Hearing Research 348 (2017) 31–43

Contents lists available at ScienceDirect

Hearing Research

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

Research Paper

Population responses in primary auditory cortex simultaneously represent the temporal envelope and periodicity features in natural speech

Daniel A. Abrams [a, *], Trent Nicol [a], Travis White-Schwoch [a], Steven Zecker [a], Nina Kraus [a, b, c]

[a] Auditory Neuroscience Laboratory, The Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL, 60208, USA
[b] Departments of Neurobiology and Physiology, Northwestern University, 2240 Campus Drive, Evanston, IL, 60208, USA
[c] Department of Otolaryngology, Northwestern University, 2240 Campus Drive, Evanston, IL, 60208, USA

ARTICLE INFO

Article history:
Received 7 October 2016
Received in revised form
4 February 2017
Accepted 13 February 2017
Available online 17 February 2017

Keywords:
Auditory thalamus
Auditory cortex
Guinea pig
Temporal coding

ABSTRACT

Speech perception relies on a listener's ability to simultaneously resolve multiple temporal features in the speech signal. Little is known regarding neural mechanisms that enable the simultaneous coding of concurrent temporal features in speech. Here we show that two categories of temporal features in speech, the low-frequency speech envelope and periodicity cues, are processed by distinct neural mechanisms within the same population of cortical neurons. We measured population activity in primary auditory cortex of anesthetized guinea pig in response to three variants of a naturally produced sentence. Results show that the envelope of population responses closely tracks the speech envelope, and this cortical activity more closely reflects wider bandwidths of the speech envelope compared to narrow bands. Additionally, neuronal populations represent the fundamental frequency of speech robustly with phase-locked responses. Importantly, these two temporal features of speech are simultaneously observed within neuronal ensembles in auditory cortex in response to clear, conversation, and compressed speech exemplars. Results show that auditory cortical neurons are adept at simultaneously resolving multiple temporal features in extended speech sentences using discrete coding mechanisms.

1. Introduction

Speech perception depends critically on temporal features of the speech signal. Rosen presented a framework for these temporal features that segregated the speech signal into three frequency ranges: the low-frequency speech envelope (2–50 Hz), periodicity cues (50–500 Hz), and the temporal fine structure (600–10,000 Hz) (Rosen, 1992). The low-frequency envelope is dominated by the syllable rate of speech and is sufficient for speech perception in quiet listening conditions (Shannon et al., 1995). Periodicity cues include the representation of the fundamental frequency (f0) of the speaker's voice and convey prosodic (Vaissiere, 2005) and, in the case of tonal languages, lexical information. The temporal fine-structure provides information about the spectrum and formant structure of speech sounds.

Our understanding of how the human central auditory system processes temporal features in speech has been facilitated by neurophysiological studies that have examined central auditory coding of specific, rudimentary acoustic features present in speech sounds (Cunningham et al., 2002; King et al., 1999; Kubanek et al., 2013; McGee et al., 1996; Nourski et al., 2009; Steinschneider et al., 1980, 1990, 1995, 1999, 2003; Warrier et al., 2011; White-Schwoch et al., 2016). This approach is based on the rationale that the speech signal is sufficiently complex to necessitate its decomposition into constituent components to describe neural mechanisms underlying each individual temporal feature. While this rationale is appealing from an experimental perspective, allowing maximum control of each temporal feature, it inherently obscures an understanding of a key element of speech processing in the auditory system: *the simultaneous processing of multiple temporal features in speech*. Specifically, what neural mechanisms enable the

---

* Corresponding author. Stanford Cognitive and Systems Neuroscience Laboratory, Stanford University, 1070 Arastradero Road, Suite 220, Palo Alto, CA 94304, USA.

E-mail address: daa@stanford.edu (D.A. Abrams).

http://dx.doi.org/10.1016/j.heares.2017.02.010
0378-5955/

simultaneous coding of concurrent temporal features in the speech signal? Given the complexity of the speech signal and well-established non-linearity of the auditory system, the neural representation of isolated acoustic features present in speech cannot predict the representation of the actual speech signal: the complex interaction of temporal features in the ongoing speech signal may result in different response characteristics than those predicted by isolated constituent parts.

Little is known regarding the simultaneous representation of concurrent temporal features in speech. The overarching goal of the current study is to investigate near-field cortical representations of the speech envelope and periodicity cues in response to naturally produced speech sentence stimuli in which these two temporal features occur simultaneously. Our primary hypothesis is that distinct temporal cues in speech are reflected by distinct neural codes in primary auditory cortex. Specifically, we hypothesize that low frequency temporal information in near-field auditory cortical activity in response to speech sentences will track the speech envelope, while, simultaneously, higher frequency components of cortical responses will track periodicity cues in speech. Below, we highlight the literature describing near-field cortical processing of speech envelope and periodicity cues in speech as well as open questions pertaining to these representations.

1.1. *Auditory cortical processing of the speech envelope*

Auditory cortical tracking of speech envelope cues, the first category of temporal features of speech described by Rosen (1992), has previously been explored in human participants; these studies have identified several features of temporal envelope representation in near-field measures of auditory cortex (Kubanek et al., 2013; Nourski et al., 2009). For example, an electrocorticography (ECoG) study in humans undergoing neurosurgical procedures showed that activity from belt areas of auditory cortex tracks the speech envelope with a high degree of fidelity (Kubanek et al., 2013). Another ECoG study in humans showed that speech envelope tracking in core auditory cortical regions using time-compressed sentences corresponded with speech comprehension abilities in those individuals (Nourski et al., 2009), consistent with results from non-invasive methods (Ahissar et al., 2001). The strength of these studies is that they described gross auditory cortical response properties in awake human participants and their relationship to perception, however they have not examined a number of specific questions regarding speech envelope tracking. For example, computing the speech envelope requires setting a number of parameters that can impact correlations with cortical activity. A key parameter is the bandwidth of the speech signal from which the temporal envelope is extracted. This is an important parameter because temporal properties extracted from narrow bands of speech can be considerably different between bands, and different relative to the broadband envelope, and it is unknown whether localized cortical activity more closely tracks temporal features from narrow or broad bands of the speech signal. This question has important implications for our understanding of auditory cortical function during speech perception: do cortical neurons track more global temporal features in speech, represented by the broadband envelope, or do they prefer temporal features in speech from narrow bands centered on the neurons' best frequency (BF)?

Most studies, including the aforementioned human studies (Kubanek et al., 2013; Nourski et al., 2009), have examined the broadband speech envelope rather than narrow bands. However, in a study of cortical responses to conspecific vocalizations in anesthetized cats, it was shown that auditory cortical activity more closely reflects patterns in narrowband temporal envelopes of vocalizations, centered on the BF of the neural population, compared to the temporal envelope of the broadband signal (Gehr et al., 2000). A limitation of this work is that the acoustic features in cat meows are relatively simple compared to the spectrotemporal fluctuations in speech, and therefore a more complete understanding of temporal envelope tracking by auditory cortical neurons requires the use of speech as the experimental stimulus.

1.2. *Auditory cortical processing of periodicity in speech*

Auditory cortical representations of periodicity cues in speech, the second category of temporal features highlighted by Rosen (1992), have also been described in the literature (Steinschneider et al., 1980, 1990, 1995, 1999, 2003). An important variable in considering auditory cortical representations of periodicity is the particular method used to characterize auditory cortical activity. For example, ECoG signals, measured with surface electrodes over auditory cortex, have not shown tracking to the fundamental frequency of speech (Kubanek et al., 2013; Nourski et al., 2009). Similarly, single-unit activity measured in response to conspecific vocalizations in primate models of the auditory systems, which share many acoustic attributes with human speech, have also failed to time-lock to the fundamental frequency of these vocalizations (Gehr et al., 2000; Gourevitch and Eggermont, 2007; Nagarajan et al., 2002; Wang et al., 1995). Rather, these studies have shown that the dominant temporal feature in speech and speech-like stimuli that is tracked by both ECoG signals and single units in auditory cortex is the low frequency temporal envelope (Gehr et al., 2000; Gourevitch and Eggermont, 2007; Kubanek et al., 2013; Nagarajan et al., 2002; Nourski et al., 2009; Wang et al., 1995).

Another method for quantifying auditory cortical function is the use of ensemble population responses measured from neurons in auditory cortex, which will be referred to here as multi-unit activity (MUA). MUAs are measured with a depth electrode, and therefore have been used predominantly in animal models of the auditory system rather than human studies, and provide an alternative method for examining neuronal representation of acoustic signals (Steinschneider et al., 1998, 2003). Importantly, previous studies have shown that MUAs track a wider range of temporal features in speech relative to single units and ECoG signals, and MUA signals measured from auditory cortical neurons can phase-lock to periodicities in speech up to ~100 Hz (Steinschneider et al., 1980). A close correspondence has been observed between periodicity representations measured with MUAs and the scalp-recorded frequency-following responses (FFR) in an animal model of the auditory system (White-Schwoch et al., 2016). Furthermore, research using non-invasive techniques has shown that the precision of phase-locking to the f0, as measured with the FFR in human listeners, is related to speech perception and language skills (Anderson et al., 2010; Kraus and White-Schwoch, 2015; Woodruff Carr et al., 2014). While MUAs have been used to examine cortical responses to periodic features in speech (Steinschneider et al., 1980, 1990, 1995, 1999, 2003), a limitation of the existing literature is that speech stimuli used to probe periodicity have tended to be brief consonant-vowel stimuli produced by a synthesizer, and to our knowledge it has not been shown to what extent cortical populations represent periodicities in naturally produced speech.

1.3. *Goals of this study*

There are three goals for this work. First, while auditory cortical tracking of both speech envelope and periodicity cues in speech have both been previously reported, to our knowledge it has never been shown that the same population of cortical neurons simultaneously tracks these two categories of temporal features. Additionally, given the importance of the speech envelope for speech