

Neural responses in human superior temporal cortex support coding of voice representations

Taylor J. Abel^{1,2}, Jasmine L. Hect¹, Madison Remick¹, Avniel Ghuman¹, Bharath Chandrasekaran³, Lori L. Holt⁴, Kyle Rupp¹

¹Department of Neurological Surgery, University of Pittsburgh

²Department of Bioengineering, University of Pittsburgh

² Department of Communication Science and Disorders, University of Pittsburgh

³Department of Psychology, Carnegie Mellon University

abeltj@upmc.edu

Abstract (max 300 words)

The ability to recognize abstract features of voice during auditory perception is an intricate feat of human audition. For the listener, this occurs in near-automatic fashion to seamlessly extract complex cues from a highly variable auditory signal. Voice perception depends on specialized regions of auditory cortex, including superior temporal gyrus (STG) and superior temporal sulcus (STS). However, the nature of voice encoding at the cortical level remains poorly understood. We leverage intracerebral recordings across human auditory cortex during presentation of voice and non-voice acoustic stimuli to examine voice encoding at the cortical level, in eight patient-participants undergoing epilepsy surgery evaluation. We show that voice-selectivity increases along the auditory hierarchy from supratemporal plane (STP) to the STG and STS. Results show accurate decoding of vocalizations from human auditory cortical activity even in the complete absence of linguistic content. These findings show an early, less-selective temporal window of neural activity in the STG and STS followed by a sustained, strongly voice-selective window. Encoding models demonstrate divergence in the encoding of acoustic features along the auditory hierarchy, wherein STG/STS responses are best explained by voice category and acoustics, as opposed to acoustic features of voice stimuli alone. This is in contrast to neural activity recorded from STP, in which responses were accounted for by acoustic features. These findings support a model of voice perception that engages categorical encoding mechanisms within STG and STS to facilitate feature extraction.