Check for updates

RESEARCH ARTICLE

Investigating the cortical mechanisms of conspecific voice perception in anesthetized marmosets using 3T fMRI

[version 1; peer review: 2 approved with reservations]

Manon Obliger-Debouche[1], Régis Trapeau[1], Lina Robinet[1], Julien Sein[1], Bruno Nazarian[1], Jean-Luc Anton[1], Emilie Rapha[1], Luc Renaud[1], Sabrina Ravel[1], Pascal Belin[2]

[1]Aix-Marseille Université, UMR 7289 CNRS, Institut de Neurosciences de la Timone, Marseille, Provence-Alpes-Côte d'Azur, 13005, France
[2]Aix-Marseille Université, Institute of Language Communication and the Brain, Aix-en-Provence, Provence-Alpes-Côte d'Azur, 13100, France

v1 **First published:** 09 May 2025, **5**:127
https://doi.org/10.12688/openreseurope.19059.1
Latest published: 09 May 2025, **5**:127
https://doi.org/10.12688/openreseurope.19059.1

Abstract

Background

Functional MRI studies have shown the existence of cortical areas selective for conspecific vocalizations ('voice patches') in the temporal lobe of both humans and macaques. Two recent studies using high-field small-animal fMRI also both evidenced anterior temporal voice patches in marmoset monkeys, suggestive of an evolutionary-old 'primate voice patch system'. Here we asked whether similar findings could be observed at lower (3T) field strength in a conventional whole-body scanner.

Methods

Fifteen (n=15) adult marmosets were scanned using a 16-channel marmoset head coil and a dedicated 'clustered-sparse' sampling design during auditory stimulation with conspecific vocalizations and scrambled control stimuli while under sevoflurane-induced anesthesia.

Results

At the group level, we observed bilateral and symmetrical activation of auditory cortical and subcortical areas when contrasting the auditory stimulation conditions to the silent baseline. At the individual level

Open Peer Review

Approval Status ? ?

	1	2
version 1 09 May 2025	? view	? view

1. **Alessandro Zanini**, University of Western Ontario, London, Canada

2. **Francesca Morfini**, Northeastern University, Boston, USA

Any reports and responses or comments on the article can be found at the end of the article.

auditory activation was more variable but sufficiently reliable to provide a valuable coincidence map of areas activated in common by seven different individuals. When contrasting the vocalizations to the scrambled controls, we found in these seven monkeys a bilateral activation of the anterior and temporal cortex.

Conclusions

Despite methodological optimizations, our study encountered several limitations, including a low signal-to-noise ratio (SNR) and reduced spatial resolution inherent to 3T fMRI, and the use of anesthesia, which decrease functional connectivity and activation reliability. However, we identified the same brain regions as those documented in the literature. Our results suggest that fMRI at 3T in clinical studies can provide clues to identify the networks involved in voice perception in marmosets – although they will be explored in much finer detail using small animal, ultra-high-filed and awake fMRI.

Plain Language Summary

This study investigated whether voice-sensitive cortex can be observed in marmoset monkeys using conventional human 3T scanners. Previous research using more advanced MRI techniques had identified such "voice patches" in primates, suggesting an ancient system for voice processing.

A large group of marmosets (15) was scanned under anesthesia while listening to natural vocalizations and their scrambled versions, using a dedicated marmoset head coil. Results show that, at the group level, auditory brain areas were consistently activated. However, individual results varied. In seven monkeys, a specific area in the anterior temporal cortex responded more to vocalizations than to scrambled sounds, similar to past studies using ultra-high-field small animal scanners.

Despite the intrinsic limitations of the 3T whole-body scanner and of the anesthesized state, the findings suggest that even human MRI scanners can help identify brain networks involved in voice perception in marmosets, though more precise studies will be needed for finer details.

Keywords
voice processing, fMRI, auditory cortex, primate voice patch system, marmosets

H2020 This article is included in the Horizon 2020 gateway.

This article is included in the European Research Council (ERC) gateway.

Corresponding author: Manon Obliger-Debouche (obliger.manon@gmail.com)

Author roles: Obliger-Debouche M: Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Trapeau R**: Conceptualization, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – Review & Editing; **Robinet L**: Investigation; **Sein J**: Investigation, Methodology, Resources, Software, Writing – Review & Editing; **Nazarian B**: Methodology, Writing – Review & Editing; **Anton JL**: Methodology; **Rapha E**: Investigation; **Renaud L**: Investigation; **Ravel S**: Supervision, Writing – Original Draft Preparation; **Belin P**: Conceptualization, Data Curation, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 788240 (project name Covoprim); and ANR- 16- CONV- 0002 (Institute for Language, Communication and the Brain) (P.B.) and ANR- 11- LABX- 0036 (Brain and Language Research Institute)
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Obliger-Debouche M *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Obliger-Debouche M, Trapeau R, Robinet L *et al.* **Investigating the cortical mechanisms of conspecific voice perception in anesthetized marmosets using 3T fMRI [version 1; peer review: 2 approved with reservations]** Open Research Europe 2025, **5**:127 https://doi.org/10.12688/openreseurope.19059.1

First published: 09 May 2025, **5**:127 https://doi.org/10.12688/openreseurope.19059.1

Introduction

In the late 1990s, the first evidence of voice-selective cortical areas in humans was obtained using functional magnetic resonance imaging (fMRI) (for review see Belin (2018)). This series of studies provided significant evidence of specialized brain regions for processing vocal sounds in humans, named temporal voice areas (TVAs) or "voice patches". Voice patches are not exclusive to humans and have also been observed in other species. Specifically, they have been observed in macaques, suggesting that they already existed in some rudimentary form in the last common ancestor of macaques and humans, indicating a long evolutionary history. Studies using fMRI and positron emission tomography (PET) have identified macaques' TVAs as those that respond preferentially to conspecific vocalizations in the superior temporal gyrus (STG) (Bodin *et al.*, 2021; Gil-da-Costa *et al.*, 2004; Gil-Da-Costa *et al.*, 2006; Joly *et al.*, 2012; Ortiz-Rios *et al.*, 2015; Petkov *et al.*, 2006; Petkov *et al.*, 2008; Poremba *et al.*, 2004).

In recent years, marmosets have gained popularity in neuroimaging studies due to their small size (Belin *et al.*, 2023; Miller *et al.*, 2016), which enables the use of high-field small animal MRI with much higher signal and spatial resolution. In 2015, a study was conducted using fMRI at 7T on six anesthetized marmosets to investigate the existence of conspecific voice sensitive areas (Sadagopan *et al.*, 2015). The results showed that regions along the lateral sulcus, near the temporal pole, exhibited a preference for conspecific vocalizations (CVs), with the most rostro-lateral region showing the highest preference. These findings are consistent with similar cortical specializations and CV sensitivity as observed in macaques, suggesting a homologous structure-function relationship in the processing of conspecific vocalizations in primates. This suggests the evolutionary conservation of cortical specialization for vocalization processing, with indications pointing to a shared ancestor approximately 40 million years ago. Additional detail was provided by a recent study (Jafari *et al.*, 2023) that used ultra-high field fMRI (9.4T) in five awake marmosets. The results reveal an extensive voice-selective network, including symmetrical regions in the temporal cortex and subcortical areas, exhibiting heightened activity during conspecific vocalizations compared to either nonvocal sounds or scrambled vocalizations. Additionally, by utilizing open-access databases, the authors uncovered robust connections between temporal and cingulate voice-selective regions. This study suggests that the marmoset's vocal brain offers a more straightforward model for understanding human cerebral voice processing, particularly with more invasive techniques like fMRI-guided electrophysiology and reinforce the notion of evolutionarily conserved cortical specialization for vocalization processing. Such findings underscore the potential of marmosets in advancing our understanding of the neural mechanisms underlying vocal processing (Belin *et al.*, 2023).

Although clinical MRI scanners are more accessible than preclinical high-field MRI scanners, facilitating their use for a variety of studies, research conducted at 3T on such MRI scanners is limited to anatomical imaging. Building on this knowledge, the aim of this study is to replicate previous results (Jafari *et al.*, 2023; Sadagopan *et al.*, 2015) in a more traditional 3T scanner with a large magnetic field. To compensate for the low magnetic field intensity, this research opts for a larger number of animals (n=15), a marmoset-specific 16-channel head antenna, a "clustered sparse" design optimized for auditory stimulation, and the use of a contrast agent (MION). This study is part of a larger project called Covoprim, which aims to compare fMRI brain responses in three distinct species: humans, macaques and marmosets (Belin *et al.*, 2018; Bodin *et al.*, 2021). Using the same scanner and the same sounds, this comparative method involving three species is necessary to obtain general conclusions about neural bases of voice perception and its evolutionary origin.

Methods

Animals

This study involved 15 adult offspring of common marmosets (*Callithrix jacchus*) (Table 1). These marmosets belonged to a colony consisting of three families and were housed in cages measuring 1.05 m × 0.85 m × 2 m in the same room. The marmosets were kept on a 12-hour light/dark cycle that began at 8 a.m., and the room was soundproofed to prevent exposure to human voices. The animals involved in the study were not reproductive parents or young marmosets, but rather adult offspring. Both the research team and the veterinary staff

Table 1. Characteristics of the 15 monkey subjects. The rightmost column indicates whether the monkeys received a contrast agent injection during the scanning process for the localizer protocol or not.

ID	Sex	Date of birth	MION
M1	F	2018/08/08	YES
M2	F	2019/01/11	YES
M3	M	2019/11/21	YES
M4	F	2021/02/13	YES
M5	F	2021/02/13	YES
M6	M	2021/07/21	YES
M7	F	2022/01/26	YES
M8	M	2022/05/02	YES
M9	F	2019/01/11	NO
M10	F	2019/11/21	NO
M11	M	2020/12/07	NO
M12	F	2021/07/21	NO
M13	M	2021/10/04	NO
M14	F	2021/10/04	NO
M15	M	2022/05/02	NO

monitored animals daily. Health was monitored by a general assessment of animal activity, and fur condition. All efforts were made to enhance the welfare of animals (i.e. social, structural and food enrichments). The ARRIVE reporting guidelines (Percie du Sert *et al.*, 2020) was used.

Auditory stimuli

A total of 48 stimuli were used, which were divided into two categories: marmoset vocalizations (n=24), and temporally matched scrambled versions of these vocalizations (n=24).

The marmoset vocalizations were unfamiliar to the monkeys and were kindly provided by Asif Ghazanfar (Ghazanfar & Liao, 2018). Within this category, the vocalizations consisted of affiliative (trill 25%, n = 6), neutral (phee 25%, n = 6, twitter 25%, n = 6), and agonistic (tsik 25%, n = 6) calls.

To generate the scrambled versions of the vocalizations their frequency distribution was randomly rearranged while maintaining the amplitude temporal envelope minimally changed (Belin *et al.*, 2002). Short-term Fourier transforms were performed with windows of 512 points (equivalent to 11.6 ms at a sampling rate of 44.1 kHz); amplitude components were scrambled within each window, only with other amplitude components, and the phase components were scrambled with other phase components; then an inverse Fourier transform was applied to reconstruct the scrambled signal (e.g., Figure 1B). The key characteristic of the scrambled sounds was that the amplitude values remained unchanged, preserving the temporal energy of the original signal. However, the spectral distribution of these sounds was altered, rendering them unrecognizable as vocalizations.

A 10-ms cosine ramp was applied to the onset and offset of all stimuli and they were normalized by their root mean square amplitude. Stimuli were processed digitally at a 48828 Hz sampling rate using TDT System 3 hardware (Tucker Davis Technologies, Alachua, FL, USA) and were delivered by an MRI-compatible loudspeaker (SPS-53-T00 Piezoelectric Miniature Speaker, Sonitron, Belgium) drived by a specific amplifier (PAA-LT3469-01, Sonitron, Belgium) at a sound pressure level of approximately 90 dB.

Scanning protocol

Functional neuroimaging studies of auditory processing face challenges, such as intense masking noise from the MR scanner (Hall *et al.*, 1999; Peelle, 2014; Petkov *et al.*, 2009). The continuous quasi-tonal stimulation created by this noise elevates

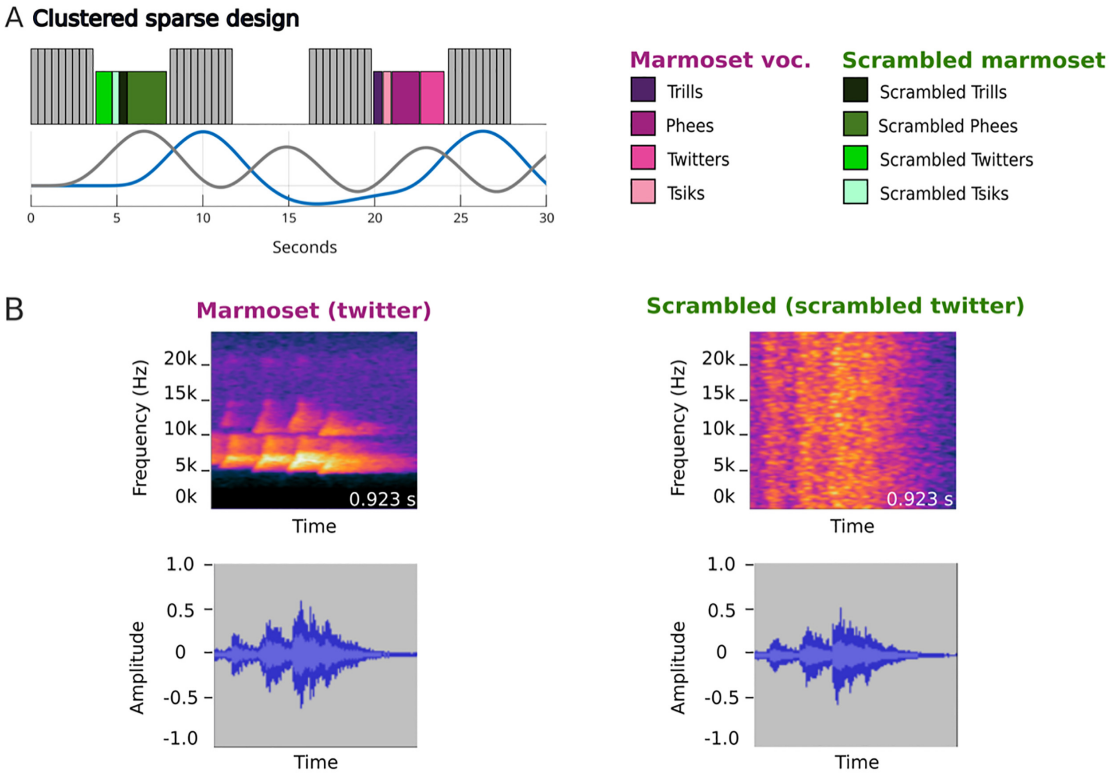


Figure 1. Marmoset Voice Localizer protocol. (**A**) Design of the acquisition protocol. Each acquisition block (in gray) was composed of seven volumes and lasted 3.92 s. Colored blocks correspond to the two main sound categories: marmoset vocalizations (in pink shade) and scrambled versions of vocalizations (in green shade). The lines represent the theoretical hemodynamic response to stimuli (in blue) or to the noise produced by the scanner (in gray). The sound categories are composed of 4 subcategories. (**B**) Two examples of sound stimuli, a twitter for the marmoset vocalization category and the scrambled version of the same twitter for the scrambled category.

the baseline activation level, hindering the detection of auditory activation in response to stimuli (Bandettini *et al.*, 1998; Shah *et al.*, 1997).

To address these challenges, we used a 'clustered-sparse' protocol aiming for an optimal compromise between number of brain volumes acquired (directly related to statistical power) and presentation of the stimuli on a silent background (to optimize voice/nonvoice perception). After pilot trials in human participants (Trapeau *et al.*, 2024), we converged to series of 7 successive brain volumes acquired in succession in 3.92s, followed by a silent interval of 4.1s during which a block of either 4 vocalizations, 4 scrambled vocalizations or silence was played. Seven volumes were acquired between blocks of stimuli for a duration of 3.92 s. (Figure 1A).

Monkeys M1 to M8 were scanned with ferrous oxide contrast agent (monocrystalline iron oxide nanoparticle, MION, 10 mg/kg), on the other hand, monkeys M9 to M15 were scanned without contrast agent and the acquisition was reduced to six volumes for a duration of 5.88 s. MION contrast agents enhance the signal-to-noise ratio (SNR) and augment the distinction between various tissue types (Gautama *et al.*, 2003). This results in a reduction of the hemodynamic response window, allowing for the acquisition of more volumes within the same session duration compared to sessions without MION.

All scans were performed under anesthesia with a protocol inspired by a previous study (Sadagopan *et al.*, 2015). Anesthesia was induced by injecting alfaxalone (7–10 mg/kg, Alfaxan, Jurox, Ireland), followed by 2% sevoflurane gas (Baxter, France) administration through a mask. During anatomical scanning, the concentration of sevoflurane gas was lowered to 1–1.5%, and during functional scanning, it was reduced to 0.5–1% by adding nitrous oxide (Air Liquide, France) at a 70/30 ratio to oxygen (Air Liquide, France) at a flow rate of 1.8 L/min. The animal's physiological characteristics, which includes heart and respiratory rate, capnometry, and oximetry, were monitored during the operation with sensors, and a warming blanket was employed to prevent hypothermia.

fMRI acquisition

Monkeys were scanned using a 3-Tesla scanner (Prisma, Siemens Healthcare, Erlangen, Germany) and a 16-channels marmoset head coil (Rogue Research, Canada). A high-resolution T1-weighted anatomical volume was acquired (MP2RAGE sequence, TE = 2.58 ms, TR = 2200 ms, flip angle = 8°, matrix size = 384 × 324 × 256, resolution 0.18 × 0.18 × 0.18 mm3) as well as a T2-weighted anatomical volume (SPACE sequence, TE = 55.8 ms, TR = 3 s, matrix size = 256 × 324 × 384, acquisition resolution 0.36 × 0.36 × 0.36 mm3, interpolated to 0.18×0.18×0.18 mm3).

Functional volumes were then acquired using EPI sequences (multiband acceleration factor: 2, TR = 588 ms, TE = 30 ms, flip angle = 48°, matrix size = 40 × 40, 18 slices, no gap, resolution 1.25 × 1.25 × 1.25 mm3). The functional volumes collected on monkeys using MION agent contrast used different EPI sequences (multiband acceleration factor: 2, GRAPPA acceleration = 2, TR = 392 ms, TE = 16.6 ms, flip angle = 42°, matrix size = 40 × 40, 18 slices, no gap, resolution 1.25 × 1.25 × 1.25 mm3).

Quantification and statistical analysis

fMRI data preprocessing. To address minor muscle spasms during anesthesia, inter-run registration, and spatial smoothing, a preprocessing of the functional data was performed. To mitigate potential motion artifacts, each functional volume was realigned and unwrapped using a reference volume. This reference volume was chosen from a stable period in the session and was spatially closest to the average across all sessions. Spatial smoothing was applied using a three-dimensional Gaussian kernel with a full width at half-maximum that was twice the size of the functional voxels (2.5 mm). Tissue segmentation and brain extraction were performed on the structural scans using a custom segmentation pipeline. This pipeline consisted of several steps: volume cropping (FSL), bias field correction (T1w × T2w debiasing; Rilling *et al.*, 2012), denoising (spatially adaptive nonlocal means; SPM), an initial brain extraction (FSL), and registration of the marmoset brain mapping template (Liu *et al.*, 2021) to the anatomical scan to provide priors for SPM's *old_segment* algorithm. A second brain extraction was then performed on the segmented tissues.

Transformation matrices between anatomical and functional data were generated using non-linear registration (ANTS). These matrices were used to align the tissue segmentation with the functional data and to register the functional results to the high-resolution anatomical scan.

fMRI data analysis. The fMRISTAT software (Worsley *et al.*, 2002) was used to calculate the general linear model estimates of the responses to two contrasts. First all sounds relative to silence (i.e., all > silence then the responses to conspecific versus scrambled vocalizations (i.e., CV > scrambled).

The GLM design matrix included several principal components of principal component analyses (PCAs) that served as covariates of no interest. Two PCAs were performed on eroded masks of the functional voxels identified as containing white matter and cerebrospinal fluid tissues separately. The number of principal components used as covariates of no interest was optimized to enhance cortical auditory activation (all > silence contrast). The first two components typically captured the MRI signal decay to steady-state equilibrium, resulting from stopping the scanner between clusters of volume acquisitions, while the remaining components captured respiratory, cardiac or motion artifacts.

For BOLD data (M9 to M15), a hemodynamic response function (HRF) with a peak at 5 s and an undershoot at 10,8 s was used during analysis (Baumann *et al.*, 2010). To analyze the MION data, a response function was designed to have a reversed sign, an earlier peak (4 s), a prolonged tail, and no undershoot, which was comparable to descriptions provided in prior literature (Leite *et al.*, 2002).

Voxel significance was determined by applying a threshold to T-maps at *p* < 0.05, with corrections for multiple comparisons based on Gaussian Random Field Theory (Worsley *et al.*, 2002). The quality of functional data in monkeys is influenced by various factors that cannot be quantitatively measured, such as anesthesia level and posture during MRI scans. Under these conditions, the overall fMRI response to sound serves as a quality assessment metric for each run, allowing for the identification and exclusion of low-quality runs (Petkov *et al.*, 2008). A jackknife procedure was used to assess the contribution of each individual run to the overall sound response. This involved systematically removing each run from the dataset and measuring both the spatial extent of significant voxels and the maximum t-value in the all > silence contrast. This approach enabled the evaluation of each run's impact on the global sound response. Only runs that contributed positively to the sound response were conserved. The study did not control for potential confounders. Additionally, there was no randomization or blinding in the experimental design.

Results

Contrast of all sounds vs. silence

We first performed individual-level analyses that revealed statistically significant ($p<0.05$ corrected) auditory activation maps (contrast: all sounds vs silence) in cortical areas for 7 out of the 15 individuals. Individual activation maps of these seven monkeys are available in the Extended data section (Obliger-Debouche *et al.*, 2025). The other 8 monkeys showed few activations, and only in subcortical areas such as the inferior colliculus (IC).

The individual statistical maps for all monkeys were then combined using a fixed-effects model, as commonly used in non-human primate studies. Figure 2 shows the group-level activation map obtained in the 15 individuals. The activation included all auditory cortex subdivisions (AuA1, AuAL, AuCL, AuCM, AuCPB, AuML, AuR, AuRM, AuRT, AuRTL and AuRPB) but also subcortical areas (IC and Thal). We also noted an unexpected activation of visual area (V1).

The coincidence map (indicating the number of participants out of 15 with significant activation in each voxel, Figure 2 C.), shows that cortical auditory activations (all > silence) were only observed in a minority of our 15 participants. This may be due to deeper levels anesthesia (activations only observed subcortically), or to non-ideal auditory stimulation in those monkeys (no activations at all). Only data from monkeys that yielded significant and substantial cortical auditory activations were selected to compute our main contrast of interest (CV > scrambled). Seven monkeys (M1, M3, M4, M5, M8, M9, and M10) were kept because of this selection.

Figure 3 shows the fixed effect contrast between vocalizations and scrambled vocalizations in these seven auditory-responsive individuals. The results demonstrate a bilateral activation of the anterior and temporal cortex. Cortical activity in the left hemisphere includes the insular cortex's granular (GI) and dysgranular (DI) areas, as well as the visual area V1. Subcortically, activity is seen in the nucleus accumbens (Acb), caudate nucleus (Cd), and hippocampal formation (HipF). Conversely, within the right hemisphere, cortical activation manifests in the orbital periallocortex (OPAl), rostrotemporal medial area of the auditory cortex (AuRTM), ventral area of the secondary somatosensory cortex (S2PV), the V1 visual area, and the gustatory cortex (Gu). Subcortically, the hippocampal formation (HipF) shows an activation.

Discussion

To engage in social interactions and communicate, primates need to perceive and process vocalizations. Studies have revealed unique brain areas, referred to as "voice patches" (Belin, 2018; Pernet *et al.*, 2015), which are specialized to processing voices, especially sounds that are conspecific to their species (Fecteau *et al.*, 2004). While human voice areas have received much attention, research into similar regions in non-human primates are currently the focus of intense investigation and may reveal important insights into the brain mechanisms and evolutionary history of voice processing. In this study, marmosets (*Callithrix jacchus*), a species known for its complex vocal repertoire and social communication, are studied to better understand the brain mechanisms of conspecific voice perception (Agamaite *et al.*, 2015; Bezerra & Souto, 2008). Due to its small size and advanced cognitive abilities, the marmoset is ideal for functional MRI (fMRI) research.

The goal of this work was to test whether voice patches could be identified in anesthetized animals using conventional 3T fMRI instead of high-field small-animal MRI. We optimized several aspects of the scanning protocol, the implementation of a specific clustered sparse design, the administration of the MION agent to half of the subjects who exhibited suboptimal resolution without it, the utilization of a 16-channel head coil to improve signal detection, and the maintenance of a low anesthesia level by employing a reduced concentration of sevoflurane in conjunction with nitrous oxide. This was the first time that this specially designed 3T marmoset head antenna has been used in a functional study. Previous studies using this antenna focused on resting-state imaging (Hayashi *et al.*, 2021).

Despite these attempts, the observed auditory activations are not satisfying. Only half of the participants showed observable activations in auditory areas. This result contrasts with that of Jafari *et al.*, whose investigation showed a stronger and more consistent auditory activation pattern in all participants. The disparities in auditory activation patterns between our study and the high-field fMRI study by Jafari *et al.* emphasize the inherent limitations of 3T large bore fMRI and of anesthesia.

One major disadvantage of 3T fMRI is a lower signal-to-noise ratio (SNR) than high-field devices. Despite efforts to increase signal detection with customized settings such as clustered sparse design and a 16-channel head coil, SNR at 3T remains dramatically lower than at higher field strengths. This decreased SNR may make it more difficult to distinguish tiny neural responses (voxel size in Jafari *et al.* (2023): 0.5mm isotropic;

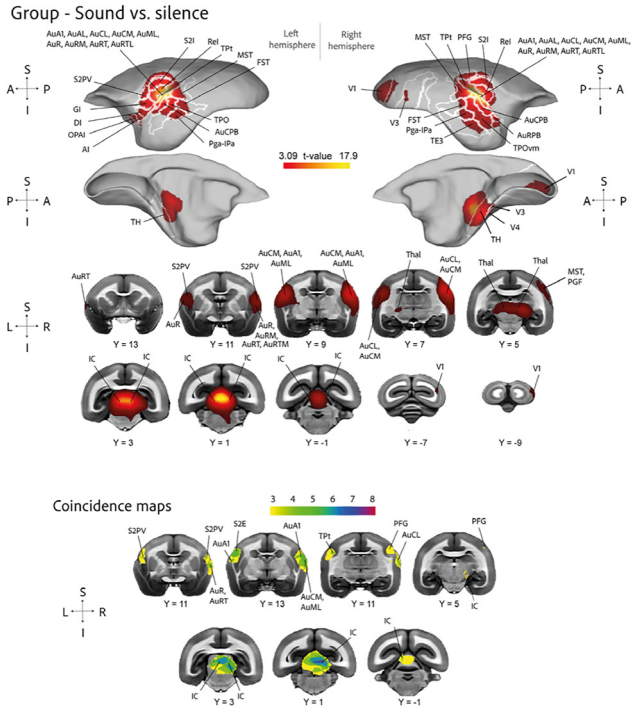


Figure 2. Fixed effect group analysis (n = 15) for the contrast sound versus silence. (A) T-maps displayed on both the left and right fiducial brain surface template (Liu *et al.*, 2021), with the lateral view presented at the top and the medial view at the bottom. The delineation of regions based on an atlas (Paxinos *et al.*, 2012) is depicted by white lines. (B) Volumetric activations are superimposed onto coronal slices of template anatomical brain (Liu *et al.*, 2021). C. Coincidence maps show the regions of the brain where the activation patterns of multiple individuals overlap.

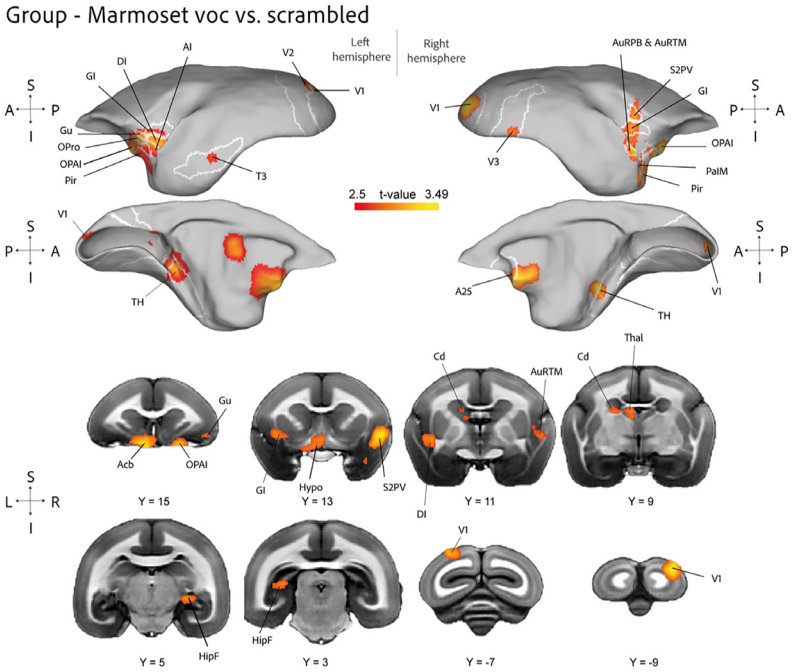


Figure 3. Fixed effect group analysis (n = 7) for the contrast marmoset vocalization versus scrambled. (A) T-maps displayed on both the left and right fiducial brain surface template (Liu *et al.*, 2021), with the lateral view presented at the top and the medial view at the bottom. The delineation of regions based on an atlas (Paxinos *et al.*, 2012) is depicted by white lines. (B) Volumetric activations are superimposed onto coronal slices of template anatomical brain (Liu *et al.*, 2021).

current study: 1.25mm), particularly in regions associated with complex cognitive activities such as auditory perception. Furthermore, the spatial resolution of functional images provided by 3T fMRI may be less precise than that achieved with higher field strengths. This limitation may impair the proper and exact localization of neuronal activations in cortical areas, thus lowering the sensitivity of activation pattern recognition.

Despite the difficulties in obtaining robust auditory activation using conventional anesthesia protocols, our fixed-effect and coincidence maps yielded valuable insights. In particular, the coincidence maps revealed common activation patterns in our cohort of subjects. Specifically, analysis of these maps showed significant activation in common subcortical brain areas, notably the inferior colliculus (IC), as well as activation in cortical auditory regions. This illustrates that low-field fMRI may reveal fundamental auditory activation patterns. Future research might focus on refining anesthetic procedures to boost the sensitivity of detecting auditory activations. The use of dexmedetomidine, a selective α2-adrenergic agonist with sedative and analgesic properties, may provide benefits compared to common anesthetics like sevoflurane used here. Dexmedetomidine-induced sedation has little impact on respiratory function and cerebral blood flow regulation, making it an attractive option for neuroimaging studies. Dexmedetomidine may improve the reliability of auditory activation signals acquired during 3T fMRI studies by lowering potential adverse effects such vasodilation and neuronal activity suppression. Using a combination of medetomidine and isoflurane proves to be more effective in eliciting visual BOLD responses compared to using isoflurane alone (Ortiz-Rios *et al.*, 2023). Moreover, the resting-state networks under medetomidine-isoflurane anesthesia closely resembled those observed in awake states.

Seven animals showed brain activity in both contrasts (all > silent and CV > scrambled), despite the unsatisfactory sedation procedure. According to previous research, voice patches were found for the CV > scrambled contrast in the anterior temporal area (Jafari *et al.*, 2023; Sadagopan *et al.*, 2015). In marmosets, conspecific vocalizations activate a frontotemporal voice perception network, as shown by Sadagopan *et al.* (2015). Like the human activations observed, this network includes both cortical and subcortical locations. Our study's activation of subcortical areas such the caudate nucleus, nucleus accumbens, and hippocampal formation is consistent with subcortical involvement according to Sadagopan *et al.*'s research. These findings imply that both cortical and subcortical brain areas are involved in the preservation of the brain architecture for vocalization processing. The activation of the nucleus accumbens and the creation of the hippocampus highlights their roles in the processing of vocalizations associated with emotions and memories, which is crucial for social interaction and communication.

In the common marmoset's auditory cortex, Jafari *et al.* (2023) found a caudal-rostral gradient, with the anterior temporal lobe showing the greatest preference for conspecific vocalizations. This study identifies a brain specialty for vocalization processing, indicating that primates may share this specialized prior to the split of their lineages. Their results are consistent with ours, showing a comparable gradient and specialization in the bilateral activity of the anterior and temporal cortex. An equivalent functional architecture for vocalization processing is suggested by the participation of areas such as the insular cortex in the left hemisphere and the rostrotemporal medial area of the auditory cortex (AuRTM) in the right hemisphere. This analogy give credits to the idea that specific cortical regions involved in vocalization may have deep evolutionary roots, serving essential roles in social communication and recognition of vocalization types and contexts.

The merging of our results with those of Sadagopan *et al.* and Jafari *et al.* highlights the significance of vocalization processing networks in evolution. The cortical and subcortical activations that we identified are consistent with the hypothesis that humans and other primates share specialized brain areas for vocalization processing, pointing to a common evolutionary path that preceded the split of New and Old World monkeys.

To conclude, our findings strongly suggest that specialized, evolutionarily conserved brain networks exist for the processing of vocalizations. These networks include subcortical structures, which are essential for affective and memory-related elements of vocalization, as well as cortical areas, which oversee intricate auditory and multisensory processing. The intricate vocal communication systems found in primates, such as marmosets, are actively supported by this coherent and consistent brain architecture.

Conclusion

Despite the challenges of using a low-field scanner and administering anesthesia, we were able to identify the same brain regions as those highlighted by the literature. This correspondence is a valuable replication of these previous studies, especially as such studies are still rare for the marmoset model. Our results thus reinforce the hypothesis that humans and other primates share brain areas specialized in conspecific voice processing. This finding is crucial as it provides further evidence of functional parallelism between the neural systems of non-human primates and those of humans, suggesting a common evolutionary basis for the processing of vocal communication. Consequently, our study makes a significant contribution to our understanding of the neurobiological mechanisms underlying vocalization in primates, while highlighting the importance of replicability of results in scientific research.

Declarations

Ethics approval

All animal experimental protocols were conducted in accordance with the European directive (2010/63/UE) guidelines and approved under protocol 2019010911313842 by the Ethics Board of Institut de Neurosciences de la Timone, on 24/03/2019. Efforts were made to ameliorate the suffering of animals using endpoints, environmental enrichment, and careful monitoring of health and well-being.

Data and software availability

Underlying data

Current Opinion in Behavioral Sciences: Constraints and flexibility during vocal development: Insights from marmoset monkeys, https://doi.org/10.1016/j.cobeha.2017.11.015 (Ghazanfar & Liao, 2018).

Zenodo: Marmoset fMRI data – Voice Localizer, https://doi.org/10.5281/zenodo.14499745 (Obliger-Debouche *et al.*, 2024).

The project contains the following underlying data:

1. Anesthetized_Marmoset_fMRI_Voice_Localizer.tar.gz
2. README.txt

Data are available under the terms of the Creative Commons Attribution 4.0 International.

Extended data

Zenodo: Extended data – Marmoset fMRI Voice localizer [Dataset], https://doi.org/10.5281/zenodo.14983863 (Obliger-Debouche *et al.*, 2025).

This project contains the following extended data:

Individual activations.pdf - brain activations of each monkey in the study

completed_ARRIVE_checklist.pdf - detailed checklist ARRIVE of the manuscript

Data are available under the terms of the Creative Commons Attribution 4.0 International.

Reporting guidelines

ARRIVE Checklist For "Investigating the cortical mechanisms of conspecific voice perception in anesthetized marmosets using 3T fMRI", https://doi.org/10.5281/zenodo.14983863

Data are available under the terms of the Creative Commons Attribution 4.0 International.

Authors' contributions

Conceptualization, M.O.D., R.T., and P.B.; Methodology, M.O.D., R.T., J.S., B.N., J.L.A., and P.B.; Data acquisition, M.O.D., R.T., L.R., and J.S.; Analyses, M.O.D., and R.T.; Funding, P.B.; Writing, M.O.D., R.T., S.R., and P.B; Reviewing, R.T., J.S., B.N., and P.B.; Supervising of the study, S.R., and P.B.

Acknowledgments

We thank Lina Boudin and Shabnam Bahramiasl for their contributions during their internships. We also thank Dr. David Meunier, Dr. Mélina Cordeau, Dr. Charly Lamothe and Dr. Margherita Giamundo for useful discussions. The authors of this paper have no competing interests to declare.

References

Agamaite JA, Chang CJ, Osmanski MS, *et al.*: **A quantitative acoustic analysis of the vocal repertoire of the common marmoset (*Callithrix jacchus*).** *J Acoust Soc Am.* 2015; **138**(5): 2906–2928.
PubMed Abstract | Publisher Full Text | Free Full Text

Bandettini PA, Jesmanowicz A, Van Kylen J, *et al.*: **Functional MRI of brain activation induced by scanner acoustic noise.** *Magn Reson Med.* 1998; **39**(3): 410–416.
PubMed Abstract | Publisher Full Text

Baumann S, Griffiths TD, Rees A, *et al.*: **Characterisation of the BOLD response time course at different levels of the auditory pathway in non-human primates.** *NeuroImage.* 2010; **50**(3): 1099–1108.
PubMed Abstract | Publisher Full Text | Free Full Text

Belin P: **The 'Vocal Brain': core and extended cerebral networks for voice processing.** In: S. Frühholz & P. Belin (Eds.), *The oxford handbook of voice perception.* Oxford University Press, 2018; 37–60.
Publisher Full Text

Belin P, Bodin C, Aglieri V: **A "voice patch" system in the primate brain for processing vocal information?** *Hear Res.* 2018; **366**: 65–74.
PubMed Abstract | Publisher Full Text

Belin P, Trapeau R, Obliger-Debouche M: **A small, but vocal, brain.** *Cell Rep.* 2023; **42**(6): 112651.
PubMed Abstract | Publisher Full Text

Belin P, Zatorre RJ, Ahad P: **Human temporal-lobe response to vocal sounds.** *Brain Res Cogn Brain Res.* 2002; **13**(1): 17–26.
PubMed Abstract | Publisher Full Text

Bezerra BM, Souto A: **Structure and usage of the vocal repertoire of *Callithrix jacchus*.** *Int J Primatol.* 2008; **29**(3): 671–701.
Publisher Full Text

Bodin C, Trapeau R, Nazarian B, *et al.*: **Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques.** *Curr Biol.* 2021; **31**(21): 4839–4844.e4.
PubMed Abstract | Publisher Full Text | Free Full Text

Fecteau S, Armony JL, Joanette Y, *et al.*: **Is voice processing species-specific in human auditory cortex? An fMRI study.** *NeuroImage.* 2004; **23**(3): 840–848.
PubMed Abstract | Publisher Full Text

Gautama T, Mandic DP, Van Hulle MM: **Signal nonlinearity in fMRI: a comparison between BOLD and MION.** *IEEE Trans Med Imaging.* 2003; **22**(5): 636–644.
PubMed Abstract | Publisher Full Text

Ghazanfar AA, Liao DA: **Constraints and flexibility during vocal development: insights from marmoset monkeys.** *Curr Opin Behav Sci.* 2018; **21**: 27–32.
PubMed Abstract | Publisher Full Text | Free Full Text

Gil-da-Costa R, Braun A, Lopes M, *et al.*: **Toward an evolutionary perspective on conceptual representation: species-specific calls activate visual and affective processing systems in the macaque.** *Proc Natl Acad Sci U S A.* 2004; **101**(50): 17516–21.
PubMed Abstract | Publisher Full Text | Free Full Text

Gil-da-Costa R, Martin A, Lopes MA, *et al.*: **Species-specific calls activate homologs of Broca's and Wernicke's areas in the macaque.** *Nat Neurosci.* 2006; **9**(8): 1064–1070.
PubMed Abstract | Publisher Full Text

Hall DA, Haggard MP, Akeroyd MA, *et al.*: **"Sparse" temporal sampling in auditory fMRI.** *Hum Brain Mapp.* 1999; **7**(3): 213–223.
PubMed Abstract | Publisher Full Text | Free Full Text

Hayashi T, Hou Y, Glasser MF, *et al.*: **The nonhuman primate neuroimaging and neuroanatomy project.** *NeuroImage.* 2021; **229**: 117726.
PubMed Abstract | Publisher Full Text | Free Full Text

Jafari A, Dureux A, Zanini A, *et al.*: **A vocalization-processing network in marmosets.** *Cell Rep.* 2023; **42**(5): 112526.
PubMed Abstract | Publisher Full Text

Joly O, Pallier C, Ramus F, *et al.*: **Processing of vocalizations in humans and monkeys: a comparative fMRI study.** *NeuroImage.* 2012; **62**(3): 1376–1389.
PubMed Abstract | Publisher Full Text

Leite FP, Tsao D, Vanduffel W, *et al.*: **Repeated fMRI using iron oxide contrast agent in awake, Behaving macaques at 3 Tesla.** *NeuroImage.* 2002; **16**(2): 283–294.
PubMed Abstract | **Publisher Full Text**

Liu C, Yen CCC, Szczupak D, *et al.*: **Marmoset brain mapping V3: population multi-modal standard volumetric and surface-based templates.** *NeuroImage.* 2021; **226**: 117620.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Miller CT, Freiwald WA, Leopold DA, *et al.*: **Marmosets: a neuroscientific model of human social behavior.** *Neuron.* 2016; **90**(2): 219–233.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Obliger-Debouche M, Trapeau R, Robinet L, *et al.*: **Extended data - marmoset fMRI voice localizer [Dataset]**. *Zenodo.* 2025.
http://www.doi.org/10.5281/ZENODO.14983863

Obliger-Debouche M, Trapeau R, Sein J, *et al.*: **Marmoset fMRI data - voice localizer.** *Zenodo.* 2024.
http://www.doi.org/10.5281/ZENODO.14499745

Ortiz-Rios M, Kuśmierek P, DeWitt I, *et al.*: **Functional MRI of the vocalization-processing network in the macaque brain.** *Front Neurosci.* 2015; **9**: 113.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ortiz-Rios M, Sirmpilatze N, König J, *et al.*: **Preserving functional network structure under anesthesia in the marmoset monkey brain.** *BioRxiv.* 2023; 2023.11.21.568138.
Reference Source

Paxinos G, Watson C, Rosa M, *et al.*: **The marmoset brain in stereotaxic coordinates**. 2012.
Reference Source

Peelle JE: **Methodological challenges and solutions in auditory functional magnetic resonance imaging.** *Front Neurosci.* 2014; **8**: 253.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Percie du Sert N, Hurst V, Ahluwalia A, *et al.*: **The ARRIVE guidelines 2.0: updated guidelines for reporting animal research.** *PLoS Biol.* 2020; **18**(7): e3000410.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Pernet CR, McAleer P, Latinus M, *et al.*: **The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices.** *NeuroImage.* 2015; **119**: 164–174.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Petkov CI, Kayser C, Augath M, *et al.*: **Functional imaging reveals numerous fields in the monkey auditory cortex.** *PLoS Biol.* 2006; **4**(7): e215.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Petkov CI, Kayser C, Augath M, *et al.*: **Optimizing the imaging of the monkey auditory cortex: sparse vs. continuous fMRI.** *Magn Reson Imaging.* 2009; **27**(8): 1065–73.
PubMed Abstract | **Publisher Full Text**

Petkov CI, Kayser C, Steudel T, *et al.*: **A voice region in the monkey brain.** *Nat Neurosci.* 2008; **11**(3): 367–374.
PubMed Abstract | **Publisher Full Text**

Poremba A, Malloy M, Saunders RC, *et al.*: **Species-specific calls evoke asymmetric activity in the monkey's temporal poles.** *Nature.* 2004; **427**(6973): 448–451.
PubMed Abstract | **Publisher Full Text**

Rilling J, Glasser M, Jbabdi S, *et al.*: **Continuity, divergence, and the evolution of brain language pathways.** *Front Evol Neurosci.* 2012; **3**: 11.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Sadagopan S, Temiz-Karayol NZ, Voss HU: **High-field functional magnetic resonance imaging of vocalization processing in marmosets.** *Sci Rep.* 2015; **5**(1): 10950.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Shah NJ, Jäncke L, Gross-Ruyken ML, *et al.*: **How does acoustic masking noise affect fMRI of the auditory cortex.** *NeuroImage.* 1997; **5**: 195.

Trapeau R, Belin P, Sein J, *et al.*: **Comparison of auditory fMRI protocols for a voice localizer.** *Open Res Eur.* 2024; **4**: 261.
Publisher Full Text

Worsley KJ, Liao CH, Aston J, *et al.*: **A general statistical analysis for fMRI data.** *NeuroImage.* 2002; **15**(1): 1–15.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 02 June 2025

https://doi.org/10.21956/openreseurope.20624.r54143

© **2025 Morfini F.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Francesca Morfini
Northeastern University, Boston, USA

The authors examined the neural correlates of listening to conspecific vocalizations in anesthetized adult marmosets (n=15) using functional magnetic resonance imaging (fMRI) in a 3T scanner. Approximately half of the marmosets received an MRI contrast agent (i.e., MION) to enhance brain tissue differentiation.

This is a strong and timely study that addresses an important question with methodological rigor. There are several strengths including the use of a 3T clinical MRI scanner—rather than more commonly used but less available high-field systems— which is relevant for broader comparative neuroscience. The results offer meaningful insights that will be valuable for both animal research and cross-species understanding of auditory processing. The manuscript is well written—clear, thorough, and accessible, and does an excellent job of grounding interpretations in the data while acknowledging limitations.

I consider this paper well-positioned to have high impact and to be of broad interest to the neuroscience community. I have three major suggestions namely, to conduct more in-depth investigation regarding why only ~half monkeys showed activity, the actual gain of employing MION, and to provide further details about the study design. I also provide minor comments for readability.

Auditory-related activation, or lack thereof.

1. Could the authors provide a more in-depth investigation into why only approximately half of the sample (n = 7 out of 15) exhibited auditory-related activation in the auditory cortex? Understanding the underlying mechanisms that distinguish successful from unsuccessful activation would significantly enhance the manuscript's contribution. Without this clarification, the proposed approach appears limited in its generalizability and potential for translation to other contexts. Conducting post-hoc exploratory analysis into the differences between these two groups would be most useful. Even a speculative discussion of possible contributing factors—such as individual differences in neural responsiveness, experimental design variability, or attentional engagement—would strengthen the impact and

interpretability of the findings.
2. The authors should include a description of the differences between these two sets of monkeys (showing vs not showing significant activation) with respect to, and not limited to, MION status, sex, age, any other relevant factor.
3. Could the authors provide more detailed information regarding the activation patterns observed in the two sets of monkeys? While activation may not have reached statistical significance for some individuals, this outcome could be closely tied to the specific thresholding criteria applied. Therefore, it would be valuable to report and describe the brain activity patterns regardless of significance (e.g., positive or negative, spatial localization, and any consistent trends across subjects).

MION vs non-MION.

1. While the authors had sound motivations for customizing their experimental design to leverage the advantages of MION, this customization introduces the possibility of systematic bias between the monkeys that received MION and those that did not (hereafter, MION vs non-MION). Specifically, the tailoring of fMRI acquisition parameters (e.g., TR, TE, flip angle) and statistical modeling choices (e.g., hemodynamic response function specifications) to each condition may limit comparability between these two groups. Given this, it would be more appropriate to treat the MION and non-MION datasets as effectively independent. Accordingly, the authors should address how they harmonized or accounted for these differences in their analyses.
2. The authors present a compelling rationale for including MION, suggesting it could enhance tissue differentiation. However, the manuscript does not follow through on this hypothesis by empirically evaluating the hypothesized benefits of using MION. Could the authors include analysis that would aim at informing whether future work would benefit from employing MION?
3. Because MION status systematically covary with multiple design elements (acquisition settings, modeling choices, etc.), group differences cannot be interpreted as clean tests of any single factor. Thus, any interpretation of findings comparing MION and non-MION groups (including in response to my previous comment) must acknowledge this bias. The manuscript should more explicitly discuss this issue throughout the introduction, results, and discussion.

Provide further details regarding study design.

1. Could the author provide the following information in the text: number of runs per subject, total functional scan duration, and the age of monkeys at the time of testing (while DOBs are listed in Table 1, ages at the time of data acquisition are not reported).
2. Please clarify the timeline between anesthesia administration and MRI acquisition.
3. Figure 2A suggests variation in stimulus duration across vocalization types (e.g., Phees vs Trills). Please clarify whether these differences were controlled or counterbalanced, and what impact, if any, this may have on the results and results' specificity to specific type of vocalizations.

Other Minor comments.

1. Could the authors rephrase the following sentence? "Although clinical MRI scanners are more accessible than preclinical high-field MRI scanners, facilitating their use for a variety of studies, research conducted at 3T on such MRI scanners is limited to anatomical imaging.".

While correct, it may result a bit confusing in its current structure.

2. It was unclear to me if the following paragraph had some repetitions regarding the number of acquired fMRI volumes (i.e., 7 volumes) and duration of acquisition (i.e., 3.92s): "*After pilot trials in human participants (Trapeau et al., 2024), we converged to series of 7 successive brain volumes acquired in succession in 3.92s, followed by a silent interval of 4.1s during which a block of either 4 vocalizations, 4 scrambled vocalizations or silence was played. Seven volumes were acquired between blocks of stimuli for a duration of 3.92 s. (Figure 1A).".* The authors could consider consolidating information or clarify what the last sentence referred to.
3. There seems to be an inconsistent use of double parentheses at times throughout the manuscript. It is not clear if this is a stylistic choice or a formatting error. E.g. (fMRI) (for review see Belin (2018)). Did the authors meant for this to be e.g.: (fMRI; for review see Belin et al., 2018)?
4. Could the authors define the acronym MION in the legend of Table 1?
5. Could the authors define the acronyms of the brain regions indicated in the brain rendering images in the legend of Figure 2?
6. There may be a typo at the end of the Conclusion section: "ultra-high-filed" for "ultra-high field"?

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: adolescence, depression, anxiety, functional magnetic resonance imaging, machine learning

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 28 May 2025

https://doi.org/10.21956/openreseurope.20624.r54149

© 2025 Zanini A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Alessandro Zanini
University of Western Ontario, London, Canada

This paper presents a study investigating vocalization processing in the common marmoset (Callithrix jacchus) using 3T functional MRI (fMRI). The authors aim to identify brain regions involved in conspecific vocalization processing by applying an optimized fMRI protocol, including a reduced concentration of sevoflurane and the use of a 16-channel head coil to improve signal detection. The study examines auditory activations in 15 marmosets, focusing on cortical and subcortical responses to vocalizations, and compares these findings with previous high-field fMRI studies. Despite the challenges of working with 3T fMRI in anesthetized animals, the results suggest that both cortical and subcortical areas are involved in vocalization processing, with common activations across subjects, although not all participants exhibited consistent auditory activations.
Given the increasing popularity of marmosets as animal model alternative to the (more expensive and rarer) macaque, this study represents a crucial step. Not all laboratories can, in fact, take advantage of small-bore high-field MRI scanner, and this can be a strong limitation to the diffusion of the marmoset as animal model. The authors made a huge effort in trying to optimize the experimental procedure and setup, and I think this article needs publication in order to advance our knowledge and help the marmoset research community develop.
I only have one major and a few minor points.

Major point:

- The fMRI data analysis does not specify how many functional runs were acquired per monkey. It would be helpful to include this information in Table 1, clarifying the actual number of MION and BOLD runs. Additionally, the authors mention that some runs were excluded due to low signal in the auditory region; it would be useful to specify the number of runs excluded after this procedure. I am also unclear about the sample size used in the group analyses: the results section indicates that only 7 animals exhibited statistically significant auditory activation maps, while the remaining 8 marmosets showed only subcortical activations. Shouldn't this imply that the runs from the 8 animals without cortical activations would have been discarded based on the selection criteria described earlier? However, Figure 2 reports an N of 15, and the 8 animals without cortical activations appear to be excluded only from the second analysis (vocalization vs scrambled vocalization). The discrepancy in sample size complicates direct comparison between the two maps. It might be clearer to present the results for the 7 monkeys in both figures, with the "all > silence" contrast for N=15 included in the supplementary material, unless the authors have a specific reason for reporting it in the main text.

Minor points:

- Abstract – conclusions: “the use of anesthesia, which decreases..” and “ultra-high-field”

- Sample: were the animals involved in the study naive to the scanner, or did they have prior scanning experience? Additionally, did they have any form of ear protection during the current study or in previous fMRI studies? This may not have been the case, given the presentation of the acoustic stimuli via loudspeaker. I'm particularly concerned about the possibility that repeated exposure to MRI noise without ear protection could have damaged the eardrums or impaired acoustic perception in some monkeys, potentially explaining the exclusion of half the sample.
- fMRI data analysis: there is a missing ")" after "all > silence"
- fMRI data analysis and Results: the authors describe the preprocessing and treatment of individual maps, but Figures 2 and 3 show group maps. It would be helpful to clarify a little more in the data analysis section how these group maps were obtained, as this is only briefly mentioned in the Results section.
- Signal-to-Noise Ratio (SNR): The authors mention the low SNR of the 3T large-bore scanner as a limitation, which is understandable given the small brain size of marmosets. I believe it would be valuable for readers to visualize the quality and spatial distribution of the signal. Would it be possible to include a map of the temporal SNR (tSNR) for the sample? A comparison of tSNR with and without MION could also provide insightful information.
- Experimental Setup (Suggestion): While the authors highlight low SNR as one of the main limitations of the study, I believe that the experimental setup might have an even greater impact. We have previously conducted auditory tasks with awake marmosets in the scanner, and the scanner noise is considerable. Despite using a sparse design, the use of a loudspeaker likely hindered the application of ear protection for the animals, which meant they were exposed to substantial noise just before the auditory stimuli. This could have influenced the subsequent activations and, over time, may lead to a loss of hearing sensitivity. In our studies, we addressed this issue by using MRI-compatible headphones with auditory tubes placed in the ear canal, surrounded by silicone earplugs, along with the sparse design. This setup, though costly, has been effective in producing strong activations, as demonstrated in previous work Jafari A et al., 2023 (Ref 1); Dureux A et al., 2024 (Ref 2), Comms Bio; Dureux A et al., 2025 (Ref 3). Given the long-term benefits, the investment in such equipment may be worthwhile.

References

1. Jafari A, Dureux A, Zanini A, Menon RS, et al.: A vocalization-processing network in marmosets. *Cell Rep*. 2023; **42** (5): 112526 PubMed Abstract | Publisher Full Text
2. Dureux A, Zanini A, Everling S: Mapping of facial and vocal processing in common marmosets with ultra-high field fMRI.*Commun Biol*. 2024; **7** (1): 317 PubMed Abstract | Publisher Full Text
3. Dureux A, Zanini A, Jafari A, Everling S: Ultra-high Field fMRI Reveals Effect of Ketamine on Vocal Processing in Common Marmosets.*J Neurosci*. 2025; **45** (15). PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and does the work have academic merit?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Awake marmoset neuroimaging

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.