Cortical voice processing in Autism Spectrum Disorder

Raphaël Gautier[1,2], Emmanuelle Houy-Durand[1,2,3], Laurent Barantin[1], Frédéric Briend[1,2], Marianne Latinus[1,2]

[1]UMR 1253, iBrain, Université de Tours, INSERM, Tours, France

[2]EXcellence Center in Autism and neurodevelopmental disorders, CHRU de Tours, Tours, France

[3]Centre Universitaire de Pédopsychiatrie, CHRU de Tours, Tours, France

Abstract

Voice processing is central to social functioning. A specific brain response to vocal sounds has been described and extensively characterized in the general population but remains critically unexplored in Autism Spectrum Disorder (ASD), a condition mainly characterized by social difficulties. The few studies conducted within the ASD population reported contradictory results with either a lack of or a typical brain response to vocal sounds in ASD. Hence, it is not clear whether at least some ASD individuals are characterized by a dysfunctional response to vocal sounds and if the discrepancies between the studies are due to sample characteristics. This registered report aims to characterize the individual brain response to vocal sounds, for ASD (n=26) and non-ASD (n=26) individuals using an fMRI block-design study contrasting vocal with non-vocal sounds. The proportion of individuals showing a specific response to vocal sounds will be compared between groups. We hypothesize a lower proportion of individuals showing a specific response to vocal sounds in the ASD group. Results from this study might resolve the discrepancies between the results described in the literature.

Table of contents

I. Introduction 1

II. Methods 3

Population 3

Material 3

Procedure 3

Power analysis 6

Study design table 7

References 8

I. Introduction

Temporal voice areas (TVAs) corresponds to a set of cortical regions which are characterized by a greater response to vocal sounds compared to non-vocal sounds (Belin et al., 2000). This pattern is observed irrespectively of whether the voice carries speech (Belin et al., 2000), and when matching non-vocal sounds for acoustical characteristics (Agus et al., 2017). A study by Pernet and collaborators (Pernet et al., 2015) suggests that this response can be reliably identified in most individuals (94%) with a high test-retest reliability ($r > .90$).

While TVAs are localized bilaterally in the Superior Temporal Sulci/Gyri (STS/STG), following a postero/anterior axis, precise individual location is characterized by a high variability (Pernet et al., 2015). TVAs are identified using a functional localizer: an fMRI block-design experiment during which vocal and non-vocal sounds are presented and which allows to subsequently identify brain regions selectively activated by voice stimuli. Three different 'voice patches' (i.e., TVA sub-clusters) have been described in each hemisphere, with different connectivity patterns, suggesting that different sub-regions of the TVAs might be differentially implicated in voice processing, although this question has not yet been resolved (Pernet et al., 2015). Apart from the TVAs, an extended voice processing network comprising the amygdala and prefrontal regions has also been identified (Pernet et al., 2015). TVAs have been suggested to be involved in the first step of voice perception (i.e., structural analysis of voices, Belin et al., 2011; Bestelmeyer et al., 2014; Charest et al., 2013; Latinus et al., 2011, 2013). More elaborated steps of voice perception, such as perception of identity information and perception of vocal affective information are hypothesized to rely on the communication and integration of information between TVAs and regions from the extended voice processing network (Belin et al., 2004, 2011; Brück et al., 2011; Maguinness et al., 2018). These results emphasize the role of this functionally defined network in social functioning.

Autism Spectrum Disorder (ASD) is characterized by difficulties in communication and social interactions and by restricted and repetitive patterns of behavior (American Psychiatric Association, 2013). Two studies (Gervais et al., 2004; Schelinski et al., 2016) investigated TVAs in ASD, reporting inconsistent results. In the first study (Gervais et al., 2004), no preferential response to voices was observed in the STS/STG for 4 out of 5 individuals, leading to the conclusion of an impaired brain response to vocal sounds in ASD. However, Schelinski et al. (2016) identified typical responses in 15 out of 16 high functioning ASD individuals, yielding no between-group differences in group-level analyses. The two studies differ according to the statistical correction and thresholds used when performing statistical inferences, but also regarding the population characteristics. In fact, Schelinski et al. (2016) recruited individuals with higher functioning level (mean ASD IQ = 110.31 (13.79)) than Gervais et al. (2004; mean ASD Intellectual Quotient (IQ) = 81 (17.8); $t(19)=3.89$, $p<.01$). Moreover, Schelinski et al. (2016) applied more stringent corrections (FWE, $p < .05$) than Gervais et al. (2004; uncorrected, $p < .001$) at the group-level analyses. While the lack of observed differences in the group level analyses in Schelinski and collaborators' article (2016) might be explained by the conservative corrections applied, the two studies both estimated individual responses at an uncorrected threshold ($p < .001$), and observed different results. These elements suggest that the discrepancies between the two studies are more likely due to differences in the ASD samples, both in terms of size and characteristics, than to a difference in statistical thresholds. Thus, considering the inconsistent results described in the literature, this study aims to further investigate vocal sounds processing in ASD by characterizing the individual brain response to vocal sounds and comparing the proportion of responders between ASD and non-ASD individuals.

Brain responses to vocal sounds will be modeled at the subject-level, allowing us to describe the proportion of individuals in each group showing a selective response to voice in the TVAs. Such a response would manifest as a higher activation to vocal than to non-vocal stimuli. At the opposite, a lack of response to vocal sounds corresponds to an equal activation level between the vocal and non-vocal conditions. As a direct consequence of the major symptoms of ASD, we hypothesize a lower proportion of individuals showing a specific response to vocal sounds in the STS/STG (i.e., a TVA activation) in the autistic sample than in the non-ASD sample. Exploratory analyses will be conducted in order to investigate links between voice brain processing and individual participants' characteristics (e.g. IQ, Autism-spectrum Quotient; AQ, …).

II. Methods

Population

A total of 52 adult participants will be recruited (ASD: 26, non-ASD: 26, matched by sex and age). The sample size has been established following funding resources, although power analyses has been conducted in order to ensure the feasibility of the planned analysis. Diagnoses will be established by a psychiatrist coauthor (E.H.D) of the study according to the DSM-5 criteria. In addition, we will report scores from the ADOS (Autism Diagnostic Observation Schedule) or ADI-R (Autism Diagnostic Interview-Revised) for descriptive purposes (these scores are not necessarily considered for the establishment of diagnoses). Participants' hearing abilities will be assessed using an audiogram in order to discard an eventual influence of hearing abilities in between group differences in voice processing.

Material

Participants will undergo a voice localizer (Pernet et al., 2015), consisting of forty 8s blocks of either vocal or non-vocal sounds (20 blocks each) intermixed with twenty 8s blocks of rest, for a total task length of 10 min 20 s. The order of the blocks was determined pseudorandomly when the experiment has been designed, but the resulting order remained fixed for all participants. Within each block, stimuli are separated from each other by a delay of at most 400ms. Sounds will be displayed using the MR Confon system at an intensity level of roughly 88dB. Vocal sounds have been obtained from different speakers of various ages from the whole life span. Vocal sounds are either speech sounds or non-speech sounds. The formers are either words, sentences or syllables from different languages (English, French, Finnish, Arabic). The latter are either emotional (e.g., laughs, sighs, cries…) or neutral (e.g., coughs, onomatopoeias…) vocal sounds. Non-vocal sounds consist of sounds from nature (e.g., wind, sea waves, ...), animals, classical music, and man-made objects (e.g., cars, clocks, ...). The task consists of passively listening to the sounds, and no behavioral output is required. Participants are asked to close their eyes and listen carefully to the sounds.

An incidental "attention/memory" task will be administered to the participants after the scanning session. During this task, participants will be presented with some of the sounds displayed during the experiment. More precisely, stimuli consist of 10 vocal and 10 non-vocal sounds, arbitrarily drawn from the set of sounds displayed during the task. For each sound, the participants will be asked to indicate whether they remember hearing it during the scanning session. This task aims to monitor and detect differences in attention level during the task. In addition, participants will be asked to rate their level of engagement with the sounds during the voice-localizer task on a 5-points Likert scale.

All participants will be evaluated on IQ using four WAIS-IV subtests (Wechsler, 1981): block design, and matrix reasoning to evaluate non-verbal IQ and similarities and vocabulary to evaluate verbal IQ. ASD specific symptoms will be evaluated for all participants using the Social Responsiveness Scale (SRS) (Chan et al., 2017; Constantino et al., 2003), and the Autism-spectrum Quotient (AQ) which quantifies general autistic traits (Baron-Cohen et al., 2001).

Procedure

fMRI scanning

Anatomical and functional images will be acquired on a 3T Siemens Prisma scanner. Anatomical images will be acquired using a T1 weighted 3D sagittal scan with the following parameters: TR = 2300ms, TE = 2.98ms, flip angle = 9°, 1mm^3 isotropic voxels. Functional images will be acquired using the following parameters: single-shot gradient-echo echo-planar imaging; FOV = 210x210mm^2; 32 slices per volume; interleaved slices order; voxel size = 3mm^3 isotropic; acquisition matrix: 70x70; flip angle = 77°; TE =

30ms; TR = 2s; TA = 2s, and 310 volumes will be acquired throughout the session. Data will be excluded in case of a structural anomaly, or if an irreparable artifact is identified based on visual inspection. Such data will be replaced (i.e., another participant will be recruited).

fMRI Preprocessing

Anatomical data. Anatomical data preprocessing will be performed using the standardized fMRIPrep pipeline (Esteban et al., 2019). The T1-weighted (T1w) images will be corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection` (Tustison et al., 2010), distributed with ANTs 2.3.3 (Avants et al., 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w reference will be skull-stripped with a Nypipe implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) will be performed on the brain-extracted T1w using `fast` (FSL 6.0.6.4, RRID:SCR_002823, Zhang et al., 2001). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) will be performed through nonlinear registration with `antsRegistration` (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* (Fonov et al., 2009, , RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym).

Functional data. First, a reference volume and its skull-stripped version will be generated using a custom methodology of *fMRIPrep*. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcflirt` (FSL 6.0.6.4, Jenkinson et al., 2002). BOLD runs will be slice-time corrected to 0.959s (0.5 of slice acquisition range 0s-1.92s) using `3dTshift` from AFNI (Cox & Hyde, 1997 , RRID:SCR_005927). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head-motion. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD reference was then co-registered to the T1w reference using `mri_coreg` (FreeSurfer) followed by `flirt` (FSL 6.0.6.4, Jenkinson & Smith, 2001) with the boundary-based registration (Greve & Fischl, 2009) cost-function. Co-registration will be configured with six degrees of freedom. Several confounding time-series will be calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD will be computed using two formulations following Power (absolute sum of relative motions, Power et al., 2014) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al., 2002). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al., 2014) . The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors will be extracted to allow for component-based noise correction (*CompCor*, Behzadi et al., 2007; Chai et al., 2012). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al., 2007 in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by thresholding the corresponding partial volume map at 0.05, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks will be resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately

within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals will be expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceed a threshold of 0.5 mm FD or 1.5 standardized DVARS are annotated as motion outliers. The BOLD time-series will be resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version will be generated using a custom methodology of *fMRIPrep*. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings will be performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings will be performed using `mri_vol2surf` (FreeSurfer). Many internal operations of *fMRIPrep* use *Nilearn* 0.9.1 (Abraham et al., 2014, RRID:SCR_001362), mostly within the functional processing workflow.

A denoising step will also be performed using the CONN Toolbox (Nieto-Castanon & Whitfield-Gabrieli, 2022; Whitfield-Gabrieli & Nieto-Castanon, 2012). First, functional data will be smoothed using spatial convolution with a Gaussian kernel of 6 mm full width half maximum (FWHM). In addition, functional data will be denoised using a standard denoising pipeline including the regression of potential confounding variables, computed in the previous steps and characterized by white matter timeseries (5 CompCor noise components), CSF timeseries (5 CompCor noise components), motion parameters and their first order derivatives (12 factors, Friston et al., 1996), outlier scans (below 26 factors; Power et al., 2014), and linear trends (2 factors), followed by bandpass frequency filtering of the BOLD timeseries (Hallquist et al., 2013) between 0.008 Hz and 0.09 Hz. CompCor (Behzadi et al., 2007; Chai et al., 2012) noise components within white matter and CSF will be estimated by computing the average BOLD signal as well as the largest principal components orthogonal to the BOLD average, motion parameters, and outlier scans within each subject's eroded segmentation masks.

TVAs ROI definition. In order to identify the TVAs in subsequent analyses, a mask resulting from the convolution of the group-level activation maps retrieved from Pernet et al. (2015) and a brain mask describing the STS & STG as defined in the Destrieux atlas (DESTRIEUX et al., 2010) was computed. The python code for computing the mask (link), and the final mask can be found in the supplementary materials (link to full repository).

Behavioral data analysis

Global scores from the "attention/memory" task will be compared between groups using a t-test.

fMRI data analysis

Activation analyses

At the subject level, denoised data will be analyzed using GLMs. Boxcar functions will be convolved with the canonical Hemodynamic Response Function (HRF) in order to model the two experimental conditions (i.e. vocal vs. non-vocal sounds). Once the model estimated, a contrast for vocal vs. non-vocal sounds will be computed and individual maps will be thresholded at the 0.02 alpha level, cluster

corrected, with a cluster defining threshold of Z = 2.05 (corresponding to an alpha level of 0.02). Any cluster within the TVA mask previously computed will be labelled as a TVA activation.

Between-group comparisons

The proportion of individuals showing a TVA activation will be compared between ASD and non-ASD individuals using a χ^2 test. Significance will be tested at the 2% α level.

Power analysis

Power analyses has been conducted in order to ensure the feasibility of the χ^2 analysis. First, we pooled the ASD groups' data from Gervais et al. (2004) and Schelinski et al. (2016) in order to describe the proportion of ASD individuals showing a TVA activation. Across the two studies, 16 ASD individuals showed a TVA response and 5 individuals did not. These proportions were compared to the theorical proportion of non-ASD individuals showing a TVA activation as described by Pernet et al. (2015) (i.e., 94%) using a χ^2 test. Results from this test suggested a different proportion of individuals showing a TVA response in ASD and non-ASD individuals ($\chi^2(1)=11.81$, $p=0.000589$). Then an effect size (w) was retrieved from this analysis using the following formula:

$$w = \sqrt{\frac{\chi^2}{n \times df}}$$

Where χ^2 corresponds to the test value (i.e., 11.81), n corresponds to the tested sample size (i.e., 16+5=21), and df corresponds to the analysis' degrees of freedom (i.e., 1). A power analysis using the obtained w effect size was then conducted for a χ^2 test with an α level of 0.02 and a power of 0.90 using the `pwr` (Champely et al., 2020) R package (R Core Team, 2022). This analysis returned a required sample size of 24 individuals per group, emphasizing the suitability of this analysis with our sample size of 26 individuals per group (see the R code for the power analysis).

Study design table

Question	Hypothesis	Sampling plan	Analysis Plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes
Is the ASD population characterized by a different proportion of individuals showing a specific response to vocal sound than the non-ASD population?	We predict a lower proportion of individual TVA activation in the ASD group than in the non-ASD group	Power analysis was conducted using an effect size estimated from the literature in order to estimate the minimum required sample size. Results indicated a minimum required sample size of n = 24 per group with a power level of .90.	A χ^2 test will be conducted in order to compare the proportion of individuals showing a TVA activation in the ASD and non-ASD groups.	Considering the effect size inferred from the literature (w=11.81), power analyses indicated that the study is powered enough to detect a between group difference at the $p < .02$ alpha level.	A different proportion of individuals showing a TVA activation in the ASD group would suggest that at least a subset of ASD individuals does not process the vocal sounds in a typical way. The failure to reject the null hypothesis will be interpreted as an absence of evidence towards either the null or the alternative hypothesis (i.e., the null hypothesis cannot be rejected).	The rejection of the null hypothesis would suggest that ASD individuals suffer from low level deficits in voice processing, which may eventually lead to higher order social dysfunction. If the statistical test fails to reject the null hypothesis, no strong conclusions can be draw considering an (absence of) impairment of voice processing in ASD.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. https://www.frontiersin.org/articles/10.3389/fninf.2014.00014

Agus, T. R., Paquette, S., Suied, C., Pressnitzer, D., & Belin, P. (2017). Voice selectivity in the temporal voice area despite matched low-level acoustic cues. *Scientific Reports*, *7*(1), 11526. https://doi.org/10.1038/s41598-017-11684-1

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation : Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26-41. https://doi.org/10.1016/j.media.2007.06.004

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ) : Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5-17. https://doi.org/10.1023/A:1005653411471

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90-101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding Voice Perception. *British Journal of Psychology*, *102*(4), 711-725. https://doi.org/10.1111/j.2044-8295.2011.02041.x

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice : Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129-135. https://doi.org/10.1016/j.tics.2004.01.008

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 6767. https://doi.org/10.1038/35002078

Bestelmeyer, P. E. G., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014). Adaptation to Vocal Expressions Reveals Multistep Perception of Auditory Emotion. *Journal of Neuroscience*, *34*(24), 8098-8105. https://doi.org/10.1523/JNEUROSCI.4820-13.2014

Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context : A neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, *8*(4), 383-403. https://doi.org/10.1016/j.plrev.2011.10.002

Chai, X. J., Castañón, A. N., Ongür, D., & Whitfield-Gabrieli, S. (2012). Anticorrelations in resting state networks without global signal regression. *NeuroImage*, *59*(2), 1420-1428. https://doi.org/10.1016/j.neuroimage.2011.08.048

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., & Rosario, H. D. (2020). *pwr : Basic Functions for Power Analysis* (1.3-0). https://cran.r-project.org/web/packages/pwr/index.html

Chan, W., Smith, L. E., Hong, J., Greenberg, J. S., & Mailick, M. R. (2017). Validating the social responsiveness scale for adults with autism. *Autism Research*, *10*(10), 1663-1671. https://doi.org/10.1002/aur.1813

Charest, I., Pernet, C., Latinus, M., Crabbe, F., & Belin, P. (2013). Cerebral Processing of Voice Gender Studied Using a Continuous Carryover fMRI Design. *Cerebral Cortex*, *23*(4), 958-966. https://doi.org/10.1093/cercor/bhs090

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., Metzger, L. M., Shoushtari, C. S., Splinter, R., & Reich, W. (2003). Validation of a Brief Quantitative Measure of Autistic Traits : Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders*, *33*(4), 427-433. https://doi.org/10.1023/A:1025014929212

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, *10*(4-5), 171-178. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L

DESTRIEUX, C., FISCHL, B., DALE, A., & HALGREN, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, *53*(1), 1-15. https://doi.org/10.1016/j.neuroimage.2010.06.010

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep : A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 1. https://doi.org/10.1038/s41592-018-0235-4

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*, S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S., & Turner, R. (1996). Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, *35*(3), 346-355. https://doi.org/10.1002/mrm.1910350312

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., Barthélémy, C., Brunelle, F., Samson, Y., & Zilbovicius, M. (2004). Abnormal cortical voice processing in autism. *Nature Neuroscience*, *7*(8), 801-802. https://doi.org/10.1038/nn1291

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63-72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Hallquist, M. N., Hwang, K., & Luna, B. (2013). The Nuisance of Nuisance Regression : Spectral Misspecification in a Common Approach to Resting-State fMRI Preprocessing Reintroduces Noise and Obscures Functional Connectivity. *NeuroImage*, *0*, 208-225. https://doi.org/10.1016/j.neuroimage.2013.05.116

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825-841. https://doi.org/10.1006/nimg.2002.1132

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143-156. https://doi.org/10.1016/S1361-8415(01)00036-6

Lanczos, C. (1964). Evaluation of Noisy Data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, *1*(1), 76-85. https://doi.org/10.1137/0701007

Latinus, M., Crabbe, F., & Belin, P. (2011). Learning-Induced Changes in the Cerebral Processing of Voice Identity. *Cerebral Cortex*, *21*(12), 2820-2828. https://doi.org/10.1093/cercor/bhr077

Latinus, M., McAleer, P., Bestelmeyer, P. E. G., & Belin, P. (2013). Norm-Based Coding of Voice Identity in Human Auditory Cortex. *Current Biology*, *23*(12), 1075-1080. https://doi.org/10.1016/j.cub.2013.04.055

Maguinness, C., Roswandowitz, C., & von Kriegstein, K. (2018). Understanding the mechanisms of familiar voice-identity recognition in the human brain. *Neuropsychologia*, *116*, 179-193. https://doi.org/10.1016/j.neuropsychologia.2018.03.039

Nieto-Castanon, A., & Whitfield-Gabrieli, S. (2022). *CONN functional connectivity toolbox : RRID SCR_009550, release 22* (22e éd.). Hilbert Press. https://doi.org/10.56441/hilbertpress.2246.5840

Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E. G., Watson, R. H., Fleming, D., Crabbe, F., Valdes-Sosa, M., & Belin, P. (2015). The human voice areas : Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *NeuroImage*, *119*, 164-174. https://doi.org/10.1016/j.neuroimage.2015.06.050

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320-341. https://doi.org/10.1016/j.neuroimage.2013.08.048

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240-256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Schelinski, S., Borowiak, K., & von Kriegstein, K. (2016). Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. *Social Cognitive and Affective Neuroscience*, *11*(11), 1812-1822. https://doi.org/10.1093/scan/nsw089

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK : Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310-1320. https://doi.org/10.1109/TMI.2010.2046908

Wechsler, D. (1981). *WAIS-R manual*. The Psychological Corporation.

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

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45-57. https://doi.org/10.1109/42.906424