, fear, anger, surprise, and disgust), and each trial was presented until a response was obtained from the participant, which consisted of a numeric key press between 1 and 4. The tasks were developed and presented using Paradigm software (v 2.4.0.191). Accuracy of emotion identification or matching served as the primary behavioural index of emotion recognition performance. All stimulus databases used across the six tasks were accessed with prior permission from the respective developers or institutions and were used strictly for research purposes in accordance with the specified usage agreements. Following is a description of the six tasks comprising the MERIT battery.

Visual Implicit Emotion Recognition Task

The *Visual implicit emotion recognition task* (alternatively called the *Visual implicit task;* Figure 1a) was designed to assess implicit (or covert) recognition of visually presented emotions. Each trial of the task comprised a facial emotion video of the upper half of a human face and four facial emotion videos of the lower half of the same face (as in the upper half), each displaying a different emotion. The upper half of the face, displaying the eye region, formed the target, while the lower halves of the face, displaying the mouth region, formed the response choices. The lower halves of the face expressed different emotions with only one correct response. The task required choosing the lower facial half, which, when combined with the upper half, would meaningfully complete the face.

To develop the upper face and lower face videos, the image sequences from the chosen four posers in the Cohn-Kanade Action Unit-Coded Facial Expression Database (Kanade et al., 2000; Lucey et al., 2010) were split horizontally at the middle dividing the face into equal-sized upper and lower halves. For each emotion sequence, all the upper face images (ranging from neutral to the full emotion display) were combined to form upper face videos, while all the lower face images were combined to form lower face videos using Windows Movie Maker. Since the participants were not explicitly told that the matching would entail recognising the emotions presented in the two facial halves, correct matching of the facial halves indicated implicit emotion recognition.

Auditory Implicit Emotion Recognition Task

The *Auditory implicit emotion recognition task* (alternatively called the *Auditory implicit task*; Figure 1b) was developed as the auditory counterpart of the visual implicit task and was intended to assess implicit (or covert) recognition of auditorily presented emotions. The sound clips used in this task were taken from the Indian Institute of Technology Kharagpur Simulated Emotion Hindi Speech Corpus (IITKGP-SEHSC; Koolagudi et al., 2011). This speech corpus was recorded by trained radio artists from the Gyanavani FM radio station in Varanasi, India. The corpus was created by eliciting emotional expressions through neutral semantic content (e.g., '*all that glitters is not gold*'), allowing emotion to be conveyed primarily through prosody rather than linguistic meaning. The database includes recordings corresponding to seven emotional categories (anger, disgust, fear, happiness,

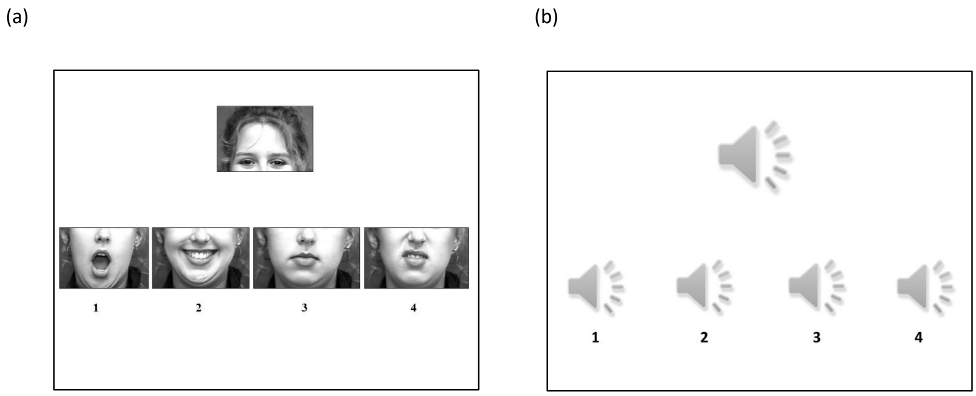

Figure 1 Diagrammatic representation of the visual implicit (a) and auditory implicit (b) emotion recognition tasks (Images © Jeffrey Cohn).

sadness, sarcasm, and surprise) and a neutral condition. As part of the original corpus validation, the emotional utterances were evaluated by postgraduate and research students from IIT Kharagpur, who identified the expressed emotion from a fixed set of labels. Overall recognition accuracy for the corpus was reported at 74%, indicating acceptable perceptual validity. For the purposes of the present study, a subset of 24 auditory stimuli was selected, comprising four exemplars for each basic emotion category.

In this task, a target sound clip appeared in the upper centre of the laptop screen with four numbered audio clips as response options given below. This task required selecting the emotion clip presented in the response that best matched the one in the target. In all five sound clips (1 target + 4 response choices) presented in each trial, the neutral statement spoken was the same, but the gender of the speaker differed for the target audio and the response clips; for instance, if the target audio was a male voice, then the response audio clips contained the same statement spoken in a female voice, and vice versa. The same statement was chosen for the five sound clips of a particular trial so that variation in the length of the sound clips and the spoken words would not affect the responses. This difference in the gender of the speaker in the target and response audio clips was created so that participants matched on the basis of emotional similarity rather than structural similarities between the target and response clips. The target sound clip was played twice, followed by a one-time presentation of each of the response audio clips from 1 to 4, and thereafter this pattern of sound clips presentation was repeated until the participants entered a response. Since the participants were not explicitly told to match the audio clips based on the similarity of the emotional information presented, correct responding (emotional match between target and test) was considered to indicate an implicit processing of emotions.

Visual Target-Auditory Response Options Cross-modal Emotion Recognition Task

The *Visual target-auditory response options cross-modal emotion recognition task* (alternatively called the *Visual target-auditory response options task*) (Figure 2a) was intended

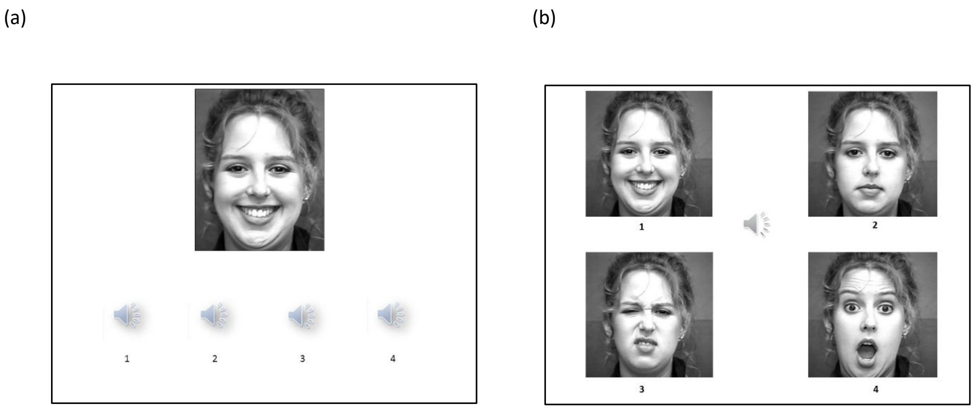

Figure 2 Diagrammatic representation of the visual target-auditory response options task and the auditory target-visual response options task (Images © Jeffrey Cohn).

to assess implicit emotion recognition through a cross-modality matching of emotional information. In this task, a full-face emotion video was presented as the target in the upper centre of the laptop screen with four numbered audio response options.

To develop the full-face emotion videos used in this task, images from each poser, ranging from neutral to the target emotion, were simply combined in Windows Movie Maker. At the beginning of every trial in this task, the audio clips started playing automatically one after the other. After the fourth audio clip had played, the audio clips started playing again, sequentially from the first audio clip. This looping of the audio clips and of the facial emotion video continued until a response was recorded from the participant. Implicit processing was inferred because no instruction was given to match stimuli based on emotional similarity, and correct responses required integration of emotional information across modalities.

Auditory Target-Visual Response Options Cross-modal Emotion Recognition Task

The *Auditory target-visual response options cross-modal emotion recognition task* (alternatively called the *Auditory target-visual response options task*) (Figure 2b) was developed to counterbalance the *Visual target-auditory response options task* and therefore entailed matching of a target audio with its corresponding facial emotion out of the given four choices. The full-face emotion videos used in this task were developed in the same manner as for the *Visual target-auditory response options task*. The audio clip and the facial videos began playing automatically at the start of each trial and looped repeatedly throughout the trial. The accuracy of emotion recognition on this task was considered to tap implicit emotion processing, as participants were not prompted to match the visual and auditory stimuli on the basis of the similarity of the emotions expressed.

Visual Explicit Emotion Recognition Task

In the *Visual explicit emotion recognition task* (alternatively called the *Visual explicit task*; Figure 3a), full-face videos displaying different emotions were presented in the upper centre of the computer screen, with four response options (Hindi names of the emotions) given below. One of the labels was the correct response option denoting the emotion expressed

(a) (b)

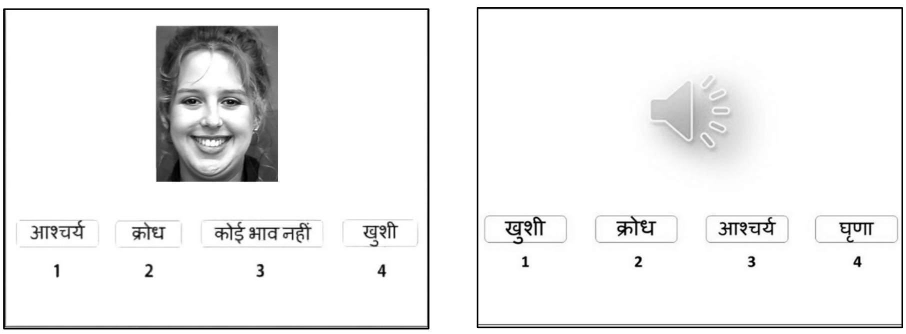

Figure 3 Diagrammatic representation of the visual explicit emotion recognition task and the auditory explicit emotion recognition task (Images © Jeffrey Cohn).

in the facial video, while the remaining three were distractor labels. The distractor labels always included a 'neutral' emotion label, while the remaining two response options included randomly selected labels from the remaining five basic emotions. This was an explicit task, as participants were instructed to label the emotions in the facial expression video.

Auditory Explicit Emotion Recognition Task

The *Auditory explicit emotion recognition task* (alternatively called the *Auditory explicit task*; Figure 3b) presented participants with a target audio clip and four emotion labels (in Hindi). One of the emotion labels in each trial was neutral, while another was the correct emotion label for the emotion presented in the sound clip. The remaining two were randomly selected from the five remaining basic emotions (excluding the correct emotion label). The audio clip played in loops throughout the trial or until a response was entered. The participants were required to label the emotion expressed in the sound clip.

Procedure

For the task-wise screening phase, participants were tested individually in the laboratory. They were asked to provide a written informed consent, following which their demographic information was collected. Participants were administered the six tasks one by one, with about a 10-minute rest period between tasks. Participants who did not wish to complete all six tasks were allowed to leave. For implicit tasks, participants were not asked to explicitly label emotions. Instead, they were instructed to make speeded matching decisions (e.g., identify which of the presented options best matched the target stimulus) using labelled response keys. Practice trials were provided before each task to ensure comprehension. For the explicit tasks, participants were instructed to carefully observe or listen to each stimulus and to identify the emotion being expressed. Participants were instructed to identify the emotion expressed in each stimulus by pressing the key corresponding to their chosen emotion label. Participants were told to respond as quickly and as accurately as possible. Practice trials were provided before each explicit task to ensure that participants understood the instructions and response format. Following completion of the tasks, participants were debriefed about the general purpose of the study (emotion recognition across tasks and modalities), and any questions were addressed.

For the performance structure analysis conducted during Part 1, data from participants who completed all six tasks were used for inter-task correlation analyses (N = 63). Item-level indices (corrected item–total correlations, item difficulty, and alpha-if-item-deleted) were computed to ensure adequate task functioning prior to performance-based analyses. Cronbach's alpha values above .60 were considered acceptable for the purposes of exploratory analysis (Herzmann et al., 2008). Inter-task Pearson correlations were computed to examine shared variance across tasks.

Results

Task Functioning and Internal Consistency

The Cronbach's alpha reliabilities of the six tasks, along with the ranges of item–total correlations and item difficulties, are shown in Table 2. Across tasks, internal consistency estimates fell within acceptable ranges with most items demonstrating moderate item–total correlations and item difficulty values within conventional bounds. A small number of items showed item difficulty values above .90, indi-

Table 2 *Internal-consistency reliabilities, ranges of item-total correlations and item difficulties for the six emotion recognition tasks*

Emotion recognition tasks	**Cronbach's alpha**	**Range of item-total correlations**	**Range of item difficulty**
Visual Implicit	.708	.126 - .507	.27 - .97
Auditory Implicit	.820	.115 - .563	.28 - .92
Visual target-Auditory response options	.713	.124 - .398	.20-.92
Auditory target-Visual response options	.685	.002 - .497	.27- .94
Visual Explicit	.700	.209 - .405	.49- .98
Auditory Explicit	.656	.085 - .468	.71- .96

cating ceiling-level performance. These items were retained to preserve coverage of the full performance range across tasks, as ceiling performance on a limited subset of items does not necessarily undermine task-level reliability in non-clinical samples. The proportion of such items was kept minimal.

Overall, the pattern of item difficulties and internal consistency estimates indicated that the six tasks functioned adequately as behavioural measures of emotion recognition performance in the present sample.

The alpha-if-item-deleted values for all the items of the six tasks were less than or equal to the overall alpha for the respective task except for item 3 of Fear (visual implicit task), item 2 of Anger and items 3 and 4 of Disgust (auditory implicit task), item 3 of Fear and 1 of Disgust (visual target-auditory response options task), item 4 of Fear, item 3 of Disgust, item 3 of Happy, and item 3 of Surprise (auditory target-visual response options task), and item 4 of Sadness and 1 and 3 of Fear (auditory explicit task). Removal of these items resulted in only negligible changes to task-level alpha values, and therefore, these items were not dropped.

Discussion

In the present study, six emotion recognition tasks varying in level of processing and sensory modality were examined to characterise task functioning across differences in processing level (implicit vs. explicit) and stimulus modality. The tasks included two unimodal implicit as well as explicit tasks in the visual and auditory modalities and two cross-modal tasks (which were implicit by design). Rather than assuming equivalence across tasks, the present approach incorporated multiple task types that differ in cognitive and perceptual requirements, allowing evaluation of task functioning across processing levels and modalities. The implicit tasks are relatively uncommon in the emotion recognition literature, as most existing measures rely on explicit emotion identification.

In the auditory domain, the present implicit task employed semantically neutral Hindi statements spoken with different emotional prosodies, allowing emotion recognition to be assessed with minimal semantic confounding and within a linguistically appropriate context. The two cross-modal implicit tasks were designed to approximate everyday emotion recognition, in which emotional information is often conveyed simultaneously across multiple sensory channels. These tasks, therefore, allowed examination of how emotional cues are matched across modalities under minimal explicit instruction. The auditory explicit task addressed a relative gap in the literature, as explicit labelling of vocal emotions (particularly using culturally appropriate, semantically neutral stimuli) has been less frequently examined. Similarly, the visual explicit task used dynamic facial expressions rather than static images (as in most previous databases), which may better reflect real-world emotion recognition.

The examination of task functioning indicated adequate internal consistency and item performance across all six tasks, suggesting that they operated as stable behavioural measures within the present sample. While further work with larger and independent samples would be necessary to establish generalisability, the present findings indicate that the tasks were sufficiently sensitive to capture individual variabilities in emotion recognition performance.

Task-level performance patterns differed as a function of processing level and modality. Implicit tasks exhibited a wider range of item difficulties, including relatively challenging items, whereas explicit tasks showed a narrower difficulty range with some items approaching ceiling performance. These patterns suggest that implicit emotion recognition may impose greater perceptual or integrative demands than explicit labelling. Modality effects were also evident: during explicit processing, auditory stimuli were generally easier to recognise than visual stimuli, whereas during implicit processing, tasks involving multiple auditory inputs were more demanding than those involving multiple visual inputs. Together, these findings highlight the importance of considering both modality and processing level when interpreting emotion recognition performance.

The task-level performance patterns observed in the present study can be considered alongside prior emotion recognition research examining performance across different task formats and modalities. Previous multimodal investigations, such as the MERT (Bänziger et al., 2009), and large-scale syntheses of emotion recognition tasks (Schlegel et al., 2017) have demonstrated that emotion recognition performance varies across task configurations, with only moderate correspondence across different measures. While these studies did not explicitly contrast implicit and explicit task difficulty, their findings underscore the broader point that task structure and stimulus characteristics influence observed performance. In this context, the wider range of item difficulty observed for implicit tasks in the present study highlights how variations in processing demands may shape emotion recognition outcomes, even when tasks are intended to assess the same underlying construct.

Part 2: Validation and exploratory analysis of performance across the MERIT battery

Method

Participants

With six observed variables, the sample size used in Study 2 (N = 175) exceeds the commonly recommended subject-to-variable ratios for exploratory factor analysis. Methodological work has shown that when communalities are moderate to high, and factor loadings are strong, stable EFA solutions can be obtained with relatively modest sample sizes (Costello & Osborne, 2005; Lorenzo-Seva & Ferrando, 2024). In the present analysis, all tasks demonstrated strong loadings on the dominant factor, supporting the adequacy of the sample for exploratory purposes. All participants (N = 175) were Asian Indians of middle-class socio-economic status, as per the BG Prasad socio-economic classification in India (Khairnar et al., 2016). Participants for Part 2 of the study were recruited from undergraduate and postgraduate colleges and from the surrounding community in the local region using convenience sampling. Recruitment advertisements were circulated on online platforms such as Facebook and WhatsApp. The same inclusion and exclusion criteria as Part 1 were applied.

Measures

The same six emotion recognition tasks described in Part 1 were administered in the present study.

Procedure

Before arriving at the experimental session, participants were requested to avoid caffeine and alcohol, as well as heavy exercise, for about two hours prior to their arrival to minimise short-term physiological arousal and the related changes in blood pressure that could influence emotion recognition performance. This instruction was guided by prior evidence linking blood pressure and related physiological functioning with individual differences in emotion processing and recognition accuracy (Jain et al., 2017; McCubbin et al., 2011, 2014; Mishra & Shukla, 2026; Pury et al., 2004; Raj et al., 2024; Shukla et al., 2017, 2018, 2019, 2024; Shukla & Acharya, 2025; Shukla & Pandey, 2024). When the participants arrived, they were asked to sit in a comfortable chair for about 10 minutes, during which time their demographic information was collected. Only such participants were allowed to continue with the study who had reported no smoking, drinking of alcoholic beverages, and who were not being treated for any ailment and were not on any kind of medication. Participants were then given all necessary information about the study and asked to fill out a consent form if they wished to participate. Consenting participants were presented with on-screen instructions on the laptop in Hindi, and simultaneously, verbal instructions were given, depending on the task being administered. Participants were first administered all the implicit tasks, followed by the two explicit tasks. After participants practised each task for three practice trials, they were asked to complete the particular task. After the task was completed, the participants were thanked, and the data were analysed. The accuracy of emotion recognition was calculated as the percentage of correct responses.

Results

To examine whether performance across the six emotion recognition tasks reflected shared variance, accuracy scores from all six tasks were subjected to an exploratory factor analysis (EFA). Principal components analysis was used as the extraction method. Bartlett's test of sphericity was found to be significant ($\chi^2(15)$ = 653.28, p < .001), indicating that the variables are sufficiently correlated and that the correlations are significantly higher than zero; thus, the correlation matrix is appropriate for conducting factor analysis. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO = .89) indicated that the variables share a high proportion of common variance, and that factor analysis may be appropriately used to uncover the nature and source of this common variance. The determinant was greater than 0.00001 (Determinant = .022), indicating that multicollinearity was not a concern.

After establishing the suitability of the data, principal components analysis was conducted. Inspection of the scree plot suggested a dominant first component; therefore, no rotation was applied. The unrotated component matrix is presented in Table 3.

Table 3 shows that all six tasks loaded strongly on a single component, accounting for 68.56% of the total variance. This pattern indicates substantial shared variance across tasks, despite differences in modality and processing level. Given that each task assessed a distinct configuration of emotional processing demands, the extracted component was interpreted as reflecting common performance variance across emotion recognition tasks rather than a definitive latent trait.

Although scree-plot inspection is informative, it may be subjective, particularly when inflection points are ambiguous (Hayton et al., 2004). To further evaluate the dimensionality of task performance, parallel analysis (PA) and the minimum average partial (MAP) procedure were conducted as complementary exploratory criteria. Parallel analysis indicated that only the first eigenvalue from the observed data (4.11) exceeded the corresponding eigenvalue obtained from random data (1.25). The MAP procedure likewise suggested retention of a single component.

Together, these converging exploratory indicators suggest that performance across the six tasks is characterised by a dominant shared component, while not precluding task-specific variance attributable to modality and processing differences.

Inter-Task Associations (Convergent Validity)

Bivariate Pearson correlations were computed to examine associations among performance accuracy across the six emotion recognition tasks in the full Study 2 sample (N = 175). These analyses were conducted to assess convergent validity by examining the extent to which performance covaried across tasks differing in modality and processing demands.

As shown in Table 4, all inter-task correlations were statistically significant (ps < .001), indicating shared variance across tasks. Overall, the pattern of correlations suggests partial overlap in performance across tasks, alongside variability that may reflect task-specific demands.

Table 3 *Unrotated component matrix displaying factor loadings for the six emotion recognition tasks*

Tasks	**Component 1** ‘General emotion recognition ability’
Visual implicit	.821
Auditory implicit	.853
Visual target-auditory response options	.883
Auditory target-visual response options	.860
Visual explicit	.736
Auditory explicit	.806

Eigenvalue: 4.113 and percent variance accounted for = 68.56%

Table 4 *Inter-correlations of accuracy scores across the six emotion recognition tasks*

	Visual Implicit	**Auditory Implicit**	**Visual target-Auditory response options**	**Auditory target-Visual response options**	**Visual Explicit**	**Auditory Explicit**
Visual Implicit	1					
Auditory Implicit	.648***	1				
Visual target-Auditory response options	.732***	.743***	1			
Auditory target-Visual response options	.630***	.704***	.751***	1		
Visual Explicit	.496***	.518***	.528***	.531***	1	
Auditory Explicit	.559***	.594***	.598***	.628***	.645***	1

Note. ***p < .001

Discussion

This study examined how emotion recognition performance varies across tasks differing in modality and processing demands, with a particular focus on the extent to which performance covaries across tasks. Exploratory factor analysis revealed a dominant component on which all six tasks loaded strongly, accounting for approximately 69% of the total variance. Rather than establishing a definitive latent construct, this pattern indicates substantial shared variance across emotion recognition tasks that differ in sensory modality and level of processing.

Converging evidence from inter-task correlations further supports this interpretation. Moderate but non-uniform associations across tasks indicate partial convergence in performance alongside task-specific variability. This pattern is consistent with the view that emotion recognition tasks draw on shared processing resources while also engaging distinct perceptual and cognitive mechanisms depending on modality and processing demands. Additionally, it provides evidence for the convergent validity of the tasks. Prior work suggests that implicit and explicit emotion recognition may rely on partly distinct neural and cognitive pathways (Ethofer et al., 2009), and that facial and vocal cues differ in the type of emotional information they convey, with facial expressions often providing stronger information about valence and vocal cues about arousal (Mauss & Robinson, 2009). Such differences may contribute to the observed task-specific dissociations.

The present findings are broadly consistent with previous research reporting supramodal overlap in emotion recognition performance (Connolly et al., 2020) and with meta-analytic evidence indicating moderate correlations across emotion recognition tasks (Schlegel et al., 2017). At the same time, the incomplete overlap observed across tasks aligns with prior work emphasising modality- and task-specific variance (Bänziger et al., 2009; Mayer et al., 2024). Thus, the present results contribute to this literature by empirically illustrating how shared and task-specific influences coexist within a single multimodal task set.

Importantly, these findings do not negate modality-specific or valence-related accounts of emotion recognition. Rather, they suggest that such distinctions may operate alongside a broader performance-related component that captures variance common to multiple task configurations. In this respect, the results complement rather than contradict earlier models proposing auditory–visual or modality-based differentiation in emotion recognition.

General Discussion

Taken together, the present findings suggest that emotion recognition performance cannot be fully characterised by a single, uniform ability across tasks. Instead, performance reflects a combination of shared and task-specific influences shaped by modality and processing demands. Rather than resolving whether emotion recognition is unidimensional or multidimensional, the present study provides empirical evidence that different task configurations yield partially overlapping but distinct performance patterns. This nuanced view aligns with contemporary perspectives that emphasise heterogeneity in emotional processing rather than strong unitary constructs.

When considered alongside prior multimodal emotion recognition research, the present findings show both convergence and extension. Similar to results obtained with established task batteries such as the MERT (Bänziger et al., 2009) and latent-variable investigations of emotion recognition performance (Connolly et al., 2020), performance across tasks in the present study exhibited substantial shared variance, with a dominant component accounting for approximately 69% of the total variance. At the same time, the presence of residual unexplained variance indicates that shared variance does not fully account for performance across all tasks. This pattern is consistent with meta-analytic evidence reporting only moderate correspondence across emotion recognition measures (Schlegel et al., 2017) and suggests that differences in task configuration, modality, or processing demands may contribute additional variance beyond a common performance component. Importantly, the present analyses do not directly isolate task-specific factors, and future research will be required to formally model and test these components.

Across two interlinked empirical studies, the present paper examined performance patterns in emotion recognition using a set of tasks that varied systematically by modality and processing level. Part 1 of the study demonstrated that the tasks functioned adequately as behavioural measures and were sensitive to individual variability. Part 2 of the study extended these findings by showing that performance across tasks shared a substantial common component, while also leaving room for task-specific variability.

Together, these findings suggest that emotion recognition performance reflects both shared and context-dependent processes rather than a strictly uniform or fully fractionated system. This interpretation aligns with contemporary perspectives in cognitive and affective science that emphasise heterogeneity in emotional processing across tasks and contexts. This pattern is broadly consistent with prior multimodal and latent-variable studies reporting substantial shared variance alongside task-specific influences in emotion recognition performance.

The present work also contributes to cross-cultural emotion research by employing linguistically and culturally appropriate auditory stimuli, thereby reducing semantic and cultural confounds in vocal emotion recognition. While the visual stimuli were not culture-matched, the inclusion of Hindi-language auditory tasks represents an important step toward broadening the cultural scope of emotion recognition research.

Taken together, the internal consistency estimates and the emergence of a coherent latent structure provide initial psychometric support for the reliability and construct validity of the MERIT battery. While the present findings do not constitute exhaustive validation, they offer foundational evidence that the tasks consistently assess a common underlying construct of emotion recognition. Rather than claiming full standardisation at this stage, the present study provides initial psychometric grounding for the task set and demonstrates its utility as an empirical framework for examining how emotion recognition performance varies across task demands. This framework may be particularly useful for future research investigating individual differences, cultural influences, and contextual modulation of emotional processing.

Limitations and Future Directions

Several limitations warrant consideration. First, the staged nature of task development and sampling limits the extent to which conclusions can be generalised. Second, the use of facial stimuli from a racial group other than the participants' introduces the possibility of cross-race effects (Elfenbein & Ambady, 2002). Future studies should use culturally matched visual stimuli to isolate emotion recognition performance from perceptual familiarity effects more precisely. Third, although the task set spans visual, auditory, and cross-modal domains, it does not encompass all channels of emotional communication, such as tactile or olfactory cues. Lastly, although the sample size used in Part 2 of the study was adequate for an exploratory factor analysis, given the small number of observed variables and the strength of the factor loadings, the present findings should be interpreted as preliminary. Future research should seek to replicate the factor structure of the MERIT battery in larger independent samples to establish further the stability and generalisability of the observed latent structure.

Conclusion

The present study provides empirical evidence that emotion recognition performance across diverse tasks is characterised by substantial shared variance, with residual variance indicating that common performance factors do not fully account for all observed differences across tasks. By examining performance across implicit and explicit tasks and multiple sensory modalities, the findings illustrate how task configuration and processing demands may shape observed emotion recognition outcomes. Rather than definitively resolving whether emotion recognition constitutes a strictly single ability, the present work clarifies how convergence across tasks can coexist with residual divergence in performance. Accordingly, the findings provide initial evidence for the reliability and construct validity of a multimodal assessment of emotion recognition, while underscoring the need for further validation across populations and contexts. Future research using larger samples, confirmatory modelling, and culturally matched stimuli will be essential for further elucidating the structure of emotion recognition.

Acknowledgement

This work was supported by Junior Research Fellowship of the University Grants

Commission [Grant number: F.15-9(JUNE, 2013)/2013(NET)].

Authors' ORCID
Meenakshi Shukla
https://orcid.org/0000-0001-9447-5987
Rakesh Pandey
https://orcid.org/0000-0001-8024-300X

References

Amaral, D. G. (2003). The amygdala, social behavior, and danger detection. *Annals of the New York Academy of Sciences*, *1000*, 337–347. https://doi.org/10.1196/annals.1280.015

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion*, *9*(5), 691–704. https://doi.org/10.1037/a0017088

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of German emotional speech. In *Ninth European Conference on Speech Communication and Technology*.

Connolly, H. L., Lefevre, C. E., Young, A. W., & Lewis, G. J. (2020). Emotion recognition ability: Evidence for a supramodal factor and its links to social cognition. *Cognition, 197*, 104166. https://doi.org/10.1016/j.cognition.2019.104166

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(7), 1–9. https://doi.org/10.7275/jyj1-4868

Denham, S. A., Blair, K. A., DeMulder, E., Levitas, J., Sawyer, K., Auerbach-Major, S., & Queenan, P. (2003). Preschool emotional competence: Pathway to social competence. *Child Development*, *74*, 238–256. http://dx.doi.org/10.1111/1467-8624.00533

Eisenberg, N., Fabes, R. A., Guthrie, I. K., & Reiser, M. (2000). Dispositional emotionality and regulation: Their role in predicting quality of social functioning. *Journal of Personality and Social Psychology*, *78*, 136–157. http://doi.org/10.1037/0022-3514.78.1.136

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Elfenbein, H. A., & Ambady, N. (2002). Is there an in-group advantage in emotion recognition? *Psychological Bulletin, 128*(2), 243–249. https://doi.org/10.1037/0033-2909.128.2.243

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience*, *21*(7), 1255–1268. https://doi.org/10.1162/jocn.2009.21099

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191–205. https://doi.org/10.1177/1094428104263675

Herzmann, G., Danthiir, V., Schacht, A., Sommer, W., & Wilhelm, O. (2008). Toward a comprehensive test battery for face cognition: Assessment of the tasks. *Behavior Research Methods*, *40*(3), 840–857. https://doi.org/10.3758/BRM.40.3.840

Jain, D., Shukla, M., & Pandey, R. (2017). Blood pressure variations and emotional dampening: Preliminary evidence for the curvilinear relationship. *Journal of Projective Psychology & Mental Health, 24*(1), 18–25.

Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00),* Grenoble, France, 46–53.

Khairnar, M. R., Wadgave, U., & Shimpi, P. V. (2016). Updated B. G. Prasad socioeconomic classification for 2016. *Journal of Indian Association of Public Health Dentistry, 14*(4), 469–470. https://doi.org/10.4103/2319-5932.195832

Klinge, C., Röder, B., & Büchel, C. (2010). Increased amygdala activation to emotional auditory stimuli in the blind. *Brain*, *133*(6), 1729–1736. https://doi.org/10.1093/brain/awq102

Koolagudi, S. G., Reddy, R., Yadav, J., & Rao, K. S. (2011, February). IITKGP-SEHSC: Hindi speech corpus for emotion analysis. In *International Conference on Devices and Communications (ICDeCom).* (pp. 1–5). Mesra, India: IEEE. https://doi.org/10.1109/ICDECOM.2011.5738540

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), e0196391. https://doi.org/10.1371/journal.pone.0196391

Lorenzo-Seva, U., & Ferrando, P. J. (2024). Determining sample size requirements in EFA solutions: A simple empirical proposal. *Multivariate Behavioral Research*, *59*(5), 899–912. https://doi.org/10.1080/00273171.2024.2342324

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. *Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010),* San Francisco, USA, 94–101.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion*, *23*(2), 209–237. https://doi.org/10.1080/02699930802204677

Mayer, J. D., Panter, A. T., & Caruso, D. R. (2024). Is emotion recognition a form of intelligence? A conceptual analysis and empirical investigation. *Personality and Individual Differences, 209*, 112468. https://doi.org/10.1016/j.paid.2023.112468

McCubbin, J. A., Loveless, J. P., Graham, J. G., Hall, G. A., Bart, R. M., Moore, D. D., Merritt, M. M., Lane, R. D., & Thayer, J. F. (2014). Emotional dampening in persons with elevated blood pressure: Affect dysregulation and risk for hypertension. *Annals of Behavioral Medicine*, *47*(1), 111–119. https://doi.org/10.1007/s12160-013-9526-2

McCubbin, J. A., Merritt, M. M., Sollers III, J. J., Evans, M. K., Zonderman, A. B., Lane, R. D., & Thayer, J. F. (2011). Cardiovascular emotional dampening: The relationship between blood pressure and recognition of emotion. *Psychosomatic Medicine*, *73*(9), 743–750. https://doi.org/10.1097/PSY.0b013e318235ed55

Mishra, A., & Shukla, M. (2026). Can cognitive impairments explain cardiovascular emotional dampening associated with elevated blood pressure? *Journal of Health Psychology.* https://doi.org/10.1177/13591053261425474

Nishijo, H., Ono, T., & Nishino, H. (1988). Single neuron responses in amygdala of alert monkey during complex sensory stimulation with affective significance. *Journal of Neuroscience*, *8*(10), 3570–3583. https://doi.org/10.1523/JNEUROSCI.08-10-03570.1988

Paquette, S., Peretz, I., & Belin, P. (2013). The "Musical Emotional Bursts": A validated set of musical affect bursts to investigate auditory affective processing. *Frontiers in Psychology*, *4*, 509. https://doi.org/10.3389/fpsyg.2013.00509

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience*, *30*(30), 10127–10134. https://doi.org/10.1523/JNEUROSCI.2161-10.2010

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, *16*(2), 331–348. https://doi.org/10.1006/nimg.2002.1087

Pury, C. L., McCubbin, J. A., Helfer, S. G., Galloway, C., & McMullen, L. J. (2004). Elevated resting blood pressure and dampened emotional response. *Psychosomatic Medicine*, *66*(4), 583–587. https://doi.org/10.1097/01.psy.0000130490.57706.88

Raj, A., Shukla, M., & Kumari, A. (2024). Dampened emotion recognition from bodily gestures: Evidence for sex-specific cardiovascular emotional dampening among prehypertensive and hypertensive individuals. *Psychologie Française, 69*(3), 287–303. https://doi.org/10.1016/j.psfr.2024.01.002

Schlegel, K., Boone, R. T., Hall, J. A., & Murphy, N. A. (2017). Individual differences in interpersonal accuracy: A meta-analysis. *Journal of Nonverbal Behavior, 41*(2), 103–137. https://doi.org/10.1007/s10919-017-0249-0

Schlegel, K., Grandjean, D., & Scherer, K. R. (2012). Emotion recognition: Unidimensional ability or a set of modality-and emotion-specific skills? *Personality and Individual Differences*, *53*(1), 16–21. https://doi.org/10.1016/j.paid.2012.01.026

Shukla, M., & Acharya, S. (2025). Cardiovascular emotional dampening and olfaction: How blood pressure affects emotional responses to odors. *Japanese Psychological Research*. https://doi.org/10.1111/jpr.12598

Shukla, M., Lau, J. Y. F, & Pandey, R. (2024). Behavioural approach-avoidance tendencies among individuals with elevated blood pressure. *Current Psychology, 43*, 698–706. https://doi.org/10.1007/s12144-023-04337-2

Shukla, M., & Pandey, R. (2024). Emotional dampening in hypertension: Impaired recognition of implicit emotional content in auditory and cross-modal stimuli. *PsyCh Journal*, *13*(1), 124–138. https://doi.org/10.1002/pchj.704

Shukla, M., Gupta, S., Upadhyaya, A., & Pandey, R. (2017). Behavioural manifestation and awareness of emotional processing deficits in individuals with hypertension: Evidence for emotional dampening hypothesis. *Indian Journal of Clinical Psychology*, *44*(2), 107–117.

Shukla, M., Pandey, R., & Lau, J. Y. (2019). Assessing emotional processing difficulties in normotensive individuals with high and isolated blood pressure elevations. *International Journal of Psychology, 54*(2), 214–222. https://doi.org/10.1002/ijop.12462

Shukla, M., Pandey, R., Jain, D., & Lau, J. Y. (2018). Poor emotional responsiveness in clinical hypertension: Reduced accuracy in the labelling and matching of emotional faces amongst individuals with hypertension and prehypertension. *Psychology & Health*, *33*(6), 765–782. https://doi.org/10.1080/08870446.2017.1401624

Sievers, B., Parkinson, C., Kohler, P. J., Hughes, J., Fogelson, S. V., & Wheatley, T. (2018). Visual and auditory brain areas share a neural code for perceived emotion. *BioRxiv*, 254961. https://doi.org/10.1101/254961