Singing in the brain: Neural representation of music and voice as revealed by fMRI

Analysis: By Aanya Gupta

INFORMATION ABOUT THE PAPER

AUTHORS	Jocelyne C. Whitehead, Jorge L. Armony
JOURNAL	Human Brain Mapping
YEAR OF PUBLICATION	2018
CITATION	Whitehead, J. C., & Armony, J. L. (2018). Singing in the brain: Neural representation of music and voice as revealed by fMRI. *Human Brain Mapping*, *39*(12), 4913–4924. https://doi.org/10.1002/hbm.24333

TABLE OF CONTENTS

01

INTRODUCTION

TERMINOLOGY:

Superior Temporal Gyrus (STG)	Site of auditory association cortex
Superior Temporal Sulcus (STS)	Main region for audiovisual integration.
Planum Polare (PP)	Involved in auditory processing and receptive language
Middle Temporal Gyrus (MTG)	Primary sources of visual motion information
Vocal Temporal Area (VTA)	You'll see in the results :)
Planum Temporale (PT)	Involved in auditory processing and receptive language as well

CENTRAL QUESTION

The aim of this study was to identify the brain's responses to vocal and instrumental music/sounds by looking at the areas of activation in the brain with an fMRI when listening to auditory stimuli.

HYPOTHESIS

- They expected to replicate previous studies showing that instrumental music, when compared with speech, activates a bilateral region in the anterior STG, particularly in the planum polare (PP), whereas speech would elicit responses along the STS.

- Critically, we hypothesized that vocal music (i.e singing) would represent an intermediate condition between these two.

- Namely, when compared with music, singing should activate STS, but when compared with speech, it should yield activations overlapping with those associated with instrumental music within the PP" (4914-4915).

SIGNIFICANCE OF WORK / BACKGROUND

- Most studies have solely employed instrumental music, and so questions still exist as to the specificity of the observed "music-preferred" areas.

- Focused on the question on whether speech and music activate distinct or overlapping regions in the brain, especially within the auditory cortex

- Taken together, these findings provide further support for a hierarchical processing of complex social acoustic stimuli along the temporal lobes, similar to what has been reported for the visual modality

- **The findings of this study contribute to the ongoing debate regarding the syntactic parallels that music has with speech, its comparable use for communicating emotional states, and their potential common evolutionary origins.**

02

METHODS

METHODOLOGY

- 24 male and female participants were involved in this study.
- The experiment consisted of three 8 minute runs, two using a continuous multiband sequence and one with an interleaved silent study state sequence.
  - Each run consisted of 90 stimuli; 30 speech, 30 singing, and 30 instrumental music excerpts
- The paper does not specify the duration of the study.
- Each of the participants passively listened to auditory stimuli while watching nature scenes. Stimuli were presented using E-prime 2.0 and delivered binaurally from MRI-compatible headphones.
- Auditory stimuli that the participants listened to belonged to one of three categories:
  - Instrumental Music
  - Speech
  - Acapella Singing

METHODOLOGY: CONTINUED

- All auditory stimuli were monaural, but presented binaurally. They were adjusted for loudness using the Moore and Glasberg Loudness model.
- Basic acoustic parameters for each of the categories were calculated using the MIRtoolbox, MATLAB scripts, and Praat Vocal Toolkit.
- High-resolution fMRI of the participants' neural responses to the auditory stimuli was recorded using a 3T Siemens TIM TRIO MRI scanner.

METHODOLOGY: CONTINUED

- Continuous acquisition: The auditory stimuli were presented in a continuous design and were jittered using a brief ISI (duration: M = 2.49 s, SD = 0.20 s)" (4915)

- Sparse acquisition: Performed using a finite impulse response (FIR) model, in the context of the general linear model, in which each of the four acquisition volumes for the two sound types was entered as a separate category

METHODOLOGY: CONTINUED

- Univariate analysis: Categories of interest (Instrumental Music, Singing, and Speech) were entered as boxcars of length equal to the stimulus duration, convolved with the canonical hemodynamic response function" (4915).
    - An ANOVA was used to identify the difference between instrumental music vs speech, instrumental music vs singing, and singing vs speech.
    - Statistical significance was determined using a voxel threshold *of* $p = .001$.
    - To identify regions commonly activated for those different categories, conjunction analyses were performed.
    - A stimulus-based analysis was also conducted, which was used for post-hoc regression analyses and multivariate analysis

- Multivariate analysis: "Parameter estimate images obtained in the stimulus-based analysis were submitted to an Independent Component Analysis (ICA)" (4916).

03

DATA + RESULTS

STATISTICAL DATA

- Continuous acquisition:
  - Univariate Analysis:
    - Instruments minus speech = Significant clusters in the PT and PP.
    - Singing minus Speech = Significant clusters in the PP and in the right PT.
    - Speech vs. Instrumental Music = Significant activity in voice-preferred areas within the STS, STG, and MTG.
    - Singing vs. Instruments = Largely overlapping activations
    - Speech versus singing = Significant clusters bilaterally in the STS, STG, and MTG
    - "Instruments minus Speech = 88% and 75% of subjects had significant clusters on the right and left hemispheres, respectively, using a significance threshold of $p = .01$" (4917)

STATISTICAL DATA

- Multivariate Analysis:
  - Smaller activation for instruments compared with speech + singing with no difference between the two vocal sounds.
  - No overlap between instruments + speech whereas singing fell in between the two.
- Sparse acquisition:
  - Instrumental music minus speech = Significant clusters in the PP and the right PT and in the STS, STG, and MTG.
  - No additional activation clusters were observed for either of the comparisons when using the silent protocol.
  - Significant correlation of the subject specific, cluster average parameter estimates for the contrast instrumental music minus speech between both runs for each of the three main clusters" (4917).

FIGURES

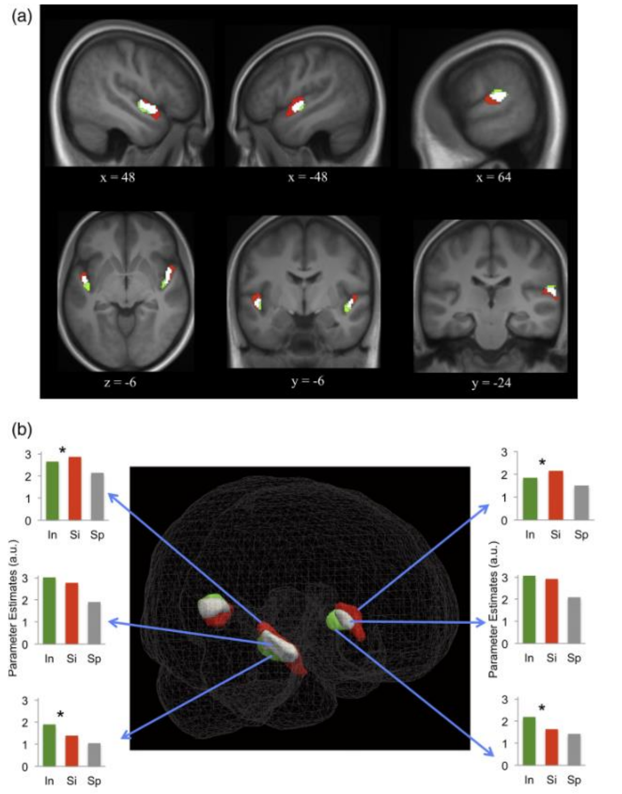

TABLE 2 Significant activations associated with contrasts of interest at the group level

Anatomical location	Left			Right			Z-score (peak voxel)	K_E
	x	y	z	x	y	z		
Continuous multiband sequence								
Instrumental music > speech								
STG (posterior)				66	−28	12	5.99	155
STG (anterior)				46	−6	−6	5.88	241
STG (anterior)	−48	−6	−4				5.37	162
Singing > speech								
STG (anterior)				50	4	−8	6.29	319
STG (posterior)				66	−26	10	5.74	213
STG (anterior)	−48	0	−6				7.43	286
[Instrumental music and singing] > speech								
STG (posterior)				66	−26	10	5.74	123
STG (anterior)				48	4	−8	5.74	154
STG (anterior)	−48	−6	−4				5.37	85
Speech > instrumental music								
STS/STG, MTG				64	−8	−4	12.63	1,570
STG/STS, MTG	−62	−12	0				15.14	2008
Singing > instrumental music								
STS/STG, MTG				62	−20	−2	11.04	1,440
STS/STG, MTG	−60	−12	2				12.31	1835
[Speech and singing] > instrumental music								
STS/STG, MTG				62	−20	−2	11.04	1,250
STS/STG, MTG	−60	−12	1				12.31	1,612
Speech > singing								
STS/STG, MTG				64	−8	−4	7.22	845
STS/STG, MTG	−60	−22	−2				9.36	1,239
Interleaved silent steady state (ISSS) sequence								
Instrumental music > speech								
STG (posterior)				48	−32	24	4.39	238
STG (anterior)				42	−12	−10	4.88	344
STG (anterior)	−48	−6	−0				4.79	253
Speech > instrumental music								
STS/STG, MTG				56	−28	−2	7.54	863
STG/STS, MTG	−64	−28	4				8.01	1,238

STG = superior temporal gyrus; STS = superior temporal sulcus; MTG = medial temporal gyrus.

TABLE 1 Mean and standard deviation values of acoustic features for each sound category

Audio features	Music	Singing	Speech
Articulation (a.u.)	.32 (.21)[a]	.27 (.15)[a]	.44 (.09)[a]
Root mean square (dB)	.13 (.05)	.16 (.04)[b]	.13 (.04)
Tempo (bpm)	125 (30)	137 (29)	126 (30)
Spectral centroid (kHz)	2.3 (1.5)	2.4 (1.0)	2.1 (1.0)
Spectral brightness (>1.5 kHz)	.44 (.26)	.42 (.16)	.37 (.15)
Spectral spread (Hz)	5.8 (3.3)	6.0 (2.6)	5.1 (1.6)
Spectral Skewness (a.u.)	.21 (.20)*	.37 (.36)*	.30 (.28)
Spectral kurtosis (a.u)	.73 (1.09)	1.5 (3.0)	.86 (1.43)
Spectral roll off 95th percentile (kHz)	4.2 (2.7)	5.0 (2.3)	4.3 (2.0)
Spectral Spectentropy (bits)	.76 (.08)[b]	.80 (.05)	.81 (.04)
Spectral flatness	.05 (.08)	.06 (.05)*	.04 (.03)*
Spectral irregularity	.78 (.32)	.95 (.38)*	.67 (.37)*
Zerocross (s^{-1})	1,335 (1,206)	97(517)	1,137 (548)
Low energy ratio	.54 (.10)	.48 (.08)[b]	.52 (.07)
Key clarity (a.u.)	6.8 (3.3)	6.0 (3.3)	7.1 (3.4)
Tonal mode (minor-major, a.u.)	−.02 (.12)	−.02 (.10)	−.02 (.08)
Pulse clarity (a.u.)	.28 (.17)[b]	.18 (.09)	.23 (.08)
Mean fundamental frequency (F0)	275 (138)	273 (90)	185 (56)[b]
Std. Dev. Fundamental frequency (F0)	47.2 (37.9)[b]	31.6 (24.9)	29.6 (15.4)
Minimum fundamental frequency (F0)	204 (109)	217 (79)	134 (45)[b]
Maximum fundamental frequency (F0)	353 (169)	327 (111)	246 (78)[b]
Fraction of locally unvoiced frames (%)	10.6 (13.3)	7.6 (9.1)	23.7 (13.1)[b]
Jitter (local) (%)	2.24 (2.57)	1.43 (1.19)[b]	2.25 (.71)
Shimmer (local) (%)	12.7 (6.8)[b]	10.1 (5.1)	10.2 (3.2)
Mean HNR	11.4 (8.0)	13.7 (5.4)	11.4 (3.1)

a.u. = arbitrary units; bpm = beats per minute. Values were calculated with MIRToolbox, except for those related to the Fundamental Frequency (http://www.tik.ee.ethz.ch/~spr/f0_detection) and the last four features (Praat).

[a] All significantly different.

[b] Significantly different from the other two.

*Significantly different from each other ($p < .05$, Bonferroni corrected).

04

DISCUSSION

CONCLUSIONS FROM DATA

HYPOTHESIS: Instrumental music would activate the PP in the anterior STG, speech would activate the STS, and singing would activate both regions.

- When compared with musical instruments, human voice, either spoken or sung, elicited significant activations in clusters along the STS in both hemispheres.

- Speech always elicited the strongest response in both hemispheres. These clusters exhibited a bias, in terms of magnitude, for human voice and responded to nonvocal sounds, confirming that the so-called vocal temporal area (VTA) should be considered as a “voice-preferring” rather than as a “voice-selective” region.

- “Music—either in instrumental or vocal form—yielded significant clusters in the anterior planum polare bilaterally and in the right planum temporale” (4919).
    - Clusters within these regions respond more strongly to both instrumental and vocal music than speech, with no significant differences between the first two categories.

CONCLUSIONS: CONTINUED

- “The overlap of the instrumental and vocal music versus speech clusters was not complete. Specifically, more posterior regions of PP responded more strongly to instrumental music than both speech and singing” (4921).

- To summarize, the authors found that vocal and musical stimuli elicited preferred responses from different parts of the temporal lobe. These included the superior temporal sulcus and gyrus for the former, and the planum polare and temporale for the latter.

- Since singing has vocal and musical properties, it used all of these areas.

WEAKNESSES: CONTINUED

- Other studies found that different cortical areas are activated depending on the language of stimulus and the participants' experiences with it. In this experiment, playing the vocal stimuli in so many different languages could lead to different areas of activation depending on those variables.

- The authors state that scanner noise might have an influence on the observed responses.
  - The authors took precautions so that the scanner noise did not influence the results of the experiment

- Previous studies attempted to control for the possible effects of general acoustic characteristics of the stimuli employed, but there are still important qualitative and quantitative differences between instrumental music and voice, which could affect the results.

WEAKNESSES OF STUDY

- “While there was no explicit task for the participants to perform, one cannot exclude the possibility that some of the participants performed some sort of stimulus categorization.

- The authors failed to identify a linear combination of acoustic parameters that correlated with activity in this region, including factors previously identified as differentiating singing from speech, such as duration, fundamental frequency floor, and vocal intensity” (4921).

- There were only 24 participants involved in this study, which means that the results might not have been accurately representative of the population.

- The authors do not specify what genre of music was played (for both the vocal and instrumental stimuli), which could have affected the areas of activation in the brain.

STRENGTHS OF STUDY

- Their methodology is strong and does not contain many flaws.
- They made sure to include people of various genders and ages in the study with a range of musical and lengual experience.
- The experiment and methods are also unbiased.
- The results are precise + accurate since the majority of the results were recorded using online software.
- The structure of their article is clear and informative, and the results are presented in a legible and informative way.
- The reasoning behind their results is clearly explained with factual scientific evidence to back it up.
- They informed the audience of the numerical evidence that they obtained, and then explained it in simpler terms + reasoning for results and then generated overall conclusions.
- They cited all of their sources

05

CONCLUSION

OVERALL CONCLUSION

- To summarize, the authors found that vocal and musical stimuli elicited preferred responses from different parts of the temporal lobe. These included the superior temporal sulcus and gyrus for the former, and the planum polare and temporale for the latter.

- Since singing has vocal and musical properties, it used all of these areas.

THANKS!

Do you have any questions?