Vocal emotion decoding in the subthalamic nucleus: An intracranial ERP study in Parkinson's disease

PERON, Julie Anne, et al.

Abstract

Using intracranial local field potential (LFP) recordings in patients with Parkinson's disease (PD) undergoing deep brain stimulation (DBS), we explored the electrophysiological activity of the subthalamic nucleus (STN) in response to emotional stimuli in the auditory modality. Previous studies focused on the influence of visual stimuli. To this end, we recorded LFPs within the STN in response to angry, happy, and neutral prosodies in 13 patients with PD who had just undergone implantation of DBS electrodes. We observed specific modulation of the right STN in response to anger and happiness, as opposed to neutral prosody, occurring at around 200–300 ms post-onset, and later at around 850–950 ms post-onset for anger and at around 3250–3350 ms post-onset for happiness. Taken together with previous reports of modulated STN activity in response to emotional visual stimuli, the present results appear to confirm that the STN is involved in emotion processing irrespective of stimulus valence and sensory modality.

Reference


DOI : 10.1016/j.bandl.2016.12.003

Available at:
http://archive-ouverte.unige.ch/unige:95989

Disclaimer: layout of this document may differ from the published version.
Vocal emotion decoding in the subthalamic nucleus: An intracranial ERP study in Parkinson’s disease

Julie Péron a,b,*, Olivier Renaud c, Claire Haegelen d,e, Lucas Tamarit b, Valérie Milesi a,b, Jean-François Houvenaghel f,g, Thibaut Dondaine f,g,h, Marc Vérin f,g, Paul Sauleau c,i,1, Didier Grandjean a,b,1

*These authors contributed equally to this work.

1. Introduction

Emotional prosody is defined as suprasegmental changes in the course of a spoken utterance, encompassing intonation, amplitude, envelope, tempo, rhythm, and voice quality (Grandjean, Banziger, & Scherer, 2006). The perception and decoding of emotional prosody has been studied through functional magnetic resonance imaging (fMRI) and patient studies, allowing researchers to delineate a distributed neural network involved in its identification and recognition (for a recent review, see Witteman, Van Heuven, & Schiller, 2012). Accordingly, models of emotional prosody processing have long postulated the existence of multiple information processing stages related to different levels of representations (Schirmer & Kotz, 2006; Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009). Following the processing of auditory information within the primary auditory regions (Bruck, Kreifelts, & Wildgruber, 2011; Wildgruber et al., 2009), including the activation of predominantly right-hemispheric primary and secondary auditory regions (for a review, see Witteman, van IJzendoorn, van de Velde, van Heuven, & Schiller, 2011), two prosody decoding steps have been identified. The first step, related to the representation of meaningful suprasegmental acoustic sequences, is thought to involve projections from the superior temporal gyrus to the anterior superior temporal sulcus. These cortical structures are thought to represent the so-called temporal voice area that encompasses voice-sensitive neuronal populations (Belin & Zatorre, 2000; Grandjean et al., 2005). In the second step, emotional information
derived at the level of the superior temporal sulcus is made available for higher-order cognitive processes mediated by the right inferior frontal gyrus (Früholz & Grandjean, 2013) and orbitofrontal cortex (Etkin et al., 2012; Grandjean, Sander, Lucas, Scherer, & Vuilleumier, 2008; Sander et al., 2005; Wildgruber et al., 2004). This step is thought to involve higher representation such as the *explicit evaluation of vocally expressed emotions*. In addition to this frontotemporal network, increased activity has been observed within the amygdala in response to emotional prosody (Früholz, Cervolo, & Grandjean, 2012; Grandjean et al., 2005; Sander et al., 2005). Furthermore, although they were not the focus of these studies, the involvement of subcortical regions (other than the amygdala) such as the thalamus (Wildgruber et al., 2004) and the basal ganglia (BG) in the processing of emotional prosody has also been reported. The involvement of the caudate and the putamen has repeatedly been observed in healthy participants by using fMRI and in brain-damaged patients by using the electroencephalographic method (Bach et al., 2008; Grandjean et al., 2005; Kotz et al., 2003; Morris, Scott, & Dolan, 1999; Paulmann & Kotz, 2008). More recently, studies exploring the effects of subthalamic nucleus (STN) deep brain stimulation (DBS) in Parkinson's disease (PD) have highlighted the involvement of the STN in the brain network subtending vocal emotions (Bruck, Wildgruber, Kreifelts, Kruger, & Wachtler, 2011; Péron, Cekic, et al., 2015; Péron, Früholz, Vérin, & Grandjean, 2013; Péron et al., 2010). From convergent evidence in the latter literature, Péron et al. (2013) have posited that the STN would coordinate neural patterns, either synchronizing or desynchronizing the activity of the different neuronal populations responsible for specific emotion components. They claim that the STN plays “the role of neural rhythm organizer at the cortical and subcortical levels in emotional processing, thus explaining why the BG are sensitive to both the temporal and the structural organization of events” (Péron, Früholz, Cervolo, & Grandjean, 2015; Péron et al., 2013). This model suggests that the BG and, more specifically, the STN, are sensitive to rhythm because of their intrinsic functional role of rhythm organizer, or coordinator of neural patterns. In a more operational way, this model hypothesizes that (i) the STN is involved in all stages of emotional processing, and (ii) the STN is involved in emotion processing irrespective of stimulus valence (positive or negative) and sensory modality (e.g., visual or auditory). To date, these two hypotheses have yet to be tested at the neurophysiological level in the vocal modality.

Studies featuring intracranial recordings of local field potentials (LFPs) in the STN in response to emotional stimuli have investigated the impact on the STN’s electrophysiological activity of processing visual stimuli that induce affective states. These studies have consistently reported event-related desynchronization of alpha activity within the STN between 1000 and 2000 ms post-onset in response to emotional stimuli in the form of both pleasant and unpleasant images drawn from the International Affective Picture System (IAPS) (Brucke et al., 2007; Huebl et al., 2011; Kühn et al., 2005). This effect has been shown to be correlated with *valence* (both positive and negative), but not *arousal* (Brucke et al., 2007). Furthermore, in the presence of depressed mood, this effect is reduced for positive emotions, but enhanced for negative emotions (Huebl et al., 2011), supporting the idea that the bias toward negative emotions in depression has an electrophysiological signature in this region (Péron et al., 2011). Taken together, these results confirm the STN’s functional role in emotion processing. However, the use of visual (emotional) material introduces several confounding factors. First, the pictures used in studies exploring the visual modality (see, for example, Benedetti et al., 2004; Kühn et al., 2005) are often not controlled for low-level physical features (brightness, contrast, spatial frequency, etc.). This is methodologically problematic, particularly for the analysis of neural oscillations in response to emotional stimuli, as these features are known to influence emotional perception and related brain activities (Delplanque, N’Diaye, Scherer, & Grandjean, 2007). Moreover, the emotional effects reported in the visual domain could be explained either by deficits in visual exploration related to oculomotor abnormalities (Fawcett, Dostrovsky, Lozano, & Hutchison, 2005), or by specific oculomotor patterns generated during the visual exploration of visual emotional stimuli (Van Reekum et al., 2007).

For all these reasons, previous research has left several questions unanswered, and it has yet to be determined whether the STN’s electrophysiological changes in response to emotional stimuli are modality-specific or supramodal, as suggested by behavioral studies exploring the emotional effects of STN DBS in PD (see Péron et al., 2013 for an exhaustive review). To the best of our knowledge, the only study in the auditory modality was recently performed by Eitan et al. (2013). They reported microelectrode recordings in the STN of 17 PD patients in response to emotional prosody stimuli (onomatopoeias from the Montreal Affective database, Bellin, Fillion-Bilodeau, & Gosselin, 2008). The authors reported ventral STN activity in response to emotional versus neutral auditory material, with increases of oscillations observed solely in the right STN; this difference has not been demonstrated in the left STN. However, similar to studies of the facial modality, this study did not control for low-level physical features known to have an impact on emotional prosody (e.g., fundamental frequency [f0], energy). Moreover, as mentioned earlier, the authors used single-unit recordings, and to the best of our knowledge, no previous study investigated LFPs in response to emotional auditory stimuli. The latter type of signal presents the advantage, as compared to single-unit recordings, to reflect the averaged dendrosomatic activity of synaptic signals of large neuronal population (Buzsaki, Anastassiou, & Koch, 2012).

In this context, the aim of the present study was thus to examine the influence of the processing of (both positive and negative) emotional prosody stimuli on event-related potentials (ERPs) in the STN, controlling for low-level physical features crucial for emotional prosody decoding. To this end, we explored the electrophysiological activity (LFPs) of the STN in response to angry, happy, and neutral prosodies, but also to acoustically matched nonhuman synthesized stimuli, in 13 PD patients.

The operational hypotheses are mainly based on Péron and colleagues’ recent model of the STN’s functional role in emotional (prosody) processing (2013), together with previous results from studies featuring intracranial recordings in the STN in response to emotional stimuli (Brucke et al., 2007; Huebl et al., 2011; Kühn et al., 2005). Péron and colleagues’ model (2013) suggested that the STN would be involved in all stages of emotional processing, irrespective of stimulus valence (positive or negative) and sensory modality. Accordingly, in the present study we expected to observe STN modulation for both positive and negative emotional prosody stimuli.

Moreover, this model hypothesizes that the STN would participate in the acquisition and expression of sequential, repetitive, motor, and cognitive behaviors, as well as emotional behaviors, elicited by external or internal triggers, recruiting and synchronizing the activity of the cortical and subcortical structures required for the relevant process, depending on the nature of the environmental cues and how the individual appraised such characteristics. As a consequence, “several different [STN] temporal dynamics [are speculated]; (i) sustained, synchronized activity, leading to the creation of a new functional neural pattern; (ii) early, transient synchronized activity, resulting in the activation of an overlearned/innate pattern; and (iii) later, transient synchronized activity as part of a mechanism that inhibits the overlearned pattern (involving the prefrontal regions)” (Péron et al., 2013, p. 370). Given that we do not explore...
the creation of a new functional pattern in this study, nor its inhibition, our present study explored the processes related to the activation of an overlearned/innate pattern. Moreover, the temporal dynamics depends on the specific process in action; for example, the activation of a pattern specific to the human voice is related to the coordination of specific brain areas and should occur at a different time, probably earlier through ascending auditory pathways (e.g., Escera, Leung, & Grimm, 2014), than the brain activation of a pattern specific to emotion through, for example, amygdala prefrontal coupling (e.g., Courtin, Karalis, Gonzalez-Campo, Wurtz, & Herry, 2014). As a consequence, we predicted two different temporal dynamics involving different brain patterns: (i) early stimulus-driven activity within the first hundred milliseconds after the onset of the human voice (vs. matched synthesized sounds) resulting in the activation of neural patterns thought to be involved in constructing the acoustic object, and (ii) later activity (after 1000 ms), as part of the mechanism driven by the emotional response related to the stimuli. Finally, on the basis of Eitan et al.'s study (2013), we predicted the observation of hemispheric specialization for emotional prosody processing in favor of the right STN.

2. Material and methods

2.1. Patients and surgery (Table 1)

Fifteen patients took part in the study. All of them met the clinical criteria of the Parkinson’s UK Brain Bank for idiopathic PD (Hughes, Daniel, Kilford, & Lees, 1992). The patients, all with medically intractable PD, underwent bilateral STN DBS at Rennes University Hospital (France); two patients underwent right unilateral STN DBS. Standard selection and exclusion criteria for surgery were applied to all of them (Welter et al., 2002). In particular, brain atrophy was excluded on the basis of a preoperative MRI.

Three months prior to the STN DBS surgery and the electrophysiological recordings, all patients were assessed in accordance with the Core Assessment Program for Intracerebral Transplantation (Langston et al., 1992) and scored on the Unified Parkinson's Disease Rating Scale I–IV (Fahn & Elton, 1987), Hoehn and Yahr scale (Hoehn & Yahr, 1967), and Schwab and England scale (Schwab & England, 1969). Patients were assessed on and off dopa both before and after surgery. In addition, all the patients were neuropsychologically assessed to rule out cognitive impairments (Péron et al., 2002). In particular, mild motor impairment in different brain areas was included in the study (see Table 1). As it has been previously shown that depression can influence emotional prosody recognition (Pérón et al., 2011), we controlled for this variable and observed that the patients did not score above the cut-off on the depression scale (see Table 1). Finally, none of the patients included in the study were hearing aids or had a history of tinnitus, and all patients were assessed by means of a standard audiometric screening procedure (AT-II-B audiometric test); two patients were excluded from the study because of presbycusis with bilateral hearing loss that was more marked at higher frequencies. Following surgery, patients were followed up clinically by a movement disorder specialist.

The study was conducted in accordance with the Declaration of Helsinki.

The macroelectrodes used were the Medtronic model 3389 (Medtronic Neurological Division, Minneapolis, MN, USA) with four platinum-iridium cylindrical surfaces (1.27 mm in diameter and 1.5 mm in length) and a contact-to-contact separation of 0.5 mm. Contact 0 was the most ventral and caudal, and Contact 3 the most dorsal and rostral. The surgical procedure consisted of the attachment to the patient’s head of a stereotactic Leksell frame under local anesthesia, and then the implantation of bilateral quadripolar DBS electrodes in the postero-lateral part of the STN in a single surgical session. The overall methodology was similar to that previously described by Benabid et al. (2000). A three-dimensional computed tomography (CT) scan of the brain with the Leksell frame was performed at the end of the surgical procedure and was registered to the preoperative MRI in order to calculate the stereotactic coordinates for positioning the two selected electrode contacts (one on the left and one on the right). The intended coordinates at the tip of Contact 0 were 10–12 mm from the midline, 0–3 mm behind the midcomissural point, and 3–5 mm below the anterior commissure (AC)-posterior commissure (PC) line. During the operation, the final course and depth of the electrode were determined by the best effect obtained on rigidity with no side effect and at the lowest voltage. A threedimensional CT brain scan performed a few days later confirmed the position of the electrodes. The pre- and postoperative image segmentation and registration workflow, the anatomical ParkMediata template used as an anatomical reference (Haegelen et al., 2013), and the electrode segmentation were performed by using pyDBS software (D’Albis et al., 2015).

2.2. Experimental setup

2.2.1. Stimuli (Fig. S1)

The vocal (prosodic) stimuli consisted of two pseudosentences spoken with different emotional prosodies (“ne kali bom sud molen!” and “kun se mina lod belam?”; mean duration = 1642 ms, range = 854–2788 ms) extracted from a previously validated data-
base, the GEneva Multimodal Emotion Portrayals corpus (Banziger & Scherer, 2010). 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 \( f_0 \). These two basic acoustic features are known to be the most correlated with emotional prosody judgments (e.g., Banse & Scherer, 1996; Grandjean et al., 2006). The first type of synthetic stimulus (synthesized intensity) consisted of a section of white noise, to which the intensity contour of the original stimulus was applied. The second synthetic stimuli (denominated as “synthesized \( f_0 \)” was a sine wave with constant amplitude. Its frequency is variable and corresponds to the

4 J. Péron et al. / Brain & Language 168 (2017) 1–11

temporal dynamics of energy and

f

original emotional and neutral sounds, in order to control for the

ness and neutral), we played synthesized stimuli, built from the

uli, and 20 trials featuring neutral stimuli, as well as 15 synthe-

ated as “synthesized \( f_0 \)” was a sine wave with constant amplitude. Its frequency is variable and corresponds to the detected \( f_0 \) of the original recording. Both synthetic stimuli had the same duration as in the original recordings. Examples of stim-

uli in each of the prosodic conditions (anger, happiness, neutral) are provided in Supplemental Fig. S1 (oscillograms and spectro-

grams). All the sounds were matched for mean energy to avoid loudness effects. Three blocks were constructed, featuring the dif-

ferent kinds of stimuli in pseudorandom order (no more than three times for the same experimental condition). Each block contained 20 trials featuring anger stimuli, 20 trials featuring happiness stim-

uli, and 20 trials featuring neutral stimuli, as well as 15 synthe-

ized intensity stimuli, 15 synthesized \( f_0 \) stimuli, and 1 section of white noise at the beginning (first stimulus) with a gradual attack to accustom the patients to the sounds. Each block contained a dif-

ferent 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: half the stimuli were pronounced by a female actor, and half the stimuli consisted of the pseudo-

sentence “ne kali bam sud molen!” The total duration of each block was \((-10 \text{ min})\), and there was a break between each one (i.e., a total of two breaks during the experiment). Each block contained pairs of identical stimuli, representing 10% of the stimuli, which were again ordered pseudorandomly. Their usefulness is explained in Section 2.2.3 (Task).

2.2.2. Trial (Fig. 1)

An example of a trial is provided in Fig. 1. In each trial, in order to avoid expectancy effects, we varied the duration of the interval between the onset of the fixation cross and the onset of the audi-

tory stimulus. In other words, the presentation of each auditory stimulus was preceded by a silent portion of pseudorandom dura-

tion, ranging from 50 to 250 ms, the so-called jitter (see jitter 1; Fig. 1). After the offset of the sound, we also included a silent por-

tion ranging from 3000 to 3500 ms (see jitter 2 + Silence; Fig. 1). 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 (see jitter 2; Fig. 1). 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.

2.2.3. Task (Fig. 1)

For each trial, the patients were asked to keep their eyes open and relaxed. They were told that they would hear meaningless speech uttered by male and female actors, as well as synthesized sounds. The loudness intensity was adjusted for each patient according to their hearing threshold at the beginning of the exper-

iment. Participants were asked to focus on these auditory stimuli and to press a button whenever they heard two identical stimuli in a row (see one-back trial; Fig. 1). These one-back trials repre-

sented only 10% of the totality of the trials and were excluded from the analyses. All the stimuli were played bilaterally through in-ear stereo headphones (Sennheiser, CX 2.00). This one-back task was administered to ensure that the patients were attentive to the sounds. Absolute silence was required for the test. Prior to the task, a pad with a push button was placed beneath the patients’ fingers. This pad generated a signal that was simultaneously recorded by the equipment. We registered the presentation of the various pros-

dic and synthesized stimuli onto the resulting electrophysiolog-

ical recordings. The hand (right or left) used to press the button was randomized across participants. Half of the patients used their dominant hand and the other half their nondominant hand. How-

ever, trials in which the patient gave a motor response were also excluded.

2.3. Recordings

All the patients were studied 2 days postoperatively in the interval between DBS electrode implantation and subsequent con-

nection to a subcutaneous stimulator. They were all taking their antiparkinsonian medication when they were assessed. Subthal-

amic LFPs were recorded through a g.BSamp (g.tec Medical Engi-

neering, Schiedlberg, Austria) biosignal amplifier connected to a PowerLab® 16/35 (ADInstruments, Dunedin, New Zealand) system. Deep brain activity was recorded bipolarly from two adjacent con-

tacts of each DBS electrode (three bipolar derivations 0–1, 1–2, and 2–3, Contact 0 being the most distal contact), amplified, and sam-

ped at a common rate of 1000 Hz. Signals were monitored online by using Labchart® (ADInstruments) software.

2.4. Data analysis

Data analysis was performed by using C++ software developed in-house by Lucas Tamarit, an engineer at the Swiss Center for Affective Sciences in Geneva (Switzerland); Cartool software developed by Denis Brunet (brainmapping.unige.ch/cartool); and the Matlab FieldTrip toolbox (http://fieldtrip.fcdonders.nl/; Oostenveld, Fries, Maris, & Schoffelen, 2011).

LFP signals were filtered offline with a 0.1–30 Hz bandpass FIR filter. Stimulus-locked epochs were then selected as follows: as the audio stimuli had a minimum duration of 854 ms, and the combined silent portions between them had a minimum duration of 3100 ms, we chose to create epochs starting 200 ms before the onset of the stimulus and ending 3700 ms post-onset. A baseline correction was applied to each individual epoch by subtracting the average level measured during the 200 ms preceding stimulus onset. These epochs served as the input for all subsequent analyses.

The selection of the STN contact pairs for the analyses was based on their anatomical position: at least one of the two contacts in the contact pair was located within the STN (\( N = 24 \) contact pairs from both sides in 13 patients, 2 right unilateral, as revealed ear-

lier; 2 patients were excluded from the LFP analyses because of basic auditory impairment). For the data analyses, we selected the first level of recordings for each patient: dipole 0–1 in both the left and the right STN. Epochs containing muscular artifacts or excessive noise were removed from the data by means of visual inspection (\(-5\% \) rejection). One-back trials and trials in which the patient gave a motor response were also excluded. ERPs were derived by averaging stimulus-locked epochs in 3074 trials: 2168 trials for the human voice (613 for anger, 791 for happiness, 764 for neutral) and 906 trials for the synthesized stimuli. Finally, to remove some of the noise in the data, and to avoid excessive wig-

gliness in the analysis, we first smoothed the epochs with a sliding window of 30 ms.

We performed several planned comparisons of the ERPs of the different conditions (see above). As the data were not binned (i.e., not average in time), analyses were carried out on all 3900 time points (i.e., one every 1 ms) of the epochs. For each of the 3900 time points, we computed a Student \( t \) statistic comparing two conditions (e.g., anger vs. neutral), but did not compute the
corresponding p value. We applied a very stringent approach to multiple testing corrections in order to control for the familywise error rate (FWER) over the 3900 tests to a level of 5%. To this end, we used the so-called cluster method (Maris & Oostenveld, 2007), which has the additional advantage of being a nonparametric method and therefore does not rely on Gaussian assumptions. This method clusters adjacent times for which the Student t statistic is above a given threshold and combines them to obtain a statistic for the cluster. To test all the time points simultaneously, we constructed N (typically = 1000) surrogate conditions by shuffling (permuting) the epochs of the two conditions to compare the values of these cluster statistics and obtain a p value for each cluster. Only clusters with p values smaller than 5% are reported in the Results section. It has been shown that these regions can be deemed significant if they are below a FWER threshold of 5% (Maris & Oostenveld, 2007). To take into account that the epochs are measured on different patients, we standardized the epochs by participant and we modified the Fieldtrip function so that epochs were permuted only within participants, in accordance with the principle that only similar epochs can be permuted. We used the 0.8 quantile as a threshold and the wcm cluster statistic (with the default power value of two).

One of the great advantages of the permutation method is the freedom to choose the statistic. To check the sensitivity of our results, we therefore repeated the analysis, first without the standardization and the per participant restriction for the permutation and second by replacing the Student t statistic with a robust version based on 10% trimmed means. The results (not shown) were close to the results based on the Student t, showing that they were not influenced by a few outlying epochs, but were actually quite robust.

Planned comparisons across the patients were performed by means of the nonparametric pairwise permutation tests described above, with the corrected cluster p value, considering the left and the right STN separately. We conducted two kinds of comparisons. First, we explored differences in the patterns of electrophysiological activity between the neutral, angry, and happy prosodies (human voice) and the synthesized (nonvoiced) sounds. They were asked to focus on these auditory stimuli and to press a button whenever they heard two identical stimuli in a row (as in the one-back trial example presented here). All the stimuli were played bilaterally through in-ear stereo headphones. This one-back task was administered to ensure that the patients were attentive to the sounds. The hand (right or left) used to press the button was randomized across participants.

Fig. 1. Trial examples and task. Legend: Jitter 1 = pre-onset silent portion of pseudorandom duration lasting between 50 and 250 ms in order to avoid an expectancy effect; Sound = auditory stimulus lasting between 854 and 2788 ms; Jitter 2 = post-offset silent portion of pseudorandom duration lasting between 50 and 500 ms in order to avoid synchronicity between the offset of the sound and the offset of the fixation cross; Jitter 2 + Silence = post-offset silent portion of pseudorandom duration lasting between 3000 and 3500 ms. For each trial, the patients were told that they would hear meaningless speech uttered by male and female actors, as well as synthesized sounds. They were asked to focus on these auditory stimuli and to press a button whenever they heard two identical stimuli in a row (as in the one-back trial example presented here). All the stimuli were played bilaterally through in-ear stereo headphones. This one-back task was administered to ensure that the patients were attentive to the sounds. The hand (right or left) used to press the button was randomized across participants.
3. Results

3.1. Anatomical location (Fig. 2)

As we selected the first level of recordings for each patient (dipole 0–1 in both the left and the right STN), we calculated the mean contact for each patient between Contact 0 and 1, and then calculated the average coordinates of this mean contact between all patients in a referential space, the ParkMedAtlis template (Haegelen et al., 2013). The mean (Talairach) coordinates of this mean contact with the AC as the origin of the coordinates were as follows: left STN, y = −10.9 ± 1.3 mm in the antero-posterior direction, x = −15.2 ± 1.6 mm in the lateral direction, z = −5.3 ± 2.0 mm under a line passing through the AC and the PC line; right STN, y = 10.6 ± 1.4 mm, x = −14.8 ± 1.6 mm, z = −3.77 ± 2.2 mm. Fig. 2 provides a view of the location of the mean contacts for the 13 patients, showing that the left and the right mean contacts were inside the STN. Individual coordinates are also provided in Supplemental Table S2.

3.2. Behavioral results

In the one-back task, all patients had a correct hit rate of 100% and an average of 4.2% false positive responses (resulting in exclusion of ERP analyses), with no missed responses.

3.3. Event-related potentials

3.3.1. Human voice versus synthesized stimuli (Fig. 3)

As emotional prosody processing is subtended by human voice processing, and as previously explained in Section 2.4 (Data Analysis), we first ran a planned contrast, comparing the human voice (anger, happiness, and neutral) with the synthesized stimuli (i.e., synthesized intensity and synthesized f0) (see Subsection 2.2.1 (Stimuli) of the Material and Methods section and Supplemental Material S1).

As shown in Fig. 3, results revealed significantly greater activity in the human voice condition than in the synthesized stimuli condition, with early activity in the left and the right STN (left STN: ~300 to ~400 ms and ~450 to ~700 ms post-onset, corrected p < 0.05; right STN: ~500 to ~600 ms post-onset, corrected p < 0.05).

3.3.2. Emotional effects

3.3.2.1. Angry versus neutral (Fig. 4). As shown in Fig. 4, results indicated significantly greater activity in the anger condition than in the neutral one (early activity ~200 to ~300 ms and later activity ~850 to ~950 ms post-onset, corrected p < 0.05) in the right STN. No significant difference was observed between the two conditions in the left STN.

3.3.2.2. Happy versus neutral (Fig. 5). As shown in Fig. 5, when we analyzed happy minus neutral prosodies, results revealed a significant difference between the two conditions in the right STN (~200 to ~300 ms and ~3250 to ~3350 ms post-onset, corrected p < 0.05). No significant difference was observed between the two conditions in the left STN.

3.3.3. Lateralization effects (right versus left STN comparisons)

When we compared electrophysiological activity in the right versus the left STN, in each of the anger and happiness conditions, we did not observe significant differences.

4. Discussion

The aim of the present study was to explore the influence of positive and negative vocal emotions (i.e., emotional prosody) on the electrophysiological activity of the STN in humans. We compared the electrophysiological activity (LFPs) of the left and the right STN in 13 PD patients in response to angry, happy, and neutral prosodies, as well as to nonhuman synthesized stimuli. We were thus able to perform two major types of analyses. First, we explored the different patterns of electrophysiological activity in response to the human voice (neutral, anger, and happy prosodies) versus the synthesized stimuli (intensity contour applied to white noise, and f0 contour applied to a pure sine oscillator; see Supplemental Material S1). Second, in order to investigate the emotional effects, we looked for differences in electrophysiological activity in response to the angry and happy prosodies versus the neutral prosody.

Before we set out our results and draw any inferences from them, it is important to acknowledge that, as in many clinical studies of this nature, the sample size was small (N = 13 patients included in the LFP analyses, 24 STN) which represents a limitation. In order to address this point, and although this is not routinely done in the literature, we applied a very stringent statistical approach to multiple testing correction to control for FWER. In addition, we investigated the sensitivity of our results to one specific participant or to a few epochs because it is known that the t test is not robust against outliers. We thus ran a robust analysis (with 10% trimmed means). As this analysis yielded similar results to those based on the t statistic, we are confident that our findings are not due just to one participant or to outlying epochs.

First of all, we would like to stress a clear event-related component in all our average ERPs, demonstrating a sensitivity of the STN for sound processing in general. This ERP complex is characterized by positivity at ~80 ms after the onset of the sound, followed by negativity at ~150 ms, itself followed by a positive fluctuation at ~250 ms. The first significant difference in our experimental
Fig. 3. Stimulus-locked left (top panel) and right (bottom panel) STN ERPs in the human voice (black) \((n = 2168\) trials) and synthesized stimuli (red) \((n = 906\) trials) response conditions from the 13 patients and from the first level of recording for each patient (dipole 0–1). The vertical black line at 0 ms indicates the onset of the auditory stimulus. Time regions where the difference between the two conditions is statistically significant, controlling for FWER at 5%, are shaded in grey. Filtering = 0.1-Hz high-pass and 30-Hz low-pass FIR filters, baseline correction ranging from 0 to 200 ms, temporal smoothing with a sliding window of 30 ms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. Stimulus-locked left (top panel) and right (bottom panel) STN ERPs in the anger (black) \((n = 613\) trials) and neutral (red) \((n = 764\) trials) response conditions from the 13 patients and from the first level of recording for each patient (dipole 0–1). The vertical grey line at 0 ms indicates the onset of the auditory stimulus. Time regions where the difference between the two conditions is statistically significant, controlling for FWER at 5%, are shaded in grey. Filtering = 0.1-Hz high-pass and 30-Hz low-pass FIR filters, baseline correction ranging from 0 to 200 ms, temporal smoothing with a sliding window of 30 ms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
and then almost simultaneously in both the early in the left voice signal (and other related processes) by controlling for f greater activity in the human voice condition than in the synthesized sounds are concerned, we observed a specific modulation of the early neural responses that were not modulated by our experimental design (see below) occurs within this third component. Further research is needed to better understand the functional role of these early neural responses that were not modulated by our experimental manipulation.

As far as the human voice versus non-voiced stimuli comparisons are concerned, we observed a specific modulation of the STN in response to the human voice. There was significantly greater activity in the human voice condition than in the synthesized stimulus condition in both the right and the left STN, occurring early in the left (~300 to ~400 ms post-onset, corrected p < 0.05) and then almost simultaneously in both the right and the left (right STN: ~500 to ~600 ms post-onset, corrected p < 0.05; left STN: ~450 to ~700 ms post-onset, corrected p < 0.05) (Fig. 3). Our comparison of human voices versus synthesized stimuli enabled us to measure the STN’s specific activity in response to human voices minus the variations in f0 and energy. In other words, when we performed this comparison, we could test the STN’s specific activity in response to the remaining spectral components of the human voice signal (and other related processes) by controlling for f0 and intensity dynamics using synthesized sounds. The involvement of cortical (auditory) structures in the processing of such spectral modulations of speech has been reported in several studies (e.g., Leaver & Rauschecker, 2010; Overath, Kumar, von Kriegstein, & Griffiths, 2008). However, to the best of our knowledge, apart from the involvement of these structures in the temporal organization of events (Kotz & Schwartz, 2010), the functional specialization of the BG and, more specifically, the STN has never before been reported in this process. It would be worthwhile for future studies to investigate the functional specialization and integration of these subcortical structures regarding the spectral components of voice processing in greater detail. Nonetheless, and based on Graybiel’s model (for an exhaustive description of her model, see Graybiel, 2008), this result is not surprising. This model proposes that the BG may be critically involved in constructing performance units of sequence representations, also called chunks. The proposition is that items of information can be organized more efficiently by recoding them to form packages (chunks) that, once learned, can be processed more automatically. The nature of these items can be emotional, as we suggested recently (Péron et al., 2013), as well as motor and cognitive (for e.g., vocal).

As far as the investigation of the emotional effects is concerned, we observed a specific modulation of the STN in response to both angry and happy prosodies only in the right STN. We observed two different temporal dynamics: (i) right STN ERPs occurring shortly after onset and at the same time for angry and for happy prosodies (Figs. 4 and 5, bottom panels; ~200 to ~300 ms post-onset, corrected p < 0.05); and (ii) later activity for angry prosody at ~850 to ~950 ms post-onset (corrected p < 0.05) and at ~3250 to ~3350 ms post-onset (corrected p < 0.05) for happy prosody. No significant difference was observed between the two conditions in the left STN (Figs. 4 and 5, top panels).

In accordance with our hypotheses, the present results seem to confirm the STN’s functional involvement in emotional prosody, whether the valence is positive (happy voices) or negative (angry voices). In addition, and as we had predicted, the STN ERP modulations observed in response to vocal emotions occurred both early and later during emotion processing. As indicated in the Introduction, studies in the visual modality have reported event-related desynchronization between 1000 and 2000 ms post-onset (Brucke et al., 2007; Huebl et al., 2011; Kühn et al., 2005). In Kühn and colleagues’ study, as well as in Huebl and colleagues’ study, these results were obtained for both positive and negative IAPS stimuli, whereas in Brücke and colleagues’ study, only positive images were presented. To the best of our knowledge, the early effects we obtained in response to anger and happy prosodies are new empirical evidence for the involvement of the STN at an early stage of emotion processing. One possible explanation for these potentially modality-driven differences is the existence of specific timing patterns for the visual versus auditory modalities.

Taken together, the present results support Péron et al. (2013) posit that the STN plays a meta-role and exhibits supramodal specialization in emotional processes. According to this model, the
STN may thus be involved in the multiple stages of emotion processing. In addition, the finding in the present results of the STN's functional role in processing vocal emotions reinforces the hypothesis whereby the BG and, in this case, the STN, are sensitive to both the temporal and structural organization of events. Péron et al. (2013) model ascribes to the STN the intrinsic functional role of neural rhythm organizer at the cortical and subcortical levels in emotion processing, which predicts its sensitivity to rhythm. In the context of vocal emotion recognition, for example, the STN would act as a marker for the transiently connected neural network (i.e., amygdala, auditory cortices, inferior frontal gyri, and orbitofrontal cortices) that subserves emotional prosody processing. If the co-activation across different neuronal populations is recurrent or functionally important, the BG-mediated synchronization presumably increases the weight of the synaptic connections within the cortical or subcortical neural network.

In addition to these responses, we also observed effects pointing to the STN's hemispheric specialization in emotional prosody processing in accordance with Eitan et al. (2013). When we tested simple effects by analyzing the right and the left STN separately, we were not able to reject the null hypothesis for the comparisons in the left STN (Figs. 4 and 5, top panels), whereas we were able to do so in the right STN for the angry versus neutral and happiness versus neutral prosodies (Figs. 4 and 5, bottom panels). In order to investigate these results in greater detail, we performed analyses comparing the electrophysiological activity of the right versus left STN for anger and for happiness. The analysis did not reveal any significant difference in electrophysiological activity between the right and the left STN for either of the conditions we tested. However, given that we were not able to test these interaction effects directly, the present results can still be regarded as reflecting an asymmetry in terms of right/left lateralization interacting with our experimental conditions taken together. Further studies are needed to carry out a systematic investigation of this obviously complex pattern, especially bearing in mind that symptoms were predominantly left-sided for 9 of the 13 patients included in the analyses, thus limiting the inferences that can be drawn from these results.

That being said, it is worth noting that models of emotional prosody perception allow for right hemispheric specialization at many stages in the process (see, Witteman et al., 2012 for a review of these models). Two recent meta-analyses, one of the lesion literature (Witteman et al., 2011), the other of the neuroimaging literature (Witteman et al., 2012), also seem to confirm this hypothesis, pointing to the conclusion that emotional prosody perception takes place mainly in the right frontotemporal network, but also involves the right BG. The meta-analysis of the lesion literature found statistically robust evidence of a greater deterioration in emotional prosodic performance following right rather than left hemispheric BG damage (Witteman et al., 2011). The meta-analysis of neuroimaging literature reported clusters in the right medial globus pallidus and caudate body when emotional prosody stimuli were compared with neutral stimuli or synthesized speech (Witteman et al., 2012). Taken together, the present results appear to indicate possible right hemispheric specialization for emotional prosody processing in the BG. Focusing on the STN, apart from the study by Eitan et al. (2013), the only information concerning this basal ganglion's potential hemispheric specialization in emotional (prosody) processing comes from an fMRI study in healthy participants undertaken by Péron, Frühholz, et al. (2016). This study yielded results that appear to contradict the right dominance hypothesis, as it was the left STN that exhibited activity during the processing of emotional (angry voices) versus neutral stimuli. This activity was observed only in a condition in which the listeners focused on a nonemotional feature of the voice (i.e., the speaker's sex; sex identification task), instead of concentrating on emotional vocal features (i.e., emotion discrimination; prosody task). In this context, we can hypothesize that hemispheric specialization differs according to whether it is driven by the stimulus and/or by the task, especially when the task requires the inhibition of prepotent responses (as was the case in Péron, Frühholz, and colleagues' study). This hypothesis of differential (left vs. right) functional roles played by the STN in emotional (but also cognitive and motor) processing deserves to be explored more thoroughly in future studies, given its fundamental, but also clinical, relevance.

In conclusion, we found both early and late modulation of STN activity in response to the human voice, as well as specific early and later activity in response to negative and positive emotions conveyed by voices. These results confirm the hypothesis that the STN is involved in emotion processing, irrespective of stimulus valence (positive or negative) and sensory modality (e.g., visual or auditory). Future studies featuring a larger number of participants, together with neuroimaging studies in healthy participants, are needed to confirm our findings.

Disclosure

The authors report no conflicts of interest. The study was carried out at the Neurology Unit of Pontchaillou Hospital (Rennes University Hospital, rue Henri Le Guilloux, 35033 Rennes, France; Prof. Marc Vérin). Data acquisition (externalization and acquisition per se) was funded by PHRC-IR Grant No. IDRCB: 2011-A00392-39 (Prof. Marc Vérin). The first author (Dr. Julie Péron) was funded by Swiss National Foundation Grant No. 105314_140622 (Prof. Didier Grandjean and Dr. Julie Péron), and NCCR Affective Sciences was funded by Swiss National Foundation Project No. 202 – UN7126 (Prof. Didier Grandjean). The funders had no role in data collection, discussion of content, preparation of the manuscript, or decision to publish.

Acknowledgements

The study was carried out at the Neurology Unit of Pontchaillou Hospital (Rennes University Hospital, rue Henri Le Guilloux, 35033 Rennes, France; Prof. Marc Vérin). Data acquisition (externalization and acquisition per se) was funded by PHRC-IR Grant No. IDRCB: 2011-A00392-39 (Prof. Marc Vérin). The first author (Dr. Julie Péron) was funded by Swiss National Foundation Grant No. 105314_140622 (Prof. Didier Grandjean and Dr. Julie Péron), and NCCR Affective Sciences was funded by Swiss National Foundation Project No. 202 – UN7126 (Prof. Didier Grandjean). The funders had no role in data collection, discussion of content, preparation of the manuscript, or decision to publish. The Cartool software (b rainmapping.unige.ch/cartool) was programmed by Denis Brunet, from the Functional Brain Mapping Laboratory, Geneva, Switzerland, and supported by the Center for Biomedical Imaging (CIBM) of Geneva and Lausanne. We would like to thank the patients for contributing their time to this study. We are also grateful to Christophe Mermoud for his help in setting up the acquisition system and to Elizabeth Wiles-Portier for revising the English style.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bandl.2016.12.003.

References


J. Péron et al. / Brain & Language 168 (2017) 1–11