ARTICLE IN PRESS

NeuroImage xxx (2017) 1–10

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

# Cortical processing of pitch: Model-based encoding and decoding of auditory fMRI responses to real-life sounds

Vittoria De Angelis [a,b], Federico De Martino [a,b,d], Michelle Moerel [c,a,b], Roberta Santoro [a,b], Lars Hausfeld [a,b], Elia Formisano [a,b,c,*]

[a] Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, The Netherlands
[b] Maastricht Brain Imaging Center (MBIC), Maastricht University, Maastricht, The Netherlands
[c] Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, The Netherlands
[d] Center for Magnetic Resonance Research, University of Minnesota Medical School, 2021 Sixth Street SE, Minneapolis, MN 55455, United States

A R T I C L E I N F O

*Keywords:*
Auditory cortex
Pitch processing
Real-life sounds
fMRI
Decoding
Encoding

A B S T R A C T

Pitch is a perceptual attribute related to the fundamental frequency (or periodicity) of a sound. So far, the cortical processing of pitch has been investigated mostly using synthetic sounds. However, the complex harmonic structure of natural sounds may require different mechanisms for the extraction and analysis of pitch. This study investigated the neural representation of pitch in human auditory cortex using model-based encoding and decoding analyses of high field (7 T) functional magnetic resonance imaging (fMRI) data collected while participants listened to a wide range of real-life sounds. Specifically, we modeled the fMRI responses as a function of the sounds' perceived pitch *height* and *salience* (related to the fundamental frequency and the harmonic structure respectively), which we estimated with a computational algorithm of pitch extraction (de Cheveigné and Kawahara, 2002). First, using single-voxel fMRI encoding, we identified a pitch-coding region in the antero-lateral Heschl's gyrus (HG) and adjacent superior temporal gyrus (STG). In these regions, the pitch representation model combining height and salience predicted the fMRI responses comparatively better than other models of acoustic processing and, in the right hemisphere, better than pitch representations based on height/salience alone. Second, we assessed with model-based decoding that multi-voxel response patterns of the identified regions are more informative of perceived pitch than the remainder of the auditory cortex. Further multivariate analyses showed that complementing a multi-resolution spectro-temporal sound representation with pitch produces a small but significant improvement to the decoding of complex sounds from fMRI response patterns.

In sum, this work extends model-based fMRI encoding and decoding methods - previously employed to examine the representation and processing of *acoustic* sound features in the human auditory system - to the representation and processing of a relevant perceptual attribute such as pitch. Taken together, the results of our model-based encoding and decoding analyses indicated that the pitch of complex real life sounds is extracted and processed in lateral HG/STG regions, at locations consistent with those indicated in several previous fMRI studies using synthetic sounds. Within these regions, pitch-related sound representations reflect the modulatory combination of height and the salience of the pitch percept.

## Introduction

Pitch plays an essential role in auditory perception, enabling us, for example, to identify distinct speakers and to perceptually organize the acoustic elements of a complex scene (Bregman, 1990; Moore, 1995). For harmonic tones, pitch is the perceptual correlate of the fundamental frequency F0, that is the sound's lowest frequency value of which all the spectral components are an integer multiple. As the same pitch can be perceived even after removal of the energy at F0 (i.e. in the case of *missing fundamental*), pitch is more generally defined in relation to the repetition rate (or periodicity) of the temporal envelope of the sound. Indeed, the energy content at the fundamental frequency does not influence the periodicity of the temporal envelope, which is solely determined by the spacing of the harmonics (de Cheveigné, 2010).

The neural mechanisms underlying pitch perception are still largely debated. The "temporal" hypothesis assumes that the periodicity is

* Corresponding author. Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6200 MD Maastricht, The Netherlands.
*E-mail address:* e.formisano@maastrichtuniversity.nl (E. Formisano).

https://doi.org/10.1016/j.neuroimage.2017.11.020
Received 21 March 2017; Accepted 11 November 2017
Available online xxxx

extracted based on the timing between successive spikes in the auditory nerve. In contrast, the “place” theory infers that pitch is determined by the harmonic template that best matches the spectral cues encoded *tonotopically* in the cochlea and throughout the ascending auditory pathway (Plack et al., 2005). Recent accounts suggest that both place and timing information are necessary in order to perceive the correct pitch (Oxenham, 2013; Oxenham et al., 2004; Shamma, 2004).

Several studies investigated the neural (fMRI) correlates of pitch processing in subcortical and cortical structures of the human auditory system by comparing the BOLD responses for a wide range of pitch evoking sounds and noise control stimuli. Using iterated ripple noise (IRN), Griffiths et al. (2001) found a positive correlation between temporal regularity and local brain activity in the cochlear nucleus (CN) and in the inferior colliculus (IC) bilaterally. Moreover, the contrast between time-varying and fixed pitch sequences revealed significant activation differences only in the auditory cortex, specifically in lateral Heschl's gyrus (HG) and in planum temporale (PT) bilaterally, as also revealed with PET (Griffiths et al., 1998). This suggested a hierarchy of pitch processing stages starting in the subcortical structures, which are sensitive to temporal regularity, and terminating at the cortical level, where perceived pitch (variations) are most likely encoded. Patterson et al. (2002) reported a selective activation in lateral HG both in response to pitch-producing IRN and melodic sounds. Barker et al. (2012) argued that the activity elicited by the IRN in lateral HG was due to the fine temporal structure of the stimuli instead of pitch *per se*, as the contrast between the responses to conventional IRN and “no pitch” IRN control sounds did not show a significant difference. A pitch-tuned region was identified in the “anterior half of the auditory cortex” in Norman-Haignere et al. (2013). The activation of this region was predominantly driven by the resolved harmonics of the stimuli, and overlapped with a low-frequency area in the tonotopy map. These results are consistent with single-unit recordings in marmoset monkeys, reporting pitch-selective neurons located in a low-frequency region near the antero-lateral border of the primary auditory cortex (Bendor and Wang, 2005), potentially corresponding to lateral HG in humans (Bendor, 2012; Bendor and Wang, 2006). In addition, selective activation in response to pitch-evoking dichotic stimuli (Huggins pitch) has been observed in PT (Garcia et al., 2010; Hall and Plack, 2007, 2009). Importantly, a covariation of neural activity and pitch salience (dissociated from the physical stimulus regularity) was revealed in a cortical area located in the antero-lateral end of HG bilaterally (Penagos et al., 2004), whereas no such relation has been found for the PT region. In summary, fMRI findings support the hypothesis that the auditory cortex is involved in pitch perception. However, the exact location of a presumed pitch processing center in the human auditory cortex remains controversial (Griffiths and Hall, 2012).

The above-mentioned studies examined pitch processing by measuring fMRI responses to synthetic stimuli. However, for sounds occurring in everyday life pitch perception is more complex than for these artificial stimuli. For instance, the pitch of complex sounds may be influenced by the sound's overall spectral content and especially by the spectral locus of maximum energy concentration, which also relates to the brightness of timbre (de Cheveigné, 2005). Moreover, the strength of the pitch percept (or *salience*) is influenced by the degree to which the spectral components of sounds are harmonic, such that inharmonic sounds tend to evoke a pitch less salient than the one evoked by harmonic tones (Houtsma, 1997). As most of the sounds originating from natural and man-made sources are not perfectly harmonic, the brain processing underlying pitch perception for real-life sounds necessarily entails computational and representational mechanisms for extracting and combining multiple dimensions of pitch, notably pitch height (i.e. the dimension of pitch specifically related to F0) and pitch salience (i.e. the dimension of pitch related to sound harmonic structure).

The aim of the present study was to investigate these mechanisms in human auditory cortex through the model-based analysis of 7 T fMRI responses to real life sounds. First, we used single-voxel encoding (Kay et al., 2008b) and modeled the fMRI responses to a large set of complex naturalistic sounds as a function of sound representation models incorporating information on pitch *height* alone, pitch *salience* alone or on a *weighted* combination of height and salience. Both pitch height and salience were estimated using the YIN algorithm (de Cheveigné and Kawahara, 2002). Then, we evaluated the capability of these various models to predict the responses to a left-out sample of stimuli. The prediction accuracy obtained for these models were compared to each other and to the accuracy obtained with models describing the sounds by their spectral energy content on the same set of features as the pitch models (i.e. frequency bins). Results showed that fMRI responses in cortical regions located bilaterally in lateral HG and adjacent STG were predicted better by the pitch-based than by the energy-based sound representations. Moreover, in the right hemisphere regions, the prediction accuracy for the model combining pitch height and salience was significantly better than for the other pitch models.

Our previous work has shown that fMRI single-voxel responses (Santoro et al., 2014) and response patterns (Santoro et al., 2017) to natural sounds can be predicted accurately by a sound representation model based on the combination of spectro-temporal modulations (Chi et al., 2005). Sound representations explicitly encoding for pitch are expected to provide complementary and relatively independent information on the sound. In fact, current models of auditory scene analysis hypothesize that the auditory system uses pitch in parallel to the multi-resolution representation for parsing the auditory objects of complex scenes (Elhilali and Shamma, 2008; Shamma et al., 2011). Thus, a final aim of the study was to test whether a sound representation model based on pitch - used as a complement to the multi-resolution model - can provide additional information for decoding complex sounds from fMRI response patterns. We addressed this question using model-based multi-voxel decoding (Miyawaki et al., 2008; Santoro et al., 2017). Results showed that pitch information contributed to sound decoding significantly only for circumstantiated regions in lateral HG and STG and not in the remainder of the auditory cortex, which supports the hypotheses on the relevance of these regions for coding pitch information.

## Materials and methods

### *Subjects and ethical statement*

Five healthy subjects that were different for the two experiments participated in Experiment 1 ($n_1 = 5$, median age = 32, three males) and Experiment 2 ($n_2 = 5$, median age = 27 years, two males). The data of Experiment 1 and Experiment 2 have been previously described (Exp. 1: De Martino et al., 2013; Moerel et al., 2013; Santoro et al., 2014; Exp. 2: Santoro et al., 2017, publicly available at https://doi.org/10.5061/dryad.np4hs) and are analyzed here using a new approach. In this section the relevant elements of experimental procedures and fMRI response estimation will be described. All subjects (Experiment 1 and Experiment 2) reported no history of hearing disorder or neurological disease, and gave informed consent before commencement of the measurements. The Institutional Review Board for human subject research at the University of Minnesota (Experiment 1) and the Ethical Committee of the Faculty of Psychology and Neuroscience at Maastricht University (Experiment 2) granted approval for the study. Procedures followed the principles expressed in the Declaration of Helsinki. Informed consent was obtained from each participant before conducting the experiments.

### *Experimental procedures and fMRI responses estimation*

Stimuli consisted of recordings of natural sounds including speech, voices, animal cries, scenes from nature, musical instruments and tool sounds (168 and 288 sounds for Experiment 1 and 2 respectively, 16 000 Hz sampling frequency, 1000 ms duration). In Experiment 1, for each subject 8 functional runs were collected; 144 sounds were presented in 6 training runs with 3 repetitions overall while the remaining 24