



# Central Nervous System Control of breathing in Natural Conversational Turn-Taking

Camilla Di Pasquasio, Lila de Pellegrin, Arthur Pineaud, Antonin Marty, Thierry Chaminade

## ► To cite this version:

Camilla Di Pasquasio, Lila de Pellegrin, Arthur Pineaud, Antonin Marty, Thierry Chaminade. Central Nervous System Control of breathing in Natural Conversational Turn-Taking. 2024. <hal-04735369>

**HAL Id: hal-04735369**

**<https://hal.science/hal-04735369v1>**

Preprint submitted on 14 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Central Nervous System Control of breathing in Natural Conversational Turn-Taking

## Authors

Camilla Di Pasquasio, Lila De Pellegrin, Arthur Pineaud, Antonin Marty, Thierry Chaminade

## Affiliations

Institut de Neurosciences de la Timone  
UMR 7289 Aix-Marseille Université - CNRS  
25 Bd Jean Moulin, 13385 Marseille cedex 5

## Abstract

Discussion is a fundamental social activity requiring coordination of speech between interlocutors. Speech production is a complex human behaviour that involves several anatomo-physiological processes, including inspiration and expiration. The aim of the present study is to investigate the neurophysiological underpinnings of speech-related respiration events in conversational turn-taking. We made use of an existing corpus of natural conversations between a participant and its interlocutor (Human or Robot) focusing on synchronised (1) behavioural (conversation turn-taking), (2) respiratory (maxima of inspiration) and (3) neurophysiological (fMRI) data. Precisely timed conversation transcripts from 25 participants were used to categorise breathing maxima based on their timing relative to the participant's speech onset. In agreement with the literature, the closest respiration time maximum to each speech turn occurred on average 200 ms prior to speech onset. The fMRI second-level contrast ( $p_{FWE} < 0.05$ , extend  $k > 5 \text{ cm}^3$ ) Resp+ (maximum respiration associated with speech) *versus* Resp-, exclusively masked to exclude speech related areas, revealed bilateral activations in the central sulcus, the brainstem and the cerebellum. The brainstem cluster comprises respiratory pattern generators, possibly the preBötzinger complex, that need to be inhibited to enslave breathing to speech production and not physiological needs, while the central sulcus cluster is likely to be located in the postcentral primary sensory cortex receiving upper torso inputs indicating that lungs are filled, and the cerebellum clusters could play a role in the timing of speech onset, 200 ms after a respiration maximum. These results show how cortical, cerebellar and brainstem coordinated control of breathing during conversational turn-taking is part of the intricate physiological mechanisms that contribute to natural communication dynamics.

## Keywords

*Natural conversation, turn-taking, respiration, fMRI, central sulcus, brainstem, cerebellum*

## Introduction

Language is a unique hallmark of human cognition, and conversations, from everyday chit chat to specific discussions, play a widely acknowledged function in social interactions (Hudson, 1996). Considering the need to investigate neural correlates of "*real-time social encounters in a truly interactive manner*" (Schilbach et al., 2013), we present here an investigation of the physiology associated with the control of respiration in natural conversations.

### Conversation, a social implementation of speech

Discussion is characterised by the exchange of spoken remarks between two or more interlocutors. This trivial everyday activity plays a very important social role in our societies. To enable this interaction, one speaker must be able to perform a phonatory exhalation to express its remarks, as well as to control the turn of speech between himself and his interlocutor. Smooth conversation entails accurate prediction of the turn-taking time, i.e. the time when one interlocutor stops speaking and therefore leaves the floor open for the other speaker. Fluid transitions between speakers typically occur within a temporal window of a few hundred milliseconds (Torreira et al., 2015). Reducing the turn-taking time increases the amount of factual information that can be exchanged between speakers in a limited amount of time. Note that a study of a variety of languages shows a universal tendency to avoid overlapping conversations and minimise silences between speech turns (Stivers et al., 2009).

Speech is a complex behaviour that requires precise coordination of multiple anatomico-physiological processes. During phonatory exhalation, a series of events are triggered to enable sound production. The vocal folds and soft palate open during this phase, allowing air from the lungs to enter the oral cavity. The subglottic pressure generated by this air flow concentrates under the vocal cords. The latter, in an open abducted position, rub against each other and vibrate, producing the phonemes that form the basis of spoken language. This phenomenon is closely linked to laryngeal pressure, a force resulting from the interaction between the vocal folds and subglottic pressure. This laryngeal pressure is essential to distinguish breathing from vocalisation, underlining the importance of muscular control in the process. The diversity of phonemes emitted then depends on the positioning of the oral and nasal cavities, contributing to the richness of speech (Demartsev et al., 2022). This control is exercised by the airway muscles, often referred to as a "valve", as well as by the lower tract, acting as a thoracic bellows. These muscular components, represented in the dorsal primary motor cortex, contribute to the precise regulation of breathing required for sound production (McKay et al., 2003). In contrast, articulation is associated by the involvement of the ventral primary motor cortex (Pulvermüller et al., 2006), that have direct access to the motor neurons that control the muscles involved in speech, such as those of the face, mouth,

tongue, larynx and pharynx (Holstege & Subramanian, 2016). Altogether, these results confirm a dissociation between the functions of dorsal and ventral motor cortices in human speech (see also Price, 2012), that would correspond to the control of respiration and articulation, respectively.

### **Central nervous system control of speech**

The neural bases of articulation in speech have been - and still are - widely investigated. Patient Louis Victor Leborgne, studied by neurosurgeon Paul Broca, was unable to express other syllables than Tan and presented a lesion in the left inferior prefrontal cortex. The link between *a* (absence) - *phasia* (speech) and a localised brain lesion is at the origin of neuropsychology, the investigation of causal links between lesions in specific brain regions and associated behavioural deficits. In contrast, little is known about the neurophysiological control of breathing in speech, despite its importance in particular in conversations. Preparing to speak indeed involves subtle coordination of interlocutors' breathing cycles. Speakers plan their speeches by synchronising their inspirations with the end of questions, thus establishing a dynamic temporal relationship between turn-taking and breathing (Rochet-Capellan & Fuchs, 2014). Other investigations revealed that inspirations before long responses cluster around the end of the question, while inspirations before short responses show greater dispersion and at earlier timings, suggesting a distinction between inspirations specifically intended for speech and those more related to physiological needs (Torreira et al., 2015). Breathing indeed adapts to communication needs during smooth conversations, underlining the dynamic and adaptive aspect of the process.

While the central nervous system control of respiration in natural conversations remains unknown, neuropsychological literature reports a significant number of cases of dissociation between automatic and voluntary breathing. A patient study on a case of severe respiratory apraxia resulting from supranuclear palsy revealed an inhibition of automatic breathing upon volitional control (Haouzi et al., 2006). Ventilation compensation by the Supplementary Motor Area (SMA) during arousal is suggested in patients with Congenital Central Hypoventilation Syndrome, who suffer from debilitating hypoventilation only during sleep (Tremoureux et al., 2014). Both previous examples illustrate complex interplay between volitional control and automatic commands of respiration, the former being able to temporarily bypass the later, an unique case for a life-sustaining physiological requirement in animals (Trevizan-Baú et al., 2024). Research in animals, further supported by one human neuroimaging investigation of volitional respiratory control (McKay et al., 2003) and a model of dyspnea (Betka et al., 2022), suggests a complex interplay between respiration-related cortical networks and the involvement of brainstem automatic respiration generation.

In the brainstem, a central respiration controller consists in a network of motoneurons located in the ventral reticular formation innervating the ribcage and upper airways (Viemari et al., 2013). Extensive animal studies has led to the identification of respiratory pattern generators, known together as the preBötzinger Complex (Smith et al., 1991), as central elements of this central controller. This complex orchestrates the three-phase - inspiration, post-inspiration and expiration - respiratory cycle (Bellingham, 1998). Because it is mainly devoted to controlling an automatic physiological reflex required for survival, respiration, it is not known how it contributes to voluntary control of breathing: actually, breathing may be an unique physiological reflex given it is usually automatic but can also be controlled, and in the case of speech, that requires to be synchronised with expiration, could be involuntary controlled but in conflict with the automatic control. Elucidating the respective role of cortical and brainstem substrates of respiration control in natural conversation is crucial to understanding how breathing can be associated either with survival or with communication in spoken social interaction.

### **Current analysis**

When we take turn during a conversation, the mechanisms discussed above harmonise in asymmetrical respiratory cycles (Rochet-Capellan & Fuchs, 2014). These cycles were measured as variations in lung amplitude and frequency over time using a respiration monitor belt, making it possible to monitor the various adaptations during a conversation. Significant differences can be observed between individuals depending on their age, gender and lifestyle, resulting in specific adjustments when breathing during speech. Studies have highlighted the influence of these factors on lung volume during spontaneous speech (Winkworth et al., 1995). Furthermore, within the same individual, intra-individual variability in lung volume during unscripted communication is also significant. Contextual factors, changing emotional states and particular cognitive demands contribute to this variability (Winkworth et al., 1995). These elements modulate the way a person adjusts lung volume during speech, creating a respiratory signature unique to each conversational context. This close coordination between the respiratory process and speech suggests that our respiratory system adapts dynamically to the demands of verbal communication, both inter-individually and intra-individually. It highlights the need to investigate this coordination taking into account sources of variability.

Here, we made use of a unique corpus of natural conversation comprising synchronised conversation, breathing and functional neuroimaging (fMRI) to contrast brain control of breathing according to whether it is associated with speech production (maximum occurring close to the participants' speech onset) or physiological survival (necessity of pulmonary exchange between O<sub>2</sub> and CO<sub>2</sub>). Conversation transcripts are used to categorise automatically breathing monitor belt maxima according to their temporal occurrence within a time window

centred on the participant's speech onset. Results indicate that cortical regions buried in the human central sulcus are significantly more active for speech-associated breathing maxima, as well as regions in the brainstem that we associate with human homologues of the respiratory rhythm generator known as PreBötzinger complex first identified in rodents (Smith et al., 1991). These results suggest that both cortical and subcortical control of breathing are involved in respiration control during turn-taking associated with natural conversation.

## Materials and Methods

### Corpus acquisition

Twenty-five native French-speaking participants (7 M, 18 F) with an average age of 28.5 years old (*s.d.* 12.4) were recorded in a fMRI protocol investigating natural conversations in Human-Human and Human-Robot Interactions. Experimental procedures were approved by the designated national ethical committee ("Comité de Protection des Personnes" CPP Sud-Marseille 1, approval number 2016-A01008-43) and respected all procedures in place. The conversational paradigm (Rauchbauer et al., 2019), its theoretical grounds (Chaminade, 2017) and the variables recorded (Rauchbauer et al., 2020), have already been described and will only be summarised. In particular, we will focus on the three variables utilised in the current analysis: brain activity (using fMRI Blood Oxygen-Level Dependent (BOLD) signals acquired continuously), respiration (focusing on inspiration-to-expiration transitions) and transcripts of the natural conversations (Hallart et al., 2021; Rauchbauer et al., 2020).

Participants were fully naive about the actual objectives of the current study, i.e. creating a multidimensional corpus of natural interactions including synchronised neurophysiological recordings. At the MRI centre were presented the cover story, that involved natural discussion about a purported neuromarketing experiment. During each functional scanning run, they underwent six trials alternating between conversation with a human and conversation with a robot. Each trial started with the presentation of one out of three images forming the neuromarketing cover story material, followed by one minute of live conversation with an interlocutor located outside of the scanner room. The bidirectional audio set-up enabled the live conversation between the scanner and the outside consisted of an active noise-cancelling MR compatible microphone (FORMI/III+ from Optoacoustics Ltd.) to reduce scanner noise and inserted earphones from Sensimetrics. Participants could see a live movie of the interlocutor projected on the scanner's stimulation screen and had their eye movements recorded.

The interlocutor was either a confederate of the experimenter or the robotic head Furhat controlled by the confederate with a Wizard of Oz setup of ~100 sentences prepared *a priori*. We recorded three trials each containing a one-minute block conversation per conversational agent (Human and Robot) in each scanning run (6 blocks total) and four runs per participant yielded an overall total of twenty-four minutes of conversation recorded. Effects of the conversational agent are irrelevant for the current analysis focusing on speech production mechanisms and won't be discussed further.

The MRI data was collected with a 3T Siemens Prisma (Siemens Medical, Erlangen, Germany) using a 20-channel head coil. BOLD sensitive functional images were acquired using

an EPI sequence in the four runs. Parameters were as follows: echo time (TE) 30 ms, repetition time (TR) 1205 ms, flip angle 65°, 54 axial slices co-planar to the anterior/posterior commissure plane, field of view 210 mm × 210 mm, matrix size 84 × 84, voxel size 2.5 × 2.5 × 2.5 mm<sup>3</sup>, with multiband acquisition factor 3. After functional scanning, structural images were acquired with a GR\_IR sequence (TE/TR 0.00228/2.4 ms, 320 sagittal slices, voxel size 0.8 × 0.8 × 0.8 mm<sup>3</sup>, field of view 204.8 × 256 × 256 mm). Data acquisition of Blood Oxygen-Level Dependent (BOLD) signal 3-dimensional images were obtained via whole brain scans every 1.205 seconds. One run totalled 385 volumes for a duration of 463.925 s (7 minutes and 425 seconds). Raw data acquired during fMRI scanning have been uploaded on OpenNeuro<sup>1</sup> including Text-format log-files containing trials onset information.

The physiological data (cardiac pulsation and respiration) in the study was recorded with the scanner's dedicated apparatus. A photoplethysmography device was placed on the left-hand index fingertip to capture pulse oximetry, while a breathing belt was positioned at chest level. Both devices were connected wirelessly through Bluetooth, and the data was acquired continuously at a frequency of 200 Hz. Synchronisation is intrinsic to the data recording apparatus, and is based on the 2.5 ms time bins constitutive of MRI recordings: time is logged in bins for the BOLD signal and the breathing data, both recorded with the MRI scanner, and the conversational data processing also include a synchronisation and resampling step (see next).

## Data processing

fMRI preprocessing with SPM12 involved slice-timing corrections, volumes realignment and magnetic field inhomogeneities correction (Penny et al., 2007). Normalisation of individual functional and anatomical data to the standard brain space of the Montreal Neurological Institute (MNI) used segmented anatomical T1 and T2 images recorded at the just after BOLD functional images and realigned with the first image of the time-series, followed by the DARTEL procedure (Ashburner, 2007). BOLD signals were extracted from masks of the Grey matter, White matter and Cerebrospinal fluid individuals' regions of interest derived from the segmentations to make detrending nuisance variable with the conn toolbox (Whitfield-Gabrieli & Nieto-Castanon, 2012), that incorporates other denoising procedures like linear detrending with a high-pass filter (threshold of 128 seconds) as well as realignment parameters for removal of artefacts due to participant movement calculated during the realignment step. Physiological recordings were processed using the PhysIO toolbox (Kasper et al., 2017) that employs model-based approaches for physiological noise correction of fMRI data. It uses physiological measurements to derive heart rate variability (referred to as the cardiac response function) and respiratory volume per time (referred to as

<sup>1</sup> <https://openneuro.org/datasets/ds001740>



the respiratory response function). Both movement-related and physiology-related possible sources of artefacts were exported as single-participant and single-run matrices containing aggregating cardiac, respiration and movement signals possibly associated with recording artefacts as well as linear and region-of-interest detrending nuisance variables. Finally, we applied a 5 mm<sup>3</sup> full-width half-maximum three-dimensional Gaussian kernel spatial smoothing.

The respiration time-series were processed in order to ensure signal integrity by replacing plateau signals corresponding to saturation of the recording belt signal. The respiration data was first smoothed using a moving average with a window size of 50. Regions of saturated signal were identified by contiguous segments where the derivative of the breathing time-series was null. Consecutive numbers in these segments are replaced by 'NaN' so that saturation is considered to represent missing data. The missing data were replaced by non-linear interpolations (*spline* method in MATLAB® R2022a's function *interp1*) taking into account the full duration of each individual run. Time series were finally resampled in milliseconds synchronised with fMRI acquisition, taking advantage of the fact that the offset of the respiration time-series is synchronised with the end of the fMRI recordings.

Conversational data from participants' audio recordings underwent preprocessing to eliminate noise during scanner acquisition, employing a noise reduction filter provided by Sox (<http://sox.sourceforge.net>). For each participant, an individual float value, ranging from 0.01 to 0.50, was assigned. The denoised data was subsequently segmented into Inter-Pausal Units (IPUs) with each participant having an individual coefficient, a float value between 0.20 and 0.95. This coefficient facilitated the automatic determination of volume thresholds for periods classified as "silence" and "IPU" within each analysis window. It was applied to the mean of the root mean square (RMS) distribution of the audio file. IPUs were defined as speech blocks occurring between silences, with minimum duration of 200 ms (Blache et al., 2009). Visual validation of successful denoising and IPU segmentation was conducted using Praat. Files were furthermore uploaded into SPPAS, version 1.9.9 (Bigi, 2015) for transcription. Automatic Text normalisation was performed using the SPPAS software tool.

### **Definition of experimental events**

Automatic processing algorithm were developed to 1) identify local respiration maxima as they correspond to transitions between inspiration and expiration associated with neuronal events (Bellingham, 1998) that we expect to differ as a function of 2) whether or not single maximum are temporally associated with speech (operationalized as IPU) onset.

First, we used the log files related to fMRI recording to identify single block onsets, then added this value to the onsets of participant's produced IPU in this block (i.e. the moment the

participant takes its turn of speech in the conversation) to obtain for each run across blocks the onset of each speech-turn from the scanned participant during the conversations. The respiration maxima temporally closest (in absolute value) from every IPU onset was categorised as "*Respiration maxima associated with speech production*" (Resp+), while all others were attributed to "*Respiration maxima not associated with speech production*" (Resp-). Attribution of single Resp+ events to every IPU event allowed us to calculate the relative time difference between these two events  $\Delta t$ .

## Analyses

Analyses were performed across the 4 runs of the 25 participants included in the corpus. For the behavioural analysis, we used MATLAB® R2020a's function *rmoutliers* to identify and remove outliers from  $\Delta t$  values, i.e. values more than three median absolute deviations from the median. The missing data is below 5%, allowing the use of all data in the analysis. Kolmogorov-Smirnoff tests of normality were run on the resulting cleaned  $\Delta t$  values on the final time values distribution and reached the significance level ( $p < 0.05$ ) for most participants and runs, indicating that the data is not normally distributed. We therefore performed skewness tests to measure the asymmetry of the distribution.

For each participant and run, BOLD data was analysed with SPM12 with three conditions, Image presentation, that will not be discussed furthermore, and the two conditions of interest, Resp+ and Resp- modelled as events (i.e. zero duration). All nuisance variables calculated with the PhysIO toolbox were introduced in these models, allowing to remove artefacts associated with physiological (breathing and heartbeat) noise, movement artefacts and detrending based on linear as well as regions-of-interest signal (Kasper et al., 2017).

Given that Resp+ events are temporally synchronised with IPUs onset, therefore hindering the distinctions between respiration and articulatory correlates of speech, we used an exclusive masking procedure to remove brain activity associated with speech production in our corpus. To do so, we performed another analysis with conditions "*IPUs from participant*" (IPU+) and "*IPUs from interlocutor*" (IPU-), describing each produced and perceived speech turn by its onset and its duration. The activation map corresponding to the contrast IPU+ versus IPU-, assessed at the second level (with the same parameters as in the main analysis presented in the next paragraph;  $p_{FWE} < 0.001$ ) was saved as a binary image, and expanded by two voxels in all directions (using MRIcron), to avoid border effects, to create a mask including brain areas involved in articulatory aspect of speech.

For the second-level analysis of interest, individual beta estimates from the two conditions of interest Resp+ and Resp- were used in a full ANOVA model using SPM12, with participants and runs as repeated measures (unequal variance for the former, equal for the later), using

the mask described previously to exclude from analysis brain regions associated with speech production. Results for the contrast Resp+ versus Resp- are provided with a threshold of  $p_{FWE} < 0.05$  at the peak level and  $p_{FDR} < 0.05$  at the cluster level (corresponding to an extent  $k > 5 \text{ cm}^3$ ).

### Central sulcus localization

We conducted a single subject and single run analysis to identify more precisely the location of the central sulcus cluster in each hemisphere, in particular to distinguish between precentral and postcentral localizations. Significant activity ( $p_{unc} < 0.001$ ) was found within a 1 cm radius around the group analysis peak in 145 out of 200 possibilities (25 participants x runs x 2 hemispheres) *i.e.* almost 3/4. We developed a procedure to minimise subjective biases to localise the respective position of the peak of individual activity in respect to the position of the central sulcus. Human intervention was minimised to two aspects of the procedure, while automatic MATLAB® and python scripts were used to draw the central sulcus and overlay the central sulcus drawing on the activity map for each participant, run and hemisphere in which activity was found.

We first saved an axial section of the normalised structural image made at the height of the peak of the identified cluster, usually around  $z = 60 \text{ mm}$  in MNI coordinates, without any activity overlaid. Three independent observers traced the central sulcus on these images. The absence of any discrepancy between the observers validated the identification of the central sulcus on each structural section. The images were then cropped manually with Photoshop to contain only the central sulcus region identified. The tracing of the central sulcus was semi-automated using *ad hoc* python scripts that draw the sulcus based on the grey value intensity of the anatomical image. To bootstrap sulcus drawing, a starting point was selected manually. As the sulcus never reaches the midline at this level of the brain, we identified the pixel that could still be identified as grey matter found at the fundus of the central sulcus closest to the midline. A python script draws a line corresponding to the central sulcus iteratively by looking for the adjacent pixel with the lowest intensity centrifugally.

Finally, a MATLAB® script overlaid the line drawing corresponding to the central sulcus to the functional image in which the peak location of the activated cluster is indicated by a cross yielding a series of images allowing us to categorise the respective location of the individual activation peaks and the central sulcus., with four possibilities: 1) the cross falls on the line, 2) at the fundus when it is close (less the one pixel) from the starting point chosen arbitrarily, and for all others, 3) anterior or 4) posterior to the central sulcus.

## Results

### Behavioural analysis

Analysis of skewness in  $\Delta t$  revealed a significant left-skewed distribution, with a maximum around -200 ms (-0.2003 ms) illustrated on [Figure 1](#). This indicates that a maximum breathing intake occurred on average 200 ms before participants speech onset. The finding of maximum breath following speech onset (in red on Figure 1) could be surprising, but it could be due to timing errors associated with both the breathing maximum estimate, as interpolation of saturated signal could fail to reproduce some aspects of the breathing signal such as the asymmetry between the dynamics of inspiration and expiration (Winkworth et al., 1995). Additionally, timing of IPU onset also suffers from possible timing errors depending on the type of phonemes produced at the beginning of the speech production. Importantly, while these are intrinsic limitations of the current investigation, they are unlikely to interfere with fMRI results given its poor temporal limitation. the statistical power of the fMRI analysis comes from the repetition of events of interest, which combine all respiration cycles recorded during the one acquisition run, *i.e.* several hundreds of breathing maxima for an approximate 7-min & 45-sec runs, with variance in the timing between events related to other behavioural factors such as the presence or absence of speech.

### Second level analysis of respiration maxima related to IPUs

Second level contrast Resp+ *versus* Resp- ( $p_{FWE} < 0.05$ , extend  $k > 5 \text{ cm}^3$ ) exclusively masked with the contrast IPU+ *versus* IPU- ( $p_{FWE} < 0.001$ ) reveal bilateral activations in the central sulcus, in the brainstem and in the dorsal and ventral cerebellum (see [Table 1](#) and [Figure 2](#)). Note that the reverse contrast revealed no activations at the thresholds used. Attribution of cerebellum clusters to the primary fissure and hemispheric lobule VIII for the dorsal and ventral clusters respectively relied on a cerebellum atlas allowing the use of MNI coordinate system with visual inspection of the cluster with regards to cerebellar folia boundaries (Schmahmann, 2000).

The precise attribution of the respiration clusters to pre- or post-central gyrus, triggered by the observation of [Figure 2](#) (top left), was analysed and results are illustrated in [Figure 3](#) (single participant and runs images are provided in [Supplementary figure 2](#)). Although these results suggest that clusters maxima are mainly found in the postcentral gyrus, a comparable number of maxima were attributed to the anterior or fundus region of the central sulcus. But it has been reported that anatomical (the fundus of the sulcus) and functional (transition for primary sensory to primary motor cortices) don't match and the boundary can extend towards the anterior precentral gyrus (see *e.g.* Fig. 10 in White et al., 1997). By extrapolation, considering that all postcentral gyrus clusters should be attributed to the primary sensory

cortex and even a fraction of the clusters associated with the other cortical regions could also belong to this functional area, we propose that the bilateral central sulcus activity reflects a neural correlate of sensory processing related to Inspiration-to-Expiration transitions.

Finally, localisation of functional areas in the brainstem is particularly challenging given the small size of brainstem nuclei compared to the fMRI voxel size, uncertainties about interindividual variability and the absence of a reference atlas. A post-mortem brainstem atlas indicated that the brainstem cluster is located between the cerebellocortical and thalamocortical bundles identified with tractography (Adil et al., 2021). Circumscribed by white fibre tracts and spanning medullary and pontine regions, this bilateral cluster, elongated along the main axis of the brainstem identified with a contrast investigating maximum respiration events, likely belongs to the human ventral reticular formation where is located the homologue of the pre-Bötzinger complex identified in rodents (Smith et al., 1991).

## Discussion

The aim of the present study is to investigate the neurophysiological underpinnings of speech-related respiration events in conversational turn-taking. We analyse an existing corpus of natural conversations between a participant and its interlocutor (Human or Robot) focusing on synchronised (1) behavioural (conversation turn-taking), (2) respiratory (maxima of inspiration and expiration) and (3) neurophysiological (fMRI) recordings. Each participant's speech production during the conversation was recorded, denoised and segmented into silence and speech turns. Participants' speech turns (or InterPausal Units, IPU) were defined as speech blocks occurring between silences, each lasting at least 200 ms (Blache et al., 2009).

### Respiration maximum and speech onset

We first investigated the temporal relation between respiration maxima and speech production. The closest respiration maximum, in absolute value, was associated with each IPU, hence referred to as "Respiration maxima associated with IPU" or "Resp+". We further investigated the distribution of temporal associations between expiration and vocal production by focusing on the relative time differences between the two categories of events. As illustrated in [Figure 1](#) with the cumulative histogram across all participants, runs, trials and turns indicated that maximum respiratory intake occurred on average 200 ms prior to speech onset. The variability could be due to the accumulation of temporal noise, both from the fMRI recordings precluding millisecond precision in the identification of participant's speech onset and from the saturation of the recording belt signal that occasionally required extrapolation. Nevertheless, [Supplementary figure 1](#) shows that the blue bar left to the 0 ms (interval [-500; 0] ms) has the highest number of observations for most participants and runs.

This result aligns with studies indicating a temporal gap between the cessation of air intake (respiration maximum) preceding the onset of speech by 200 ms (MacIntyre & Scott, 2022). Other studies have considered various factors. For example, one study found that inhalation depth correlates with the speech envelope's total power and relates to upcoming speech, suggesting internal models for enhanced communication (Abbasi et al., 2023). Another study highlighted that respiration and pauses are crucial for turn-taking, showing that successful turns often occur right after a new inhalation, indicating that speakers coordinate their breathing with turn-taking (Rochet-Capellan & Fuchs, 2014).

Altogether, the temporal relationship between speech and respiration appears to be preserved despite the lack of naturalness experienced when speaking while lying supine in a scanner during a noisy fMRI acquisition. For the fMRI analysis, we defined "Resp-" events as "Respiration maxima not associated with IPU". Together, Resp+ and Resp- identify all respiration maxima, and are used as events for the fMRI analysis. Note that the physiological

data (cardiac and respiratory cycles) were processed using the PhysIO toolbox (Kasper et al., 2017) and used to derive nuisance variables describing noise induced by heart pulsation, respiratory volume as well as individual participants' movements derived from functional magnetic resonance images (fMRI) realignment. Physiological artefacts being removed with this denoising procedure, the results obtained reflect central nervous system respiratory control associated with conversational speech onset and not artefacts derived from peripheral physiology (Raitamaa et al., 2021).

The physiological aspects of speech breathing have shown muscular components, represented in the motor cortex, that contribute to the regulation of breathing required for voluntary control and sound production (McKay et al., 2003; Winkworth et al., 1995), underlining its clear differentiation from metabolic breathing. The unique data used here, a natural conversation, allows us to further the understanding of breathing control in a social interaction. To identify brain correlates of speech-related breathing maximum, we compared events described as Resp+ to Resp- using a classical preprocessing of data and two-stage analysis in SPM12 (Penny et al., 2007). As this contrast yielded results clearly associated with speech production (in premotor cortices in particular), a masking procedure was used. A mask describing the main effect of speech production was created from an independent analysis of the same data defining each IPU as a block to identify. Using this mask to exclude speech associated brain regions at the population level, allowed us to identify areas significantly more activated in the contrast Resp+ *versus* Resp- ( $p_{FWE} < 0.001$ , extend  $k > 5 \text{ cm}^3$ ) and not associated with speech production. This analysis revealed bilateral activations in the dorsal and ventral cerebellum, in the brainstem and in the central sulcus, all these clusters being further described in [Table 1](#) and illustrated in [Figure 2](#).

### **Speech-associated respiration maxima and central sulcus response**

Several regions around the central sulcus are involved in modulating respiratory activities. Studies have shown that specific regions are activated during the initiation and coordination of voluntary respiratory movements. These regions send signals to brainstem structures involved in regulating respiratory rhythm such as the inferior olive bilaterally (Sears et al., 1995). The central sulcus is also involved in speech production. Specifically, the primary motor cortex is responsible for the control muscles necessary for producing speech sounds. It sends signals that are transmitted through regions of the brainstem (though the exact pathways remain elusive, see Belyk et al., 2021) to control precise movements of the laryngeal muscles, vocal folds, and articulations involved in speech. Motor cortices are therefore essential in coordinating and regulating speech-producing movements.

The exact location of the cluster at the fundus of the central sulcus (see Figure 2) made it difficult to attribute the response across participants to the pre- or post-central gyri,



especially given that the boundary between cytoarchitectonic Brodmann Areas 3 (primary somatosensory cortex in the postcentral gyrus) and 4 (primary motor cortex in the precentral gyrus) is at the fundus of the central sulcus (White et al., 1997). Hypothesising that the data smoothing associated with DARTEL normalisation and group-level analysis could explain the impossibility of a clear attribution, we analysed the localization of the cluster in single participants and single runs with unsmoothed anatomical and functional data in a semi-automatic fashion. Results showing the relative position individual participants' cluster of activity in the central sulcus drawn (shown in [Supplementary figure 2](#) and summarised in [Figure 3](#)) reveal that a majority of activated clusters for the contrast Resp+ *versus* Resp- are found in the post-central sulcus. As this would locate them in primary sensory cortices, the central sulcus cluster would be related to sensory processing, and not motor control. While speculative, these results suggest that sensory information from the chest upper body area, for example the diaphragmatic aspects of respiratory functions, could be processed in this primary sensory region.

### **Speech-associated respiration maxima response in the brainstem**

This sensory component found in the central sulcus might be associated with "perceiving" the maximum intake of oxygen in the lungs required for producing subsequent speech. This could in turn also explain increased response for Resp+ events in brainstem regions associated with control of breathing. The cluster was attributed to the ventral respiration controller preBötzing complex, which contains both inspiration and expiration neurons. This complex has multiple subcortical, but very limited cortical, projections in rodents (Yang & Feldman, 2018).

The main contribution of preBötzing complex is to generate rhythmic respiratory patterns (Smith et al., 1991). Studies suggest that connections indeed exist between the motor cortex and brainstem regions involved in respiratory regulation that include the preBötzing complex as well as other interconnected pontine and medullary nuclei such as the nucleus ambiguus, the inferior olive or the periaqueductal grey matter also involved in vocal production (Zhang & Ghazanfar, 2020). These connections enable bidirectional communication between the cortex and the brainstem to coordinate respiratory movements necessary for speech production (Belyk et al., 2021). The interactions may occur through descending projections from the motor cortex to the brainstem, regulating preBötzing complex activity and influencing respiratory control. Reciprocally, signals from several brainstem nuclei project back to the cortex and can carry sensory information about the current phase of the respiratory cycle given the level of inspiration.

### **Speech-associated respiration maxima and cerebellar response**



This temporal dynamic is reflected in the third result of the fMRI analysis regarding the attribution of cerebellum clusters in the bilateral primary fissure (V, VI) and right hemispheric lobule VIIa and VIIb more activated in the condition Resp+.

The cerebellum, a brain structure traditionally associated with motor coordination and balance, has increasingly drawn attention for its significant role in speech production and perception (Ackermann et al., 2007). They suggest that the cerebellum plays a crucial role in regulating speech rate, with evidence indicating its involvement in pushing speech rates beyond 3 Hz, emphasising its importance in the temporal organisation of verbal utterances. Neuropsychological studies confirm its involvement in various speech-related conditions, such as agrammatism and transcortical motor aphasia in patients (Ackermann et al., 2007). Transcranial stimulation of the right cerebellum improved naming abilities in individuals with aphasia (Sebastian et al., 2020). Moreover, ataxia, a condition characterised by cerebellar impairment, has been associated with significant deficits in spontaneous speech reducing intelligibility and naturalness (Hilger et al., 2022). In addition, damage to cerebellar fibre tracts in the lobules VI of the cerebellum may further compromise speech motor control, contributing to ataxic dysarthria, a speech disorder characterised by difficulty in articulation and pronunciation (Lehman et al., 2020). Other studies highlight the functional convergence of motor and social processes in lobule IV/V, suggesting a broader role in cognitive and social behaviours (Chao et al., 2021). Lobule VIII is implicated in language processing, with studies reporting correlations between greater grey matter volume in lobules VII and VIII and better performance on measures of language (Stoodley & Schmahmann, 2018).

Importantly, one function repeatedly associated with the cerebellum is the timing associated with motor control (Tanaka et al., 2021). Given the importance of the relative timing between the events under investigation, the maximum respiration intake, and the relative timing of a specific action, speech onset, it is proposed that the cerebellar responses that we report associated with maximum respiratory events are linked to the preparatory components of vocal production.

### **A cortico-brainstem-cerebellar network for speech turn-taking in natural conversation**

We therefore have three actors involved in the coordination between respiratory maximum and speech onset during natural conversation: the posterior part of the central sulcus dorsally, a region of the brainstem likely to be associated with brainstem respiratory pattern generator, and regions of the cerebellum that could be involved in temporal coordination of motor control. Increased response in the brainstem cluster, supposedly involved in the generation of breathing rhythms, likely reflects the active inhibition of the respiratory pattern generators that is required to enslave expiration to vocal production. Had this inhibition been passive, we would have expected an increased response in the brainstem

in the opposite contrast, focusing on maximum respiratory intake not associated with speech production, which yielded no significant response in the current analysis. The source of this inhibition could be the cerebellum clusters found in the same contrast, as it has been reported a direct inhibition of respiration by electric stimulation of Purkinje cells (reviewed in Xu & Frazier, 2002). Such interpretation is in agreement with the role of the cerebellum in the temporal aspects of motor control. Finally, the central sulcus cluster, attributed to the primary sensory cortex, could detect by afferents from upper airways the moment when the lungs are full and therefore ready to initiate speech. This important piece of information could also be integrated by brainstem and/or cerebellar components. Altogether, the relatively complex, but fully automatic, synchronisation of breathing with speech production involves the combination of several information reflected in the activation of regions belonging to different regions of the central nervous system.

## Conclusion

Our study confirmed a temporal dynamic between respiration and conversational dynamics, and identified several key actors involved in this dynamic at different levels of the central nervous system. These actors are the dorsal central sulcus, possibly corresponding to the primary sensory cortex encoding information coming from the chest level that can be associated with the filling of the lungs, the cerebellum, likely involved in computing the temporal dynamic between lung filling and speech onset, and the brainstem, where respiratory rhythm generators of the preBötzinger complex located in the ventral reticular formation must be inhibited to enslave respiration to speech production. While speculative, these interpretations are based on a large but parcellar literature that investigated the different actors in isolation or by pairs, especially in animal studies, complemented by neuropsychological results of speech impairments in humans reviewed in introduction. But to our knowledge, it is the first time that a corpus of natural conversation by humans including synchronised functional neuroimaging is used to identify the actors in relation to breathing and speech turn-taking. While further research will be necessary to investigate the dynamics of this central nervous system network, these results underscore the importance of integrating peripheral physiological signals to investigate complex behaviours in ecological settings.

## Acknowledgements

CDP is supported by the Institute for Language, Communication and the brain (ILCB ANR-16-CONV-0002). The research is supported by grants ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) and AAP-ID-17- 46-170301-11.1 by the Excellence Initiative of AixMarseille University (A\*MIDEX), a French "Investissement d'Avenir" programme and the Institute for Language, Communication and the Brain.

## Tables and figures

Table 1. Significant activations associated with the contrast Resp+ versus Resp- ( $p_{FWE} < 0.001$ , extend  $> 5 \text{ cm}^3$ ). MNI coordinates (in mm) are used to locate the clusters of significant increased activity.

<i>Location</i>		<i>Extend (voxels)</i>	<i>peak T-values</i>	<i>MNI coordinates (mm)</i>		
<i>Anatomy</i>	<i>Side</i>			<i>x</i>	<i>y</i>	<i>z</i>
<i>Central Sulcus (fundus, dorsal part)</i>	<i>R</i>	<i>819</i>	<i>16.13</i>	<i>20</i>	<i>-27</i>	<i>60</i>
	<i>L</i>	<i>681</i>	<i>15.28</i>	<i>-21</i>	<i>-30</i>	<i>63</i>
<i>Cerebellum (centred on Primary Fissure)</i>	<i>R</i>	<i>5676</i>	<i>21.36</i>	<i>14</i>	<i>-60</i>	<i>-20</i>
	<i>L</i>	<i>*</i>	<i>20.96</i>	<i>-15</i>	<i>-60</i>	<i>-18</i>
<i>Cerebellum (hemispheric lobule VIII)</i>	<i>L</i>	<i>763</i>	<i>14.38</i>	<i>-12</i>	<i>-66</i>	<i>-50</i>
	<i>R</i>	<i>1141</i>	<i>14.16</i>	<i>10</i>	<i>-68</i>	<i>-48</i>
<i>Brainstem (ventral reticular formation)</i>	<i>L</i>	<i>706</i>	<i>10.82</i>	<i>-8</i>	<i>-38</i>	<i>-45</i>
	<i>R</i>	<i>*</i>	<i>10.47</i>	<i>8</i>	<i>-38</i>	<i>-45</i>

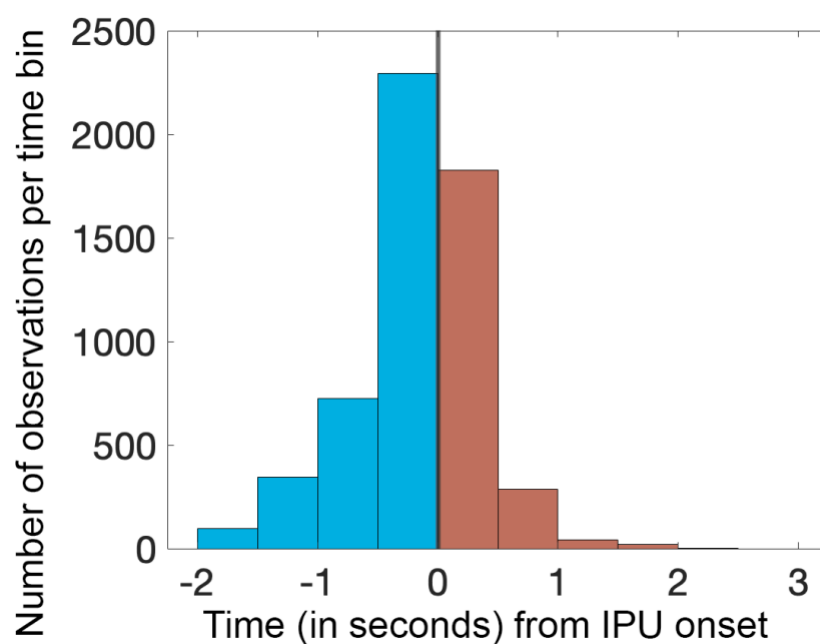


Figure 1. Time difference  $\Delta t$  between IPU onset and max breathings indicates a left skewness peaking at 200 ms, indicating that maximum of breath occurred on average 200 ms preceding participants speech onset.

Figure 1

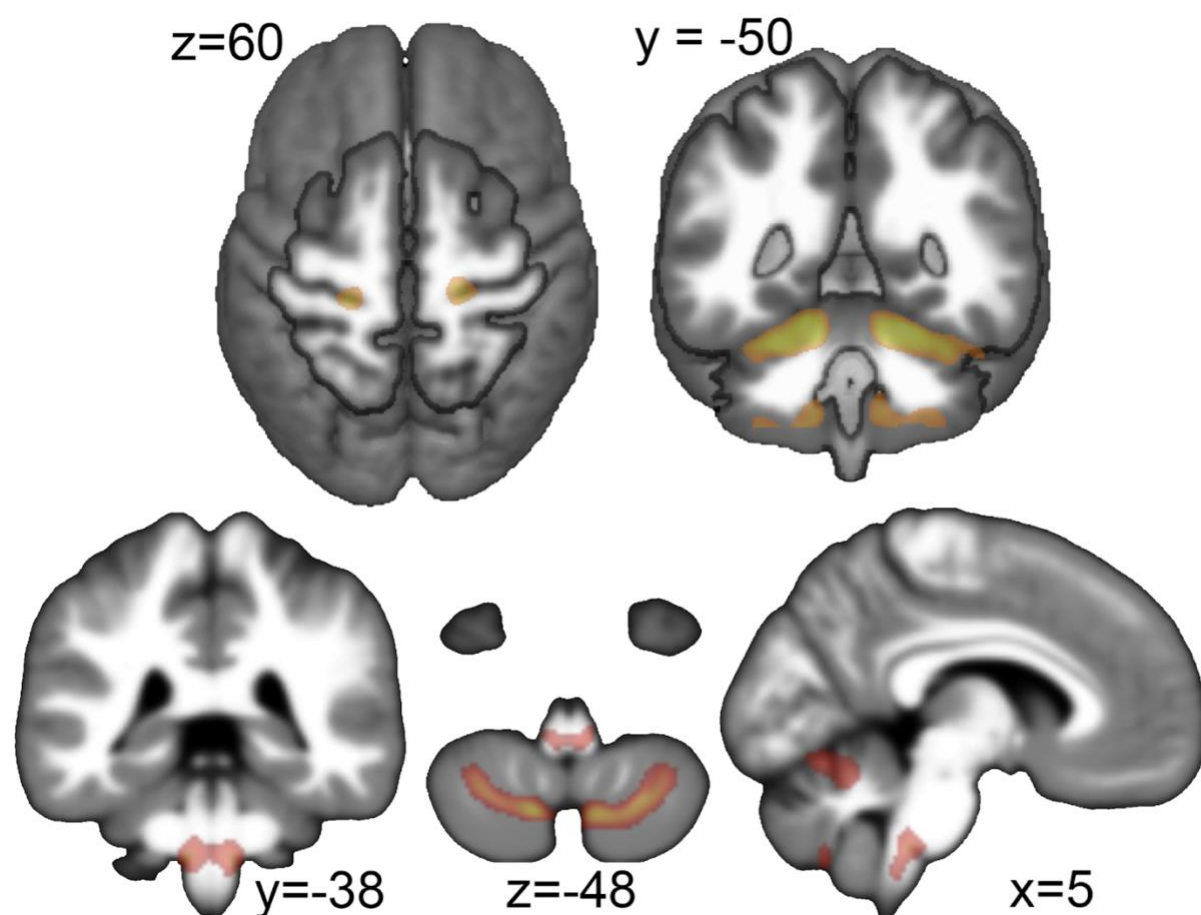


Figure 2: Brain activations for maximum breath intake associated with a speech onset (Resp+ versus Resp- [ $p_{FWE} < 0.001$ , extend  $> 5 \text{ cm}^3$ ] exclusively masked by IPU+ versus IPU-). Top left: the bilateral central sulcus fundus clusters. Top right: Dorsal and ventral (also see bottom, centre) cerebellar bilateral clusters. Bottom: coronal, axial and sagittal views of the brainstem cluster.

Figure 2

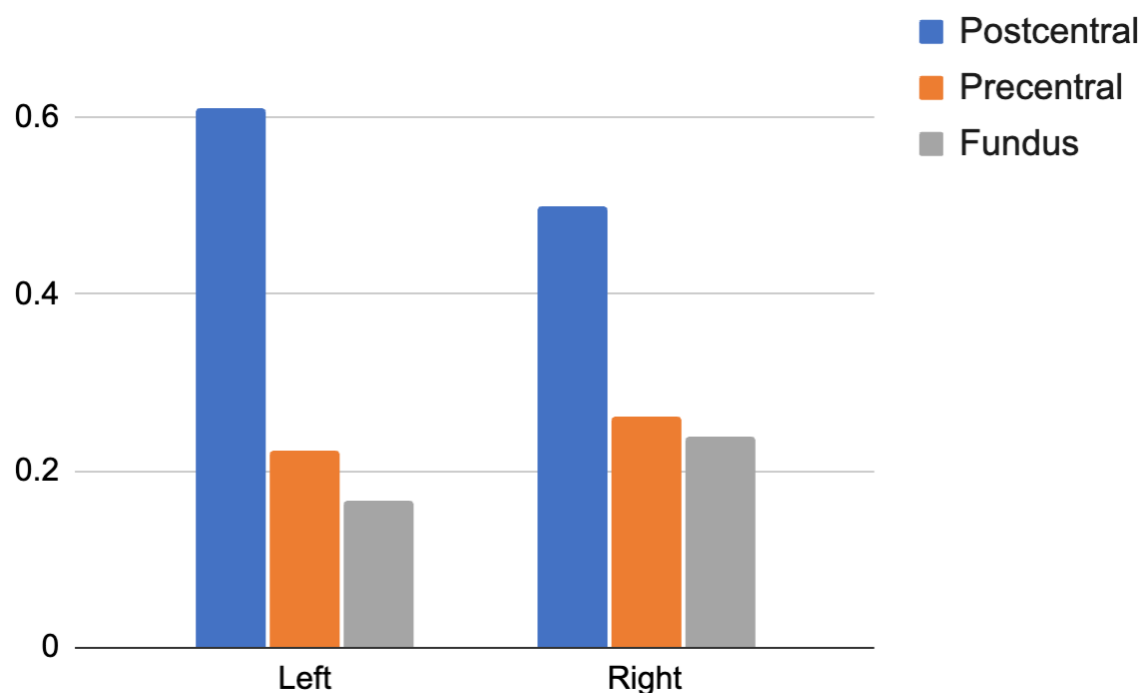


Figure 3: Individual participants and sessions contrast Resp+ *versus* Resp- ( $p_{\text{unc}} < 0.001$ ) within a 1-cm radius sphere centred on the group central cluster peak of activity were semi-automatically attributed to the fundus of the sulcus and the pre- and the post-central gyri. The figure presents the number of observations for the three possible cortical assignments.

Figure 3

## References

- Abbasi, O., Kluger, D. S., Chalas, N., Steingraber, N., Meyer, L., & Gross, J. (2023). Predictive coordination of breathing during intra-personal speaking and listening. *iScience*, 26(8), 107281. <https://doi.org/10.1016/j.isci.2023.107281>
- Ackermann, H., Mathiak, K., & Riecker, A. (2007). The contribution of the cerebellum to speech production and speech perception: Clinical and functional imaging data. *The Cerebellum*, 6(3), 202–213. <https://doi.org/10.1080/14734220701266742>
- Adil, S. M., Calabrese, E., Charalambous, L. T., Cook, J. J., Rahimpour, S., Atik, A. F., Cofer, G. P., Parente, B. A., Johnson, G. A., Lad, S. P., & White, L. E. (2021). A high-resolution interactive atlas of the human brainstem using magnetic resonance imaging. *NeuroImage*, 237, 118135. <https://doi.org/10.1016/j.neuroimage.2021.118135>
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1), 95–113. <https://doi.org/10.1016/j.neuroimage.2007.07.007>
- Bellingham, M. C. (1998). Driving respiration: The respiratory central pattern generator. *Clinical and Experimental Pharmacology and Physiology*, 25(10), 847–856. <https://doi.org/10.1111/j.1440-1681.1998.tb02166.x>
- Belyk, M., Brown, R., Beal, D. S., Roebroek, A., McGettigan, C., Guldner, S., & Kotz, S. A. (2021). Human larynx motor cortices coordinate respiration for vocal-motor control. *NeuroImage*, 239, 118326. <https://doi.org/10.1016/j.neuroimage.2021.118326>
- Betka, S., Adler, D., Similowski, T., & Blanke, O. (2022). Breathing control, brain, and bodily self-consciousness: Toward immersive digiceuticals to alleviate respiratory suffering. *Biological Psychology*, 171, 108329. <https://doi.org/10.1016/j.biopsycho.2022.108329>
- Bigi, B. (2015). SPPAS - MULTI-LINGUAL APPROACHES TO THE AUTOMATIC ANNOTATION OF SPEECH. *The Phonetician*, 111–112(ISSN:0741-6164), 54–69.
- Blache, P., Bertrand, R., & Ferré, G. (2009). Creating and Exploiting Multimodal Annotated Corpora: The ToMA Project. In M. Kipp, J.-C. Martin, P. Paggio, & D. Heylen (Eds.), *Multimodal Corpora: From Models of Natural Interaction to Systems and Applications* (pp. 38–53). Springer. [https://doi.org/10.1007/978-3-642-04793-0\\_3](https://doi.org/10.1007/978-3-642-04793-0_3)
- Chaminade, T. (2017). An experimental approach to study the physiology of natural social interactions. *Interaction Studies*, 18(2), 254–275.
- Chao, O. Y., Zhang, H., Pathak, S. S., Huston, J. P., & Yang, Y.-M. (2021). Functional Convergence of Motor and Social Processes in Lobule IV/V of the Mouse Cerebellum. *The Cerebellum*, 20(6), 836–852. <https://doi.org/10.1007/s12311-021-01246-7>
- Demartsev, V., Manser, M. B., & Tattersall, G. J. (2022). Vocalization-associated respiration patterns: Thermography-based monitoring and detection of preparation for calling. *Journal of Experimental Biology*, 225(5), jeb243474. <https://doi.org/10.1242/jeb.243474>



Hallart, C., Maes, J., Spatola, N., Prévot, L., & Chaminade, T. (2021). Comparaison linguistique et neuro-physiologique de conversations humain humain et humain robot. *Traitement Automatique Des Langues*, 61(3), 25.

Haouzi, P., Chenuel, B., & Barroche, G. (2006). Interactions between volitional and automatic breathing during respiratory apraxia. *Respiratory Physiology & Neurobiology*, 152(2), 169–175. <https://doi.org/10.1016/j.resp.2005.08.004>

Hilger, A., Cloud, C., & Fahey, T. (2022). Speech Impairment in Cerebellar Ataxia Affects Naturalness More Than Intelligibility. *The Cerebellum*, 22(4), 601–612. <https://doi.org/10.1007/s12311-022-01427-y>

Holstege, G., & Subramanian, H. H. (2016). Two different motor systems are needed to generate human speech. *Journal of Comparative Neurology*, 524(8), 1558–1577. <https://doi.org/10.1002/cne.23898>

Hudson, R. A. (1996). *Sociolinguistics* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139166843>

Kasper, L., Bollmann, S., Diaconescu, A. O., Hutton, C., Heinzle, J., Iglesias, S., Hauser, T. U., Sebold, M., Manjaly, Z.-M., Pruessmann, K. P., & Stephan, K. E. (2017). The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. *Journal of Neuroscience Methods*, 276, 56–72. <https://doi.org/10.1016/j.jneumeth.2016.10.019>

Lehman, V. T., Black, D. F., DeLone, D. R., Blezek, D. J., Kaufmann, T. J., Brinjikji, W., & Welker, K. M. (2020). Current concepts of cross-sectional and functional anatomy of the cerebellum: A pictorial review and atlas. *The British Journal of Radiology*, 93(1106), 20190467. <https://doi.org/10.1259/bjr.20190467>

MacIntyre, A. D., & Scott, S. K. (2022). Listeners are sensitive to the speech breathing time series: Evidence from a gap detection task. *Cognition*, 225, 105171. <https://doi.org/10.1016/j.cognition.2022.105171>

McKay, L. C., Evans, K. C., Frackowiak, R. S. J., & Corfield, D. R. (2003). Neural correlates of voluntary breathing in humans. *Journal of Applied Physiology*, 95(3), 1170–1178. <https://doi.org/10.1152/japplphysiol.00641.2002>

Penny, W., Friston, K., Ashburner, J., Kiebel, S., & Nichols, T. (2007). Statistical Parametric Mapping: The Analysis of Functional Brain Images. In *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. <https://doi.org/10.1016/B978-0-12-372560-8.X5000-1>

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, 62(2), 816–847. <https://doi.org/10.1016/j.neuroimage.2012.04.062>

Pulvermüller, F., Huss, M., Kherif, F., Moscoso Del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, 103(20), 7865–7870. <https://doi.org/10.1073/pnas.0509989103>

Raitamaa, L., Huotari, N., Korhonen, V., Helakari, H., Koivula, A., Kananen, J., & Kiviniemi, V. (2021). Spectral analysis of physiological brain pulsations affecting the BOLD signal. *Human Brain Mapping*, 42(13), 4298–4313. <https://doi.org/10.1002/hbm.25547>

Rauchbauer, B., Hmamouche, Y., Bigi, B., Prevot, L., Ochs, M., & Chaminade, T. (2020). Multimodal corpus of bidirectional conversation of human-human and human-robot interaction during fMRI scanning. *Proceedings of The 12th Language Resources and Evaluation Conference*, 661–668.

Rauchbauer, B., Nazarian, B., Bourhis, M., Ochs, M., Prévot, L., & Chaminade, T. (2019). Brain activity during reciprocal social interaction investigated using conversational robots as control condition. *Philosophical Transactions of the Royal Society B*, 374(1771), 20180033.

Rochet-Capellan, A., & Fuchs, S. (2014). Take a breath and take the turn: How breathing meets turns in spontaneous dialogue. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658), 20130399. <https://doi.org/10.1098/rstb.2013.0399>

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36(4), 393–414. <https://doi.org/10.1017/s0140525x12000660>

Schmahmann, J. D. (Ed.). (2000). *MRI atlas of the human cerebellum*. Academic Press.

Sears, T. A., Magzoub, S., & Seers, C. (1995). The Representation of Breathing Movements in the Cat Inferior Olive. In A. Taylor, M. H. Gladden, & R. Durbaba (Eds.), *Alpha and Gamma Motor Systems* (pp. 356–364). Springer US. [https://doi.org/10.1007/978-1-4615-1935-5\\_76](https://doi.org/10.1007/978-1-4615-1935-5_76)

Sebastian, R., Kim, J. H., Brenowitz, R., Tippet, D. C., Desmond, J. E., Celnik, P. A., & Hillis, A. E. (2020). Cerebellar neuromodulation improves naming in post-stroke aphasia. *Brain Communications*, 2(2), fcaa179. <https://doi.org/10.1093/braincomms/fcaa179>

Smith, J. C., Ellenberger, H. H., Ballanyi, K., Richter, D. W., & Feldman, J. L. (1991). Pre-Bötzinger Complex: A Brainstem Region that May Generate Respiratory Rhythm in Mammals. *Science*, 254(5032), 726–729. <https://doi.org/10.1126/science.1683005>

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592. <https://doi.org/10.1073/pnas.0903616106>

Stoodley, C. J., & Schmahmann, J. D. (2018). Functional topography of the human cerebellum. In *Handbook of Clinical Neurology* (Vol. 154, pp. 59–70). Elsevier. <https://doi.org/10.1016/B978-0-444-63956-1.00004-7>

Tanaka, M., Kunimatsu, J., Suzuki, T. W., Kameda, M., Ohmae, S., Uematsu, A., & Takeya, R. (2021). Roles of the Cerebellum in Motor Preparation and Prediction of Timing. *Neuroscience*, 462, 220–234. <https://doi.org/10.1016/j.neuroscience.2020.04.039>

Torreira, F., Bäckström, S., & Levinson, S. C. (2015). Breathing for answering: The time course of response planning in conversation. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00284>

Tremoureux, L., Raux, M., Hudson, A. L., Ranohavimparany, A., Straus, C., & Similowski, T. (2014). Does the Supplementary Motor Area Keep Patients with Ondine's Curse Syndrome Breathing While Awake? *PLoS ONE*, 9(1), e84534. <https://doi.org/10.1371/journal.pone.0084534>

Trevizan-Baú, P., Stanić, D., Furuya, W. I., Dhingra, R. R., & Dutschmann, M. (2024). Neuroanatomical frameworks for volitional control of breathing and orofacial behaviors. *Respiratory Physiology & Neurobiology*, 323, 104227. <https://doi.org/10.1016/j.resp.2024.104227>

Viemari, J.-C., Menuet, C., & Hilaire, G. (2013). Complexe de pré-Bötzinger et automatisme respiratoire: Identification électrophysiologique, moléculaire et génétique d'une structure cruciale pour la respiration. *Médecine/Sciences*, 29(10), 875–882. <https://doi.org/10.1051/medsci/20132910015>

White, L. E., Andrews, T. J., Hulette, C., Richards, A., Groelle, M., Paydarfar, J., & Purves, D. (1997). Structure of the human sensorimotor system. I: Morphology and cytoarchitecture of the central sulcus. *Cereb Cortex*, 7(1), 18-30.

Whitfield-Gabrieli, S., & Nieto-Castanon, A. (2012). Conn: A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks. *Brain Connectivity*, 2(3), 125–141. <https://doi.org/10.1089/brain.2012.0073>

Winkworth, A. L., Davis, P. J., Adams, R. D., & Ellis, E. (1995). Breathing Patterns During Spontaneous Speech. *Journal of Speech, Language, and Hearing Research*, 38(1), 124–144. <https://doi.org/10.1044/jshr.3801.124>

Xu, F., & Frazier, D. T. (2002). Role of the cerebellar deep nuclei in respiratory modulation. *The Cerebellum*, 1(1), 35–40. <https://doi.org/10.1080/147342202753203078>

Yang, C. F., & Feldman, J. L. (2018). Efferent projections of excitatory and inhibitory preBötzinger Complex neurons. *Journal of Comparative Neurology*, 526(8), 1389–1402. <https://doi.org/10.1002/cne.24415>

Zhang, Y. S., & Ghazanfar, A. A. (2020). A Hierarchy of Autonomous Systems for Vocal Production. *Trends in Neurosciences*, 43(2), 115–126. <https://doi.org/10.1016/j.tins.2019.12.006>