Differential Engagement of Associative-Limbic and Sensorimotor Regions of the Cerebellum and Basal Ganglia in Explicit vs. Implicit Emotional Processing

Leonardo Ceravolo[1#], Marine Thomasson[2,3#], Ioana Medeleine Constantin[4], Émilie Chassot[2,4], Jordan Pierce[5], Alexandre Cionca[6], Didier Grandjean[1], Frédéric Assal[2,7] & Julie Péron[2,4]*

[#] *These authors contributed equally to this work*

[1] Neuroscience of Emotion and Affective Dynamics Laboratory, Department of Psychology and Swiss Centre for Affective Sciences, University of Geneva, Switzerland.

[2] Cognitive Neurology Unit, Department of Neurology, University Hospitals of Geneva, Geneva, Switzerland.

[3] School of Physical and Occupational Therapy, McGill University, Montréal, QC, Canada.

[4] Clinical and Experimental Neuropsychology Laboratory, Department of Psychology and Swiss Centre for Affective Sciences, University of Geneva, Switzerland.

[5] Cognitive and Affective Neuroscience Lab, Department of Psychology, University of Nebraska-Lincoln, USA

[6] Medical Image Processing Laboratory, Neuro-X Institute, École Polytechnique Fédérale De Lausanne (EPFL), Geneva, Switzerland

[7] Faculty of Medicine, University of Geneva, Switzerland.

*Corresponding author:

Professor Julie Péron, Clinical and Experimental Neuropsychology Laboratory, Faculté de Psychologie et des Sciences de l'Education, Université de Genève, 40 bd du Pont d'Arve, 1205 Geneva, Switzerland. Tel.: +41 22 379 94 55; email: julie.peron@unige.ch

Disclosure: The authors report no conflicts of interest.

Abstract

Emotional prosody processing involves multiple brain regions, but the specific roles of the cerebellum and basal ganglia in explicit (conscious) and implicit (incidental) tasks are not well known or understood. This study investigated how the cerebellum and basal ganglia contribute to explicit (emotion categorization) and implicit (gender categorization) processing of emotional prosody. Twenty-eight healthy French-speaking participants underwent high-resolution functional MRI while performing a vocal emotion processing task under such implicit and explicit conditions. Behavioral data analyses indicated greater accuracy in the gender discrimination task (implicit processing). Neuroimaging partially supported our hypothesis according to which explicit emotion processing yielded increased activations in associative-limbic regions (e.g., inferior frontal gyrus, Crus I and caudate) linked to higher-order functions, while implicit emotion processing engaged sensorimotor regions (primary motor cortex, primary somatosensory cortex) and areas associated with automatic processing (putamen, posterior insula, cerebellar lobules VIIIa-b and IX). Unexpected activity during task conditions suggest motor preparation effects and more complex brain network dynamics. These results challenge modular views of brain function and highlight the need to consider emotional processing as complex, dynamic, network-based interactions.

Introduction

Recent studies about non-motor roles of the basal ganglia and cerebellum provide a holistic view of the neural basis of emotional stimuli processing (Manto et al., 2024). At the behavioral level, it suggests that organisms must not only respond to their current conditions but also anticipate future changes in their environment to survive and evolve effectively, by constructing what are known as internal models (Koziol et al., 2014). These predictions are then compared with the outcomes of their actions. Such predictive mechanisms allow organisms to prepare adaptive responses even before environmental changes occur. Neuroanatomically, studies suggested that the cerebellum is involved in this form of predictive adaptation, especially in relation to the temporal structure of unfolding events (Cabaraux et al., 2020; Ito, 2008). Over time, experiences are grouped into functional units known as "chunks" (Graybiel, 1998). These chunks are cognitive structures that facilitate the processing and management of information. By consolidating internal models and creating chunks, organisms establish habits that reduce the need for conscious processing in routine tasks (Pierce & Péron, 2020). It is suggested that this metacognitive function applies to sensorimotor, cognitive, but also emotional processes (Manto et al., 2024) explaining why neuroanatomical (Adamaszek et al., 2017; Schutter & Van Honk, 2005), cerebellar patient observation and clinical studies (Thomasson et al., 2022) have suggested the involvement of the cerebellum in emotional processing (Ceravolo et al., 2021; Thomasson et al., 2021). The shift from a goal-directed and controlled mode to a habitual mode is supported by the transition from the associative to the sensorimotor pathway (Graybiel, 2008). The associative-limbic loops include the associative and limbic parts of the cerebral cortex, as well as associative and limbic parts of the basal ganglia and cerebellum (notably, the caudate nucleus, ventral striatum, ventral part of the subthalamic nucleus, dorsomedial globus

pallidus, lateral cerebellar hemispheres: Crus I-II, as well as the fastigial and dentate nuclei) (Habas et al., 2009). The sensorimotor loop includes the sensorimotor, somatosensory and motor parts of the cerebral cortex, as well as the sensorimotor parts of the basal ganglia and cerebellum (notably, the putamen, dorsolateral globus pallidus, dorsal part of the subthalamic nucleus, anterior cerebellum, interposed nuclei, anterior part of the vermis, as well as lobules VIIIa-b) (Krack et al., 2010; Middleton & Strick, 2000). In the domain of emotion processing, certain vocal patterns, such as tone or intonation used to convey emotions (i.e., emotional prosody), may shift from a consciously controlled to an automatic response, illustrating a potential transition from associative-limbic to sensorimotor circuits—though this shift in the emotional domain has yet to be empirically studied.

Exploring such a holistic proposition through the lens of emotional auditory perception appears to be a promising approach, as emotional prosody involves processing temporal structures in events. This requires both sensory brain regions for analyzing low-level acoustic features and cognitive regions for evaluative judgments (Frühholz et al., 2012; Frühholz et al., 2016; Grandjean, 2021; Steiner et al., 2022). Additionally, manipulating attentional focus during emotional prosody processing could offer a nuanced way to compare brain activity when participants consciously process emotions versus when emotion processing happens automatically or incidentally, without the need for conscious recognition of affective states. The former, known as 'explicit emotion processing', occurs when participants are asked to make decisions about the emotional aspects of what they hear or see. In contrast, 'implicit emotion processing' involves tasks in which attention is directed away from the emotional content, focusing on another aspect of the stimulus (e.g., speaker's gender), while emotional tone remains present (Bach et al., 2008; Grandjean et al., 2005).

Therefore, it can be reasonably hypothesized that large-scale cortico-subcortico-cerebellar networks involved in sensorimotor loops (including the cerebellum and basal ganglia) would be particularly engaged during implicit emotional processing, where attention is not directly focused on the emotion itself. In contrast, explicit emotional processing, which requires detailed and conscious evaluation, should involve the cortico-subcortico-cerebellar associative-limbic loop that is crucial for integrating emotions within a goal-directed cognitive context (Pierce & Péron, 2022). By conducting a retrospective analysis of studies that have investigated the brain bases of emotional prosody (by distinguishing between implicit and explicit processing), we can observe that literature aligns with this perspective, even if reported results were not originally interpreted within this theoretical framework. Ethofer et al. (2009) highlighted that explicit tasks tend to activate the associative regions of the basal ganglia, such as the caudate nucleus and the anterior putamen. Bach et al. (2008) showed differential activations within the basal ganglia, where the associative parts are more active during tasks that require explicit judgments of emotional content. Wildgruber et al. (2005) found that explicit emotional tasks activated not only cortical areas but also associative regions of the basal ganglia, supporting the hypothesis that these regions play a role in tasks requiring conscious evaluation of emotional stimuli. Interestingly, a recent meta-analysis (Pierce et al., 2023) reported for explicit processing of emotional stimuli that the right cerebellar lobule VI, left vermis/lobule VI, and left lobule VI were more specifically involved. In contrast, implicit processing was processed indirectly and showed a specific activation in a small cluster in the left cerebellar lobule VI.

With the abovementioned literature in mind, we hypothesized that activity within the cerebellum and the basal ganglia would be sensitive to the level of processing required by the task—here, explicit and implicit vocal emotion processing tasks. Specifically, we expected

that associative-limbic regions (e.g., the caudate nucleus, ventral striatum, ventral part of the subthalamic nucleus, dorsomedial globus pallidus, lateral cerebellar hemispheres: Crus I-II, as well as the fastigial and dentate nuclei) would be more active during explicit emotional processing, while sensorimotor-related regions (e.g., the putamen, dorsolateral globus pallidus, dorsal part of the subthalamic nucleus, anterior cerebellum, interposed nuclei, anterior part of the vermis, as well as lobules VIIIa-b) would be predominantly engaged during implicit emotional processing. We tested this hypothesis by using high-resolution fMRI on twenty-eight healthy participants to explore the involvement of specific parts of the cerebellum and basal ganglia, as well as their relationship with cortical networks involved in goal-directed (explicit) versus habit-driven (implicit) emotional prosody processing.

Materials and Methods

Participants

Twenty-eight healthy participants were recruited for the study. They were all French speakers with a mean age of 65.4 years (SE = 8.9, range = 48-77). According to the Edinburgh Handedness Inventory criteria (Oldfield, 1971), 26 participants were right-handed and 2 were left-handed. Their mean education level was 16.4 years (SE = 3.0, range = 12-22). Exclusion criteria for the recruitment were: 1) history of neurological disorders, 2) head trauma, 3) anoxia, stroke or major cognitive deterioration, as attested by their score on the Montreal Cognitive Assessment (Nasreddine & Patel, 2016) (mean score = 28, SD = 1.3, range = 25-30). Moreover, none of them wore hearing aids or had a history of tinnitus or a hearing impairment, as attested either by their PEGA score (mean = 28.7, SD = 1.4, range = 25-30). All participants gave their written informed consent and the study was approved by the local ethics committee. Using a two-tailed testing method with an effect size of 0.5, an alpha of

0.05 and our sample of N=28, we obtain a power of 72.27% as calculated in G*Power 3.1.9.7 (Faul et al., 2007).

Vocal emotion recognition task procedure

The vocal stimuli consisted of two speech-like but semantically meaningless sentences extracted from Banse and Scherer's validated database (Banse & Scherer, 1996). These pseudo-sentences were spoken in a neutral, happy or angry tone by two male and two female speakers, resulting in 18 different stimuli. During scanning, these binaurally recorded auditory stimuli were played through MRI-compatible headphones and displayed using E-Prime (http://www.pstnet.com/eprime.cfm). Participants performed a categorization task with two conditions for the Task factor (implicit, explicit) and three for the Emotion factor (anger, happiness, neutral). This paradigm was divided into six counterbalanced blocks (16 trials per block). For three of these blocks (implicit), participants had to do a gender decision (participants had to indicate "male" or "female" by pressing a key with their index or middle finger). For the other three blocks (explicit), participants performed an emotion categorization task (they had to indicate "happiness" or "neutral" / "anger" or "neutral" by pressing a key with their index or middle finger).

We assumed that during the gender decision task, participants would focus on the gender of the voice, but the emotional tone of the voice still would be processed implicitly. Before each block, a message indicating "gender" or "emotion" was displayed on the screen for 1 sec as an instruction. The vocal stimuli were then played (duration: 2.5 s) and a fixation cross was presented at the center of the screen, and the two possible responses displayed on the screen for 2 s. Each of the two runs contained all 96 trials and lasted approximatively 10 minutes while functional MRI was acquired.

Behavioral data analysis

Behavioral results illustrate the probability of accurately recognizing voice gender or emotion (Task factor) as a function of the Emotion factor. These analyses were computed using mixed effects, regression analyses for the response accuracy (of-interest) and for the associated reaction times (of-no-interest) in R studio (Allaire, 2012).

For the 'response' dependent variable, we used a logistic regression using the lme4 'glmer' package (Bates et al., 2015) with Voice emotion (Emo) interacting with Task and Run as fixed effects and participant identity (ParticipantID) and gender (ParticipantGender), stimuli, stimuli order (StimuliOrder), voice actor identity (ActorID) and Gender (ActorGender) as random factors. The model formula was the following:

$$response \sim Emo * Task + Run + (1|ParticipantID) + (1|ParticipantGender) + (1|Stimuli) + (1|StimuliOrder) + (1|ActorID) + (1|ActorGender).$$

For the 'reaction times' dependent variable, we used a linear regression using the lme4 'lmer' package with the exact same formula as above.

For our planned comparisons, we used a confidence interval of 97.5% and a Bonferroni correction for multiple comparisons. Specific contrasts were computed using the 'phia' package (De Rosario-Martinez et al., 2015).

No difference was observed between models including or excluding participant gender as a random effect (p>.1).

MRI data acquisition

All functional imaging data were recorded on a 3-T Siemens Trio System (Siemens, Erlangen, Germany) scanner equipped with a 32-channel antenna. A 3D sequence was used to acquire

185 a high-resolution T1-weighted image (1mm isotropic voxels; TR=1900msec; TE=2.27ms;
186 matrix resolution=256 x 256; FA=9degrees; 192 total slices).
187 For the task, functional images were acquired in descending order using a multi-band echo-
188 planar imaging (EPI) sequence of 54 slices aligned along the anterior-posterior commissure
189 (voxel size: 2.5mm isotropic; TR=1300msec; TE=20ms; FOV=205 x 205mm; matrix
190 resolution=84 x 84; FA=64degrees; BW=1952Hz/px). Functional image acquisition was
191 continuous throughout each run of the task.

192 MRI data analysis

193 Data were preprocessed using a mix of SPM12 (SPM12, Wellcome Trust Centre for
194 Neuroimaging, London, UK), the CONN toolbox (Whitfield-Gabrieli & Nieto-Castanon, 2012)
195 and the Artifact detection tools (ART) toolbox
196 (https://www.nitrc.org/projects/artifact_detect/). Functional data were first converted to 4D
197 Nifti in SPM12 for each run separately to get one single file per run, for more efficient
198 computation performance and decrease storage volume. Preprocessing steps included the
199 following: realignment and unwarp, slice-timing correction, artifact detection followed by an
200 additional denoising (scrubbing including functional regression and functional bandpass
201 [0.01-0.1Hz]). At this stage, direct segmentation and normalization into the Montreal
202 Neurological Institute (MNI) space (Collins et al., 1994) was performed for improved
203 comparison between the brain tissues of our sample of participants and with regards to
204 other existing fMRI studies. Finally, data were spatially smoothed with an isotropic Gaussian
205 filter of 8 mm full width at half maximum. Following an initial step consisting in the creation
206 of a participant-specific matrix including the onset and reaction times of each trial of each
207 run, preprocessed, smoothed functional images were then analyzed with SPM12 both at the
208 participant-level ('first-level') as well as at the group-level ('second-level'). A single general

linear model was used to compute first-level statistics, using a convolution with the Hemodynamic Response Function (HRF). Each event was therefore modelled using HRF function, locked to the onset of each voice stimulus using continuous MRI data acquisition. The design matrix included three columns per run representing the Emotion factor (anger, happiness, neutral) as well as six motion parameters—included as regressors of no interest to account for movement in the data—as well as trial-level reaction times, also as a regressors of no interest—since accuracy was of interest, not response times. Trials in which an evaluation error was made for each task was also included—regressors of no interest. Regressors of interest were used to compute simple contrasts for each participant, leading to separate main effects of Angry, Neutral and Happy voices for each task, for a total of six simple contrasts (Implicit task: Angry, Neutral and Happy voices; Explicit task: Angry, Neutral and Happy voices).

This model yielded two flexible factorial second-level analyses, with factors Task and Emotion. We also had the Participants factor (Factor 1 in each analysis) to consider interindividual variability, with data independence set to 'yes', variance to 'unequal' while for the within factors of interest data independence set to 'no', variance to 'unequal'. Second-level models focused on: the Task factor (Model 1) and on the interaction between Task and Emotion (Model 2, Task * Emotion).

Activations were ultimately thresholded in SPM12 by using a voxel-wise FDR correction at $p<.05$. Thresholded contrast activations were then rendered on brains from the CONN toolbox (Whitfield-Gabrieli & Nieto-Castanon, 2012). Regions were labelled using the latest version of the 'automated anatomical labelling' ('aal3') atlas (Rolls et al., 2020; Tzourio-Mazoyer et al., 2002).

Results

This section reports the behavioral and neuroimaging results of our sample of participants (N=28), who were asked to categorize the gender (implicit emotion processing) or the emotion (explicit emotion processing) of angry, happy or neutral voices presented to them via headphones while lying in the fMRI scanner. The explicit and implicit tasks illustrate the Task factor while the emotion expressed by each voice stimulus (anger, happiness, neutral) represents the levels of the Emotion factor. Main effects as well as partial and full interaction effects are described below.

Behavioral data

Behavioral data included one dependent variable of interest—response accuracy (Fig.1)—and one of no-interest, namely reaction times (Fig.S1). These results are presented below and summarized in Table S1.

Response patterns and accuracy

Response data of each task were treated as 'correct' or 'incorrect' and coded as '1' or '0', respectively. Logistic regression was used to better characterize data and allow for the interpretation of the probability of a correct response as a function of our Task and Emotion factors (fixed effects)—including also the split into two runs, and taking into consideration relevant random effects (see Methods).

The probability of a correct response was predicted by the Task with higher accuracy for the gender task ($\chi^2(1)=93.33$, $p<.001$), voice Emotion ($\chi^2(2)=19,11$, $p<.001$), Run (with higher accuracy in run 2, $\chi^2(1)=8.81$, $p<.01$) and the interaction between Task and Emotion ($\chi^2(2)=85.73$, $p<.001$). Across tasks, neutral and angry voices yielded a higher probability of

correct response compared to happy voices for the Emotion factor ($\chi^2(1)=5.54$, $p<.05$). The interaction between Task and Emotion was mainly explained by an inverted pattern of response between tasks, in which higher accuracy probability was observed for angry compared to neutral and happy voices in the emotion discrimination task. The opposite trend occurred in the gender task paired with a general higher accuracy for the gender task ([Task Gender > Task Emotion * Angry > Neutral+Happy voices]: $\chi^2(1)=84.88$, $p<.001$). Additionally, differences were observed in accurate response probability between angry vs happy voices in the Emotion task ($\chi^2(1)=32.59$, $p<.001$)—but not in the Gender task ($\chi^2(1)=1.74$, $p=.19$)—and also between tasks, with higher accuracy for angry vs happy in the Gender task as opposed to the inverse in the Emotion task ($\chi^2(1)=71.97$, $p<.001$). Response probability for angry vs neutral voices also differed in both tasks (Gender: $\chi^2(1)=6.49$, $p<.05$; Emotion: $\chi^2(1)=9.23$, $p<.01$) and between tasks ($\chi^2(1)=53.46$, $p<.001$) as well as for happy vs neutral voices in the Emotion task only ($\chi^2(1)=8.99$, $p<.01$; Gender task: $\chi^2(1)=1.21$, $p>.10$; Gender vs Emotion task: $\chi^2(1)=3.15$, $p<.10$). Variance explained by the fixed effects was 38.24% (R^2c) while the full model including random effects (R^2c) explained 61.67% of the variance of the response probability data. See Fig.1 for details and illustration.

Reaction times data

Reaction times data of each task were analyzed next. Linear regression was used as a function of the Task and Emotion factors (fixed effects)—including the split into two runs, and also taking into consideration relevant random effects (see Methods). Since the tasks are rather 'easy' to perform and the fact that no instruction concerning response time was given to the participants, reaction times data were of no interest in the present paper, but were examined anyway for completeness.

Reaction times were predicted by the Task with faster values for the gender task ($\chi^2(1)=19.96$, $p<.001$), voice Emotion ($\chi^2(2)=7.03$, $p<.05$), Run ($\chi^2(1)=23.06$, $p<.001$) and the interaction between Task and Emotion ($\chi^2(2)=17.53$, $p<.001$). Across tasks, angry and happy voices yielded slower reaction times compared to neutral voices for the Emotion factor ($\chi^2(1)=5.92$, $p<.05$). The interaction between Task and Emotion was again mainly explained by a partly inverted pattern of data between tasks, in which happy and neutral compared to angry voices led to slower reaction times in the emotion discrimination task while the inverse was observed in the gender task ([Task Gender > Task Emotion * Angry > Neutral + Happy voices]: $\chi^2(1)=16.53$, $p<.001$). Additionally, differences were observed between angry vs happy voices with slower reaction times for happy voices in the Emotion task ($\chi^2(1)=5.21$, $p<.05$) but not in the Gender task ($\chi^2(1)=0.94$, $p>.10$)—but differences were also observed between tasks (slower reaction times for angry vs happy in the Gender task as opposed to the inverse in the Emotion task ($\chi^2(1)=16.83$, $p<.001$). Reaction times were also slower for angry vs neutral voices the Gender task ($\chi^2(1)=7.01$, $p<.01$) but not in the Emotion task ($\chi^2(1)=0.26$, $p>.10$), and between tasks ($\chi^2(1)=9.46$, $p<.01$) as well as slower reaction times for happy vs neutral voices in the Emotion task only ($\chi^2(1)=8.40$, $p<.01$; Gender task: $\chi^2(1)=2.50$, $p>.10$; Gender vs Emotion task: $\chi^2(1)=2.56$, $p<.10$). Variance explained by the fixed effects was 31.25% (R^2c) while the full model including random effects (R^2c) explained 48.33% of the variance of the reaction times data. See Fig.S1 for details and illustration.

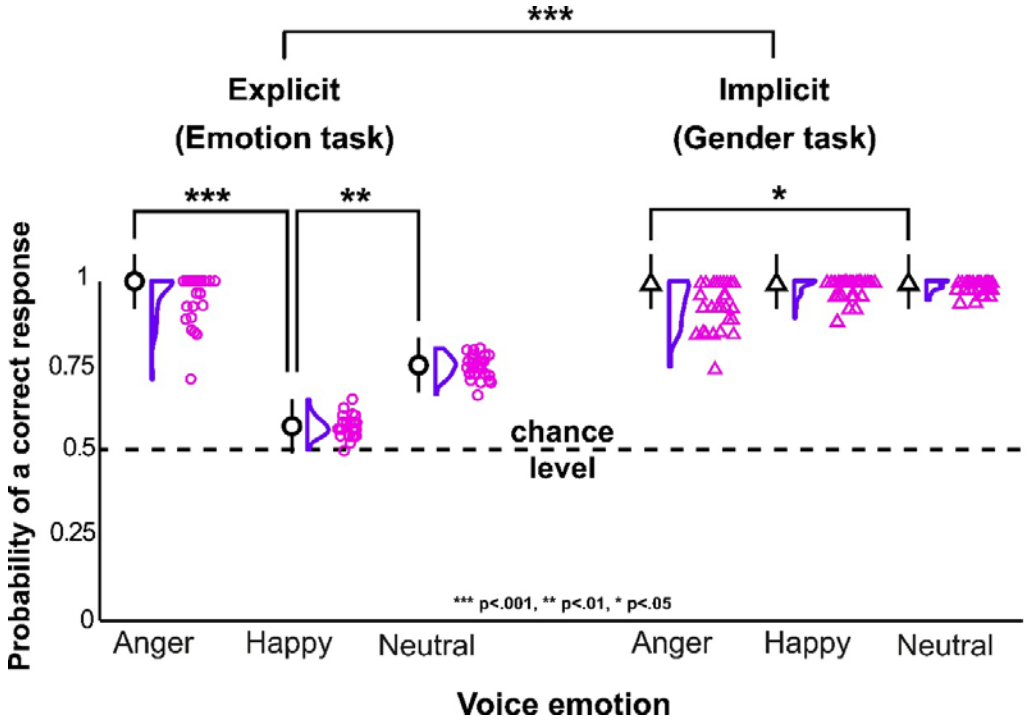

Fig.1: Behavioral results for response probability in the Gender (Implicit) and the Emotion (Explicit) discrimination task. Circles and triangles represent individual values for each participant (N=28; circles: Emotion task, triangles: Gender task), with data curves as violin distribution and mean values with error bars representing the standard error of the mean (Y axis) as a function of voice emotion (X axis). Chance level for response probability is 50% (0.5, X axis, blue) since all tasks involved a 2-alternate forced choice type of response. *** p<.001, ** p<.01, * p<.05. Descriptive values are reported in Table S1.

Neuroimaging data

Functional images were acquired continuously during the session of each participant to assess brain correlates across tasks (explicit and implicit vocal emotion processing), emotions when contrasting the tasks (angry, happy and neutral voices for explicit > implicit and implicit > explicit) and the other interactions between tasks and emotions. Wholebrain voxel-wise statistics are presented below with correction for multiple tests (false-discovery rate, 'FDR'). Since reaction times were of no-interest and participants were not instructed to respond as fast as possible, reaction times were used as a trial-level covariate of no interest in the

neuroimaging statistical models. Error trials—although accounting for a very small percentage of the trials—were also concatenated per task into a regressor of no interest.

Effects of the type of task (explicit, implicit emotion processing)

Both tasks were designed to highlight and analyzed with respect to the levels of vocal emotion processing: focusing on voice gender or on expressed emotion to implicitly or explicitly process vocal emotion, respectively.

The contrast between explicit vs implicit processing of vocal emotions showed enhanced activations in the lateral temporal and inferior frontal cortex, especially in the STG, STS and anterior insula (Fig.2AC) and in all subparts of the IFG (Fig.2ABCH). Additional activations were observed in the pre-supplementary motor area (preSMA, Fig.2GI), the basal ganglia (caudate nucleus, putamen, Fig.2M), the thalamus and in widely spread regions of the cerebellum (Crus I and II, Fig.2N). See also Table S2.

In the inverse contrast, implicit vs explicit vocal emotion processing, we observed a very distinct network of activations involving the primary motor and somatosensory cortices (Fig.2DEF), the ventromedial (Fig.2JL) and dorsolateral (Fig.2DEF) prefrontal cortices, the orbitofrontal cortex (Fig.2K), the posterior insula (Fig.2DF) as well as the parahippocampal gyrus (Fig.2K), basal ganglia (putamen, caudate nucleus; Fig.2O) and the cerebellar lobules IX, VIIIa-b and Crus II (Fig.2P). See Table S3.

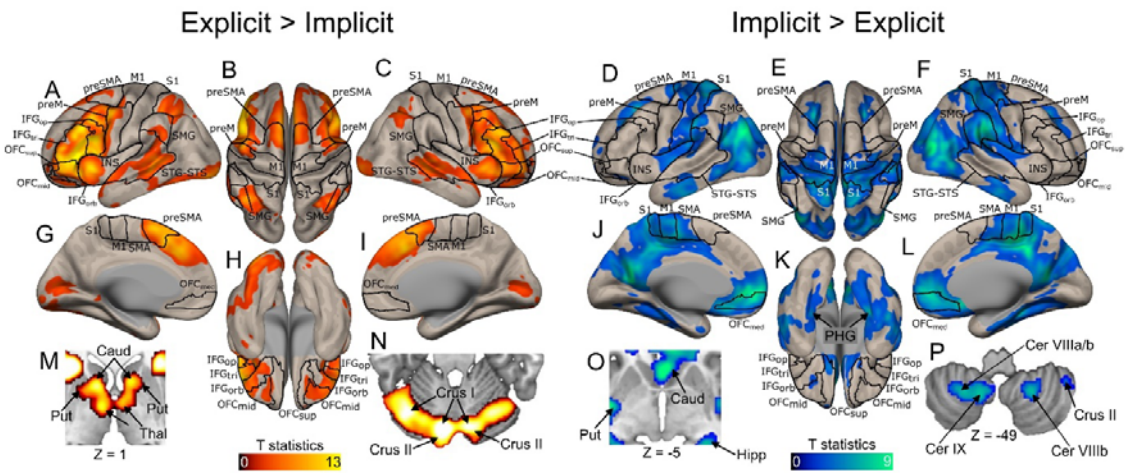

Fig.2: Whole brain activations contrasting the levels of the Task factor. Lateral and medial cortical as well as basal ganglia and cerebellar activations were identified when contrasting explicit to implicit emotion processing tasks (Explicit>Implicit: red-to-yellow activations, lateral cortex 'A,B,C,H'; medial cortex 'G,I'; basal ganglia 'M'; cerebellum 'N'; Implicit>Explicit: blue-to-green activations, lateral cortex 'D,E,F,K'; medial cortex 'J,L'; basal ganglia 'O'; cerebellum 'P'). All activations reported using an FDR corrected voxel-wise threshold of $p<.05$. Color bars represent T-values. IFG: inferior frontal gyrus; INS: insula; STG: superior temporal gyrus; STS: superior temporal sulcus; preSMA: pre-supplementary motor area; SMA: supplementary motor area; preM: premotor cortex; M1: primary motor cortex; S1: primary somatosensory cortex; SMG: supramarginal gyrus. Caud: caudate nucleus; Put: putamen; Thal: thalamus; OFC: orbitofrontal cortex; Cer: cerebellar lobule; Hipp: hippocampus. Suffixes: orb, pars orbitalis; tri, pars triangularis; op, pars opercularis; sup, superior; mid, middle; med, medial.

Task effects for each emotion (angry, happy, neutral voices)

We then contrasted tasks and looked at each emotion individually. These results are overlaid in Fig.3 (Table S4) for the Explicit > Implicit contrast and reported in Fig.S2-4 (Tables S5-7) for the Implicit > Explicit contrast.

When contrasting Explicit to Implicit emotion processing, we observed a large overlap for angry, happy and neutral voices, especially in the bilateral inferior frontal cortex *pars triangularis* (Fig.3A-G), bilateral pre-supplementary motor area (Fig.3H-J, L-N) and in the cerebellum—specifically in the right Crus I and cerebellar lobule VI (Fig.3O-Q). Specific

overlaps between angry and neutral voices for this contrast were observed in the bilateral STG and STS (Fig.3ACFG), orbitofrontal cortex (Fig.3K), right Crus II (Fig.3O-Q) and in the basal ganglia (caudate nucleus, Fig.3K). Overlaps shared by happy and neutral voices in this contrast were mainly observed in the left IFG (Fig.3BC), preSMA (Fig.3HI, LM) and in the cerebellum, especially in the left hemisphere Crus I and Crus II and in the vermis Crus II (Fig.3O-Q).

Other results pertaining to the Task * Emotion interactions—happy vs neutral voices in both tasks as well as angry vs happy (and inverse) in the explicit task—are reported in Fig.S5 (Tables S8 and S9) and in Fig.S6 (Tables S10 and S11), respectively.

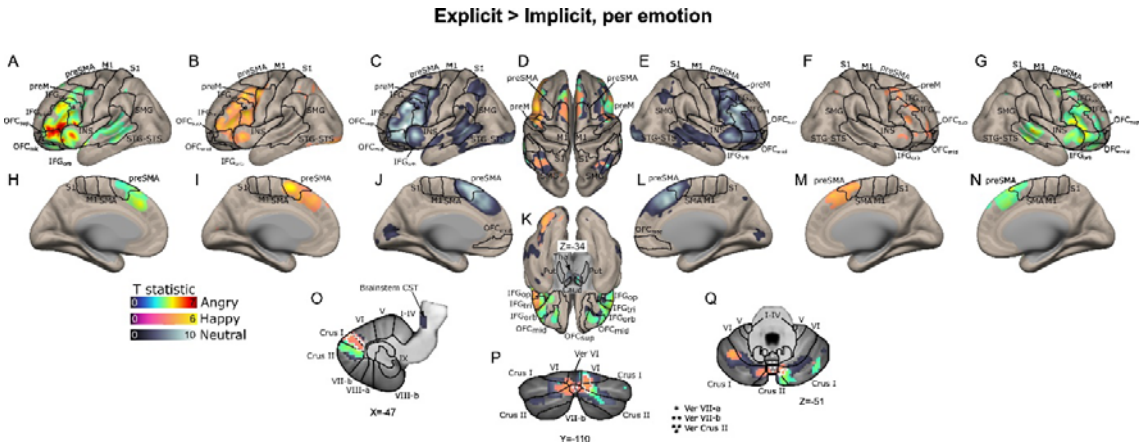

Fig.3: Whole brain activations contrasting the levels of the Task factor for each emotion. Lateral and medial cortical as well as basal ganglia and cerebellar activations were highlighted when contrasting explicit to implicit emotion processing tasks (lateral cortex 'A-G, K'; medial cortex 'H-J, L-N'; basal ganglia 'K'; cerebellum 'O-Q') for angry ('Ang'; blue-to-red colormap), happy ('Hap'; pink-to-yellow colormap) and neutral ('Neu'; black-to-white colormap) voices. All activations reported using an FDR corrected voxel-wise threshold of $p<.05$. Color bars represent T-values. White outline in panels 'O-Q': cerebellar overlap between all three emotions. IFG: inferior frontal gyrus; INS: insula; STG: superior temporal gyrus; STS: superior temporal sulcus; preSMA: pre-supplementary motor area; SMA: supplementary motor area; preM: premotor cortex; M1: primary motor cortex; S1: primary somatosensory cortex; SMG: supramarginal gyrus. Caud: caudate nucleus; Put: putamen; Thal: thalamus; OFC: orbitofrontal cortex; Brainstem CST: cortico-spinal tract of the

brainstem; Ver: Vermis. Suffixes: orb, pars orbitalis; tri, pars triangularis; op, pars opercularis; sup, superior; mid, middle; med, medial.

Discussion

The present study partially sustains the hypothesis that the level of emotional processing required by a task influences brain activation patterns in regions within the cerebellum, basal ganglia, and associated cortical networks. Activation in some other brain areas are however challenging the present hypotheses. Based on our detailed interpretation below, these results are not mutually exclusive but provide further evidence that cognitive and emotional processes arise from the dynamic interplay of large-scale brain networks, where the cerebellum and basal ganglia play roles as crucial as the cerebral cortex, amygdala, or other widely recognized regions of the "affective brain" (Damasio, 1994; LeDoux, 1998).

The results for the explicit task showed enhanced activations in associative-limbic regions, such as the lateral temporal cortex, including the superior temporal gyrus (STG) and superior temporal sulcus (STS) and the inferior frontal cortex, specifically the inferior frontal gyrus (IFG). Additionally, activations were observed in the cerebellum, including Crus I, Crus II, and the vermis, as well as the basal ganglia, notably the caudate nucleus, and the thalamus. These areas are associated with complex emotional processing (Kringelbach & Rolls, 2004), decision-making (Bechara et al., 2000), and social cognition (Adolphs, 2009), which aligns with the idea that they are more engaged when explicit emotional processing is required. The results for the implicit task showed activations in sensorimotor regions, such as the primary motor cortex (M1), primary somatosensory cortex (S1), and the cerebellum, specifically lobules IX and VIIIa-b. Additionally, activations were observed in the basal ganglia, including the putamen, and in regions involved in automatic processing, such as the

408 posterior insula (Bamiou et al., 2003) and precuneus (Cavanna & Trimble, 2006). This fits
409 with the hypothesis that sensorimotor-related regions are more engaged during implicit
410 tasks where the emotional content of the voice is not the focus but is processed incidentally
411 (Pessoa & Adolphs, 2010) and / or when integrating sensory and emotional information
412 (Craig, 2009). This task-dependent engagement may reflect the brain's adaptive mechanism
413 for managing emotional information efficiently. For explicit tasks, the engagement of higher-
414 order cognitive and associative regions facilitates deliberate analysis and conscious decision-
415 making. For implicit tasks, the engagement of sensorimotor and more automatic processing
416 regions suggests a streamlined pathway for handling emotional information when the
417 primary task does not demand conscious focus on emotions. This distinction enhances our
418 understanding of how different types of emotional processing can be integrated into broader
419 cognitive processes and how emotional 'chunk-like' habits can develop in response to
420 repeated experiences. Such habits could be particularly relevant in real-life scenarios where
421 emotional information is processed in the background while attention is focused on another
422 task.

423 Some results are however surprising with respect to our formulated hypotheses. First, results
424 also showed activations in sensorimotor regions for explicit processing. For example, the
425 "explicit > implicit" contrast shows activations in the pre-supplementary motor area
426 (preSMA), which is unexpected as these regions should be less active during explicit
427 processing (Fuster, 2008). Those activations are, however, coherent with a higher
428 involvement of such regions in the embodiment of vocal happiness, since preSMA underlies
429 its production (Selosse et al., 2024). This could also be due to motor preparation or response
430 mechanisms (Nachev et al., 2008; Picard & Strick, 1996) associated with the explicit task.
431 Participants might have used motor strategies or mental imagery processes that activate

motor areas, even when the task requires explicit emotional processing. Second, we observed activations spread across several cerebellar regions (Crus I and II for explicit, lobules IX, VIIIa-b, and Crus II for implicit) without a clear distinction between associative-limbic and sensorimotor regions. The cerebellum is known to be involved in a wide range of cognitive and emotional functions (Schmahmann, 2019), and such diffuse cerebellar activations may reflect an integration of these functions rather than a segregation between associative-limbic and sensorimotor functions. This is in accordance with previous literature that has shown a cerebellar involvement in both the sensory and cognitive processing of emotions, suggesting a broader metacognitive role of this structure (Pierce & Péron, 2020). Another possibility is that the sensitivity of sensorimotor processes to the implicit and explicit conditions may not be sufficient. Third, some cortical (lateral occipital cortex, superior parietal cortex) and subcortical regions (caudate nucleus) showed significant activations in both types of tasks (explicit and implicit). While these activations are not clearly explained by the hypothesis of differentiation between associative-limbic and sensorimotor networks, they might represent additional attentional, cognitive control, or perceptual processing required by the emotional stimuli used in both tasks, rather than strictly emotional processes (Corbetta & Shulman, 2002). They might also suggest that the stimuli have multisensory or complex characteristics that require integration into broader sensory and attentional networks.

According to dynamic systems theories, brain activity reflects smooth transitions between different network states rather than static and isolated processing modules (Bassett & Sporns, 2017). Widespread activation in cerebellar regions without a clear distinction between associative-limbic and sensorimotor regions suggest that the cerebellum plays a more integrated and context-dependent role in emotional processing (Schmahmann, 1997).

This supports the idea that brain functions are not segregated but rather that different cerebellar regions can dynamically engage depending on the specific demands of the task, thereby integrating cognitive, emotional, and sensorimotor processing. Classical cognitive theories often propose a dichotomy between explicit (associative-limbic) and implicit (sensorimotor) processing, with each associated with distinct neural circuits. However, our findings challenge this view by showing overlapping activations across different networks, such as the lateral occipital cortex and superior parietal cortex, for both explicit and implicit tasks. These overlaps are not necessarily contradictory; instead, they may reflect the brain's flexible use of resources to adapt to complex stimuli that require both emotional and perceptual processing. Thus, rather than being strictly separated, explicit and implicit processing may coexist and interact, dynamically influencing each other based on the context and the individual's strategies. The brain dynamically reorganizes itself based on task demands, participant strategies, and contextual factors. Such implications challenge the way we understand emotional processing in the brain, not as a linear or modular process, but as a result of complex, dynamic interactions among various neural networks (Rizzolatti & Sinigaglia, 2010).

While our study provides valuable insights into the neural correlates of explicit and implicit emotional processing, methodological limitations need to be considered when interpreting the results. The small sample size and narrow age range (mean age 65.4) limit generalization across broader and younger populations. The use of a restricted set of emotions (anger, happiness, neutral) may reduce ecological validity, while task design assumptions about implicit processing may not hold true for all participants, potentially confounding results. Temporal fMRI resolution may also obscure more subtle neural dynamics. Furthermore, experimental settings, task constraints, and potential biases related to task-switching further

challenge the interpretability results. These aspects were however kept constant and proven to be identical across participants. Finally, individual strategies used by our participants could be different and should not be neglected. Addressing these limitations in future research would enhance the robustness and applicability of the findings.

This study investigated the neural correlates of explicit versus implicit emotional processing, focusing on the cerebellum, basal ganglia, and associated cortical networks. Our results confirm that different levels of emotional processing engage distinct large scale brain networks involving the basal ganglia and the cerebellum. However, overlapping activations also emphasize that parts of these networks are not strictly polarized between the two emotional processes, with different regions interacting based on task demands and individual strategies. The results highlight the importance of considering the cerebellum in models of emotion functions and going beyond traditional modular theories to embrace more integrated and dynamic frameworks.

Data availability

Data and codes will be made available on the free repository YARETA, which meets the FAIR guidelines, upon acceptance of the manuscript for publication.

Funding

The present research was supported by Swiss National Science Foundation (SNSF) grants to JAP (PI) and FA (Co-PI) within the project "Influence of top-down mechanisms on cerebellar activity during vocal emotion decoding" (2023-2026), Grant N°: 105314_215015 through the

University of Geneva. The funders had no role in data collection, discussion of content, preparation of the manuscript, or decision to publish.

Acknowledgments

We warmly thank the team of the Brain and Behavioral Laboratory of the University of Geneva, Switzerland, for their help with setting-up MRI sequences and with MRI data acquisition.

Ethics

This study was conducted following the written informed consent of all participants (N=28), and according to the strict regulations of the University of Geneva and the declaration of Helsinki. Participants were informed they could exit the study without justification any time they wanted or felt it was necessary for them. The state-wise ethics committee of Geneva University Hospital (Comité Cantonal d'Ethique en Recherche, CCER) reviewed and accepted our ethics application (number is 2024-00174).

References

Adamaszek, M., D'Agata, F., Ferrucci, R., Habas, C., Keulen, S., Kirkby, K., Leggio, M., Mariën, P., Molinari, M., & Moulton, E. (2017). Consensus paper: cerebellum and emotion. *The Cerebellum*, *16*, 552-576.

Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annual review of psychology*, *60*(1), 693-716.

Allaire, J. (2012). RStudio: integrated development environment for R. *Boston, MA*, *770*(394), 165-171.

Bach, D. R., Grandjean, D., Sander, D., Herdener, M., Strik, W. K., & Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *Neuroimage*, *42*(2), 919-927.

Bamiou, D.-E., Musiek, F. E., & Luxon, L. M. (2003). The insula (Island of Reil) and its role in auditory processing: literature review. *Brain research reviews*, *42*(2), 143-154.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, *70*(3), 614.

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature neuroscience*, *20*(3), 353-364.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2015). Package 'lme4'. *convergence*, *12*(1), 2.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral cortex*, *10*(3), 295-307.

Cabaraux, P., Gandini, J., Kakei, S., Manto, M., Mitoma, H., & Tanaka, H. (2020). Dysmetria and errors in predictions: the role of internal forward model. *International journal of molecular sciences*, *21*(18), 6900.

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, *129*(3), 564-583.

Ceravolo, L., Frühholz, S., Pierce, J., Grandjean, D., & Péron, J. (2021). Basal ganglia and cerebellum contributions to vocal emotion processing as revealed by high-resolution fMRI. *Scientific reports*, *11*(1), 10645.

Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of computer assisted tomography*, *18*(2), 192-205.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201-215.

Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*(1), 59-70.

Damasio, A. (1994). Descartes' Error. Emotion, Reason and the Human Brain. New York (Grosset/Putnam) 1994.

De Rosario-Martinez, H., Fox, J., Team, R. C., & De Rosario-Martinez, M. H. (2015). Package 'phia'. *CRAN Repos. Retrieved*, *1*, 2015.

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.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191.

Frühholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral cortex*, *22*(5), 1107-1117.

Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions—Towards a unifying neural network perspective of affective sound processing. *Neuroscience & Biobehavioral Reviews*, *68*, 96-110.

Fuster, J. (2008). The Prefrontal Cortex, Academic Press. *New York.[Google Scholar]*.

Grandjean, D. (2021). Brain networks of emotional prosody processing. *Emotion Review*, *13*(1), 34-43.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature neuroscience*, *8*(2), 145-146.

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of learning and memory*, *70*(1-2), 119-136.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.*, *31*(1), 359-387.

Habas, C., Kamdar, N., Nguyen, D., Prater, K., Beckmann, C. F., Menon, V., & Greicius, M. D. (2009). Distinct cerebellar contributions to intrinsic connectivity networks. *Journal of neuroscience*, *29*(26), 8586-8594.

Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience*, *9*(4), 304-313.

Koziol, L. F., Budding, D., Andreasen, N., D'Arrigo, S., Bulgheroni, S., Imamizu, H., Ito, M., Manto, M., Marvel, C., & Parker, K. (2014). Consensus paper: the cerebellum's role in movement and cognition. *The Cerebellum*, *13*, 151-177.

Krack, P., Hariz, M. I., Baunez, C., Guridi, J., & Obeso, J. A. (2010). Deep brain stimulation: from neurology to psychiatry? *Trends in neurosciences*, *33*(10), 474-484.

Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Progress in Neurobiology*, *72*(5), 341-372.

LeDoux, J. E. (1998). *The emotional brain: The mysterious underpinnings of emotional life*. Simon and Schuster.

Manto, M., Adamaszek, M., Apps, R., Carlson, E., Guarque-Chabrera, J., Heleven, E., Kakei, S., Khodakhah, K., Kuo, S.-H., & Lin, C.-Y. R. (2024). Consensus Paper: Cerebellum and Reward. *The Cerebellum*, 1-24.

Middleton, F. A., & Strick, P. L. (2000). Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain research reviews*, *31*(2-3), 236-250.

Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience*, *9*(11), 856-869.

Nasreddine, Z. S., & Patel, B. B. (2016). Validation of Montreal cognitive assessment, MoCA, alternate French versions. *Canadian Journal of Neurological Sciences*, *43*(5), 665-671.

Oldfield, R. (1971). Edinburgh handedness inventory. *Journal of Abnormal Psychology*.

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a'low road'to'many roads' of evaluating biological significance. *Nature Reviews Neuroscience*, *11*(11), 773-782.

Picard, N., & Strick, P. L. (1996). Motor areas of the medial wall: a review of their location and functional activation. *Cerebral cortex*, *6*(3), 342-353.

Pierce, J. E., & Péron, J. (2020). The basal ganglia and the cerebellum in human emotion. *Social cognitive and affective neuroscience*, *15*(5), 599-613.

Pierce, J. E., & Péron, J. A. (2022). Reward-based learning and emotional habit formation in the cerebellum. In *The Emotional Cerebellum* (pp. 125-140). Springer.

Pierce, J. E., Thomasson, M., Voruz, P., Selosse, G., & Péron, J. (2023). Explicit and implicit emotion processing in the cerebellum: a meta-analysis and systematic review. *The Cerebellum*, *22*(5), 852-864.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience*, *11*(4), 264-274.

Rolls, E. T., Huang, C.-C., Lin, C.-P., Feng, J., & Joliot, M. (2020). Automated anatomical labelling atlas 3. *Neuroimage*, *206*, 116189.

Schmahmann, J. D. (1997). *The cerebellum and cognition*. Academic Press.

Schmahmann, J. D. (2019). The cerebellum and cognition. *Neuroscience letters*, *688*, 62-75.

Schutter, D. J., & Van Honk, J. (2005). The cerebellum on the rise in human emotion. *The Cerebellum*, *4*, 290-294.

Selosse, G., Grandjean, D., & Ceravolo, L. (2024). Neural correlates of embodied and vibratory mechanisms associated with vocal emotion production. *bioRxiv*, 2024.2005. 2014.594073.

Steiner, F., Fernandez, N., Dietziker, J., Stämpfli, P., Seifritz, E., Rey, A., & Frühholz, S. (2022). Affective speech modulates a cortico-limbic network in real time. *Progress in Neurobiology*, *214*, 102278.

Thomasson, M., Benis, D., Saj, A., Voruz, P., Ronchi, R., Grandjean, D., Assal, F., & Péron, J. (2021). Sensory contribution to vocal emotion deficit in patients with cerebellar stroke. *NeuroImage: Clinical*, *31*, 102690.

Thomasson, M., Benis, D., Voruz, P., Saj, A., Vérin, M., Assal, F., Grandjean, D., & Péron, J. (2022). Crossed functional specialization between the basal ganglia and cerebellum during vocal emotion decoding: insights from stroke and Parkinson's disease. *Cognitive, Affective, & Behavioral Neuroscience*, *22*(5), 1030-1043.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*(1), 273-289.

Whitfield-Gabrieli, S., & Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain connectivity*, *2*(3), 125-141.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., & Ackermann, H. (2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage*, *24*(4), 1233-1241.