Talking in Fury: The Cortico-Subcortical Network Underlying Angry Vocalizations

Sascha Frühholz[1,2], Hannah S. Klaas[1], Sona Patel[2] and Didier Grandjean[1,2]

[1]Neuroscience of Emotion and Affective Dynamics Laboratory (NEAD), Department of Psychology, University of Geneva, Geneva, Switzerland and [2]Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

Sascha Frühholz and Hannah S. Klaas contributed equally to this study.

Address correspondence to Sascha Frühholz, University of Geneva, Swiss Center for Affective Sciences, 9 Chemin des Mines, CH-1202 Geneva, Switzerland. Email: sascha.fruehholz@unige.ch

Although the neural basis for the *perception* of vocal emotions has been described extensively, the neural basis for the *expression* of vocal emotions is almost unknown. Here, we asked participants both to repeat and to express high-arousing angry vocalizations to command (i.e., evoked expressions). First, repeated expressions elicited activity in the left middle superior temporal gyrus (STG), pointing to a short auditory memory trace for the repetition of vocal expressions. Evoked expressions activated the left hippocampus, suggesting the retrieval of long-term stored scripts. Secondly, angry compared with neutral expressions elicited activity in the inferior frontal cortex IFC and the dorsal basal ganglia (BG), specifically during evoked expressions. Angry expressions also activated the amygdala and anterior cingulate cortex (ACC), and the latter correlated with pupil size as an indicator of bodily arousal during emotional output behavior. Though uncorrelated, both ACC activity and pupil diameter were also increased during repetition trials indicating increased control demands during the more constraint production type of precisely repeating prosodic intonations. Finally, different acoustic measures of angry expressions were associated with activity in the left STG, bilateral inferior frontal gyrus, and dorsal BG.

Keywords: ACC, amygdala, basal ganglia, fMRI, vocal emotions

Introduction

Emotional prosody is crucial for signaling affective states during social interactions. An impairment in the expression of emotional prosody can have severe consequences in social contexts and for social and emotional development (Bell et al. 1990). Early evidence for the neural basis of emotional vocal expressions comes from patient studies. Brain lesions leading to affective aprosodia point to a dominant role of the right frontal cortex in emotional prosody production (Borod et al. 2002; Ross and Monnot 2008). These studies also point to an involvement of the basal ganglia (BG; Cohen et al. 1994).

The importance of the right frontal cortex and of the BG has also been supported by recent neuroimaging studies on emotional vocal productions. They reported activations in the inferior frontal gyrus (IFG; Aziz-Zadeh et al. 2010; Laukka et al. 2011; Pichon and Kell 2013) and several subregions of the BG (Laukka et al. 2011; Pichon and Kell 2013). In addition, activations were reported in the STG (Dogil et al. 2002; Aziz-Zadeh et al. 2010; Laukka et al. 2011; Pichon and Kell 2013). These findings of STG involvement have been interpreted as arising from phonological feedback processing (Dogil et al. 2002), as well as from the provision of articulatory maps (Aziz-Zadeh et al. 2010) in relation to acoustic features of emotional vocal output, especially in the right STG (Pichon and Kell

2013). The IFG is supposed to have a role in articulatory monitoring and modulation of vocal expressions (Dogil et al. 2002; Aziz-Zadeh et al. 2010; Laukka et al. 2011). The BG are assumed to be responsible for volitional control of vocally expressed affect (Laukka et al. 2011), especially in their dorsal part, while the ventral part seems to add the emotional component during the preparation of vocal expressions (Pichon and Kell 2013). These results together provide evidence for an extended cortico-subcortical network involved in the production of emotional prosody that partly overlaps with a neural model of mammalian vocalizations (Jurgens 2009; Hage 2010). Some inconsistencies remain in the previous findings, however, as well as some open questions, especially concerning the BG, amygdala, anterior cingulate cortex (ACC), STG, and IFG, and also concerning different types of vocal productions.

In this study, we addressed many of these open questions concerning these brain regions by investigating the neural basis of angry vocalizations. First, though the BG are an integral part of the mammalian vocalizations network (Jurgens 2009; Hage 2010), the importance of the BG for the expression of emotional prosody in human vocalizations is still under discussion (Ross and Monnot 2008). Laukka et al. (2011) claimed that vocal affect is regulated by the BG by showing that increasing BG activation was inversely related to levels of nervousness in the voice. Nervousness, however, is only one aspect of vocal affect among several other important vocal acoustic features. Pichon and Kell (2013) found an involvement of the BG, especially of the ventral parts, during an emotional induction phase prior to vocalizations, which showed a strong connectivity to the dorsal BG during the production phase of vocalizations that followed. This is indicative of a functional segregation in the BG related to emotional and sensorimotor processing in the ventral and dorsal BG, respectively (Yelnik 2008; Péron et al. 2013). However, Pichon and Kell (2013) were not able to properly validate the accuracy and validity of vocal production during the experimental phase. Thus, their results only indirectly provide evidence for a role of different BG subregions for the production of vocal emotions.

Hence, the results of previous studies provide only limited evidence for direct involvement of the BG in producing vocal affect. There are reasons, however, for a specific and important role of the BG during vocal output behavior and specifically for prosody production (Péron et al. 2013). It has been recently suggested that the BG have a specific role during propositional speech production, with particular involvement in the sequencing of speech units and their decoding (Riecker et al. 2002; Kotz and Schwartze 2010; Paulmann et al. 2011). This role has currently been proposed only for propositional, but not for emotional, speech. For emotional prosody and especially for

anger prosody (Banse and Scherer 1996; Patel et al. 2011), the sequencing of speech units is especially relevant (Péron et al. 2013). Emotional compared with neutral prosody can be described by a change in the timing of speech sequences indicated by the dynamics of acoustical features. Hence, we expected to find especially activations in the dorsal BG in our study according to the demands of dynamic speech patterning, because the dorsal BG seem to be strongly linked to the sensorimotor output components of vocal expressions (Pichon and Kell 2013).

Besides the BG, the amygdala is another important brain structure that is strongly involved in the processing of emotional stimuli (LeDoux 2012). This structure is particularly involved in the processing of high-arousal vocal emotions (Grandjean et al. 2005; Sander et al. 2005; Wiethoff et al. 2009; Frühholz and Grandjean 2013a). In addition, it is also important for emotional output behavior by regulating the autonomous nervous system (Coccaro et al. 2011; LeDoux 2012). The amygdala also regulates autonomic reactions that support motor execution, especially in emotional contexts (LeDoux 2000). Thus, it should also be involved in emotional output behavior, such as vocal emotions, but a strong link to the amygdala in a recent model of mammalian vocalizations is largely missing yet (Jurgens 2009; Hage 2010). One recent study in humans reported amygdala activity during emotional vocalizations, but only during an emotion induction phase and not during the production of vocal expressions (Pichon and Kell 2013). The relative lack of support for amygdala activations during the production of vocal expressions may have been due to the use of low-arousal emotions (such as neutral and sad) (Dogil et al. 2002; Aziz-Zadeh et al. 2010) or of less distinguished and vaguely defined emotions (such as nervousness) (Laukka et al. 2011).

Here, we expected to find amygdala activation during wrathful vocal expressions of anger, a vocal expression of negative valence and of high arousal. Being a phylogenetically old emotion that is negative in valence and high in arousal and power (Banse and Scherer 1996; Patel et al. 2011), vocal wrath or "hot anger" should be especially conducive to elicit activations in subcortical structures during its expression. In addition, "hot" specifically compared with "cold" anger and generally compared with other emotional vocalizations can be reliably analyzed for acoustic and voice quality features (Patel et al. 2011). Vocal anger usually involves a strong activation in the autonomous nervous system. A brain structure engaged in arousal and the generation of autonomous reactions during emotional states is the ACC. Together with the amygdala, this region is implicated in a system of emotional control and affective autonomic response generation (Critchley 2009). The ACC is also supposed to volitionally and motivationally control the initialization of primate vocalizations in general (Jurgens 2009; Hage 2010). Thus, along with activation in the amygdala during the expression of high-arousal and negative vocal expressions of anger, we expected activation in the ACC.

The final and crucial question we addressed here was whether different types of emotional prosody production, specifically repetition (i.e., imitation) and evoked production (see Fig. 1), activate different neural regions. This has not been studied yet using functional neuroimaging, but patient studies provide some evidence for a neuronal dissociation of these production types (Heilman et al. 2004; Ross and Monnot 2008). All the reported patients had lesions in right frontal areas, with medial frontal lobe lesions leading to stronger

impairments in evoked expressions of prosody (Heilman et al. 2004), while especially small focal lesions in the lateral frontal operculum (fOP) can lead to stronger impairments in repeating prosody compared with larger posterior fOP lesions leading to evoked production deficits (Ross and Monnot 2008). However, lesion studies were not able to precisely locate the 2 different production types, because there was much variation in the size and location of lesions in the right frontal areas (Ross and Monnot 2008) and in the brain regions additionally involved, such as the ACC (Heilman et al. 2004), the insula, or the BG (Ross and Monnot 2008). For the latter we especially might expect higher dorsal BG activity during the evoked production of prosody, since this mode more strongly requires self-generation of prosodic sequences and sensorimotor control (Pichon and Kell 2013).

Materials and Methods

Participants

Fifteen healthy, native French-speaking and right-handed volunteers (8 females, mean age 23.67 years, SD 3.50 years) participated in the experiment. All participants had normal or corrected-to-normal vision and normal hearing, and no history of psychiatric or neurologic incidents. Participants gave their informed and written consent for their participation in the experiment. The study was approved by the local ethics committee in accordance with ethical and data security guidelines of the University of Geneva. After a postevaluation of the stimulus recordings (see below), it was determined that <40% of the angry recordings of 2 participants were categorized as being angry and <40% of their neutral recordings were categorized as being neutral. Therefore, these 2 participants were excluded from analyses, resulting in 13 participants in the final sample (7 females, mean age 23.85 years, SD 3.69 years, age range 19–32 years).

Stimulus Material

During the main experiment, participants had to express neutral and angry prosody using 2-syllable, 5-letter pseudowords consisting of a C–V–C–V–C combination (C = consonant and V = vowel) as stimulus material. Pseudowords were chosen to avoid any semantic and emotional meaning that might influence the production of emotional intonations. Four different pseudowords ("belam," "lagod," "minad," and "namil") were selected, which were similar to pseudowords that are already incorporated in the Geneva Multimodal Emotion Portrayal corpus (Bänziger and Scherer 2010). The 4 pseudowords were chosen from a sample of different pseudowords (2-syllable pseudowords, voiced sounds, no fricatives) spoken by 2 male and 2 female actors in a neutral and angry tone before the experiment. Thirty-two pseudowords (2 male actors/2 female actors × 4 pseudowords × 2 emotions) were selected after a behavioral evaluation of the database by 12 participants (9 females, mean age 27.17 years, SD 4.39 years). All selected words spoken in an angry tone were significantly evaluated as being angry ($F_{4,44} = 65.099$, $P < 0.001$). All words spoken in a neutral tone were significantly rated as being neutral ($F_{4,44} = 148.731$, $P < 0.001$). Angry words were judged as higher in arousal than in neutral words ($F_{4,44} = 159.415$, $P < 0.001$). The selected stimuli were then normalized for the mean energy across all stimuli.

Task Procedure

Prior to the experiment, each participant was trained with a short version of the experiment using pseudowords that were not included in the main experiment. Participants were especially trained not to move their heads while speaking. To further reduce head movements during scanning, participants' heads were tightly fixed in the functional magnetic resonance imaging (fMRI) scanner.

The main experiment consisted of 4 experimental blocks (2 repetition blocks and 2 evoked production blocks), each consisting of 38
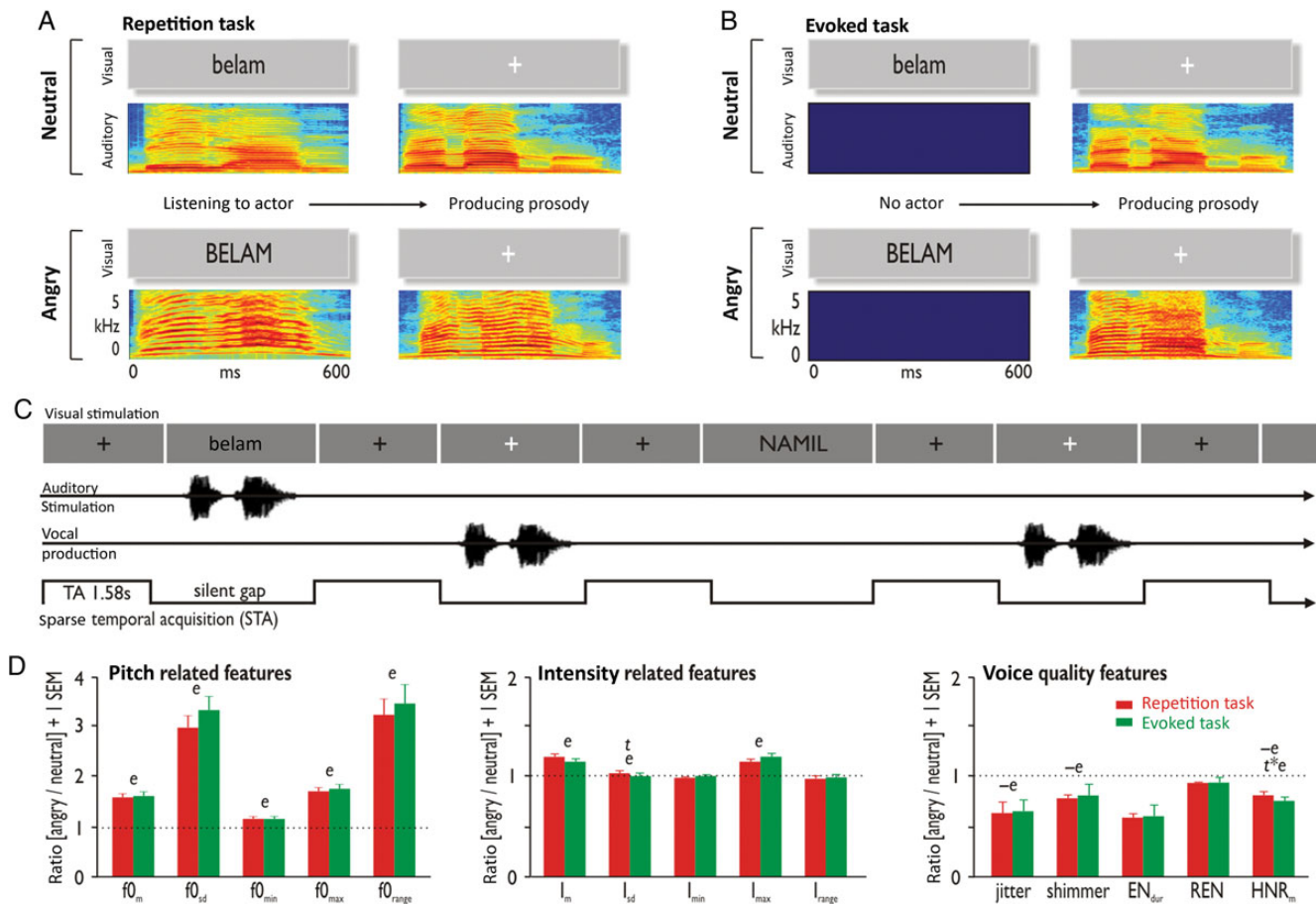
Figure 1. (*A*) During the repetition task, participants had to repeat and imitate the vocal production of neutral and angry prosody on 4 two-syllable pseudowords ("belam," "lagod," "minad," and "namil"), which were presented visually on a screen and which they had heard immediately beforehand when listening to recordings of actors ("listening to actor"). The production of neutral prosody was indicated by words written in lowercase letters (e.g., "belam") and the production of angry prosody was indicated by words written in uppercase letters (e.g., "BELAM"). The spectrograms show example recordings of the actors (left column in *A* and *B*) and of the participants (right column in *A* and *B*) for producing prosody on the pseudoword "belam." (*B*) During the evoked task, participants were asked to freely produce prosodic intonations without hearing the recording of an actor beforehand ("no actor"). This task included evoked production of prosody to the degree that participants could freely produce acted angry prosody without the constraint to imitate a certain prosodic style. (*C*) Shown are 2 example trials. The first trial depicts a repetition trial for the word "belam," which had to be spoken in neutral tone (indicated by visually presented lower case letters) as previously heard from listening to an actor recording (auditory stimulation). The second trial depicts an evoked trial for the word "namil" spoken in an angry tone (indicated by visually presented upper case letters). Both auditory stimulation and vocal recording occurred during the silent gap between image acquisitions. (*D*) Results of the acoustical feature analysis of prosody productions during the repetition task (red) and the evoked task (green), separately for pitch-related acoustic features (left panel), for intensity-related features (middle panel), and for other voice quality features (right panel). The bars represent the ratio between feature scores for angry trials divided by the same feature scores for neutral trials. The "e" indicates significant effects of the factor *emotion* (angry > neutral) for a 2 × 2 repeated-measures ANOVA, "-e" indicates the reverse effect (neutral > anger); "t" indicates a significant effect for the factor *task*, and "t*e" indicates a significant task * emotion interaction. Acoustical features: $f0_m$, mean of pitch; $f0_{sd}$, standard deviation of pitch; $f0_{min}$, minimum of pitch; $f0_{max}$, maximum of pitch; $f0_{range}$, range of pitch; I_m, mean of intensity; I_{sd}, standard deviation of intensity; I_{min}, minimum of normalized intensity; I_{max}, maximum of normalized intensity; I_{range}, range of intensity; EN_{dur}, relative energy normalized to total duration; REN, relative energy in 0–500 Hz band to 0–8 kHz band; HNR_m, mean of the harmonic-to-noise ratio.

trials. Repetition and evoked production blocks alternated across the experiment and the block sequence was counterbalanced across participants. The 38 trials of each block consisted of 32 trials that included the production of prosody and 6 null events without auditory stimulation and vocal productions. During null event trials, no stimulus would appear and participants were told to rest. The order of the trials was randomized for each participant.

In repetition blocks (Fig. 1*A*), participants were asked to repeat the prosodic intonations, which they had immediately heard spoken beforehand by the actors. In evoked production blocks (Fig. 1*B*), participants had to produce the prosody freely. This evoked production task included a freely acted production of prosody with no constraint of imitating or repeating a certain prosodic style. We have to note, however, that the evoked task did not represent the production of vocalizations resulting from really experiencing the underlying emotion or feeling of anger, but rather a relatively unconstraint production of

vocalizations on demand. In both the repetition and the evoked production blocks, the pseudoword was first presented on a gray screen for 800 ms starting 250 ms after the last volume acquisition (Fig. 1*C*). It was presented either in uppercase letters (indicating angry prosody production), or in lowercase letters (indicating neutral prosody production). For the repetition trials, the word was presented together with the voice of the actors. Afterwards, a black cross appeared on the screen during one volume acquisition (TA = 1580 ms, see below). After the volume acquisition, the black cross turned into a white cross. The white cross indicated that participants should produce the prosody asked for. The white cross stayed on the screen for 1580 ms, after which the cross turned black again and the next volume was acquired. Participants' prosody productions were recorded in the silent gap during volume acquisitions using an fMRI-compatible Sennheiser optical microphone (16 bits and 44.1 kHz) and a digital voice recorder.

Functional Voice Localizer Scanning

We used 8 s sound clips taken from an existing database (see http://vnl.psy.gla.ac.uk/) (Belin and Zatorre 2000) to determine human voice-sensitive regions in the bilateral superior temporal cortex. The sound clips consisted of 20 sequences of nonemotional human voices and 20 sequences of animal or environmental sounds. Each sound clip was presented once with a fixation cross on the screen and a 4-s gap between each clip. The scanning sequence also contained twenty 8 s silent events. Participants listened passively to the stimuli.

Mouth Movement Localizer Scanning

We were interested only in the activations related primarily to the production of emotional prosody, not to the movement of the vocal apparatus during speaking. Thus, to be able to exclude sensorimotor regions showing activations due to mouth movement only, we conducted a movement localizer scanning in the experiment. The movement localizer consisted of 8 movement blocks and 8 resting blocks. In each block, the same word appeared 10 times, alternating with a cross. The word and the cross each appeared for 1 s on the screen. In movement blocks, the color of the words and crosses was green, and participants were instructed to form the word with their lips as soon as it appeared on the screen. In resting blocks, words and crosses were red and participants were instructed not to move their lips and watch. Between each block, there was a 5-s gap indicated by a blank screen. The 4 stimulus words of the main experiment were used. Each word was used in 2 movement blocks and in 2 resting blocks.

Image Acquisition and Image Processing

All functional imaging data were recorded on a 3-T Siemens Trio System (Siemens, Erlangen, Germany) using a T_2*-weighted gradient echo-planar imaging sequence [time to repetition (TR) = 3290 ms, time of acquisition TA = 1580 ms, time to echo (TE) = 30 ms, flip angle FA = 90°, 28 slices, slice thickness 4 mm, distance factor = 20%, 64 matrix (3 × 3 mm)]. We used a sparse temporal acquisition protocol for the main experiment, which allowed presentation of auditory stimuli in the silent gap between volume acquisitions. It also allowed us to record the prosody productions of the participants (see below), which are unaffected by the background scanner noise. A high-resolution magnetization-prepared rapid acquisition gradient echo, T_1-weighted sequence [1 mm slices, TR = 1900 ms, TE = 2.27 ms, time to inversion (TI) = 900 ms, FoV 296 mm, in-plane 1 × 1 mm] was obtained in sagittal orientation to obtain structural brain images from each subject.

Images from the main experiment and from both localizer scans were preprocessed and analyzed using the Statistical Parametric Mapping software SPM8 (Welcome Department of Cognitive Neurology, London, UK). Functional images were realigned and coregistered to the anatomical image. During realignment we ensured that head motion in any spatial dimension of each participant was <1.5 mm, which is less than half of the voxel size used for image acquisition. Segmentation of the anatomical image revealed warping parameters that were used to normalize the functional images to the Montreal Neurological Institute (MNI) stereotactic template brain. Functional images were resampled to a 2-mm³ voxel size and spatially smoothed using an isotropic Gaussian kernel of 8 mm³ full-width at half-maximum.

Pupil Diameter Measurement and Analysis

We recorded the pupil diameter of each participant continuously throughout the main experiment by using an MRI-compatible long-range eye tracker system (EyeTrac 6, Applied Science Laboratories, USA) at a sampling rate of 60 Hz. Eye blinks in the pupil data were interpolated. The pupil diameter was supposed to be an indicator of the bodily arousal states (Partala and Surakka 2003) during the emotional vocalizations of the participants. For the cases in which blinks affected >20% of a trial, the entire trial was excluded from further statistical analyses. The average percentage of valid trials was 86.95% (SD = 7.24). For valid trials, the time course of the pupil diameter was extracted for a window of −1000– to 3000 ms, time locked to the appearance of the white fixation cross (the signal to the participants to produce prosody). The time courses were baseline corrected

according to the mean signal in the baseline period −1000 to 0 ms. The mean pupil diameter was scored in the time period 0–1580 ms. This was the silent gap interval during which participants were asked to produce prosody. The mean pupil size was determined separately for each experimental condition. Two participants had to be excluded from this analysis because of bad or missing pupil data due to acquisition problems during the experiment. The mean scores were subsequently subjected to a 2 × 2 repeated-measures analysis of variance (ANOVA) with the within-subject factors *task* (repeated and evoked) and *emotion* (neutral and anger). A statistical threshold of $P < 0.05$ was used for this analysis.

Statistical Analyses of Functional Data

Main Experiment

Each trial was modeled with a boxcar function defined by the onset of auditory stimulation and the onset of vocal productions, including the duration of each event. The boxcar function was convolved with a standard hemodynamic response function (HRF) on a single-subject level taking into account the temporal sparse acquisition pattern (see Kumar et al. 2007; Frühholz et al. 2012). For the main experiment, separate regressors were created for each of the 4 experimental conditions (2 tasks × 2 emotions) for correct trials as defined by the perceptual evaluation of the prosody productions (see below). Only trials were included in the first 4 regressors, which were reliably classified as neutral or angry prosody in the perceptual evaluation of the recorded prosody productions. A fifth regressor modeled all trials with unreliable classification. Six motion correction parameters were finally included as regressors of no interest to minimize false-positive activations that were due to task-correlated motion. Simple contrasts for each experimental condition for each participant were taken to a second-level random-effects ANOVA group analysis.

To obtain the differences in blood oxygen level-dependent responses, the following contrasts were computed for the group analysis using a single ANOVA (i.e., the flexible factorial design option in SPM) including all experimental conditions. To reveal the main effect of *task*, we compared repetition trials with evoked trials and vice versa. To reveal the main effect of *emotion*, we compared angry production trials with neutral production trials. The effect of angry compared with neutral trials was also computed separately for each the repetition and evoked task. Finally, we also computed interaction contrasts to find specific activation for angry trials during the repetition and during the evoked task. All contrasts were thresholded at $P < 0.001$ and a cluster extent of $k = 33$. This combined voxel and cluster threshold corresponds to $P < 0.05$ corrected at the cluster level and was determined by the 3DClustSim algorithm implemented in the AFNI software (http://afni.nimh.nih.gov/afni) using the estimated smoothness of the data across all contrasts computed. Across all contrasts, this procedure resulted in a maximum $k = 33$, and this was set as cluster threshold for all contrasts.

We extracted beta estimates in several ROIs, including the bilateral IFG (left IFG, right IFG, and right IFGor), the bilateral STG [left mSTG, left posterior STG (pSTG), and right pSTG], the bilateral amygdala, left hippocampus (HC), and bilateral BG (left putamen and right caudate nucleus), and the ACC. Peak activations for these ROIs were taken from the analysis of experimental main effects to directly investigate experimental interaction effects in these ROIs additionally to the interaction analyses described above. The ROIs were created in regions which were of primary interest in this study. Multiple ROIs were created in the IFG and the STG, because it was previously shown that subregions in these areas can have different functional roles during the processing and production of vocalizations (Frühholz and Grandjean 2013b, c). For these ROI analyses, we scored mean beta estimates in 3 mm radius spheres around peak activations. Since we did not find peak activations for the amygdala in the main analysis, we extracted mean beta estimates in an anatomical amygdala mask as defined by the AAL brain atlas (Tzourio-Mazoyer et al. 2002), because the amygdala was one of our main ROIs. Beta estimates were subjected to a 2 × 2 repeated-measures ANOVA with the within-subject factors *task* (repeated and evoked) and *emotion* (neutral and anger), especially to determine interaction effects in these ROIs between the experimental factors.

We additionally performed a correlation analysis on the extracted beta estimates. For each participant, we separately computed the mean values for each acoustical feature, the mean values of the beta estimates for each ROI, and the mean value for the pupil diameter. This was done separately for each of the 4 experimental conditions. Correlations were computed on difference scores between conditions for which we found functional activation according to the main effects of *emotion* and *task*. We computed correlations between each of the ROI difference scores and the difference scores of the acoustical features. The ROI difference scores and the difference scores of the acoustical features were also correlated with the pupil diameter difference scores.

Voice Localizer

Trials with vocal and nonvocal stimuli were modeled with a boxcar function aligned to the onset of each stimulus, 8 s in duration. The boxcar function was convolved with a standard HRF. We compared vocal against nonvocal animal and nonvocal environmental stimuli for each subject. These single-subject contrasts were taken to a group level. The group results were again thresholded at $P < 0.001$ and a cluster extent of $k = 33$. Using this contrast, we determined voice-sensitive regions along the STG and the STS in both hemispheres (Supplementary Fig. 1).

Movement Localizer

Blocks with mouth movement and resting blocks were modeled with a boxcar function aligned to the onset of each block, 20 s in duration. The boxcar function was convolved with a standard HRF. We compared movement blocks against resting blocks for each subject. These single-subject contrasts were taken to a group level. The group results were thresholded at $P < 0.001$ and a cluster extent of $k = 33$. Using this contrast, we determined regions in the primary motor cortex involved in mouth movements during speaking (Supplementary Fig. 1).

Vocal Production Analysis

An acoustic analysis was performed on the recorded voice samples of the participants to validate the emotionality of the samples and to determine whether there were differences between the evoked and repetition conditions. Rather than computing a large number of acoustic features, a selected group of parameters were examined in PRAAT (Boersma 2001). These included the duration, the mean and standard deviation of fundamental frequency ($f0$) and intensity, and the harmonics-to-noise ratio (HNR; the proportion of the periodic and aperiodic components that are present in the signal). These features were specifically chosen as a substantial amount of evidence has shown that anger can be differentiated from neutral prosody using these features. Specifically, the mean $f0$ and intensity, as well as the SD of $f0$ and intensity, were predicted to be greater for anger than for neutral prosody, and HNR was predicted to be smaller for anger than for neutral prosody (Banse and Scherer 1996; Patel et al. 2011).

Perceptual Evaluation of Prosody Production

We conducted a post-experimental perceptual evaluation of all vocalizations in order to include trials for functional analysis that were reliably perceived as neutral and angry expressions. The test included 48 participants (24 females, mean age 24.2 years, SD 4.2 years, age range 18–35 years). Each recording had to be judged on a continuous scale ranging from 100% neutral to "0" to 100% angry. Participants were asked to judge the emotion and its magnitude by using a visual-analog scale with 3 labeled anchors ("100% neutral," "100%" angry, and "0" as the midpoint). Judgments were made by moving a bar to any point on the scale (shown on a computer screen), indicating participants' impression of how angry or neutral the emotion was based on the speaker's prosody. Participants were told to choose "100% neutral" for an unambiguous perception of a neutral voice and to choose "100% angry" for the unambiguous perception of an angry voice. The "0" represented the midpoint of the scale for the case that the prosody could not reliably be identified either as angry or as neutral (i.e., in case of an ambiguous perception of the vocal prosody). Participants were also told that they can rate the voices on any point on

the scale indicating the level of unambiguity or perceptual clearness of the vocal expression according to a dimensional approach to emotional perception (Fontaine et al. 2007). For example, a rating of 75% on the scale between "0" and "100% angry" would indicate a less ambiguous angry vocalization compared with a rating of 25%. The same applied to the scale between "0" and "100% neutral."

From the perceptual evaluation results, we defined threshold values to classify angry and neutral prosody productions as *angry* or *neutral*. Therefore, for each production, the mean of the judgment values was computed. A bimodal distribution of mean values resulted in 2 peaks. The 2 peaks were computed by the median values for the neutral and the angry stimuli distributions. With reference to these 2 peak values ($Md_{neutral} = 55.75\%/SD = 18.10$ on the scale between "0" and "100% neutral"; $Md_{anger} = 48.77\%/SD = 23.61$ on the scale between "0" and "100% angry"), cutoffs were defined for the neutral and the angry stimuli productions separately by adding one standard deviation in the direction of the zero point (i.e., "0" as the midpoint of the scale). The neutral productions were included in the analyses above a threshold of 37.65% for neutral trials (i.e., trials with ratings between 37.65 and 100% on the scale toward the "100% neutral" anchor point), and only angry productions that lay over a threshold of 25.16% were included for anger trials (i.e., trials with ratings between 25.16 and 100% on the scale toward the "100% anger" anchor point). According to these cutoff values, it revealed that 84.32% of the angry prosody productions were judged as angry and 80.91% of the neutral prosody productions as neutral. Follow-up Wilcoxon tests for paired samples revealed no significant ($Z = 0.637$, $P = 0.524$) median differences of evaluations between the repeated ($Md_{rep} = 53.25$, SD = 22.78) and the evoked productions ($Md_{spo} = 52.33$, SD = 22.78).

Results of the perceptual evaluation indicated that 84.32% of the angry prosody productions were judged as angry and 80.91% of the neutral prosody productions as neutral, thus indicating a similar task-related effort to produce neutral and angry vocalizations. We identified "bad" or unreliable speakers as those for whom <40% of their angry recordings were judged as being angry or <40% of their neutral recordings were categorized as being neutral. Using these criteria, 2 participants were excluded from the functional analyses, resulting in a final sample of 13 participants (see the Materials and Methods section above). Furthermore, only trials that were identified as angry or neutral were included in the functional analysis. This resulted in the inclusion of 82.61% of the trials.

Results

Vocal Production Analysis

Separate 2×2 repeated-measures ANOVAs, including the within-subject factors *task* (repeated and evoked) and *emotion* (neutral and anger), were conducted on acoustical features (Fig. 1D and Supplementary Table 1) including a statistical threshold of $P < 0.05$. The duration of vocal productions did not differ across emotions ($F_{1,12} = 1.613$, $P = 0.228$) and across tasks ($F_{1,12} = 3.420$, $P = 0.089$). There were significant *emotion* effects for features related to fundamental frequency ($f0$) (all $F_{1,12} > 111.112$, all $P < 0.001$), such as mean $f0$, standard deviation of $f0$, minimum and maximum $f0$, and $f0$ range. These effects indicated higher values for angry than for neutral prosody productions.

Furthermore, significant *emotion* effects were revealed for intensity-related features (i.e., loudness) (all $F_{1,12} > 53.290$, all $P < 0.001$), such as mean intensity (I_m), minimum intensity (I_{min}), and standard deviation of intensity (I_{sd}), indicating higher values for angry than for neutral trials. The emotion effect for the maximum intensity (I_{max}) and the intensity range (I_{range}) did not reach significance (all $F_{1,12} < 2.593$, $P > 0.133$).

For some acoustical features, there was a reverse effect of *emotion* pointing to higher values for neutral than for angry

trials. The acoustical measures of jitter ($F_{1,12} = 15.236$, $P = 0.002$) and shimmer ($F_{1,12} = 20.672$, $P = 0.001$) revealed such an effect. Furthermore, also the HNR revealed higher values during neutral than during angry trials ($F_{1,12} = 37.036$, $P < 0.001$). For this feature, also the interaction between the factors *emotion* and *task* reached significance ($F_{1,12} = 6.972$, $P = 0.022$), indicating a higher mean HNR during the repetition of neutral prosody compared with that of angry prosody ($t_{12} = 5.159$, $P < 0.001$). There were no other significant interactions (all $F_{1,12} < 3.360$, all $P > 0.090$).

There was one *task* effect for I_{sd} indicating higher values for repetition than for evoked trials ($F_{1,12} = 4.893$, $P = 0.047$). All other task effects for the remaining acoustical features did not reach significance (all $F_{1,12} < 4.386$, all $P > 0.058$).

Functional Brain Data

Because we were only interested in functional activity which was specifically related to the production of angry prosody and not to general activation due to mouth movements, we only report functional activations here, which were located outside the primary sensorimotor regions as determined by the mouth movement localizer scan (Supplementary Fig. 1).

Repeated and Evoked Production of Prosody

For repeated compared with evoked prosody production, we found significant activation in the left mSTG (Fig. 2A; see Supplementary Table 2A and B). The mSTG activation was located within the temporal voice area (TVA), as defined by the voice localizer scan (Supplementary Fig. 1B). When comparing evoked with repetitive prosody productions, we found activity in the left HC (see Fig. 2B).

Functional Activations for Angry Compared with Neutral Trials

For angry trials, we found increased activations in the left and right STG, bilateral IFG, and left ACC, and in several subregions of the BG, such as the putamen and the caudate nucleus (Fig. 2D; Supplementary Table 2C). Activations extended from the left mSTG to the pSTG, and an activation peak was also found in the right mSTG. STG activation was again located within the TVA. A repeated-measures ANOVA on beta estimates including the factors *task* (repeated and evoked) and *emotion* (neutral and angry) indicated a significant main effect for the factor *task* ($F_{1,12} = 10.418$, $P = 0.007$) for the ROI within the right mSTG, indicating increased activity for repetition
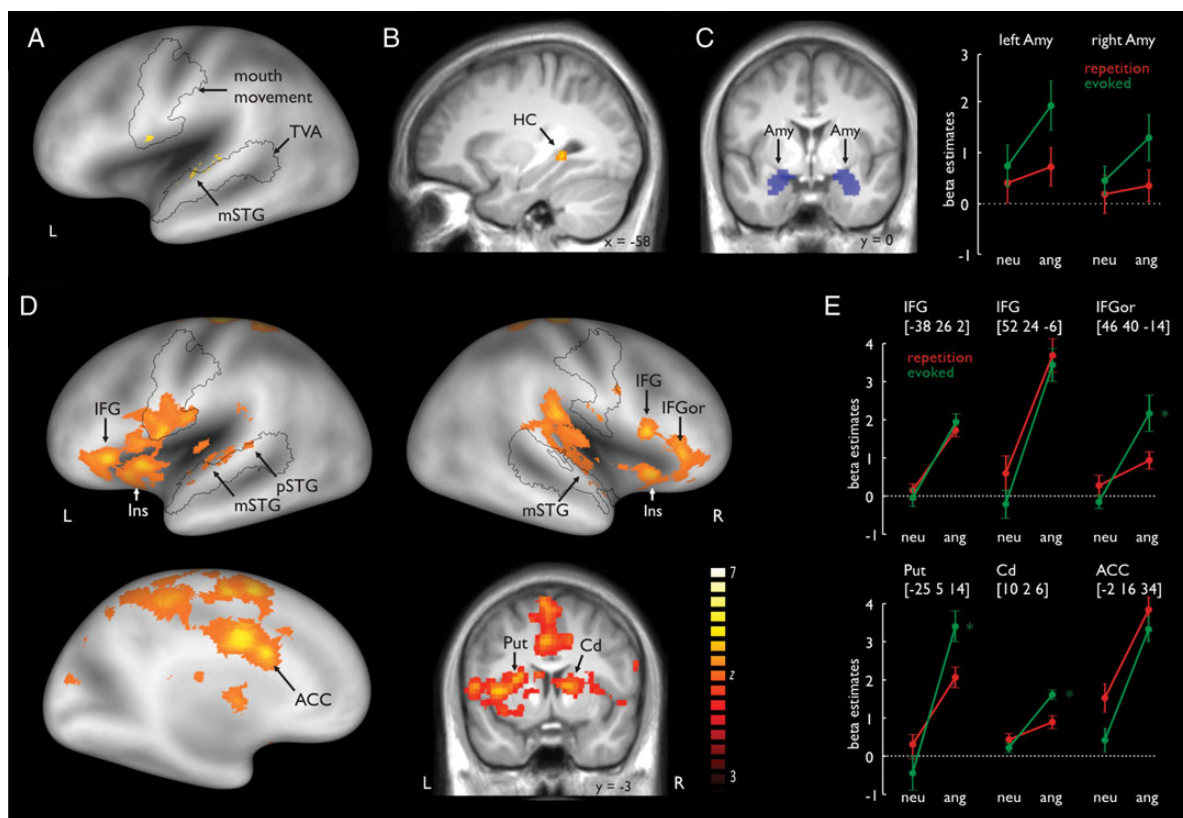

Figure 2. (A) Repeated compared with evoked production of prosody revealed activity in the left mSTG located within the TVA. The black outline represents the frontal and temporal areas related to silent mouth movement according to the functional mouth movement localizer scan and the voice-sensitive temporal cortex (see Supplementary material). (B) Evoked compared with the repeated production of prosody revealed activity in the left HC. (C) Mean activity in the bilateral amygdala was extracted by using a brain mask defined by the AAL atlas (blue inserts in the right panel) (Tzourio-Mazoyer et al. 2002). Functional activity in the left amygdala revealed a significant signal increase for angry (ang) compared with neutral (neu) voices, as well as for the evoked compared with the repetition condition. (D) Functional activations for the production of angry compared with neutral prosody revealed activity in the left mSTG and pSTG, right STG, ACC, bilateral BG (left putamen and right caudate nucleus), and bilateral IFG. Additional activity was also found in the right pars orbitalis of the IFG (IFGor). (E) Signal plots for the main regions of interest. Numbers in brackets represent peak MNI coordinates. Asterisks indicate a significant signal increase in the bilateral BG and IFGor for angry compared with neutral prosody, especially for the evoked condition. ACC, anterior cingulate cortex; Cd: caudate nucleus; HC: hippocampus; IFG: inferior frontal gyrus; IFGor: inferior frontal gyrus pars orbitalis; Ins: insula; MNI: Montreal Neurological Institute; mSTG: middle temporal gyrus; pSTG: posterior temporal gyrus; Put: putamen; TVA: temporal voice area.

compared with evoked trials. There was no significant interaction ($F_{1,12} = 0.087$, $P = 0.774$). The reverse contrasts of neutral compared with angry trials did not reveal any significant functional activations.

Furthermore, activations were found in the bilateral IFG for angry productions (Fig. 2D). One peak was located in the left IFG and 2 peaks were found in the right IFG. For the right IFG, one peak of activity was located in the pars orbitalis of the IFG (IFGor) and the other peak was located more posterior in the pars opercularis of the IFG. Only in the IFGor peak did we find a significant interaction between the factors *emotion* and *task* ($F_{1,12} = 12.437$, $P = 0.004$) (Fig. 2E). Here, the activation difference was especially pronounced during evoked angry compared with evoked neutral prosody production ($t_{12} = 4.381$, $P < 0.001$). No other post hoc tests revealed a significant difference between experimental conditions (all t's < 1.941, all P's > 0.076).

Regarding the left ACC (Fig. 2D), there was a significant main effect for emotion, indicating an increase of activity in this region when prosody was produced angrily compared with neutral productions. An ANOVA on ROI beta estimates in the ACC also revealed a significant main effect for *task* ($F_{1,12} = 9.938$, $P = 0.008$), indicating greater involvement of the ACC in the repetition compared with the evoked production. We also found activity in the right insula in angry compared with neutral production.

Furthermore, for angry productions, we found activity in the bilateral dorsal BG. There were activation peaks in the left putamen and the right caudate nucleus (Fig. 3C). According to the ROI analysis, this effect was especially pronounced for evoked angry prosody production in both the left putamen ($F_{1,12} = 26.538$, $P < 0.001$) and the right caudate nucleus ($F_{1,12} = 10.902$, $P = 0.006$), as indicated by significant interaction effects. For the left putamen, paired post hoc t-tests revealed a significant difference between angry and neutral productions for evoked prosody ($t_{12} = 8.664$, $P < 0.001$). For the right caudate nucleus, post hoc tests revealed a significant difference between angry and neutral productions for evoked prosody ($t_{12} = 6.926$, $P < 0.001$). No other post hoc test revealed a significant difference between experimental conditions (all t's < 1.465, all P's > 0.169).

Functional Activations for Angry Compared with Neutral Trials Within Each Task

For angry trials in the repetition task, we obtained activations in the bilateral IFG, bilateral insula, and middle and left dorsal ACC (Supplementary Table 3B). For angry trials in evoked tasks, there were activation peaks in the right IFG, the pars orbitalis of the IFG (IFGor), the right insula, the left middle cingulate gyrus, the right STG, the bilateral putamen, and the right caudate nucleus (Supplementary Table 3A). We furthermore computed interaction contrasts to find activity during the production of angry prosody that was specific to each task. No significant activation was revealed in this analysis.

Functional Amygdala Activity

Our whole-brain analysis did not reveal significant activity in the bilateral amygdala. However, since the amygdala was one of our primary ROIs, we performed a ROI analysis on mean beta estimates extracted from the bilateral amygdala masks (Fig. 2C). Repeated-measures ANOVAs on the beta estimates revealed a significant effect for the factor *emotion* for the left amygdala ($F_{1,12} = 4.899$, $P = 0.047$), indicating higher activation in the amygdala during angry than during neutral prosody production. There was also a significant effect for the factor *task* ($F_{1,12} = 2.957$, $P = 0.031$), indicating a greater involvement of the amygdala during evoked prosody production than during repetition of prosody. An interaction did not reach significance ($F_{1,12} = 2.290$, $P = 0.278$). No significant main effects or interactions were found in the right amygdala (all $F_{1,12} < 2.947$, all $P > 0.112$).

Correlations of Functional Activations with Features of Vocal Productions

To find out which functional activations in our primary ROIs are related to specific acoustic features of vocal productions, we computed correlations between functional activations in our ROIs and vocal features of the emotional prosody productions. Significant correlations are reported in Table 1, all other correlations were not significant (all r's < |0.533|, all P's > 0.061). Correlations were computed on difference scores that resulted from comparing the different experimental conditions. During the repetition compared with the evoked
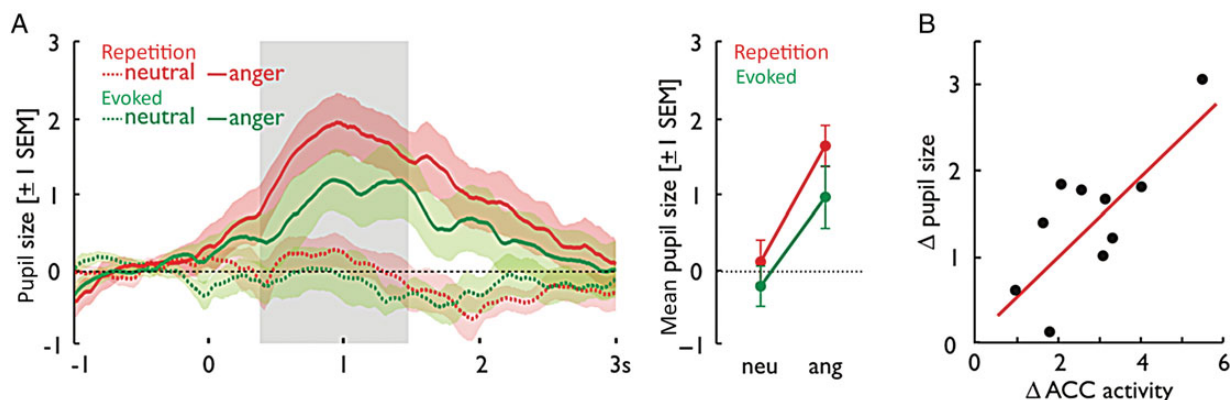

Figure 3. (A) Grand average time course (left panel) of the pupil size time locked to the onset of the silent gap (time "0") during the sparse temporal acquisition sequence. Mean pupil (right panel) size was scored in a time window of 500–1500 ms after the onset of the silent scanning gap (gray underlay in the right panel). (B) Significant positive signal relationship between the pupil size difference, comparing angry with neutral prosody productions, and the respective signal in the anterior cingulate cortex (ACC).

Condition	Brain region	Acoustical feature	*r*	*P*-values
(A) Repetition trials	Left mSTG	$f0_{max}$	0.606	0.028
		$f0_{sd}$	0.712	0.006
		$f0_{range}$	0.594	0.032
		I_{min}	0.567	0.006
	Left IFG	I_m	0.586	0.035
	Right IFG	I_m	0.783	0.002
(B) Evoked trials	Left pSTG	$f0_{sd}$	0.663	0.013
	Right Ins	I_{sd}	0.600	0.030
(C) Evoked angry trials	Left Put	HNR_m	0.562	0.046

(**A**) Repetition compared with evoked trials. (**B**) Evoked compared with repetition trials. (**C**) Angry productions during the evoked trials. For abbreviations, see Figure 1.

production of prosody, an increase in the maximum *f*0, standard deviation of *f*0, and *f*0 range was associated with an increase of activity in the left mSTG. Similarly, activation in the left pSTG was associated with an increase in the standard deviation of *f*0 for repetition trials. The bilateral IFG correlated positively with mean intensity for repetition trials. During evoked prosody production trials, the minimum intensity level was correlated with activity in the left mSTG. The standard deviation of intensity correlated positively with activity in the right insula. The mean HNR was positively associated with activation in the left putamen during evoked angry prosody production.

Pupil Diameter

Besides the functional brain activity, we also recorded the pupil diameter of participants as an indicator of the bodily arousal (Fig. 3). The ANOVA for the mean pupil diameter data revealed a main effect both for the factor *task* ($F_{1,10} = 7.641$, $P = 0.020$) and for the factor *emotion* ($F_{1,10} = 32.647$, $P < 0.001$), indicating increased diameter (i.e., arousal) for repetition ($M = 0.88$, SD = 0.26) compared with evoked trials ($M = 0.414$, SD = 0.414), as well as for angry ($M = 1.33$, SD = 0.36) compared with neutral productions ($M = -0.03$, SD = 0.19), respectively. There was no interaction between the factors ($F_{1,10} = 2.617$, $P = 0.137$). A positive correlation was found between pupil diameter and activity in the left dorsal ACC ($r = 0.706$, $P = 0.015$) for the comparison of angry versus neutral trials, but this effect might have been driven by one participant showing considerably higher values on both measures. Excluding this participant from the analysis resulted in a nonsignificant correlation ($r = 0.385$, $P = 0.271$), indicating that the original effect might have to be taken with some caution. However, the Kolmogorov–Smirnoff test including all participants revealed that both variables (all *P*'s > 0.132) were normally distributed. Furthermore, the data of the specific participant were close the commonly defined threshold of data outliers (i.e., 1.5 times the interquartile range), thus indicating some validity of the original data including all participants. No correlation was found between the pupil diameter and the ACC activity for the comparison of repetition with evoked trials ($r = 0.106$, $P = 0.756$).

Discussion

This study aimed to determine the neural network and the functional role of several cortical and subcortical brain regions involved in the production of wrathful vocalizations of anger. We revealed 3 main findings. First, besides activity in the STG

and IFG, we found activity in several parts of the dorsal BG, the dorsal ACC, and the amygdala in response to angry compared with neutral prosody productions. Secondly, we found that the repetition and evoked production of prosody relied on different brain networks supporting previous notions from patient studies about a neural separation for these different types of vocalizations. Specifically, the evoked production of hot anger revealed an extended and specific cortical and subcortical brain network. Finally, we found that many of the vocal features of prosody production were directly associated with activity in specific brain regions, indicating that several acoustic features of emotional vocalizations are directly controlled by specific brain regions.

Producing Angry Prosody

We expected to find a widespread cortico-subcortical network, which we supposed to be directly and functionally involved in the production of angry prosody, consisting of subcortical regions, such as the BG and the amygdala, but also of cortical regions, such as the STG, the ACC, and the IFG.

We found bilateral IFG activity for the production of angry prosody, including several IFG subregions in the right hemisphere. Left hemispheric IFG activation could be attributed to different articulatory plans for vocal tract coordination while producing angry prosody (Ackermann and Riecker 2010). The right IFG might be associated with monitoring and regulation functions in emotional expressive behavior (Phillips et al. 2008) and aggressive behavior (Potegal 2012). In our study, the right IFG, in close interaction with the phonological feedback processing in the STG (as discussed in the next section), could serve higher-order controlling functions in the expression of emotional prosody. Whereas a more posterior activation in the right IFG during angry prosody production was similarly found for both production tasks, activation in the anterior IFGor was more specifically enhanced when angry prosody was produced during the evoked condition. This might resemble a general posterior-to-anterior organization within the IFG, from first-order monitoring in the posterior IFG to complex monitoring functions in the anterior IFG (Petrides 2005). The evoked production of prosody might involve 2 levels: First creating a production script for the expression of prosody and then monitoring the production process. Thus, regulation and coordination demands were more complex during evoked prosody production, pointing to a more anterior activation in the IFGor. This result is also in accordance with a distinction within the orbitofrontal cortex, in which the processing of more complex stimuli is located more anteriorly (Kringelbach 2004).

Along with greater activation in the IFGor, activation of the dorsal BG (i.e., putamen and the caudate nucleus) was found for angry prosody production, which was again more pronounced in the evoked task. Besides a general role of the BG in motor planning and (emotional) output behavior (Yelnik 2008; Péron et al. 2013), the dorsal BG might have a specific role in the dynamic and temporal sequencing of speech in terms of sensorimotor control (Kotz and Schwartze 2010; Pichon and Kell 2013). Although this function can be seen as a common feature of both linguistic and affective prosody, temporal dynamics are essential features of angry prosody, indicated by an increased mean and variation in *f*0 and intensity (Banse and Scherer 1996; Patel et al. 2011). Both *f*0 and

intensity dynamics determine the temporal and rhythmic unfolding of emotional prosody, and this rhythmic aspect during vocal productions can be impaired in individuals with BG lesions (Péron et al. 2010). In our study, the repetition task involved actor recordings as an external cue, thereby minimizing the need to self-generate the temporal dynamics for prosody production. In contrast, the dorsal BG showed stronger activity in the evoked task for producing angry prosody, which may reflect stronger self-generated prosody dynamics.

While the BG might be associated with the temporal and dynamic sequencing of emotional speech output, the ACC might have 2 different functions during emotional vocalizations. First, the emotion effect that we observed in the ACC might point to its function of regulating the arousal level and thus in the control of the autonomous nervous system for emotional output behavior (Critchley et al. 2003). Activity in the ACC was positively correlated with the individual's pupil diameter as a physiological measure of bodily arousal (Partala and Surakka 2003) during the production of angry compared with neutral prosody. This result is further supported by our finding of activation in the right insula together with the ACC during angry prosody production, since the insula is similarly associated with the generation of autonomic responses (Ullsperger et al. 2010). Thus, both the ACC and the insula seem to regulate autonomic arousal when people speak in an aggressive angry tone.

Though uncorrelated, the increased activity of the ACC and the increased pupil size for repetition compared with evoked trials together might point to the second function of ACC serving increased performance monitoring during the more demanding repetition/imitation of emotional vocalizations. The pupil size has been shown to be an indicator of cognitive load especially during language-related tasks (Hyona et al. 1995), and the ACC seems a central brain area serving performance and error monitoring during increased demands of cognitive control (Kerns et al. 2004). The more constraint production type of exactly repeating prosodic intonation should have involved increased cognitive load and performance and error monitoring demands. Thus, beyond the functions of the ACC for volitional initiation of vocalizations (Jurgens 2009; Hage 2010), our data seem to suggest that the ACC is also involved regulating the arousal level as well as monitoring the vocal performance depending on the production type.

The present study also found activation in the amygdala during the production of angry prosody and provides strong evidence to extend the general vocalization network (Jurgens 2009; Hage 2010) by the limbic brain system, which plays an important role during emotional vocalizations. Our results emphasize the importance of the amygdala underlying emotional output behavior, but contradict the prevailing view of the amygdala as a structure mainly involved in the detection of stimuli and conditioning (Cardinal et al. 2002). One important role of the amygdala is its involvement in the expression of emotionally relevant behavior. The amygdala is known to be involved in the expression and regulation of angry behavior, probably in the experience of an aggressive impulse (Coccaro et al. 2011). The amygdala was also more activated in our study during evoked productions, indicating that evoked prosody might be associated with stronger emotional regulatory effects by the amygdala. This interpretation is supported by recent findings of amygdala involvement during emotional

prosody preparation (Pichon and Kell 2013). One alternative interpretation might be that, instead of regulating emotional output behavior, amygdala activity might also originate from auditory feedback processing of own emotional vocalizations. While this might partly explain higher amygdala activity during angry compared with neutral vocalizations, evoked compared with repetition trials also revealed stronger amygdala activity, and both are balanced in terms of expressing neutral and angry vocalizations. A considerable proportion of the amygdala activity thus might be modulated by the vocal production type. In the present study, evoked trials were less constraint than repetition trials in terms of how to vocalize. This provides some evidence for its differential regulatory role underlying emotional and especially angry vocalizations (Coccaro et al. 2011) under conditions of less constraint production modes, which might also have been a more stressful and thus more "emotional" production mode. This is also supported by the important regulatory role of the amygdala in pathological expressions of vocal emotions (Lauterbach et al. 2013).

The final brain structure found for producing angry prosody was the STG. During the production of prosody, specific prosodic speech elements have to be derived and translated into autonomous motor output. Evidence suggests that the STG supports sensory-motor integration (Hickok 2009; Peschke et al. 2009), the gating of amygdala connections (Pehrs et al. 2013), and phonological feedback processing (Zheng et al. 2010), which are necessary for accurate production of angry prosody. This is supported by our finding that left STG activity correlated with pitch features of vocal productions, supporting the online adjustments of vocal output behavior by auditory feedback processes (Aziz-Zadeh et al. 2010). The STG might also serve as a phonological short-term store facilitating prosody production, especially in cases when there is a delay between the perception and production of prosody, such as during the repetition task.

Repeated and Evoked Vocal Expressions

Besides our first experimental questions about the brain network underlying angry vocalizations, the second questions concerned the differential modulation of brain activity depending on the type of vocal productions. We thus compared brain activity during prosody production for the repetition and the evoked production of prosody. The repetition task elicited activity in the left mSTG and pSTG. The STG seems to function as a phonological store during the repetition of prosody, which is corroborated by recent findings for a short-term storage system in the STG (Ravizza et al. 2010; Acheson et al. 2011) and the closely located planum temporale (Hickok et al. 2009). The planum temporale was found in a recent study during the nonaffective repetition of pseudowords (McGettigan et al. 2011), and the same region has been proposed for sensory-motor integration that maps perceived speech directly to motor speech output. During the repetition task in the present study, the actors' intonations had to be perceived and stored for subsequent repetition. Hence, it is likely that left STG activation is due to phonological short-term storage of the pseudoword before its production.

The evoked task elicited activity in the left HC. The HC is generally involved in long-term memory functions, as well as emotional regulation functions (Fanselow and Dong 2010).

Our finding of HC activity might indicate the retrieval of long-term stored production rules during evoked prosody production for speech. During the evoked task, participants had to produce prosody naturally without relying on a prosody template that was immediately heard beforehand. For this evoked production, the participants had to retrieve a prototype script from long-term memory following the production rules for prosody intonations. This might be also explained by the fact that the evoked task did not involve vocal productions triggered by underlying emotional states, but rather required relatively unconstraint vocal productions on demand, which makes the use of prototype scripts more likely. Another explanation for the HC activation might be that the pseudowords were learned by being processed and constantly repeated (Paulesu et al. 2009). Yet, as the same words were used in both the repetition and evoked production conditions, learning should have occurred in both conditions and not only in the evoked production of prosody. This, however, is not the case since there was no HC activation during the repeated production of prosody.

Linking Brain Activity in the STG, BG, and IFG with Specific Voice Features

Our final experimental question concerned the involvement of specific brain areas, which underlie the production of specific acoustic features during angry prosody. We accordingly observed that some of the neural activations mentioned earlier were associated with specific acoustical features of the vocal expressions. Activation within the STG was positively associated with pitch-related features, such as the maximum, variation, and range of the $f0$. As the STG is responsible for sensory-motor integration (Hickok 2009; Peschke et al. 2009) and phonological feedback processing (Zheng et al. 2010), this association might indicate regulatory effort and online adjustments for accurate $f0$ production in angry prosody. Thus, emotional vocalizations strongly depend on auditory feedback and perceptual processing of own-vocalizations in the auditory cortex. Our results strongly suggest to extend recent models of mammalian vocalizations (Jurgens 2009; Hage 2010) by additionally including auditory-motor loops as an integral part of vocal expressions. Activation in the IFG was positively associated with the intensity of vocal productions, potentially due to the monitoring and regulation functions of the IFG. Intensity was higher in angry than in neutral trials. The production of angry prosody probably required to a greater extent the regulation and monitoring of intensity, resulting in an association of intensity and IFG activation. Finally, dorsal BG activations were positively correlated with the HNR of the produced speech, confirming results from BG lesion studies (Van Lancker Sidtis et al. 2010).

Conclusions

The production of high-arousing emotional prosody comprises a cortical–subcortical network encompassing not only the bilateral IFG and STG, but also the bilateral dorsal BG, the left amygdala, and the ACC. Several of these structures directly regulate the bodily arousal or activation level as well as the acoustic properties underlying emotional vocalization. The results implicate both left and right hemispheric structures in the production of emotional prosody, which contrasts with a prominent model stating that affective prosody is processed dominantly by the right hemisphere and that its organization is analogous to propositional prosody in the left hemisphere (Ross and Monnot 2008). The results about the central role of the BG might inspire research on the neural basis of impairments in the production of emotional prosody such as they occur in individuals with Parkinson disease (Péron et al. 2010, 2013). The results might also inspire future research on the neural basis of angry vocalizations that are based on real experiences of emotions or feelings instead of vocalizations produced on command.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxford-journals.org/.

Funding

Notes

Conflict of Interest: None declared.

References

Acheson DJ, Hamidi M, Binder JR, Postle BR. 2011. A common neural substrate for language production and verbal working memory. J Cogn Neurosci. 23:1358–1367.

Ackermann H, Riecker A. 2010. The contribution(s) of the insula to speech production: a review of the clinical and functional imaging literature. Brain Struct Funct. 214:419–433.

Aziz-Zadeh L, Sheng T, Gheytanchi A. 2010. Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. PLoS ONE. 5. e8759.

Banse R, Scherer KR. 1996. Acoustic profiles in vocal emotion expression. J Pers Soc Psychol. 70:614–636.

Bänziger T, Scherer KR. 2010. Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) Corpus. In: Bänziger T, Scherer KR, Roesch EB, editors. Blueprint for affective computing: a sourcebook Oxford. UK: Oxford University Press. p. 271–294.

Belin P, Zatorre RJ. 2000. Voice-selective areas in human auditory cortex. Nature. 403:309.

Bell WL, Davis DL, Morgan-Fisher A, Ross ED. 1990. Acquired aprosodia in children. J Child Neurol. 5:19.

Boersma P. 2001. Praat, a system for doing phonetics by computer. Glot Int. 5:341–345.

Borod JC, Bloom RL, Brickman AM, Nakhutina L, Curko EA. 2002. Emotional processing deficits in individuals with unilateral brain damage. Appl Neuropsychol. 9:23–36.

Cardinal RN, Parkinson JA, Hall J, Everitt BJ. 2002. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci Biobehav Rev. 26:321–352.

Coccaro EF, Sripada CS, Yanowitch RN, Phan KL. 2011. Corticolimbic function in impulsive aggressive behavior. Biol Psychiatry. 69:1153–1159.

Cohen MJ, Riccio CA, Flannery AM. 1994. Expressive aprosodia following stroke to the right basal ganglia: a case report. Neuropsychology. 8:242–245.

Critchley HD. 2009. Psychophysiology of neural, cognitive and affective integration: fMRI and autonomic indicants. Int J Psychophysiol. 73:88–94.

Critchley HD, Mathias CJ, Josephs O, O'Doherty J, Zanini S, Dewar BK, Cipolotti L, Shallice T, Dolan RJ. 2003. Human cingulate cortex and

autonomic control: converging neuroimaging and clinical evidence. Brain. 126:2139–2152.

Dogil G, Ackermann H, Grodd W, Haider H, Kamp H, Mayer J, Riecker A, Wildgruber D. 2002. The speaking brain: a tutorial introduction to fMRI experiments in the production of speech, prosody and syntax. J Neurolinguistics. 15:59–90.

Fanselow MS, Dong H-W. 2010. Are the dorsal and ventral hippocampus functionally distinct structures? Neuron. 65:7–19.

Fontaine JR, Scherer KR, Roesch EB, Ellsworth PC. 2007. The world of emotions is not two-dimensional. Psychol Sci. 18:1050–1057.

Frühholz S, Ceravolo L, Grandjean D. 2012. Specific brain networks during explicit and implicit decoding of emotional prosody. Cereb Cortex. 22:1107–1117.

Frühholz S, Grandjean D. 2013a. Amygdala subregions differentially respond and rapidly adapt to threatening voices. Cortex. 49:1394–1403.

Frühholz S, Grandjean D. 2013b. Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: a quantitative meta-analysis. Neurosci Biobehav Rev. 37:24–35.

Frühholz S, Grandjean D. 2013c. Processing of emotional vocalizations in bilateral inferior frontal cortex. Neurosci Biobehav Rev. 37:2847–2855.

Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P. 2005. The voices of wrath: brain responses to angry prosody in meaningless speech. Nat Neurosci. 8:145–146.

Hage SR. 2010. Neuronal networks involved in the generation of vocalization. In: Brudzynski SM, editor. Handbook of mammalian vocalization—an integrative neuroscience approach. Oxford: Academic Press. p. 339–349.

Heilman KM, Leon SA, Rosenbek JC. 2004. Affective aprosodia from a medial frontal stroke. Brain Lang. 89:411–416.

Hickok G. 2009. The functional neuroanatomy of language. Phys Life Rev. 6:121–143.

Hickok G, Okada K, Serences JT. 2009. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. J Neurophysiol. 101:2725–2732.

Hyona J, Tommola J, Alaja AM. 1995. Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. Q J Exp Psychol A. 48:598–612.

Jurgens U. 2009. The neural control of vocalization in mammals: a review. J Voice. 23:1–10.

Kerns JG, Cohen JD, MacDonald AW 3rd, Cho RY, Stenger VA, Carter CS. 2004. Anterior cingulate conflict monitoring and adjustments in control. Science. 303:1023–1026.

Kotz SA, Schwartze M. 2010. Cortical speech processing unplugged: a timely subcortico-cortical framework. Trends Cogn Sci. 14:392–399.

Kringelbach M. 2004. The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. Prog Neurobiol. 72:341–372.

Kumar S, Stephan KE, Warren JD, Friston KJ, Griffiths TD. 2007. Hierarchical processing of auditory objects in humans. PLoS Comput Biol. 3:e100.

Laukka P, Åhs F, Furmark T, Fredrikson M. 2011. Neurofunctional correlates of expressed vocal affect in social phobia. Cogn Affect Behav Neurosci. 11:413–425.

Lauterbach EC, Cummings JL, Kuppuswamy PS. 2013. Toward a more precise, clinically—informed pathophysiology of pathological laughing and crying. Neurosci Biobehav Rev. 37:1893–1916.

LeDoux J. 2012. Rethinking the emotional brain. Neuron. 73:653–676.

LeDoux JE. 2000. Emotion circuits in the brain. Annu Rev Neurosci. 23:155–184.

McGettigan C, Warren JE, Eisner F, Marshall CR, Shanmugalingam P, Scott SK. 2011. Neural correlates of sublexical processing in phonological working memory. J Cogn Neurosci. 23:961–977.

Partala T, Surakka V. 2003. Pupil size variation as an indication of affective processing. Int J Hum Comput Studies. 59:185–198.

Patel S, Scherer KR, Bjorkner E, Sundberg J. 2011. Mapping emotions into acoustic space: the role of voice production. Biol Psychol. 87:93–98.

Paulesu E, Vallar G, Berlingeri M, Signorini M, Vitali P, Burani C, Perani D, Fazio F. 2009. Supercalifragilisticexpialidocious: how the brain learns words never heard before. Neuroimage. 45:1368–1377.

Paulmann S, Ott DVM, Kotz SA. 2011. Emotional speech perception unfolding in time: the role of the basal ganglia. PLoS ONE. 6: e17694.

Pehrs C, Deserno L, Bakels JH, Schlochtermeier LH, Kappelhoff H, Jacobs AM, Fritz TH, Koelsch S, Kuchinke L. 2013. How music alters a kiss: superior temporal gyrus controls fusiform-amygdalar effective connectivity. Soc Cogn Affect Neurosci. doi: 10.1093/scan/nst169.

Péron J, Frühholz S, Verin M, Grandjean D. 2013. Subthalamic nucleus: a key structure for emotional component synchronization in humans. Neurosci Biobehav Rev. 37:358–373.

Péron J, Grandjean D, Le Jeune F, Sauleau P, Haegelen C, Drapier D, Rouaud T, Drapier S, Verin M. 2010. Recognition of emotional prosody is altered after subthalamic nucleus deep brain stimulation in Parkinson's disease. Neuropsychologia. 48:1053–1062.

Peschke C, Ziegler W, Kappes J, Baumgaertner A. 2009. Auditory–motor integration during fast repetition: the neuronal correlates of shadowing. Neuroimage. 47:392–402.

Petrides M. 2005. Lateral prefrontal cortex: architectonic and functional organization. Philos Trans R Soc Lond B Biol Sci. 360:781–795.

Phillips ML, Ladouceur CD, Drevets WC. 2008. A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. Mol Psychiatry. 13:833–857.

Pichon S, Kell CA. 2013. Affective and sensorimotor components of emotional prosody generation. J Neurosci. 33:1640–1650.

Potegal M. 2012. Temporal and frontal lobe initiation and regulation of the top-down escalation of anger and aggression. Behav Brain Res. 231:386–395.

Ravizza SM, Hazeltine E, Ruiz S, Zhu DC. 2010. Left TPJ activity in verbal working memory: implications for storage-and sensory-specific models of short term memory. Neuroimage. 55:1836–1846.

Riecker A, Wildgruber D, Dogil G, Grodd W, Ackermann H. 2002. Hemispheric lateralization effects of rhythm implementation during syllable repetitions: an fMRI study. NeuroImage. 16:169–176.

Ross ED, Monnot M. 2008. Neurology of affective prosody and its functional-anatomic organization in right hemisphere. Brain Lang. 104:51–74.

Sander D, Grandjean D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P. 2005. Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. Neuroimage. 28:848–858.

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. 15:273–289.

Ullsperger M, Harsay HA, Wessel JR, Ridderinkhof KR. 2010. Conscious perception of errors and its relation to the anterior insula. Brain Struct Funct. 214:629–643.

Van Lancker Sidtis D, Rogers T, Godier V, Tagliati M, Sidtis JJ. 2010. Voice and fluency changes as a function of speech task and deep brain stimulation. J Speech Lang Hear Res. 53:1167.

Wiethoff S, Wildgruber D, Grodd W, Ethofer T. 2009. Response and habituation of the amygdala during processing of emotional prosody. Neuroreport. 20:1356–1360.

Yelnik J. 2008. Modeling the organization of the basal ganglia. Rev Neurol (Paris). 164:969–976.

Zheng ZZ, Munhall KG, Johnsrude IS. 2010. Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. J Cogn Neurosci. 22:1770–1781.