High-level neural categorization of human voices as revealed by fast periodic auditory stimulation

Francesca M. Barbero[1], Roberta P. Calce[1], Bruno Rossion[1,2,3], Olivier Collignon[1,4]

1. Institute of research in Psychology (IPSY) & Institute of Neuroscience (IoNS), University of Louvain (UCL), Louvain-la-Neuve, Belgium
2. Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France 3. Université de Lorraine, CHRU-Nancy, Service de Neurologie, F-54000, France 2. Centre for Mind/Brain Sciences, University of Trento, Trento, Italy

Introduction

Several studies have corroborated the existence of voice preferring regions in the human brain[1,2]. Whether this preference is driven by low level acoustic properties peculiar of voices, or whether it reflects a higher-level categorical response is still under debate.

We combined EEG recording with a Fast Periodic Auditory Stimulation (FPAS) oddball paradigm to investigate whether categorically responses to voices partially abstracts from some low-level acoustic features.

EEG - Fast auditory periodic stimulation

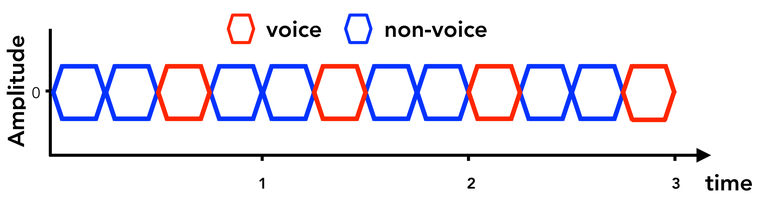
Sixteen participants (19-26 years) listened to three types of periodic sequences:

- **standard** → voice selectivity
- **scrambled** → control for frequency content
- **harmonic** → control for harmonicity, pitch

Voices were presented each third sound in 60 s-long sequences

- 4 repetition per sequence type
- Orthogonal non-periodic task

Standard sequence



Vocal sounds (55 stimuli):

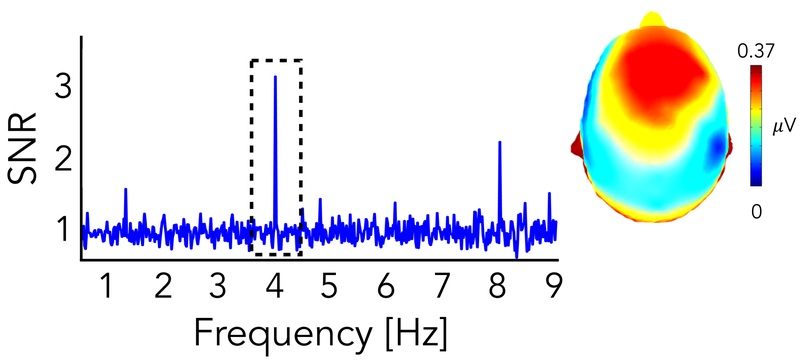
- Speech and non-speech vocalizations
- Speakers of different age, sex, emotional state

Non-vocal sounds (137 stimuli):

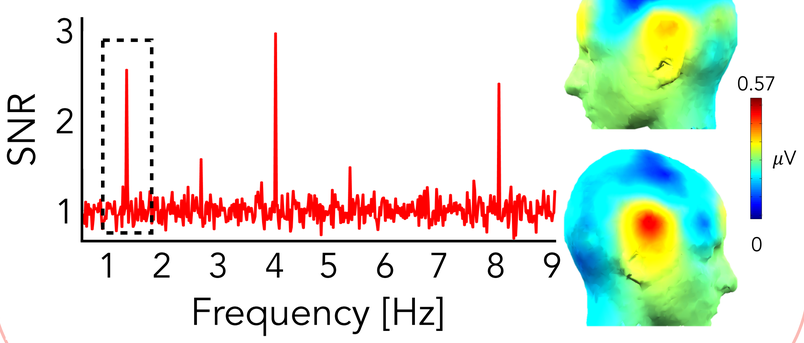
- Naturalistic sounds (6)
- Instruments (32)
- Objects (89)
- Auditory scenes (10)

250 ms-long sounds

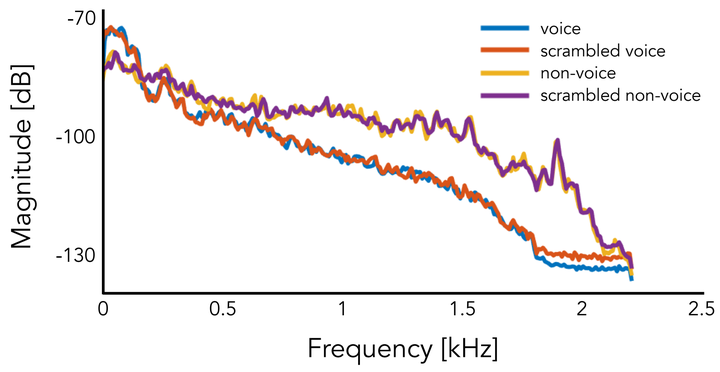
Base frequency 4 Hz



Target frequency 1.33 Hz



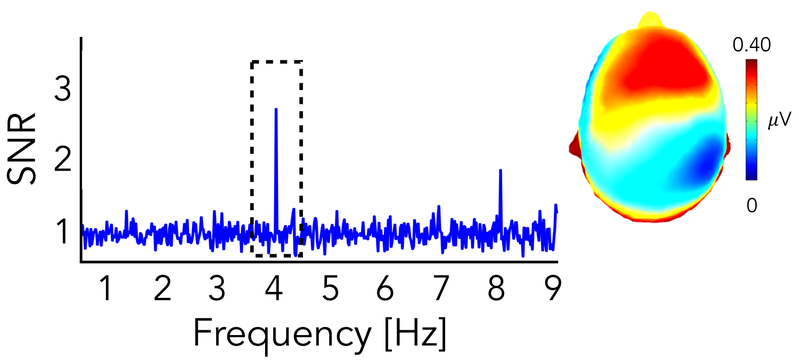
Scrambled sequence



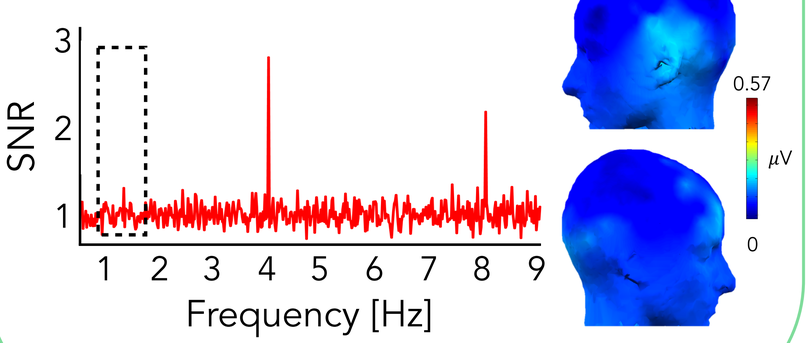
Stimuli as in standard sequence, frequency scrambled[4]:

- Frequency content and temporal structure close to that of the original stimuli
- Altered harmonicity
- Sounds are not recognisable anymore

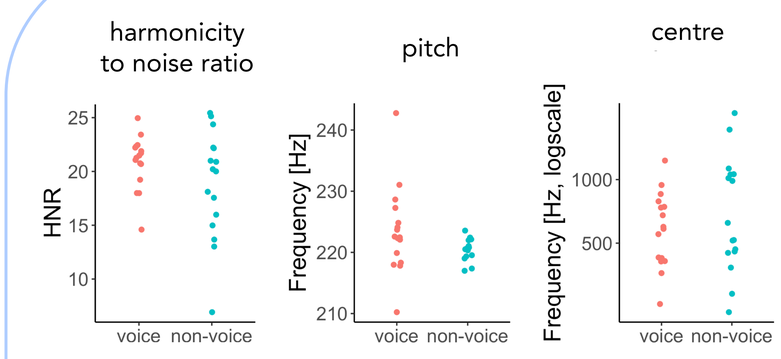
Base frequency 4 Hz



Target frequency 1.33 Hz



Harmonic sequence

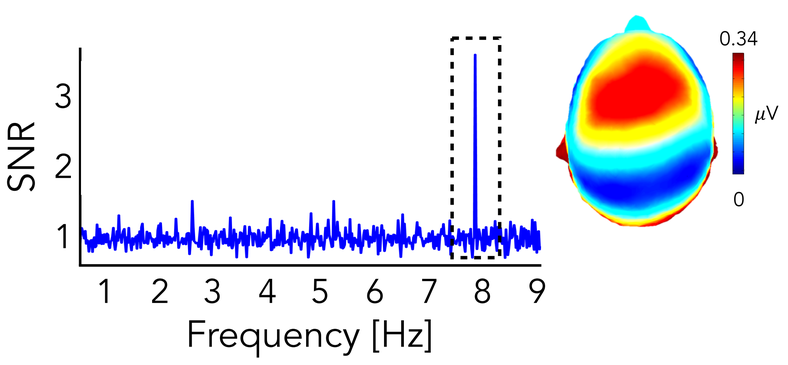


Vocal sounds (singing voices, 16 stimuli) and non-vocal sounds (instruments, 16 stimuli)[5,6] matched for:

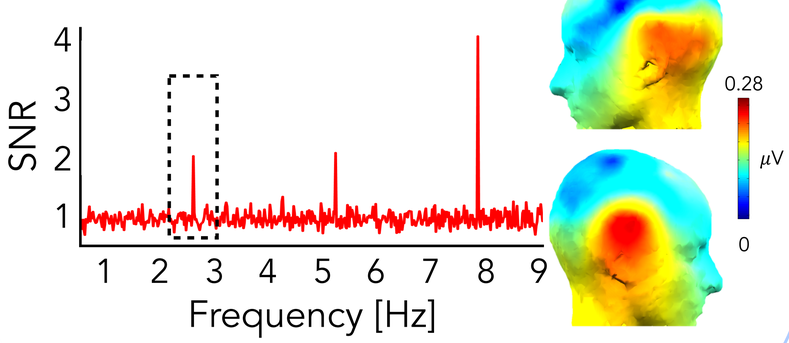
- Harmonicity-to-noise ratio (HNR)
- Pitch, spectral centre of gravity

128 ms-long sounds

Base frequency 7.8 Hz

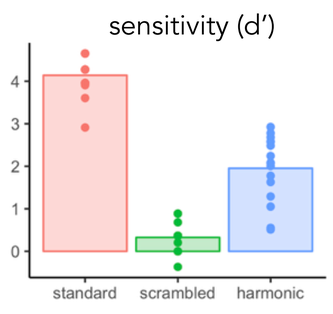


Target frequency 2.6 Hz



Behavioral experiment

- Sixteen participants (18-26) listened to short sequences (5 sounds) of the three types
- Participants had to indicate whether each sequence contained a voice or not (50% occurrences)



Conclusion

- Our results show robust voice selective brain responses over superior temporal electrodes that cannot be explained by frequency content nor harmonicity typical of voice samples alone.
- FPAS paradigm allowed us to characterize voice selective responses with a high signal to noise ratio in a very short acquisition time (4 minutes only).

[1] Belin et al. 2000, Nature; [2] Belin et al. 2002, *Cogn Brain Res*; [3] Liu-Shuang et al. 2014, *Neuropsychologia*; [4] Dormal et al. 2018, *J. Cogn. Neurosci.*; [5] Agus et al. 2017, *Sci. Rep.*; [6] Goto et al. 2003, *Proc. Int. Conf. Music Inf. Retr.*

francesca.barbero@uclouvain.be