NeuroImage 166 (2018) 60–70

Contents lists available at ScienceDirect

NeuroImage

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

Encoding of natural timbre dimensions in human auditory cortex

Emily J. Allen[a,*], Michelle Moerel[b,c], Agustín Lage-Castellanos[b,d], Federico De Martino[b,e], Elia Formisano[b,c], Andrew J. Oxenham[a]

[a] Department of Psychology, University of Minnesota, Minneapolis, MN 55455, United States
[b] Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6200 MD, Maastricht, The Netherlands
[c] Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, 6200 MD, Maastricht, The Netherlands
[d] Department of NeuroInformatics, Cuban Center for Neuroscience, Street 190 e/25 and 27 Cubanacán Playa Havana, CP 11600, Cuba
[e] Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Minneapolis, MN 55455, United States

ARTICLE INFO

Keywords:
Auditory cortex
Encoding models
Music
Perception
Timbre

ABSTRACT

Timbre, or sound quality, is a crucial but poorly understood dimension of auditory perception that is important in describing speech, music, and environmental sounds. The present study investigates the cortical representation of different timbral dimensions. Encoding models have typically incorporated the physical characteristics of sounds as features when attempting to understand their neural representation with functional MRI. Here we test an encoding model that is based on five subjectively derived dimensions of timbre to predict cortical responses to natural orchestral sounds. Results show that this timbre model can outperform other models based on spectral characteristics, and can perform as well as a complex joint spectrotemporal modulation model. In cortical regions at the medial border of Heschl's gyrus, bilaterally, and regions at its posterior adjacency in the right hemisphere, the timbre model outperforms even the complex joint spectrotemporal modulation model. These findings suggest that the responses of cortical neuronal populations in auditory cortex may reflect the encoding of perceptual timbre dimensions.

Introduction

Timbre, the perceptual quality or color of a sound, is defined as everything by which a listener can distinguish between two sounds with the same loudness, pitch, spatial location, and duration (ANSI, 2013). For instance, it is differences in timbre that allow us to distinguish a violin from a guitar, or one vowel sound from another. Among the typical adjectives that fall under the category of timbre are “brightness”, “clarity”, “harshness”, “fullness”, and “noisiness” (Stepanek, 2006). Efforts have been made to identify and quantify the most salient aspects of timbre through the use of multidimensional scaling (MDS) techniques (e.g., Grey, 1977; Elliott et al., 2013). MDS utilizes subjective measures to determine how perceptually similar a selection of sounds are to one another, thereby creating a geometric representation that derives the subjective distances between a diverse set of stimuli using as few dimensions as possible (Grey, 1977). After collecting similarity ratings for musical instrument sounds with unique timbres, Grey (1977) used MDS to identify three dimensions that best represented the distribution of timbres. The first dimension was related to the spectral energy distribution of the sounds (ranging from a low to high spectral centroid, corresponding to timbral descriptors ranging from dull to bright), and the other two related to temporal patterns, such as whether the onset was rapid (like a struck piano note or a plucked guitar string) or slow (as is characteristic of many woodwind instruments) and the synchronicity of higher harmonic transients.

Grey's influential study contained only sixteen instrumental sounds from three instrument families, placing some limits on the generalizability of the outcomes, and used sounds that may not have all had exactly the same fundamental frequency (F0), which itself may have affected some aspects of timbre judgments (e.g., Moore and Glasberg, 1990; Warrier and Zatorre, 2002; Allen and Oxenham, 2014). Elliott et al. (2013) extended Grey's approach by using 42 natural orchestral instruments from five instrument families, all with the same F0 (311 Hz, the E♭ above middle C). After collecting similarity and semantic ratings, they performed multiple analyses, including MDS. They consistently found five dimensions to be both necessary and sufficient for describing the timbre space of these orchestral sounds.

The aim of the current study was to determine whether similar dimensions can be identified in the cortical representations of timbral differences. Although the literature on the neural representations of

* Corresponding author.
E-mail address: prac0010@umn.edu (E.J. Allen).

https://doi.org/10.1016/j.neuroimage.2017.10.050
Received 31 March 2017; Received in revised form 19 October 2017; Accepted 24 October 2017

timbre is limited, there is some evidence to suggest it is processed in both primary and secondary auditory cortical regions including superior temporal sulcus (STS), posterior Heschl's gyrus (HG), and planum temporale (PT), bilaterally, with possible hemispheric asymmetries (Casey et al., 2012; Halpern et al., 2004; Menon et al., 2002; Staeren et al., 2009; Warren et al., 2005). However, previous studies have not attempted to differentiate the neural representations of different timbral dimensions, and have not explored the possibility that a subjectively based model of timbre could predict patterns of cortical activation in response to sound. In the present study, we use fMRI encoding (Kay et al., 2008; Moerel et al., 2012; Santoro et al., 2014) to determine whether neural populations in the cortex can represent the timbre dimensions identified by Elliott et al. (2013), and compare this model's performance with that of models based on the spectral and temporal characteristics of the sounds.

Materials and methods

Ethics statement

The experimental procedures were approved by the Institutional Review Board (IRB) for human subject research at the University of Minnesota. Written informed consent was obtained from each participant before starting the measurements.

Participants

Ten right-handed subjects (mean age of 28.6 years, standard deviation [STD] = 8.6 years; five females, five males) participated in this study. All subjects had normal hearing, defined as audiometric pure-tone thresholds of 20 dB hearing level (HL) or better, at octave frequencies between 250 Hz and 8 kHz, and were recruited from the University of Minnesota community. Musical experience of subjects ranged from zero to 18 years, with eight of the 10 subjects having at least 10 years of musical experience.

Stimuli and procedure

The stimulus set consisted of 42 professionally recorded natural Western orchestral instrument sounds, taken from the study of Elliott et al. (2013). The sounds were originally obtained from the McGill University Master Samples collection (Opolko and Wapnick, 2006) and were manipulated to all have the same F0 of 311 Hz (E♭), and a subjective duration of 1 s, as described in Elliott et al. (2013). Spectrograms for a subset of these sounds are shown in Fig. 1. Instrument families included strings, flutes, brass, single reeds, and double reeds. When the rms of the stimuli was normalized, the perceptual loudness of the sounds at the level of 75 dB SPL varied noticeably. In order to equalize the perceived loudness of the stimuli, we processed them using a loudness model (Chen et al., 2011; Moore, 2014), and scaled the sounds to produce roughly equal predicted loudness for each sound. This resulted in perceptually equal loudness for 41 of the 42 sounds. One of the sounds, a muted C trumpet, required manual adjustment to subjectively match the perceptual loudness of the other sounds, presumably because certain aspects of the sound (e.g., sharp attack and broad spectrum) were not adequately captured by the loudness model. The adjusted level was selected by four raters (inter-rater differences were no more than 2 dB).

After the loudness adjustments, the average level of the sounds was 74 dB SPL and the range was 62–81 dB SPL (STD = 3.2 dB). Sounds were presented via MRI-compatible Sensimetrics (Malden, MA) S14 earphones with custom filters.

Magnetic resonance imaging

Images were acquired in a 3T MR scanner (Siemens Prisma) at the

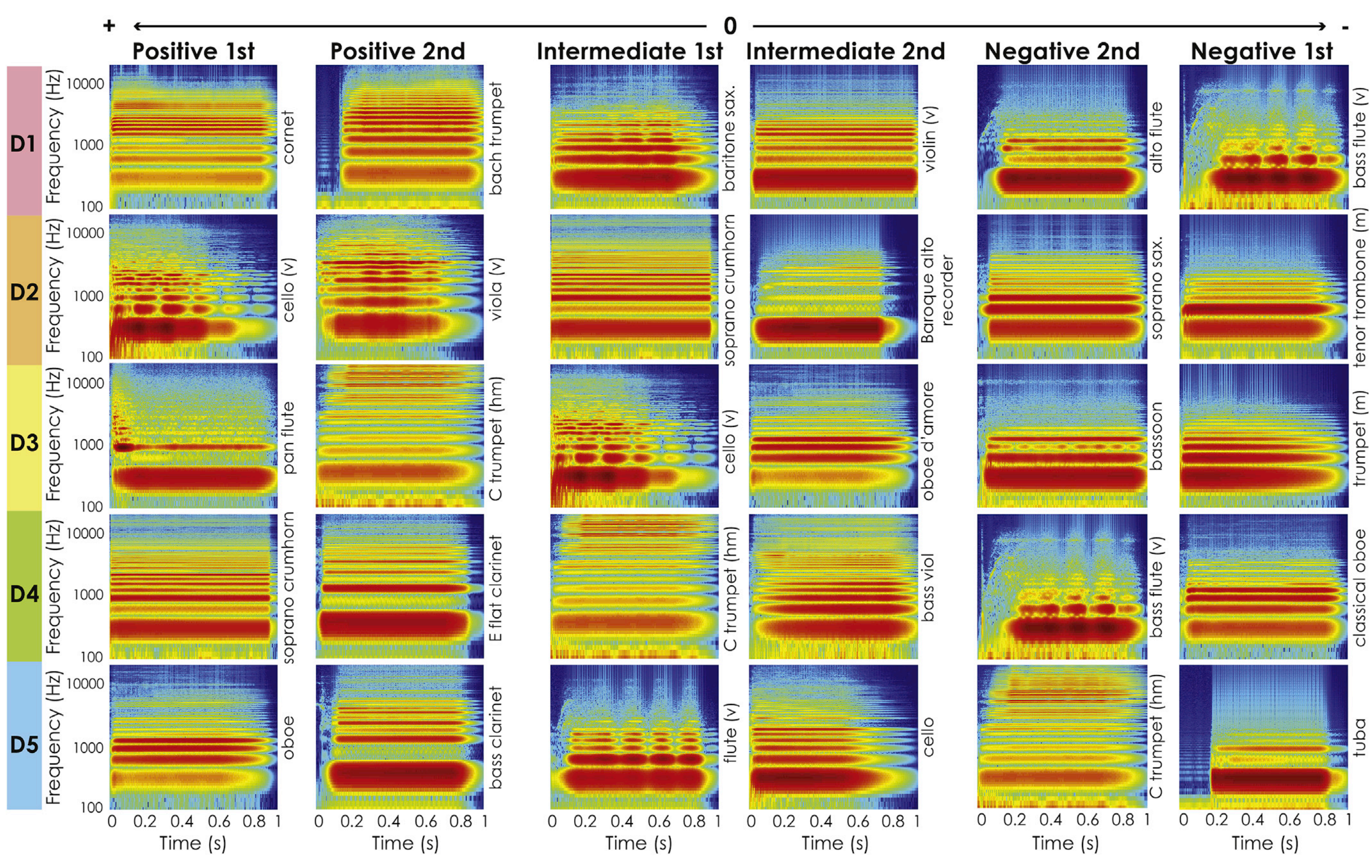

Fig. 1. Spectrograms of the sounds with (columns from left to right) the two most positive, two intermediate, and the two most negative values on each of the five timbre dimensions (rows). Abbreviations: v = vibrato, m = muted, h = harmonic.