scientific reports

OPEN

Basal ganglia and cerebellum contributions to vocal emotion processing as revealed by high-resolution fMRI

Leonardo Ceravolo[1,2✉], Sascha Frühholz[3,4,5], Jordan Pierce[6], Didier Grandjean[1,2,8] & Julie Péron[6,7,8]

Until recently, brain networks underlying emotional voice prosody decoding and processing were focused on modulations in primary and secondary auditory, ventral frontal and prefrontal cortices, and the amygdala. Growing interest for a specific role of the basal ganglia and cerebellum was recently brought into the spotlight. In the present study, we aimed at characterizing the role of such subcortical brain regions in vocal emotion processing, at the level of both brain activation and functional and effective connectivity, using high resolution functional magnetic resonance imaging. Variance explained by low-level acoustic parameters (fundamental frequency, voice energy) was also modelled. Wholebrain data revealed expected contributions of the temporal and frontal cortices, basal ganglia and cerebellum to vocal emotion processing, while functional connectivity analyses highlighted correlations between basal ganglia and cerebellum, especially for angry voices. Seed-to-seed and seed-to-voxel effective connectivity revealed direct connections within the basal ganglia—especially between the putamen and external globus pallidus—and between the subthalamic nucleus and the cerebellum. Our results speak in favour of crucial contributions of the basal ganglia, especially the putamen, external globus pallidus and subthalamic nucleus, and several cerebellar lobules and nuclei for an efficient decoding of and response to vocal emotions.

Social communication through voice entails semantic as well as prosodic meaning, the latter being generally defined as the melody of the human voice. The processing of human voice prosody leads to widespread changes in multiple cerebral regions, especially in the superior temporal and inferior frontal cortices[1–4]. Given their tripartite functional compartmentalization, whereby each basal ganglia (BG) is linked to either the motor, associative or limbic cortex[5,6], there is every reason to suppose that the BG might play a major role in emotional processing in humans. This assertion is reinforced by both the BG's intrinsic function and their functional and effective connectivity with the rest of the brain[7], revealed by functional magnetic resonance imaging (fMRI)[8], electrophysiological data[9], lesion studies[10], as well as by deep brain stimulation of the BG[11]. There is growing evidence for the involvement of the BG in vocal emotional processing, not only directly, but also through their connections with structures known to be involved in emotional processing, such as the superior frontal and temporal gyri, the amygdala and the cerebellum[12].

Evidence gathered from fMRI and lesion models has led to the hypothesis that the BG play a critical and potentially direct role in vocal emotion processing, by promoting efficient decoding of emotional information from vocal cue sequences and rhythmic aspects of speech[13,14]. The highly connected, closed loop nature of the BG make them perfectly situated to coordinate activity in other cortical and subcortical regions related to emotional voice perception. The subthalamic nucleus (STN) may synchronize neural oscillations within a broader limbic network in order to facilitate efficient processing of auditory and emotion information[11]. This synchronization

[1]Neuroscience of Emotion and Affective Dynamics Laboratory, Department of Psychology and Educational Sciences, University of Geneva, 40 bd du Pont-d'Arve, 1205 Geneva, Switzerland. [2]Swiss Centre for Affective Sciences, University of Geneva, Geneva, Switzerland. [3]Department of Psychology, University of Zürich, Zurich, Switzerland. [4]Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland. [5]Department of Psychology, University of Oslo, Oslo, Norway. [6]Clinical and Experimental Neuropsychology Laboratory, Department of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland. [7]Cognitive Neurology Unit, Department of Neurology, University Hospitals of Geneva, Geneva, Switzerland. [8]These authors contributed equally: Didier Grandjean and Julie Péron. ✉email: leonardo.ceravolo@unige.ch

would strengthen cortical representations of repeated stimulus–response pairings to form "chunks" of behavioural/cognitive response patterns that could be processed more automatically over learning[15]. Simultaneously, these chunks may be modified by the cerebellum to minimize the prediction error of an internal model based on its representation of the current sensory state and expected outcome of ongoing auditory processing[16,17]. Furthermore, the BG and cerebellum may analyse temporal patterns in acoustic stimuli to extract salient emotional cues to feedback to cortex. Nevertheless, the way in which these subcortical and cortical structures exhibit coupling (or decoupling) in order to allow the emergence of a cognitive process such as emotional prosody recognition (i.e., functional integration) remains largely unexplored in affective neuroscience, especially the patterns of connectivity between the BG and the cerebellum in vocal emotion decoding[7,12].

As for the subthalamic nucleus, the BG can be divided in at least three functional compartments relative to their cortical efferences: motor, associative and limbic[5–7]. In the present study, BG regions of interest were the striatum, the globus pallidus (internal and external parts) and the STN[7,18–21]. These regions also play a critical role in selecting a relevant response pattern—and inhibiting irrelevant ones—and in reward feedback and anticipation[7]. BG efferences also connect them more directly to the cerebellum, which can also be separated into motor, associative/cognitive, and limbic subparts[22] that were recently highlighted by resting state functional connectivity[23], specific task-based parcellation[24] and cerebellar topography[22]. In the scope of the present study, the cerebellum would help fine-tune the selected response initiated in the BG, generate an internal model of current goal states and somehow close the loop of reward encoding[7,25] in addition to simultaneously assessing auditory timing for further iterations of vocal emotion decoding across time (lesion studies)[26–28]. Specific areas of the cerebellum associated with (vocal) emotion processing are the cerebellum crus of ansiform lobule I and II (Crus I,II), cerebellar lobules IV, V, VI, VIIb, VIII and IX, Vermis[7,12,22,29–32] and deep cerebellar nuclei, especially the dentate[7] and fastigial nucleus[33,34].

Recent neuroimaging studies helped gain new insights into the role(s) of the BG in emotion processing but still presented shortcomings that need to be overcome. These studies failed: to take advantage of high-resolution scanning of the BG; to investigate the functional and effective connectivity among the BG and between the BG and different subparts of the temporal regions[35] that sustain emotional prosody processing, and more crucially between the BG and the cerebellum; to assess the impact of low-level acoustic parameters on voice prosody processing in the BG or cerebellum, despite their impact in temporal and frontal brain regions[36,37].

Considering the abovementioned literature, the present study was designed to improve our current understanding of the functional integration of the BG and cerebellum during emotional prosody processing in humans, taking into account low-level acoustic parameters of interest such as synthesized fundamental frequency (f0) and energy, using high resolution fMRI in healthy participants. We therefore hypothesized: (i) an increase of BOLD signal in the STN, striatum, globus pallidus (internal, GPi; external, GPe) and cerebellum (Crus I-II, Vermis, cerebellar lobules IV-IX) during the processing of emotional (angry and happy) voices, as opposed to emotionally neutral voice prosody and (ii) similarly for emotional voices when removing variance explained by low-level acoustics (synthesized energy and f0); (iii) enhanced BOLD signal in the BG (STN, striatum, globus pallidus) for angry voice envelope (synthesized energy); (iv) functional connectivity between the BG, especially in the STN and GPi/GPe, the cerebellum (Vermis and cerebellar lobules IV-IX, dentate nucleus) and temporal (superior temporal gyrus) and frontal voice areas (inferior frontal cortex, orbitofrontal cortex) when contrasting emotional to neutral voices (independently of synthesized energy and f0); (v), enhanced effective coupling within the BG (striatum, STN, GPi/GPe) for angry and/or happy voices.

Results

Fifteen (8 female, 7 male) participants were included in the final analysis of the present study. Their task was a one-back task on neutral, happy and angry sentences of pseudowords ('*ne kali bam sud molen!*') presented binaurally through MR-compatible headphones. Both the original voices and synthesized versions of them—synthesized energy and synthesized f0 voices—were included as stimuli across two runs of about 10 min each, in pseudorandom order. The factors of interest were therefore the Emotion and the type of voice (Acoustic Parameters factor) and the interaction between these two factors. More details on the task and paradigm can be found in the Methods section.

Wholebrain results. We performed voxel-level general linear analyses subdivided into three different models in order to find enhanced brain activity related to the factorial design of our data. The models of interest were model 1 and 2, in which we modelled the Emotion factor and the two-way interaction between Emotion and Acoustic Parameters factors. The former analysis revealed emotion-specific enhanced patterns of activity that are presented in this section (for the general effect of Emotion, see Fig.S1), while the full interaction between factors did not yield any significant results. We present, however, one significant result of interest, as part of our hypotheses, for the rhythmicity of angry voices (synthesized energy of angry > neutral prosody). Finally, results for model 3 – the main effect of Acoustic Parameters – are reported in the supplementary data (Supplementary Table 1–3).

Main effect of emotion factor. Wholebrain results for the Emotion factor revealed significant enhanced activity for both angry > neutral voices (Supplementary Table 4) and happy > neutral voices (Supplementary Table 5) contrasts. Enhanced activations for emotional (angry and happy) compared to neutral voices were also significant especially in the superior temporal cortex and inferior frontal cortex, bilaterally (see Supplementary Table 6). Brain activity specific to angry voices (angry > neutral voices) replicated the involvement of the temporal cortex for processing such stimuli, especially in the anterior part of the middle temporal cortex (aMTG) and the posterior superior temporal gyrus and sulcus (pSTG and pSTS, respectively), bilaterally (Fig. 1a,b,g).
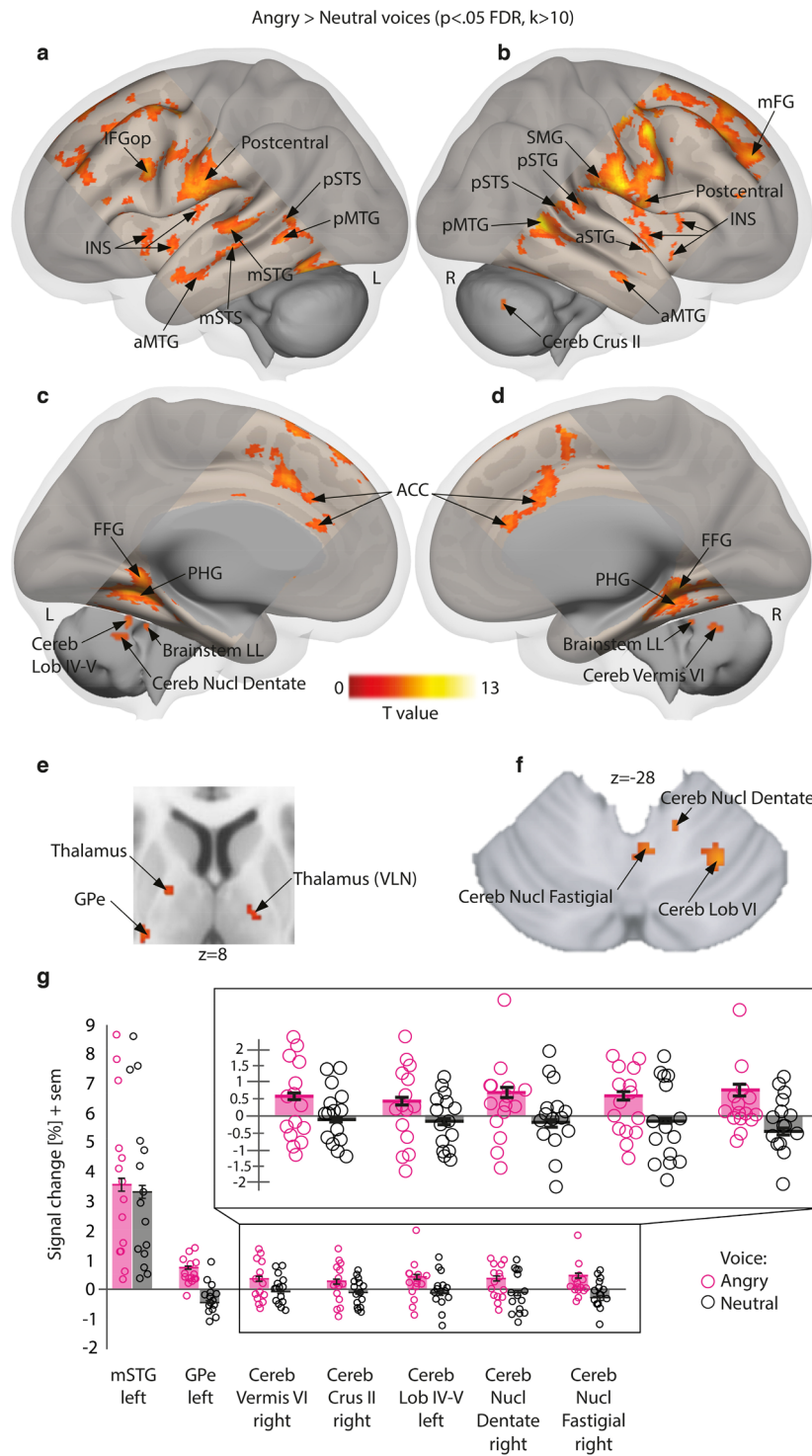
Figure 1. Enhanced brain measures for implicitly processing angry compared to neutral voices, corrected for multiple comparisons (wholebrain voxel-wise $p < .05$ FDR, $k > 10$ voxels). (**a–b**) Lateral activations rendered on a sagittal image highlighting middle and superior temporal regions. (**c–d**) Medial activations of the anterior cingulate cortex, parahippocampal cortex and cerebellum. (**e**) Subcortical activity in the thalamus and globus pallidus displayed on an axial slice. (**f**) Cerebellar activations displayed on an axial slice. G, Percentage of signal change extracted using singular value decomposition on 9 voxels around each peak in a subset of regions with individual values (circles), mean values (bars) and standard error of the mean (error bars) for angry and neutral voices. Pink circles: angry voices; Black circles: neutral voices. L: left; R: right; IFGop: inferior frontal gyrus pars opercularis; STG: superior temporal gyrus; STS: superior temporal sulcus; MTG: middle temporal gyrus; INS: insula; SMG: supramarginal gyrus; FG: frontal gyrus; FFG: fusiform gyrus; PHG: parahippocampal gyrus; ACC: anterior cingulate cortex; Cereb: cerebellum; Cereb Lob: cerebellum lobule; Cereb Nucl Dentate: dentate nucleus of the cerebellum; Cereb Nucl Fastigial: fastigial nucleus of the cerebellum; Brainstem LL: lateral lemniscus of the brainstem; Thalamus VLN: ventral lateral nucleus of the thalamus; GPe: external globus pallidus; Cereb Crus: cerebellum crus of ansiform lobule; ACC: anterior cingulate cortex. 'a' prefix: anterior part; 'm' prefix: mid part; 'p' prefix: posterior part.

Enhanced activity was also observed in medial brain areas such as the anterior cingulate cortex (ACC), the para-hippocampal gyrus and the fusiform gyrus (Fig. 1c,d). Activity in the basal ganglia was restricted to the external globus pallidus (GPe) while we also observed enhanced activity in several parts of the thalamus (Fig. 1e). Finally, large parts of the cerebellum were also more active (Fig. 1g) during angry as opposed to neutral voice processing, namely the Crus II area (Fig. 1b), lobules IV-V and VI (Fig. 1c,f), Vermis area VI (Fig. 1d) as well as deep nuclei such as the dentate (Fig. 1c,f) and fastigial nucleus (Fig. 1f). More details are available in Supplementary Table 4.

As for angry voices, brain activity specific to normal happy voices (happy > neutral voices) highlighted the anterior, mid and posterior portions of the temporal cortex (aSTS, aMTG; mSTS, mSTG; pSTS, pSTG, pMTG, respectively), bilaterally (Fig. 2a,b,g). Enhanced activity was medially observed in the ACC, parahippocampal gyrus and fusiform gyrus (Fig. 2c,d). Increase of activity in the basal ganglia was observed in the GPe and bilateral putamen, and in the ventral lateral nucleus of the thalamus (Fig. 2e). Multiple subparts of the cerebellum showed significant differences. Cerebellum areas were more activated (Fig. 2g) during happy as opposed to neutral voice processing, especially in the lateral Crus I area, bilaterally (Fig. 2a,b,f), in lobules VI, VIIb and VIII (Fig. 2c,d,f), in Vermis areas III and IV-V (Fig. 3CD) as well as in the dentate nucleus (Fig. 2f). More details are available in Supplementary Table 5.

Interaction effect between Emotion and Acoustic Parameters factors.

The full, two-way interaction between our Emotion and Acoustic Parameters factors did not reveal significant results when contrasting angry or happy voices to neutral voices while taking into account normal compared to synthesized voices. We, however, had a specific hypothesis concerning the rhythmicity of angry voices, namely the impact of the 'envelope' of such voices on basal ganglia regions. We therefore used model 3 to compute a contrast dedicated to highlighting brain regions sensitive to the envelope of angry compared to neutral, synthesized energy voices [synthesized energy for angry > neutral voices]. The contrast revealed enhanced activity in the left ventral lateral and lateral posterior nucleus of the thalamus, putamen, substantia nigra, right caudate head, thalamus as well as in the bilateral insula, left amygdala and right mid-to-anterior and posterior STG (Supplementary Table 7). Similar regions, especially large parts of the STG and STS, were also more active for the synthesized energy of happy voices, namely for the [synthesized energy for happy > neutral voices] contrast (Supplementary Table 8).

Functional connectivity results.

Wholebrain analyses revealed significant results for both of our factors (Emotion, Acoustic Parameters) but their interaction did not yield any above-statistical-threshold activations. Computing functional/effective connectivity analyses (both seed-to-seed and seed-to-voxel), however, did reveal several coupled and anti-coupled networks underlying such two-way interaction between the Emotion and the Acoustic Parameters factors. While functional connectivity results were primarily used to further compute effective connectivity, we kept them in the present section due to their specificity and general meaning. These results are presented below.

Seed-to-seed functional connectivity.

Computed using 137 ROI composed of 58 'aal' regions within our field of view, 23 brainstem regions, 22 basal ganglia regions and 34 cerebellum regions, seed-to-seed analyses revealed significant results for the interaction between Emotion and Acoustic Parameters factors, for each emotion of interest. Our contrasts of interest therefore included angry or happy compared to neutral voices when spoken normally as opposed to synthesized $f0$ and energy voices. Seed-to-seed functional connectivity specific to angry original voices were therefore computed with the [angry > neutral voices * original > $f0$ & energy synthesized voices] contrast, revealing coupled networks. As predicted, we observed coupling between the basal ganglia and the cerebellum, more specifically between the left GPe and right cerebellum lobule X (Fig. 3). Coupled functional connectivity was also observed between the left pSTG and right frontal operculum and in the brainstem between major motor (right parieto-occipito-temporo-pontine tract) and sensory tracts (bilateral spinothalamic tract). Detailed results are reported in Supplementary Table 9.

Looking at positive emotion stimuli, happy voices yielded coupled and anti-coupled seed-to-seed functional connectivity results, as seen in the [happy > neutral voices * original > $f0$ & energy synthesized voices] contrast (Fig. 4). Coupled functional connectivity revealed three distinct networks: (1) Internal globus pallidus (GPi) and aSTG in the right hemisphere; (2) Left pMTG and right central operculum cortex; (3) Right corticospinal tract (major motor tract) and right lateral lemniscus (major sensory tract). Happy voices also led to two separate anti-coupled networks involving the right paracingulate cortex and subcalcarine cortex as well as in posterior temporal areas, namely between the left pMTG and right pSTG (Fig. 5). Details reported in Supplementary Table 10.

Seed-to-voxel effective connectivity with the basal ganglia as seeds.

In order to determine the direct relations between BG regions and the rest of the brain, namely each voxel, we computed seed-to-voxel analyses using multivariate regressions and took as seeds only the BG (N = 22 ROI; Fig. 6). We only observed significant effective connectivity specific to angry—but not happy—voices through the interaction with the Acoustic Parameters factor [angry > neutral voices * original > $f0$ & energy synthesized voices]. This multivariate analysis revealed a direct coupling between the left STN (seed) and the ipsilateral cerebellum crus II of ansiform lobule (MNI xyz −4 −86 −42; $t_{14} = 4.14$, k = 26 voxels; $p = 0.031$ FDR corrected, two-tailed; Fig. 5a). We also observed an anti-coupling between the left GPe (seed) and left temporo-occipital MTG (MNI xyz -60 -50 -2) and MFG (MNI xyz −44 34 20; for both contrasts, $t_{14} = 4.14$, k = 29 and 20 voxels, respectively; $p = 0.018$ and 0.048 FDR corrected, two-tailed, respectively; Fig. 5b). Finally, direct coupling was observed between the left caudate nucleus (seed) and voxels covering part of the right primary auditory cortex and planum temporale (MNI xyz 54 −12 0; $t_{14} = 4.14$, k = 64 voxels, $p = 0.00009$ FDR corrected, two-tailed; Fig. 5c).
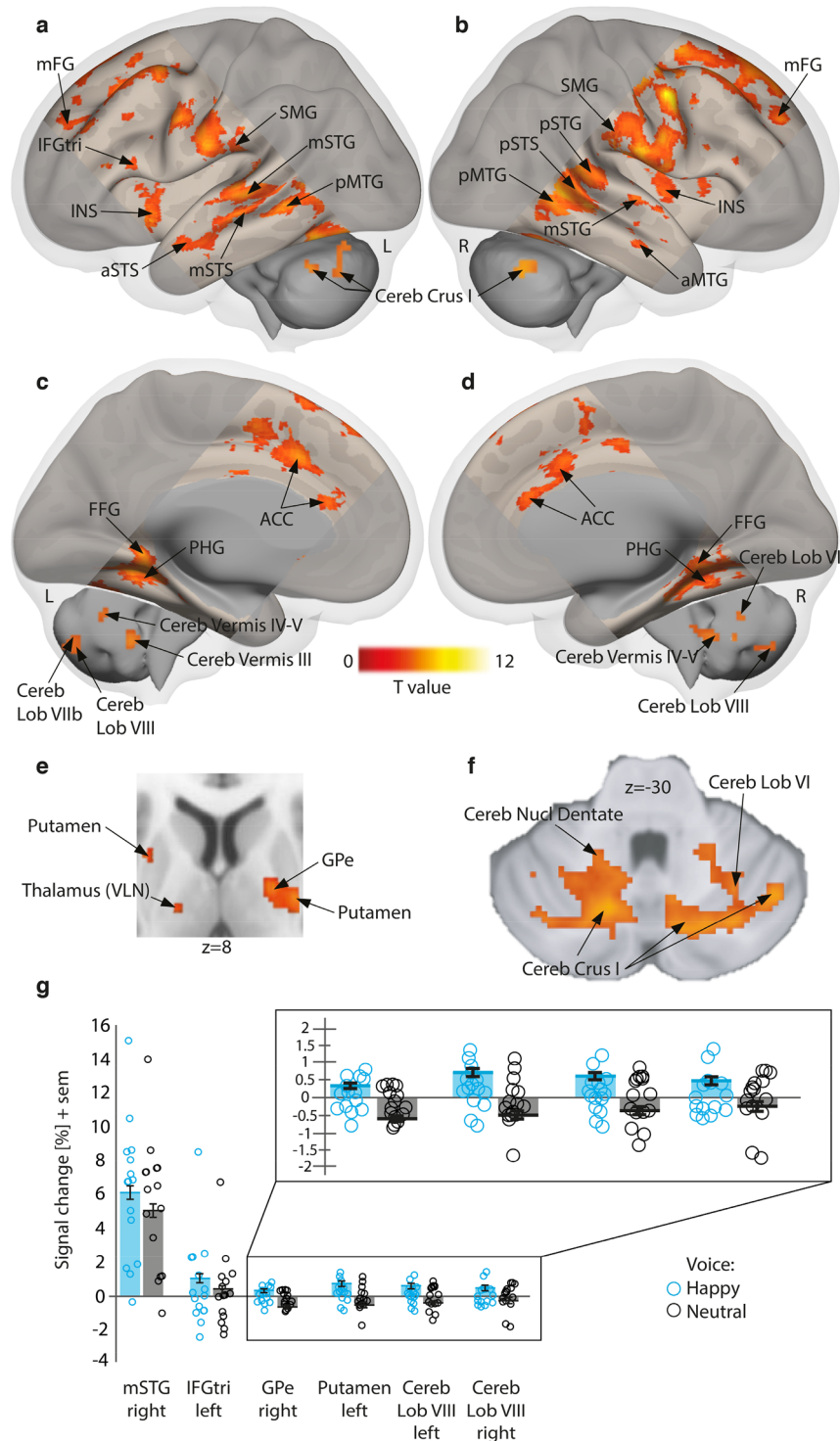
Figure 2. Enhanced brain measures for implicitly processing happy compared to neutral voices, corrected for multiple comparisons (wholebrain voxel-wise $p < .05$ FDR, k > 10 voxels). (**a**–**b**) Lateral activations rendered on a sagittal image highlighting middle, superior temporal and cerebellar regions. (**c**–**d**) Medial activations of the anterior cingulate cortex, parahippocampal cortex and cerebellum. **e** Subcortical activity in the putamen, thalamus and globus pallidus displayed on an axial slice. (**f**) Cerebellar activations displayed on an axial slice. (**g**) Percentage of signal change extracted using singular value decomposition on 9 voxels around each peak in a subset of regions with individual values (circles), mean values (bars) and standard error of the mean (error bars) for happy and neutral voices. Blue circles: happy voices; Black circles: neutral voices. L: left; R: right; IFGtri: inferior frontal gyrus triangularis part; STG: superior temporal gyrus; STS: superior temporal sulcus; MTG: middle temporal gyrus; INS: insula; SMG: supramarginal gyrus; FG: frontal gyrus; FFG: fusiform gyrus; PHG: parahippocampal gyrus; ACC: anterior cingulate cortex; Cereb: cerebellum; Cereb Lob: cerebellum lobule; Cereb Nucl Dentate: dentate nucleus of the cerebellum; Brainstem LL: lateral lemniscus of the brainstem; Thalamus VLN: ventral lateral nucleus of the thalamus; GPe: external globus pallidus; Cereb Crus: cerebellum crus of ansiform lobule; ACC: anterior cingulate cortex. 'a' prefix: anterior part; 'm' prefix: mid part; 'p' prefix: posterior part.
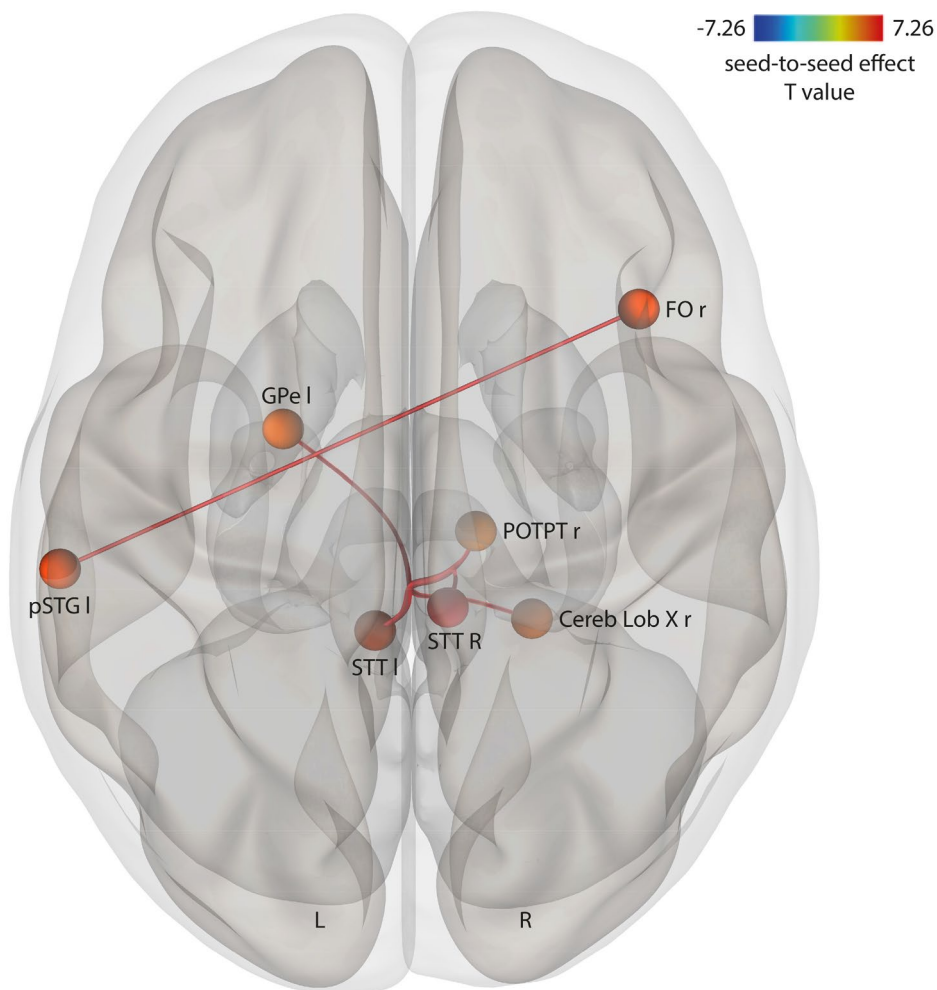
Figure 3. Coupled seed-to-seed, gPPI functional connectivity for the interaction between the Emotion and the Acoustic parameter factors contrasting angry > neutral voices * original > f0 & energy synthesized voices, corrected for multiple comparisons ($p < .05$ FDR). l and L: left; r and R: right; FO: frontal operculum; GPe: external globus pallidus; pSTG: posterior superior temporal gyrus; STT: spinothalamic tract of the brainstem; POTPT: parieto-occipito-temporo-pontine tract of the brainstem; Cereb Lob: cerebellum lobule.

Seed-to-seed effective connectivity within the basal ganglia. We were ultimately interested in the effective connectivity within the basal ganglia when processing emotional (angry, happy) voices and independently of low-level acoustic parameters (synthesized f0, energy). We therefore used multiple regression analyses within the BG for our interaction contrasts to highlight direct relations between BG regions. The anger specific contrast [angry > neutral voices * original > f0 & energy synthesized voices] did not reveal any effective connectivity in BG regions whereas the happiness specific contrast [happy > neutral voices * original > f0 & energy synthesized voices] revealed coupling between the left putamen and GPi ($t_{14} = 3.78$, $p = 0.030$ FDR corrected, two-tailed) as well as anti-coupling between the left GPi and the ipsilateral nucleus accumbens ($t_{14} = -3.65$, $p = 0.039$ FDR corrected, two-tailed).

Discussion

The present study aimed at determining the functional role of both the basal ganglia and the cerebellum according to an integrative neural model of vocal emotion perception, decoding and integration using focal, high-resolution fMRI. It was assumed that connectivity–functional and/or effective–between the BG and the cerebellum would underlie the differential processing of emotion, namely angry and/or happy compared to neutral voices, especially when constraining our data by the use of low-level acoustic parameters of no-interest (synthesized f0 and synthesized energy voices). Our results confirmed the hypothesized involvement of subparts of the BG and cerebellum in processing vocal emotions. The interaction between emotion and acoustical parameters yielded significant results only for connectivity analyses. Functional connectivity data revealed coupled and anti-coupled
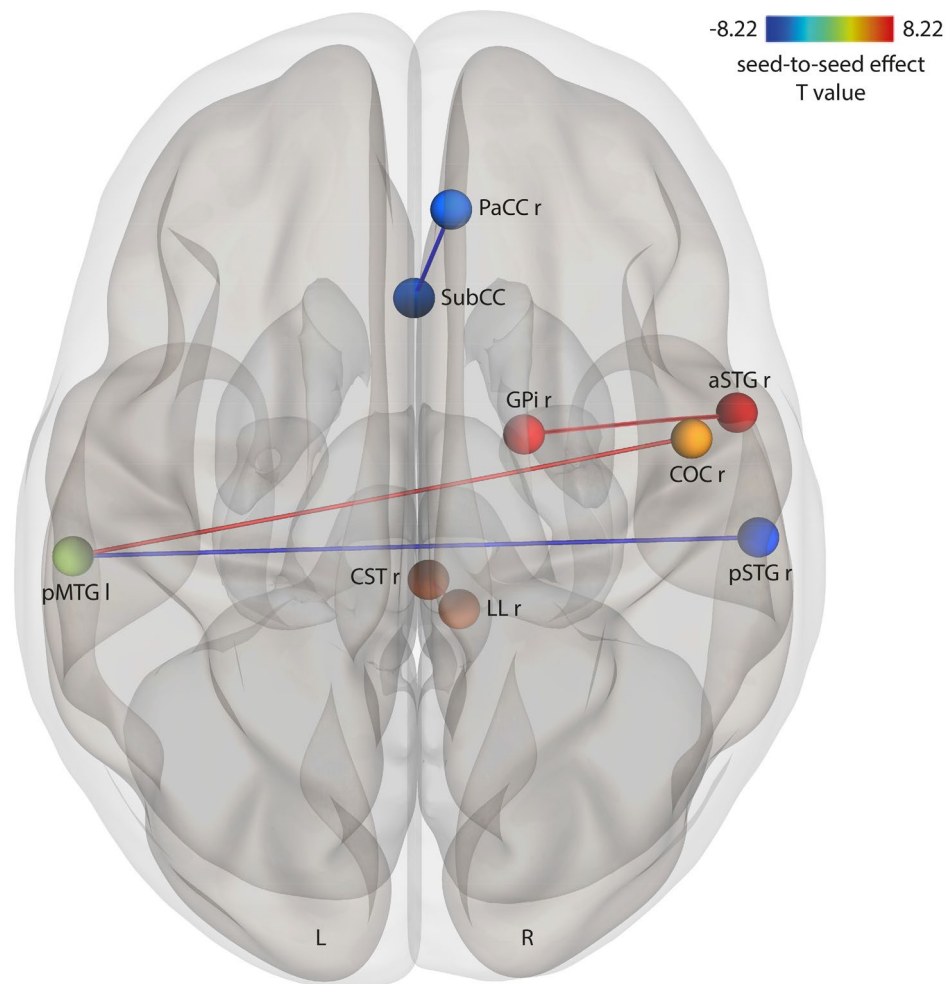
Figure 4. Coupled and anti-coupled seed-to-seed, gPPI functional connectivity for the interaction between the Emotion and the Acoustic parameter factors contrasting happy > neutral voices * original > f0 & energy synthesized voices, corrected for multiple comparisons ($p < .05$ FDR). l and L: left; r and R: right; PaCC: paracingulate cortex; SubCC: subcalcarine cortex; GPi: internal globus pallidus; COC: central operculum cortex; aSTG: anterior superior temporal gyrus; pSTG: posterior superior temporal gyrus; pMTG: posterior middle temporal gyrus; CST: corticospinal tract of the brainstem; LL: lateral lemniscus of the brainstem.

networks involving the BG and cerebellum, while effective connectivity within the BG and with the BG as seeds, shed new light on the involvement of the internal and external globus pallidus, putamen, left STN and caudate nucleus in vocal emotion processing.

The implication of subcortical structures other than the amygdala involved in emotion processing was only recently emphasized[21,38] and through deep brain stimulation in the STN as a neurosurgical treatment for Parkinson's disease and obsessive–compulsive disorder, a new research window opened[11]. According to Péron and colleagues' model (2013)[11] and in line with existing literature and our results, the processing of emotion would rely on both the direct ('hyperdirect pathway') and indirect coupling between STN subterritories (motor, associative and limbic) and the neocortex, especially the orbitofrontal cortex (OFC) and modality-specific primary and secondary cortices. Indirect coupling would transit from the STN to the OFC through the BG, especially the GPi and GPe, thalamus, substantia nigra and ventral tegmental area, and/or through the amygdala that exhibits some direct connections with the BG as well[11]. The STN could synchronize oscillations in relevant areas across the brain including the cerebellum to shape cortical learning and facilitate habitual, overlearned processing of familiar stimuli types[7]. Our results fit well with such model and constrain it by adding some nuance to the expected synchronized regions across the brain. In fact, we observed enhanced activity in several subparts of the BG and in different territories of the cerebellum. More specifically, we observed for angry—similarly for happy—voice processing the involvement of the GPe and thalamus as well as of several lobules (IV, V, VI), nuclei (fastigial, dentate) and areas (Vermis area VI) of the cerebellum and posterior, mid and anterior temporal regions within the voice-sensitive areas. GP activity fits with a more accurate recognition of vocal emotion in healthy compared

Effective connectivity analysis (seed-to-voxel), seeds within the basal ganglia, p<.05 FDR
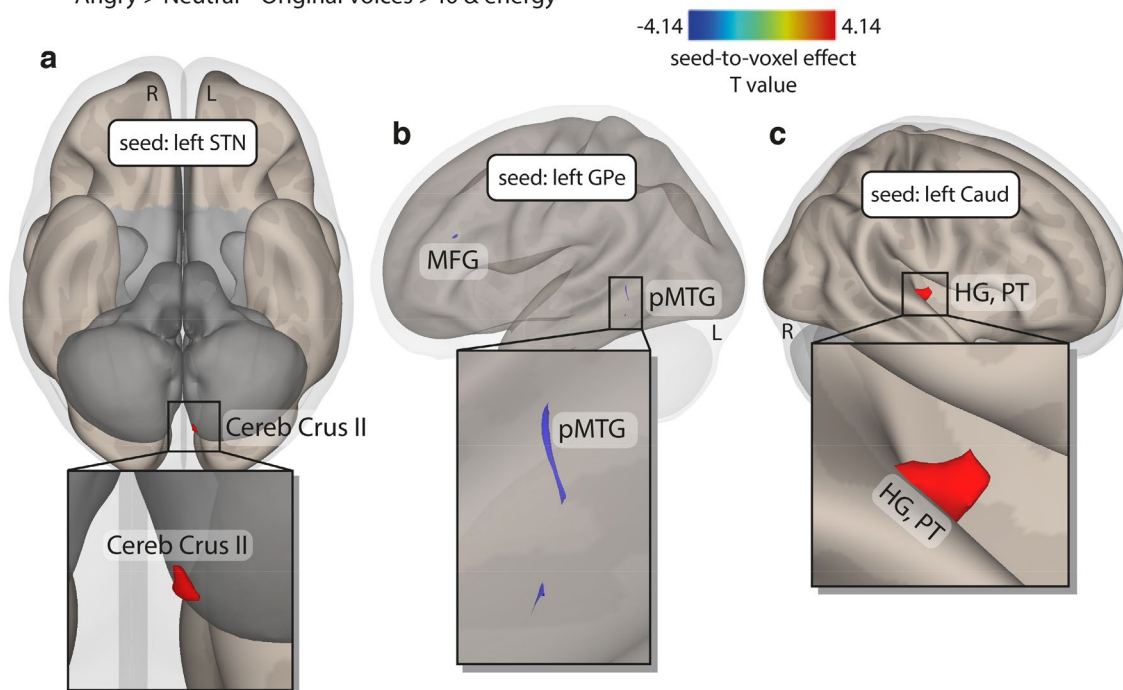
Angry > Neutral * Original voices > f0 & energy

Figure 5. Coupled and anti-coupled seed-to-voxel, gPPI effective connectivity for the interaction between the Emotion and the Acoustic parameter factors contrasting angry > neutral voices * original > f0 & energy synthesized voices, corrected for multiple comparisons ($p < .05$ FDR). (**a**) Inferior view showing direct coupling between the left STN (seed) and the ipsilateral Cerebellum Crus II. (**b**) Sagittal view showing direct anti-coupling between the left GPe (seed) and the left MFG and pMTG. (**c**) Sagittal view showing direct coupling between the left caudate nucleus (seed) and the right primary auditory cortex (Heschl's gyrus) and planum temporale. L: left; R: right; STN: subthalamic nucleus; GPe: external globus pallidus; Caud: caudate nucleus; Cereb Crus II: cerebellum crus II of ansiform lobule; pMTG: posterior middle temporal gyrus; MFG: middle frontal gyrus; HG: Heschl's gyrus; PT: planum temporale.

to BG-lesioned patients[39], and with a general role of the more dorsal BG for the sequencing and anticipation of acoustic temporal variations[18]. The BG would therefore be crucial to detect and classify auditory patterns, subsequently synchronizing activity in other regions for selecting the appropriate response.

The limbic cerebellum (predominately the vermis) and associative regions of the cerebellum (including posterior hemispheric lobules[24,40–42]), present in our wholebrain and connectivity results are in line with the general role of the cerebellum in auditory perception[43] and more specifically emotion recognition and perception[44]. These areas of the cerebellum then could modulate cortical oscillations based on prediction error feedback relative to the given context[45,46]. By continuously monitoring incoming stimuli for deviations from expected emotional structure (e.g., an angry voice) and fine-tuned interval timing[47], the limbic and associative territories of the cerebellum–in our results, Vermis IV and VI and hemispheric lobules IV–VI, VIII, respectively–could signal the need for greater attentional control of sensory cortical responses. Cerebellum activity in our results would also fit well with response adaptation and motor control[48], preparing a response following vocal emotion decoding and processing[49], especially when the voice or sound is perceived as aversive[50]. Input to the limbic cerebellum (Vermis and fastigial nucleus) from OFC or the BG regarding the salience of emotional stimuli would shape internal models about how an emotional response would affect the individual in their current state, and, thus, how the cerebellum modifies limbic responses, especially in the temporal domain[27].

The idea of temporal pattern analysis in the associative territory of the cerebellum has been proposed, especially when patterns are irregular and not rhythmic[26], which includes temporal patterns of vocal emotion and emotional prosody. Specifically, a double dissociation between patients with a BG or cerebellum lesion confirmed that cerebellar lesions alter non-rhythmic–but not rhythmic–temporal prediction while BG lesions showed the opposite pattern[27]. Additionally, misattributions in emotion recognition between surprise and fear correlated with lesions in lobules VIIb, VIII and X of the cerebellum[12], regions that overlap with our results for angry and happy voices in both the wholebrain activation and connectivity analyses and are in line with previous evidence of emotional processing within these specific regions[22,31,32]. Therefore, these cerebellar lobules may have a crucial function in emotion recognition in voices, notably in temporal pattern analysis and critical low-level acoustics integration such as f0 or pitch.

The importance of BG-cerebellum connections in vocal emotion processing, especially for anger, was further emphasized by our functional connectivity data for angry, but not happy, original voice processing (removing the variance explained by synthesized f0 and energy), which revealed coupling between the GPi and putamen

with lobule X of the cerebellum. These results are consistent with a coupling of BG and cerebellum activity in time for autonomic emotional reaction and prediction generation[51] or interval timing[47] and motor prediction[48] but cerebellar lobule X is more rarely observed in emotion-related tasks. This cerebellar lobule, however, was recently integrated in the 'triple nonmotor representation' and evidence shows its limbic ties with the neocortex[52]. It is also important to note here that many cerebellar sub-regions often labelled as 'motor' (for example, linked to hand or eye movements) are also significantly involved in cognitive or emotional tasks[53,54], such as lobules V, VI, VIII[24]. Our results therefore converge toward a critical role of the cerebellum in coordination with the BG for both the decoding of vocal emotion—in the temporal, voice-sensitive areas—and the conversion to a motor response[48] as an output behaviour following a subjective feeling of emotion[7,49].

Furthermore, our effective connectivity results strongly emphasized within-BG direct relations between the putamen and GPi (coupling) and between the GPi and nucleus accumbens (anti-coupling) as well as between BG seeds and frontal and superior temporal regions. Additionally, effective seed-to-voxel connectivity revealed direct coupling between the left STN and ipsilateral cerebellum crus II of the ansiform lobule. While the role of the STN in emotion processing[20,55–58] and vocal emotion recognition[11,19,49,59,60] has gathered strong interest in the recent years, the crus II area of the cerebellum also subserves cognition and emotion processes[29,44,61]. Direct coupling was also observed between the left caudate nucleus and the primary auditory cortex and planum temporale, fitting well again with the direct coupling between the BG and modality-specific sensory cortex[11] with the caudate playing a critical role in voice arousal[62] and emotion processing[63].

We interestingly also observed direct anti-coupling between the left GPe, involved in the explicit recognition of emotional prosody[39], and ipsilateral posterior MTG and MFG, superior to and slightly overlapping with the triangularis part of the IFG. Activity modulations in these latter lateral brain areas were repeatedly observed in voice processing in general[64] and vocal emotion[65,66], especially when contrasting happy to angry voices[67]. The fact that posterior MTG activity was previously linked to happy vs. angry voice processing therefore could explain the coupling we observed that is specific to happy voices, especially since GP functioning relates to explicit vs. implicit emotion recognition[39].

While our data depict a relatively clear image of the importance of the BG and cerebellum for vocal emotion processing and further output response, some limitations should be mentioned. First, sample size was limited and even though we were strict with the correction of p values in our statistical analyses, a sample size closer to 25 participants would have been better for reliable data generalization and reproducibility. Second, p values for wholebrain data analyses were corrected for multiple comparisons using voxel-wise False Discovery Rate (FDR), namely by dividing the p value by the number of activated voxels rather by the total number of voxels in the brain—namely Family-Wise Error (FWE) correction. While FDR is widely used in the functional MRI literature, we cannot exclude more voxels with false positives as compared to FWE correction. Third, and as often observed in the literature, we included happy, angry and neutral emotions as vocal stimuli but other critical emotions such as fear, surprise, sadness or several others were not included, therefore restricting our conclusions. Fourth, although we did include low-level acoustic parameters to control for emotion-specific activity, other meaningful ones should be used in the future, for instance the spectral domain related to voice quality perception, which is thought also important for emotional voice recognition. Fifth, we used high-resolution fMRI, greatly improving spatial resolution with, however, the added cost of a truncated field of view. We cannot therefore exclude the fact that frontal and parietal regions, excluded at data acquisition, would play a role in vocal emotion processing, in terms of both activation and connectivity using the same task. It is, however, worth mentioning that the focus of the present study was on cerebellar and basal ganglia contributions to vocal emotion processing. Sixth, we did not divide the STN and other BG or cerebellar regions into their known associative, motor and limbic subparts. A more precise understanding of the specific role of each subpart of the BG nuclei is therefore unfortunately not possible at this stage. Such concern should be addressed in the future by the use of subject-level delineation of BG sub-territories and/or by using even higher fMRI resolution, such as with a 7-T scanner. Finally, while our functional connectivity results were consistent with existing literature, we cannot rule out that other regions may mediate the correlations between ROI, so these should be taken with more caution than the effective connectivity results that used more direct mathematical association calculations (multiple regressions). In addition to these limitations, future studies should try to highlight emotional substrates within the BG and cerebellum pertaining to sub-components of emotion[44], such as for example perception and/or decoding, subjective feeling, response output, behavioural response to emotion, as well as giving more importance to task designs allowing for a clearer topography and parcellation of the affective BG and cerebellum. Future studies should also include patients with known alterations and/or lesions of basal ganglia and cerebellar brain regions such as Parkinson disease—or any relevant lesion within these regions of interest[48]—or with biases in emotion recognition and processing[44] such as in depression or schizophrenia and compare them to healthy, matched controls.

In conclusion, the present study aimed at a better understanding of the implications of basal ganglia and cerebellum involvement in vocal emotion processing. Through the combination of wholebrain analysis, functional and effective connectivity analyses and with the partial exclusion of low-level acoustics of interest (voice $f0$, energy) our data depict a clearer role of the STN, GP and putamen in vocal emotion processing, especially for auditory pattern detection and synchronization across cortical and subcortical limbic networks. The current results add weight to the assertion that both direct and indirect coupling between these BG regions and the cortex is modulated by BG and cerebellum connections. Our results also favour a framework in which the brain could use temporal regularities ('patterns') to analyse and anticipate the timing of future events, and constrain attention and action accordingly. Further work could use a dedicated task and focus on BG and cerebellum subterritories since their specific role(s) is of the highest interest for affective and social neuroscience research.

Material and methods

Participants.　We initially included 19 healthy participants but excluded four of them from the analyses because of MRI signal artifacts (N = 2) or psychiatric disorder (N = 2). The remaining sample consisted of seven males and eight females (N = 15), with a mean age of 30.5 years (SD = 3.48, range 27–37 years; mean age (SD) for female participants was 30.25 (3.24) and for male participants 30.85 (3.98)). All included participants were right-handed, native French speakers, and had normal or corrected-to-normal vision and normal hearing. None of them had a history of neurological disease or psychiatric disorder.

Ethics declarations.　Participants gave written informed consent for their participation in accordance with the ethical and data security guidelines of the University of Geneva. The study was approved by the local ethics committee and conducted according to the Declaration of Helsinki.

Experimental setup

One-back task.　*Stimuli.*　The vocal (prosodic) stimuli consisted of two pseudosentences spoken with different emotional prosodies ("*ne kali bam sud molen!*" and "*kun se mina lod belam?*"; mean duration = 1642 ms, range = 854–2788 ms) extracted from a previously validated database, the GEneva Multimodal Emotion Portrayals (GEMEP) corpus[68]. Alongside these prosodic stimuli (anger, happiness and neutral), we played synthesized stimuli, built from the original emotional and neutral sounds, in order to control for the temporal dynamics of energy and $f0$. These two basic acoustic features are known to be the most correlated with emotional prosody judgments[69,70]. The first type of synthetic stimulus (synthesized *intensity*) consisted of a section of white/pink noise, to which the intensity contour of the original stimulus was applied. The second type of synthetic stimulus (synthesized $f0$) was a series of pure sine waves (with constant amplitude), the frequency of which corresponded to the $f0$ of the original vocal stimulus, allowing us to maintain the temporal dynamics of the $f0$. Both synthetic stimuli had the same duration as in the original recordings. All sounds were matched for mean energy to avoid too strong loudness effects. Two runs were constructed, featuring the different kinds of stimuli in pseudorandom order (no more than three times for the same experimental condition). Each run contained 20 trials featuring anger stimuli, 20 trials featuring happiness stimuli, and 20 trials featuring neutral stimuli, as well as 15 synthesized intensity stimuli, 15 synthesized $f0$ stimuli, and one section of white noise at the beginning (first stimulus) with a gradual onset to accustom the participants to the auditory material. Each run contained a different list of stimuli. In each prosodic condition, we controlled for the pseudo-sentence being pronounced and the sex of the actor who pronounced the utterances: a female actor pronounced half the stimuli, half of them consisting of the pseudo-sentence "*ne kali bam sud molen!*". The total duration of each run was ~ 10 min, and there was a short break between them. Each run contained pairs of identical subsequent stimuli, representing 10% of the total stimuli (pseudorandom order) to allow a one-back task to be performed by the participants, therefore forcing them to carefully attend each stimulus.

Experimental procedure, paradigm.　In order to avoid expectancy effects, we varied in each trial the duration of the interval between the onset of the fixation cross and the onset of the auditory stimulus. In other words, the presentation of each auditory stimulus was preceded by a silent portion of pseudorandom duration, ranging from 50 to 250 ms, the so-called jitter (Fig. 6). After the offset of the sound, we also included a silent portion ranging from 3000 to 3500 ms. In order to avoid the offset of the sound and the offset of the fixation cross being synchronous, we varied the duration of the interval between these two offsets. Finally, in order to minimize any retinal afterimage, we ensured that the color of the fixation cross did not contrast too greatly with the color of the desktop background.

For each trial, the participants were asked to keep their eyes open and relaxed. They were told they would hear meaningless speech uttered by male and female actors, as well as synthesized sounds. The binaurally recorded auditory stimuli were played through MR-compatible headphones (MR Confon GmbH, Magdeburg, Germany). Loudness intensity was adjusted for each participant according to her/his hearing threshold at the beginning of the experiment. Participants were asked to focus on these auditory stimuli and to press a button whenever they heard two identical stimuli in a row. These one-back trials represented only 10% of all trials and were excluded from the analyses. The one-back task[71] was administered to ensure that the patients were paying attention to the stimuli. Prior to the task, an MR-compatible response box (Current Designs Inc., Philadelphia, PA, USA) was placed beneath the participant's fingers. A similar task, greatly overlapping with the one used here, was previously used by Julie Péron[60].

Image acquisition.　Imaging was conducted at the Brain and Behaviour Laboratory (BBL) of the University of Geneva. For the main task, high-resolution imaging data was acquired on a 3 T Siemens Trio System (Siemens, Erlangen, Germany) using a T2*-weighted gradient echo planar imaging sequence with 440 volumes per run (EPI; 1.5 × 1.5x2.2 mm voxels, slice thickness = 2 mm, gap = 0.2 mm, 31 slices, RT = 2320 ms, TE = 33 ms, flip angle = 90°, matrix = 128 × 128, field of view = 192 mm). The acquired volumes, representing a truncated field of view compared to standard wholebrain acquisition, were almost perpendicular to the anterior commissure-posterior commissure (AC/PC) line to cover all regions of interest, especially the basal ganglia, cerebellum and the temporal lobe (see Fig.S2). Therefore, the term 'wholebrain' in this manuscript refers exclusively to our truncated field of view, not to volumes covering the wholebrain. The total number of volumes for our fifteen participants was 13′200 for a total number of slices of 409′200. A T1-weighted, magnetization- prepared, rapid-acquisition, gradient echo anatomical scan (slice thickness = 1 mm, 176 slices, RT = 2530 ms, TE = 3.31 ms, flip angle = 7°, matrix = 256 × 256, FOV = 256 mm) was also acquired.
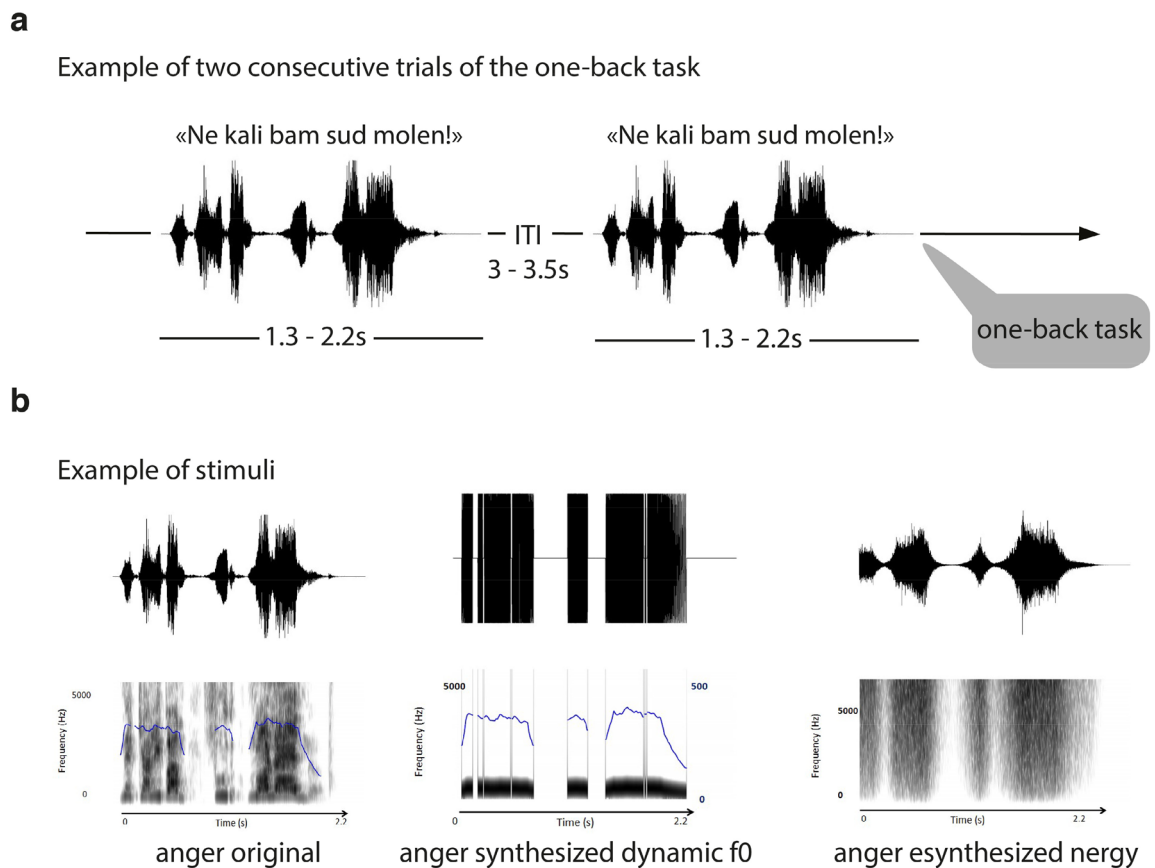
a

Example of two consecutive trials of the one-back task

b

Example of stimuli

anger original   anger synthesized dynamic f0   anger esynthesized nergy

Figure 6. Experimental timeline and details of stimuli for the one-back task. (**a**) Following technical scans (localizer and field map), the first run started for 10 min during which participants had to perform a one-back task on the voice presented auditorily to them using an MRI-compatible button box. The second run followed similarly for 10 more minutes and the session ended with the acquisition of an anatomical image for 5 min. During the complete session, the participant laid down in the scanner and had to pay attention to auditorily presented vocal stimuli and do a one-back task (10% of all trials). All stimuli had a duration of 1.3–2.2 s and an inter trial interval of 3–3.5 s. (**b**) Voice stimuli consisted of pseudowords arranged in sentences with either original vocal signal, synthesized dynamic $f0$ manipulation or synthesized energy.

Image analysis.   *Wholebrain analyses.*   Functional images analysis was carried out using Statistical Parametric Mapping software 12 (SPM12, Wellcome Trust Centre for Neuroimaging, London, UK). Preprocessing steps included realignment to the first volume of the time series, slice timing, iterative normalization into the Montreal Neurological Institute space[72] using the DARTEL toolbox[73] and spatial smoothing with an isotropic Gaussian filter of 6 mm full width at half maximum. To remove low-frequency components, we used a high-pass filter with a cutoff frequency of 128 s. Anatomical locations were defined using a standardized coordinate database using the Automated Anatomical Labelling atlas[74] incorporated in the xjView toolbox (http://www.alivelearn.net/xjview), an atlas of the brainstem[75], basal ganglia[76] and cerebellum[77,78] displayed in FMRIB Software Library v6.0 (FSL)[79] through FSLeyes.

A general linear model was used to compute first-level statistics, in which each run was modelled as a distinct session and each trial was convolved with the hemodynamic response function, time-locked to the onset of each stimulus. Separate regressors were created for each condition, namely for the Emotion and the Acoustic Parameters factors (Design matrix columns for each run (N = 9): anger original, anger $f0$, anger energy, happy original, happy $f0$, happy energy, neutral original, neutral $f0$, neutral energy). Finally, regressors of no-interest included the repetition trials of the one-back task that were concatenated across conditions and added as an additional regressor together with six motion parameters for each run to account for movement. Regressors of interest were used to compute nine simple contrasts (one per column of the design matrix, across runs) for each participant (across runs), leading to a main effect of each condition cited above at the first-level of analysis. Simple contrasts were then used in three distinct flexible factorial, second-level analyses. In model 1, the effect of the Emotion (angry, happy, neutral voices, acoustically untouched or 'original') factor was modelled with one Participant factor and one Emotion factor. In model 2, factors Participant, Emotion (angry, happy, neutral voices) and Acoustic Parameters (original, $f0$ synthesized, energy synthesized parameters) were included to model the two-way interaction between our main factors (Emotion*Acoustic Parameters). Model 3 included the main effect of the Acoustic Parameters (normal, $f0$ synthesized, energy synthesized parameters) factor, modelled with one Participant factor and one Acoustic Parameters factor. For each model, independence of the Participant factor

www.nature.com/scientificreports/

was set to 'true', variance to 'unequal' and the Emotion, Acoustic Parameters and Emotion*Acoustic Parameters factors with independence as 'false', variance as 'unequal'.

All neuroimaging activations were thresholded in SPM12 by using a wholebrain voxel-wise false discovery rate (FDR) correction at $p < 0.05$ with an arbitrary cluster extent of k > 10 voxels.

Functional and effective connectivity analysis. Functional and effective connectivity analyses were performed using the CONN toolbox[80] version 18.b implemented in Matlab 9.0 (The MathWorks, Inc., Natick, MA, USA) for the two-way interaction between our factors, namely Emotion and Acoustic Parameters (design matrix identical to wholebrain analyses). As in wholebrain data analysis, repetition trials of the one-back task were modelled as a single column including a concatenation of all their onset times across conditions (regressor of no-interest). Functional connectivity analyses were mainly carried out to orient further effective connectivity analysis and we decided to report both types of connectivity for a clear overview of the results. Functional connectivity analyses were computed using as seeds each region of interest (ROI) of the following atlases: the Automated Anatomical Labelling ('aal') atlas[74] (N = 58 ROI), an atlas of the brainstem[75] (N = 23 ROI), basal ganglia[76] (N = 22 ROI) and cerebellum[77,78] (N = 34 ROI). All ROI (N = 137; Supplementary Table 11) were within the bounds of our truncated field of view. Frontal, parietal and occipital areas outside the bounds of our field of view, specifically of the 'aal' atlas, were isolated through CONN time-course visualization and removed from the analyses when a region had a flat time-course. For effective connectivity analyses and according to our hypotheses, seed regions were limited to the basal ganglia[76] (N = 22 ROI). Spurious sources of noise were estimated and removed using the automated toolbox preprocessing algorithm, and the residual BOLD time-series was band-pass filtered using a low frequency window ($0.008 < f < 0.09$ Hz). Correlation maps were then created for each condition of interest by taking the residual BOLD time-course for each condition from atlas regions of interest and computing bivariate Pearson's correlation coefficients between the time courses of each voxel of each ROI of the atlas, averaged by ROI ('functional connectivity' analyses). 'Effective connectivity' was approached using multivariate regressions between each seed ROI and all other ROI—or all brain voxels for seed to voxel analysis—and a model was generated and used to characterize the direct connectivity between pairs. For both types of connectivity, we used generalized psychophysiological interaction (gPPI) measures, representing the level of task-modulated (often labelled 'effective') connectivity between ROI or between ROI and voxels. gPPI is computed using a separate multiple regression model for each target (ROI/voxel). Each model includes three predictors: (1) task effects convolved with a canonical hemodynamic response function (psychological factor); (2) each seed ROI BOLD time series (physiological factor) and (3) the interaction term between the psychological and the physiological factors, the output of which is regression coefficients associated with this interaction term. Finally, group-level analyses were performed on these regression coefficients to assess for main effects within-group for contrasts of interest in seed-to-seed and seed-to-voxel analyses. Therefore, 'functional connectivity' is defined in the present study as a gPPI analysis using bivariate correlations between ROI, while 'effective connectivity' defines the gPPI analysis using multivariate regressions between ROI/voxels. Connectivity analyses were computed using methods in line with most recent best practices[81]. For both analyses, type I error was controlled by the use of seed-level (seed-to-seed analyses) and cluster-level (seed-to-voxel analysis) false discovery rate correction with $p < 0.05$ FDR to correct for multiple comparisons.

Data availability
All data and codes, batches used in the present study are available on request to the corresponding author.

References
 1. Ethofer, T. *et al.* Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *Neuroimage* **30**, 580–587 (2006).
 2. Frühholz, S. & Grandjean, D. Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: a quantitative meta-analysis. *Neurosci. Biobehav. Rev.* **37**, 24–35 (2013).
 3. Frühholz, S. & Grandjean, D. Processing of emotional vocalizations in bilateral inferior frontal cortex. *Neurosci. Biobehav. Rev.* **37**, 2847–2855 (2013).
 4. Grandjean, D. Brain networks of emotional prosody processing. *Emot. Rev.* 1754073919898522 (2020)
 5. Alexander, G. E. & Crutcher, M. D. Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci.* **13**, 266–271 (1990).
 6. Lambert, C. *et al.* Confirmation of functional zones within the human subthalamic nucleus: patterns of connectivity and sub-parcellation using diffusion weighted imaging. *Neuroimage* **60**, 83–94 (2012).
 7. Pierce, J. E. & Péron, J. The basal ganglia and the cerebellum in human emotion. *Soc. Cogn. Affect. Neurosci.* **15**, 599 (2020).
 8. Péron, J., Frühholz, S., Ceravolo, L. & Grandjean, D. Structural and functional connectivity of the subthalamic nucleus during vocal emotion decoding. *Soc. Cogn. Affect. Neurosci.* **11**, 349–356. https://doi.org/10.1093/scan/nsv118 (2016).
 9. Péron, J. *et al.* Electrophysiological activity of the subthalamic nucleus in response to emotional prosody: an intracranial ERP study in Parkinson's disease. *Mov. Disord.* **29**, S132 (2014).
 10. Cohen, M. J., Riccio, C. A. & Flannery, A. M. Expressive aprosodia following stroke to the right basal ganglia: a case report. *Neuropsychology* **8**, 242 (1994).
 11. Péron, J., Frühholz, S., Vérin, M. & Grandjean, D. Subthalamic nucleus: a key structure for emotional component synchronization in humans. *Neurosci. Biobehav. Rev.* **37**, 358–373. https://doi.org/10.1016/j.neubiorev.2013.01.001 (2013).
 12. Thomasson, M. *et al.* Cerebellar contribution to vocal emotion decoding: Insights from stroke and neuroimaging. *Neuropsychologia* **132**, 107141 (2019).
 13. Kotz, S. A. & Schwartze, M. Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cogn. Sci.* **14**, 392–399 (2010).

14. Pell, M. D. & Leonard, C. L. Processing emotional tone from speech in Parkinson's disease: a role for the basal ganglia. *Cogn. Affect. Behav. Neurosci.* **3**, 275–288 (2003).
15. Graybiel, A. M. Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* **31**, 359–387. https://doi.org/10.1146/annurev.neuro.29.051605.112851 (2008).
16. Sokolov, A. A., Miall, R. C. & Ivry, R. B. The cerebellum: adaptive prediction for movement and cognition. *Trends Cogn Sci* **21**, 313–332. https://doi.org/10.1016/j.tics.2017.02.005 (2017).
17. Bostan, A. C. & Strick, P. L. The basal ganglia and the cerebellum: nodes in an integrated network. *Nat. Rev. Neurosci.* **19**, 338–350. https://doi.org/10.1038/s41583-018-0002-7 (2018).
18. Kotz, S. A., Schwartze, M. & Schmidt-Kassow, M. Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. *Cortex* **45**, 982–990 (2009).
19. Péron, J., Frühholz, S., Ceravolo, L. & Grandjean, D. Structural and functional connectivity of the subthalamic nucleus during vocal emotion decoding. *Soc. Cogn. Affect. Neurosci.* **11**, 349–356 (2015).
20. Schneider, F. *et al.* Deep brain stimulation of the subthalamic nucleus enhances emotional processing in Parkinson disease. *Arch. Gen. Psychiatry* **60**, 296–302 (2003).
21. Wager, T. D. *et al.* The neuroimaging of emotion. *Handbook Emot.* **3**, 249–271 (2008).
22. Leggio, M. & Olivito, G. Topography of the cerebellum in relation to social brain regions and emotions. In *Handbook of clinical neurology*, Vol 154, 71–84 (Elsevier, 2018).
23. Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C. & Yeo, B. T. The organization of the human cerebellum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 2322 (2011).
24. King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B. & Diedrichsen, J. Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nat. Neurosci.* **22**, 1371–1378 (2019).
25. Larry, N., Yarkoni, M., Lixenberg, A. & Joshua, M. Cerebellar climbing fibers encode expected reward size. *Elife* **8**, e46870 (2019).
26. Breska, A. & Ivry, R. B. Taxonomies of timing: Where does the cerebellum fit in?. *Curr. Opin. Behav. Sci.* **8**, 282–288 (2016).
27. Breska, A. & Ivry, R. B. Double dissociation of single-interval and rhythmic temporal prediction in cerebellar degeneration and Parkinson's disease. *Proc. Natl. Acad. Sci.* **115**, 12283–12288 (2018).
28. Grube, M., Cooper, F. E., Chinnery, P. F. & Griffiths, T. D. Dissociation of duration-based and beat-based auditory timing in cerebellar degeneration. *Proc. Natl. Acad. Sci.* **107**, 11597–11601 (2010).
29. Baumann, O. & Mattingley, J. B. Functional topography of primary emotion processing in the human cerebellum. *Neuroimage* **61**, 805–811 (2012).
30. Habas, C. *et al.* Distinct cerebellar contributions to intrinsic connectivity networks. *J. Neurosci.* **29**, 8586–8594 (2009).
31. Stoodley, C. J. & Schmahmann, J. D. Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage* **44**, 489–501 (2009).
32. Stoodley, C. J. & Schmahmann, J. D. The cerebellum and language: evidence from patients with cerebellar degeneration. *Brain Lang.* **110**, 149–153 (2009).
33. Wang, J., Dong, W. W., Zhang, W. H., Zheng, J. & Wang, X. Electrical stimulation of cerebellar fastigial nucleus: mechanism of neuroprotection and prospects for clinical application against cerebral ischemia. *CNS Neurosci. Ther.* **20**, 710–716 (2014).
34. Zhang, X.-Y., Wang, J.-J. & Zhu, J.-N. Cerebellar fastigial nucleus: from anatomic construction to physiological functions. *Cerebellum Ataxias* **3**, 9 (2016).
35. Frühholz, S., Ceravolo, L. & Grandjean, D. Specific brain networks during explicit and implicit decoding of emotional prosody. *Cereb. Cortex* **22**, 1107–1117 (2011).
36. Frühholz, S., Ceravolo, L. & Grandjean, D. Specific brain networks during explicit and implicit decoding of emotional prosody. *Cereb. Cortex* **22**, 1107–1117 (2012).
37. Schirmer, A. & Kotz, S. A. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.* **10**, 24–30 (2006).
38. Tamietto, M. & De Gelder, B. Neural bases of the non-conscious perception of emotional signals. *Nat. Rev. Neurosci.* **11**, 697 (2010).
39. Paulmann, S., Pell, M. D. & Kotz, S. A. Functional contributions of the basal ganglia to emotional prosody: evidence from ERPs. *Brain Res.* **1217**, 171–178 (2008).
40. Guell, X., Gabrieli, J. D. & Schmahmann, J. D. Triple representation of language, working memory, social and emotion processing in the cerebellum: convergent evidence from task and seed-based resting-state fMRI analyses in a single large cohort. *Neuroimage* **172**, 437–449 (2018).
41. Guell, X., Schmahmann, J. Cerebellar functional anatomy: a didactic summary based on human fMRI evidence. *Cerebellum* **19**, 1–5 (2020).
42. Schmahmann, J. D. & Sherman, J. C. The cerebellar cognitive affective syndrome. *Brain: J. Neurol.* **121**, 561–579 (1998).
43. Baumann, O. *et al.* Consensus paper: the role of the cerebellum in perceptual processes. *Cerebellum* **14**, 197–220 (2015).
44. Adamaszek, M. *et al.* Consensus paper: cerebellum and emotion. *Cerebellum* **16**, 552–576 (2017).
45. Schmahmann, J. D. The cerebellum and cognition. *Neurosci. Lett.* **688**, 62–75. https://doi.org/10.1016/j.neulet.2018.07.005 (2019).
46. Booth, J. R., Wood, L., Lu, D., Houk, J. C. & Bitan, T. The role of the basal ganglia and cerebellum in language processing. *Brain Res.* **1133**, 136–144. https://doi.org/10.1016/j.brainres.2006.11.074 (2007).
47. Diedrichsen, J., Ivry, R. B., & Pressing, J. Cerebellar and basal ganglia contributions to interval timing. *Functional and neural mechanisms of interval timing*, 457–481 (2003).
48. Lungu, O. V. *et al.* Trial-to-trial adaptation: parsing out the roles of cerebellum and BG in predictive motor timing. *J. Cogn. Neurosci.* **28**, 920–934 (2016).
49. Frühholz, S., Trost, W. & Kotz, S. A. The sound of emotions—towards a unifying neural network perspective of affective sound processing. *Neurosci. Biobehav. Rev.* **68**, 96–110 (2016).
50. Zald, D. H. & Pardo, J. V. The neural correlates of aversive auditory stimulation. *Neuroimage* **16**, 746–753 (2002).
51. Annoni, J. M., Ptak, R., Caldara-Schnetzer, A. S., Khateb, A. & Pollermann, B. Z. Decoupling of autonomic and cognitive emotional reactions after cerebellar stroke. *Ann. Neurol.* **53**, 654–658 (2003).
52. Guell, X., Schmahmann, J. D., Gabrieli, J. D. & Ghosh, S. S. Functional gradients of the cerebellum. *Elife* **7**, e36652 (2018).
53. Stoodley, C. J. & Schmahmann, J. D. Evidence for topographic organization in the cerebellum of motor control versus cognitive and affective processing. *Cortex* **46**, 831–844 (2010).
54. Stoodley, C. J., Valera, E. M. & Schmahmann, J. D. Functional topography of the cerebellum for motor and cognitive tasks: an fMRI study. *Neuroimage* **59**, 1560–1570 (2012).
55. Kühn, A. *et al.* Activation of the subthalamic region during emotional processing in Parkinson disease. *Neurology* **65**, 707–713 (2005).
56. Mallet, L. *et al.* Stimulation of subterritories of the subthalamic nucleus reveals its role in the integration of the emotional and motor aspects of behavior. *Proc. Natl. Acad. Sci.* **104**, 10661–10666 (2007).
57. Péron, J. The role of the subthalamic nucleus in emotional processing. *Clin. Neurophysiol.* **127**, e39 (2016).
58. Sieger, T. *et al.* Distinct populations of neurons respond to emotional valence and arousal in the human subthalamic nucleus. *Proc. Natl. Acad. Sci.* **112**, 3116–3121 (2015).
59. Péron, J. *et al.* Recognition of emotional prosody is altered after subthalamic nucleus deep brain stimulation in Parkinson's disease. *Neuropsychologia* **48**, 1053–1062 (2010).

60. Péron, J. *et al.* Vocal emotion decoding in the subthalamic nucleus: an intracranial ERP study in Parkinson's disease. *Brain Lang.* **168**, 1–11 (2017).
61. Schmahmann, J. D. The cerebrocerebellar system: anatomic substrates of the cerebellar contribution to cognition and emotion. *Int. Rev. Psychiatry* **13**, 247–260 (2001).
62. Bestelmeyer, P. E., Kotz, S. A. & Belin, P. Effects of emotional valence and arousal on the voice perception network. *Soc. Cogn. Affect. Neurosci.* **12**, 1351–1358 (2017).
63. Grandjean, D. Brain mechanisms in emotional voice production and perception and early life interactions. In *Early Vocal Contact and Preterm Infant Brain Development*, 71–87 (Springer, Cham, 2017).
64. Aglieri, V., Chaminade, T., Takerkart, S. & Belin, P. Functional connectivity within the voice perception network and its behavioural relevance. *Neuroimage* **183**, 356–365 (2018).
65. Leitman, D. I. *et al.* " It's not what you say, but how you say it": a reciprocal temporo-frontal network for affective prosody. *Front. Hum. Neurosci.* **4**, 19 (2010).
66. Witteman, J., Van Heuven, V. J. & Schiller, N. O. Hearing feelings: a quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia* **50**, 2752–2763 (2012).
67. Johnstone, T., Van Reekum, C. M., Oakes, T. R. & Davidson, R. J. The voice of emotion: an FMRI study of neural responses to angry and happy vocal expressions. *Soc. Cogn. Affect. Neurosci.* **1**, 242–249 (2006).
68. Banziger, T. & Scherer, K. R. in *A blueprint for an affectively competent agent: Cross-fertilization between Emotion Psychology, Affective Neuroscience, and Affective Computing* (eds T. Banziger, K. Scherer, & E Roesch) (Oxford University Press, 2010).
69. Banse, R. & Scherer, K. R. Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* **70**, 614–636 (1996).
70. Grandjean, D., Banziger, T. & Scherer, K. R. Intonation as an interface between language and affect. *Prog. Brain Res.* **156**, 235–247 (2006).
71. Kirchner, W. K. Age differences in short-term retention of rapidly changing information. *J. Exp. Psychol.* **55**, 352 (1958).
72. Collins, D. L., Neelin, P., Peters, T. M. & Evans, A. C. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* **18**, 192–205 (1994).
73. Ashburner, J. A fast diffeomorphic image registration algorithm. *Neuroimage* **38**, 95–113 (2007).
74. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273–289 (2002).
75. Fonov, V. *et al.* Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage* **54**, 313–327 (2011).
76. Amunts, K. *et al.* BigBrain: an ultrahigh-resolution 3D human brain model. *Science* **340**, 1472–1475 (2013).
77. Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E. & Ramnani, N. A probabilistic MR atlas of the human cerebellum. *Neuroimage* **46**, 39–46 (2009).
78. Diedrichsen, J. *et al.* Imaging the deep cerebellar nuclei: a probabilistic atlas and normalization procedure. *Neuroimage* **54**, 1786–1794 (2011).
79. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, S208–S219 (2004).
80. Whitfield-Gabrieli, S. & Nieto-Castanon, A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect.* **2**, 125–141 (2012).
81. Reid, A. T. *et al.* Advancing functional connectivity research from association to causation. *Nat. Neurosci.* **22**(11), 1751–1760 (2019).

Acknowledgements

Author contributions

L.C. wrote and edited the manuscript, analysed the data and acquired a small part of the data. S.F. collected the data and analysed the data, edited a final version of the manuscript. J.Pierce helped improve the manuscript and edited the final version of the manuscript. D.G. and J.Péron designed the study, edited the manuscript and supervised the study. J.Péron additionally collected the data with S.F. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/ 10.1038/s41598-021-90222-6.

Correspondence and requests for materials should be addressed to L.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.