



## Research Paper

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



Daniel A. Abrams<sup>a,\*</sup>, Trent Nicol<sup>a</sup>, Travis White-Schwoch<sup>a</sup>, Steven Zecker<sup>a</sup>,  
Nina Kraus<sup>a, b, c</sup>

<sup>a</sup> Auditory Neuroscience Laboratory, The Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL, 60208, USA

<sup>b</sup> Departments of Neurobiology and Physiology, Northwestern University, 2240 Campus Drive, Evanston, IL, 60208, USA

<sup>c</sup> 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.

© 2017 Elsevier B.V. All rights reserved.

## 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 ( $f_0$ ) of the speaker's voice and convey prosodic (Vaisiere, 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](mailto:daa@stanford.edu) (D.A. Abrams).

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 (Kubaneck 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 (Kubaneck 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 (Kubaneck 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 (Kubaneck 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; Kubaneck 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  $f_0$ , 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

perception, a second goal of the current work was to examine the representation of the speech envelope with a focus on investigating whether cortical neurons prefer wide or narrow bands of temporal features in the speech signal. This latter question is important as it is unknown whether cortical ensembles track more global envelope features in the speech signal, as reflected by the broadband envelope of speech, or narrowband temporal features associated with the neural population's best frequency. A third goal of this work was to examine the representation of periodic aspects of the speech signal using naturally-produced speech sentences, building on findings from studies using isolated consonant-vowel stimuli.

To address the representation of speech envelope and periodicity in cortical neurons, we measured near-field cortical activity in primary auditory cortex of anesthetized guinea pig in response to a speech sentence produced in three different modes of speech: conversational speech, "clear" speech (Abrams et al., 2008, 2009, 2010; Uchanski, 2005) and compressed speech (Abrams et al., 2008, 2009; Watson et al., 1990). We used these three modes of speech as a means of varying speech envelope cues in an ecologically meaningful fashion: clear speech is a mode of speech naturally produced by talkers in noisy environments that increases speech understanding (Bradlow et al., 2003) and, among other acoustic variations, is characterized as having enhanced amplitude modulation depth relative to conversational speech (Bradlow et al., 2003; Uchanski, 2005). Compressed speech, which approximates rapidly-produced speech, has faster speech envelope cues relative to conversational speech. Furthermore, as a means to vary periodicity cues in our stimulus set, we used relatively long sentence stimuli (1.5 s), which exhibited natural fluctuations in  $f_0$  within and across utterances.

## 2. Methods

The research protocol was approved in advance by the Animal Care and Use Committee of Northwestern University.

### 2.1. Animal preparation

The experimental materials and procedures were similar to those reported previously (Cunningham et al., 2002; McGee et al., 1996). Seven pigmented guinea pigs of either sex, weighing between 325 and 650 g, were used as subjects. Animals were initially anesthetized with ketamine hydrochloride (100 mg/kg) and xylazine (8 mg/kg). Smaller supplemental doses (25 mg/kg ketamine; 4 mg/kg xylazine) were administered hourly throughout the rest of the experiment. Following the induction of anesthesia, the animal was mounted in a stereotaxic device for the duration of the experiment. Body temperature was maintained at 37.5 °C by using a thermistor-controlled heating pad on the guinea pig's abdomen (Harvard). Normal hearing sensitivity was confirmed by auditory brainstem response (ABR). The ABR was elicited by a click stimulus at 70 and 40 dBHL (referenced to normal guinea pig click thresholds) from a recording site located at the posterior vertex/midline of the scalp using an EMG needle electrode. A rostral-caudal incision was made along the scalp surface and the tissue was retracted to expose the skull. Holes were drilled in the skull under an operating microscope. The dura was removed with a cautery to prevent damage to the recording electrode, and the cortical surface was coated with mineral oil.

### 2.2. Anatomical structures

We measured neural activity from primary auditory cortex, which in guinea pig consists of two adjacent areas, A1 and the dorso-caudal field (DC). Primary cortical responses are

characterized by tonotopic organization, sharp frequency tuning, a preference for tonal stimuli and short response latencies (Redies et al., 1989; Wallace et al., 2000).

### 2.3. Acoustic stimuli

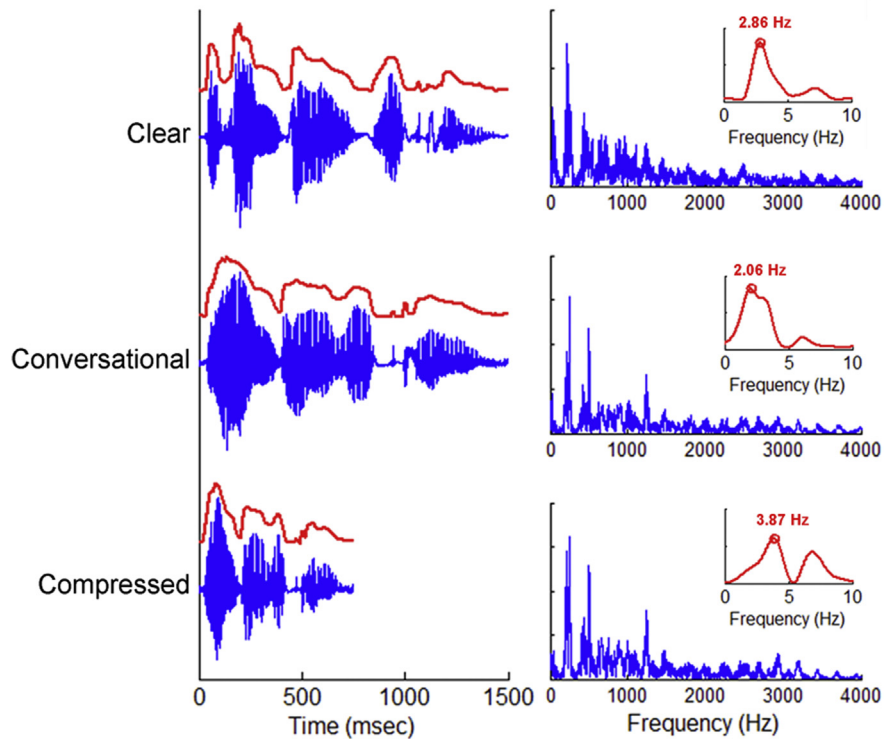
Stimuli consisted of the sentence "The young boy left home" produced in three modes of speech: conversational, clear, and compressed (Fig. 1). These stimuli have been published in previous studies of auditory cortical processing of speech in typically developing children (Abrams et al., 2008, 2010) and children with reading impairments (Abrams et al., 2009, 2010). These three modes of speech have different envelopes and were used to elicit a variety of cortical activation patterns. Conversational speech is defined as speech produced in a natural and informal manner. Clear speech is a well-described mode of speech resulting from greater diction; clear speech enables greater speech intelligibility in noisy acoustic environments relative to conversational speech in part due to greater amplitude modulation (i.e., speech envelope cues) (Bradlow et al., 2003; Uchanski, 2005). Compressed speech replicates rapidly-produced speech and is characterized by a higher-frequency speech envelope. Compressed speech is more difficult to perceive compared to conversational speech (Beasley et al., 1980).

Conversational and clear sentences were recorded in a sound-proof booth by an adult male speaker at a sampling rate of 16 kHz. Conversational and clear speech sentences were normalized for overall duration to control for slower speaking rates in clear speech (Uchanski, 2005). Normalization was achieved by compressing the clear sentence by 23% and expanding the conversational sentence by 23%. To generate the compressed sentence stimulus, we doubled the rate of the conversational sample using a signal-processing algorithm in Adobe Audition (Adobe Systems Inc.). This algorithm does not alter the fundamental frequency or perceived pitch of the signal. The duration of the clear and conversational speech sentences was 1500 msec, and the duration of the compressed sentence was 750 msec.

30 repetitions of each speech stimulus were delivered to the contralateral ear using Etymotic insert earphones (ER2) through the hollow earbars of the stereotaxic device. The sound pressure level (SPL, expressed in dB re 20 mPa) was calibrated over a frequency range of 20 to 20,000 Hz using a condenser microphone (Brüel and Kjaer). Sentence stimuli were delivered at a level of 60 dBA. Stimuli were presented with alternating polarities to remove any possibility of a stimulus artifact within the response. The delivery system output the signal through a 16-bit converter at a sampling rate of 16 kHz. That system triggered the PC-based collection computer. All stimuli were presented in quiet in a sound-treated booth (IAC). Third-octave tone-pips were used to map auditory cortex. Tones were 100 msec in duration with a rise-fall time of 10 msec. BFs for the recording sites ranged from 200 Hz to 3150 Hz, and included the following BFs: 200 Hz (3 sites), 250 Hz (2 sites), 500 Hz (2 sites), 630 Hz (2 sites), 1000 Hz (1 site), 2000 Hz (3 sites), and 3150 Hz (1 site).

### 2.4. Neurophysiologic recording

Cortex was accessed with a vertical approach using tungsten microelectrodes (Micro Probe) with impedance between 1 and 2 M $\Omega$  at 1 kHz. An electrode was advanced perpendicular to the surface of cortex using a remote-controlled micromanipulator (Märzhäuser-Wetzlar). The coordinates of the electrode were determined at a point slightly above cortex at the first penetration, and these coordinates were kept for the remainder of the experiment. For recording primary areas A1 and DC, locations were



**Fig. 1. Stimuli.** Left column: Stimulus waveforms (blue) and broadband speech envelopes (red) for clear (top), conversational (center) and compressed (bottom) conditions. Right column: Frequency spectra for the three stimulus conditions (blue). Insets: Frequency spectra of the broadband speech envelopes (red) with the modal frequencies noted. All speech stimuli are normalized in amplitude. The broadband speech envelope was calculated by extracting the absolute value of the Hilbert transform of the raw stimulus waveform, then low-pass filtering this signal at 50 Hz (Rosen, 1992). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

approximately 3 mm caudal to bregma and 10 mm lateral of the sagittal suture. Recordings were made at depths of 500–900  $\mu\text{m}$ , corresponding to cortical layers III and IV in guinea pig.

The electrode signal was amplified using Grass amplifiers with filters set between 1 and 20,000 Hz. The analog signal was digitized at 33.3 kHz by an A-D card (MCC) attached to a PC. Responses were logged and stored using Matlab routines designed by our lab. Recorded brain responses were averaged and off-line filtered between 500 and 3000 Hz to isolate multiunit activity (MUA; e.g., (Steinschneider et al., 1990). MUA envelopes were extracted by calculating the absolute value of the Hilbert transform of the band-pass filtered MUAs. The resulting MUA waveforms were used as the basis for all subsequent analyses of the neural data. A caveat to using this measure of MUA is that there is no way to determine the different cell types contributing to the overall response (Fishman et al., 2000). While the goal is to measure exclusively the contribution of cortical pyramidal cells, it is possible that MUA recordings also include contributions from stellate cells and thalamocortical axons, the latter of which are known to phase-lock to faster stimulus rates than pyramidal neurons.

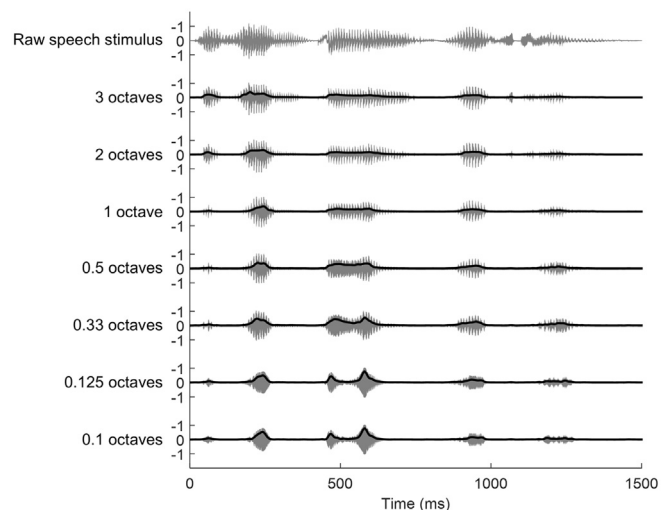
## 2.5. Data analysis

All data analyses were performed using software written in Matlab (The Mathworks, Inc).

### 2.5.1. Speech envelope, bandwidth analysis

A goal of this work was to investigate whether auditory cortical neurons, across the tonotopic map, phase-lock to temporal features in narrow or wide-band segments of the stimulus. To address this question, we extracted the temporal envelope from the stimulus across a number of stimulus bandwidths including: (1) 3 octaves;

(2) 2 octaves; (3) 1 octave; (4) 1/2 octave; (5) 1/3 octave; (6) 1/8 octave and (7) 1/10 octave (Fig. 2). An important consideration is that we used each cortical site's best frequency (BF) as the center frequency for these filtering conditions. The steps involved in



**Fig. 2. Filtered stimuli and envelope extraction for the “clear” stimulus condition.**

This is an illustrative example of the stimulus filtering performed in the speech envelope, cross-correlation analysis. In this figure, the center frequency for envelope extraction was chosen to be 1,000 Hz, as an example of the procedure used. However, in all analysis, the center frequency for envelope extraction was chosen to be the best frequency of cortical MUAs. Since narrower pass-bands (i.e., plots at the bottom of the figure) are inherently smaller in amplitude, all traces in the figure were amplitude normalized between  $-1$  and  $1$  to allow a more detailed view of each of the filtered stimuli.

extracting the envelope across these bandwidth conditions were the following. First, the BF of the cortical site was identified using the results from the pure-tone probes. The raw stimulus waveform was then filtered at the seven above listed bandwidths using the BF of the cortical site as the center frequency. The band-pass filters were Butterworth filters with 30 dB/octave rolloffs. The upper cutoff frequency for each band-width condition was calculated according to the equation  $2^{\text{bandwidth}/2} \times \text{BF}$  and the lower cutoff frequency was calculated according to the equation  $\text{BF}/2^{\text{bandwidth}/2}$  (Everest, 2001). The temporal envelope was then extracted from each of the band-filtered stimuli by calculating the absolute value of the Hilbert transform. The resulting waveforms were then low-pass filtered at 50 Hz to extract the slowly-varying temporal envelopes of the speech signal (Rosen, 1992).

### 2.5.2. Speech envelope, cross-correlation analysis

Prior to cross-correlation, MUA waveforms were low-pass filtered at 50 Hz using the same filter settings as those used to extract temporal envelopes of the speech signal (Butterworth filters with 30 dB/octave rolloffs). Low-pass filtering the MUA waveforms removed peaks in the response corresponding to periodicity coding and other high-frequency events in the MUA, leaving only the low-frequency content corresponding to speech envelope frequencies (Rosen, 1992). The rationale for this filtering step is that we wanted to isolate and quantify the correspondence between slowly varying aspects of the stimuli and MUAs, while controlling for rapid aspects of the stimuli and responses (e.g., periodicity). We then performed cross-correlations between low-pass filtered MUAs and the seven variations of the speech envelopes described above. Peaks in the resulting correlograms were auto-detected in Matlab for lags between 10 and 50 msec to account for conduction and propagation delays in the auditory system.

A final goal of the speech envelope MUA analysis was to examine whether specific MUA response properties, including BF and MUA envelope response amplitude, were related to bandwidth preferences of the MUA. Therefore, we performed Pearson's correlations between BF of the MUA and bandwidth preferences of the MUA, based on whether the MUA preferred (i.e., showed the greatest correlation with) the 3 octave, 2 octave, 1 octave, 1/2 octave, 1/3 octave, 1/8 octave or 1/10 octave stimulus envelope. We also performed a Pearson's correlation between the amplitudes of the MUA envelopes and bandwidth preferences of the MUAs. Amplitudes of MUA envelopes were calculated as the RMS amplitude across the entire response period (clear and conversational speech conditions: 0–1500 msec; compressed speech condition: 0–750 msec).

### 2.5.3. Periodicity analysis, frequency domain

We calculated the spectra of MUAs for specific time segments of the response based on the presence of pronounced periodicity in MUAs. To allow for direct comparison with periodic aspects of the speech stimuli, the spectra of corresponding time segments of the stimuli were also computed. Spectra were calculated using a 33,333 point fast Fourier transform (FFT) and a rectangular window. In all stimulus and MUA FFTs, there were peaks at ~100 Hz corresponding to the  $f_0$  of the stimulus, and the frequency and magnitude of the FFT maxima were recorded. A one-way ANOVA was performed to determine if there were statistical differences in MUA periodicity representation between clear, conversational and compressed speech conditions.

### 2.5.4. Periodicity analysis, time domain

For those segments of MUAs that indicated pronounced periodicity, we analyzed the latency difference between local maxima in the MUAs. Local maxima were auto-detected in Matlab and then

verified visually. A similar procedure was performed for corresponding segments of the speech stimuli. The mean of the inter-maximum periods was converted to Hz for subsequent analysis.

## 3. Results

### 3.1. Auditory cortical MUAs and the speech envelope

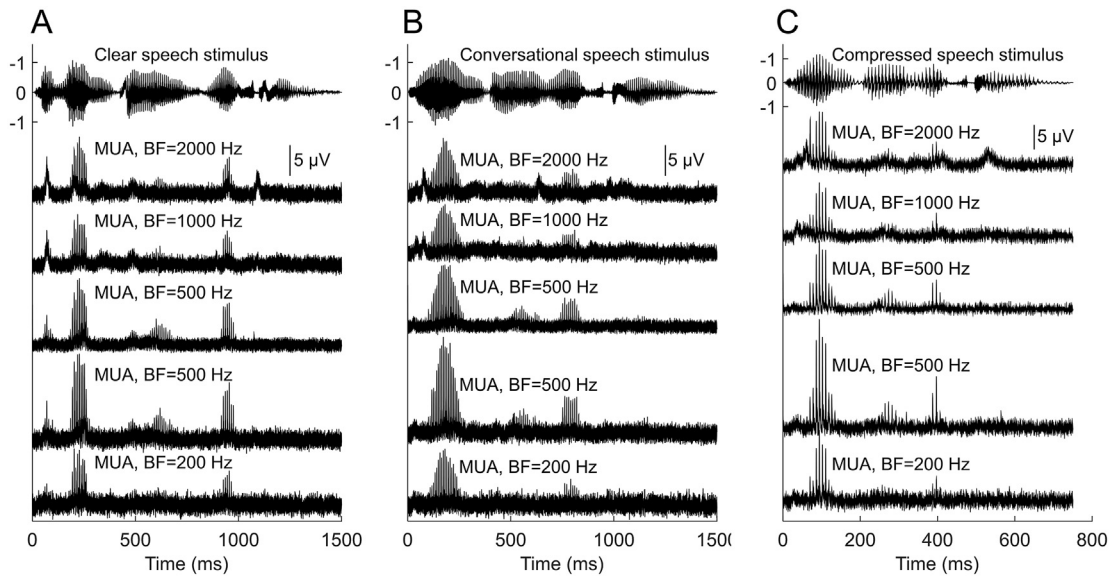
In all speech conditions (clear, conversational, and compressed), low frequency temporal information in auditory cortical MUAs (Fig. 3) appeared to track the low-frequency temporal envelope of the speech stimuli (i.e., the speech envelope). To investigate this possibility, the envelope of all MUAs was computed and compared to the stimulus envelope (Fig. 4). A characteristic of the speech envelope is that it varies between different pass-bands in a given stimulus, and it is not known if cortical neurons represent the envelope of narrow (Gourevitch and Eggermont, 2007) or wide-bands (Nagarajan et al., 2002) of speech. We investigated this question by extracting seven variations of the speech envelope from increasingly wide bands of the stimulus, using the best frequency of each MUA as the center frequency (Fig. 4, bottom color plots). We performed cross-correlations between each envelope iteration and the wideband cortical MUA envelope and identified the maximum in the correlogram between 10 and 50 msec for each stimulus bandwidth. The representative MUA from the recording site of Fig. 4A showed a clear preference for wider bands of the speech envelope, with the largest stimulus-response correlations for stimulus envelopes greater than 1 octave (Fig. 4A, inset). Note that a 1  $\mu\text{V}$  scale bar is included in the figure, however absolute amplitudes of MUAs are reduced as a result of LP filtering for this particular analysis (see Methods, 2.5.1 Speech envelope, bandwidth analysis).

Not all MUAs showed a preference for wide bands of the stimulus envelope. Another representative MUA showed no preference for any of the bandwidth conditions (Fig. 4B). Preference for envelope bandwidth in the three stimulus conditions for all cortical MUAs is plotted in Fig. 5A–C. Nearly all bandwidth-correlation functions are either flat or upward sloping, indicating either no preference or preference for wider bandwidths. Very few of these functions are downward sloping, which would indicate a preference for narrow-band envelopes. Mean correlation functions indicate slight increased preference for wide bandwidths in all three stimulus conditions (Fig. 5D), and this preference appears to be greatest in the clear stimulus condition. The histogram in Fig. 5E shows that across all speech conditions and recording locations, the most common bandwidth preference for MUAs in primary auditory cortex is 1 octave. Furthermore, maximum correlations between primary cortical MUAs and “preferred” speech envelopes were often moderate to strong correlations ( $r \geq 0.4$ ).

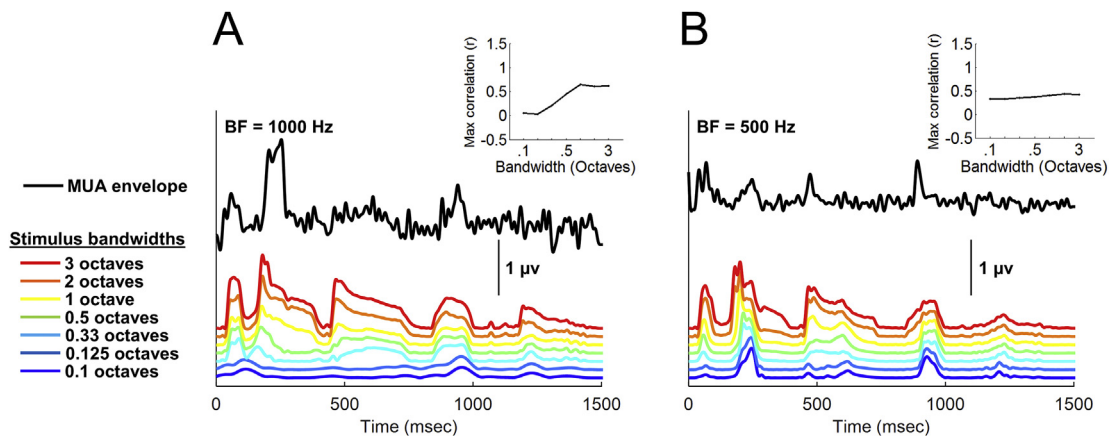
The next goal of our analysis was to examine whether specific MUA response properties, including BF and MUA envelope response amplitude, were related to bandwidth preferences of the MUAs. First, Pearson's correlation results indicate that the BF of the MUA is not related to bandwidth preference ( $R = 0.25$ ,  $P > 0.10$ ). We then examined the relationship between the RMS amplitude of the MUAs and their preferred bandwidths and found a statistically marginal relationship between these factors ( $P = 0.041$ ), however when responses from a single MUA were removed from the analysis, the correlation was no longer significant. We therefore conclude that there is no relationship between these factors.

### 3.2. Auditory cortical MUAs and periodicity

The clear speech stimulus (top) and five representative MUAs measured in response to this speech sample are plotted in Fig. 6. For this analysis, cortical MUAs were computed by band-pass filtering



**Fig. 3. Auditory cortical MUAs for clear, conversational, and compressed speech conditions.** (A) The clear speech stimulus waveform is plotted on the top row, followed by five representative MUAs measured from different recording sites in auditory cortex with different best frequencies in response to the clear speech stimulus. Panels (B) and (C) are organized the same as (A), however for the conversational and compressed speech conditions, respectively. Cortical MUAs were computed by first band-pass filtering raw MUAs (500–3000 Hz) and then extracting the envelope of this signal by calculating the absolute value of the Hilbert transformed band-pass signal. These traces show that auditory cortical MUAs appear to track the low-frequency temporal envelope of the speech stimuli.

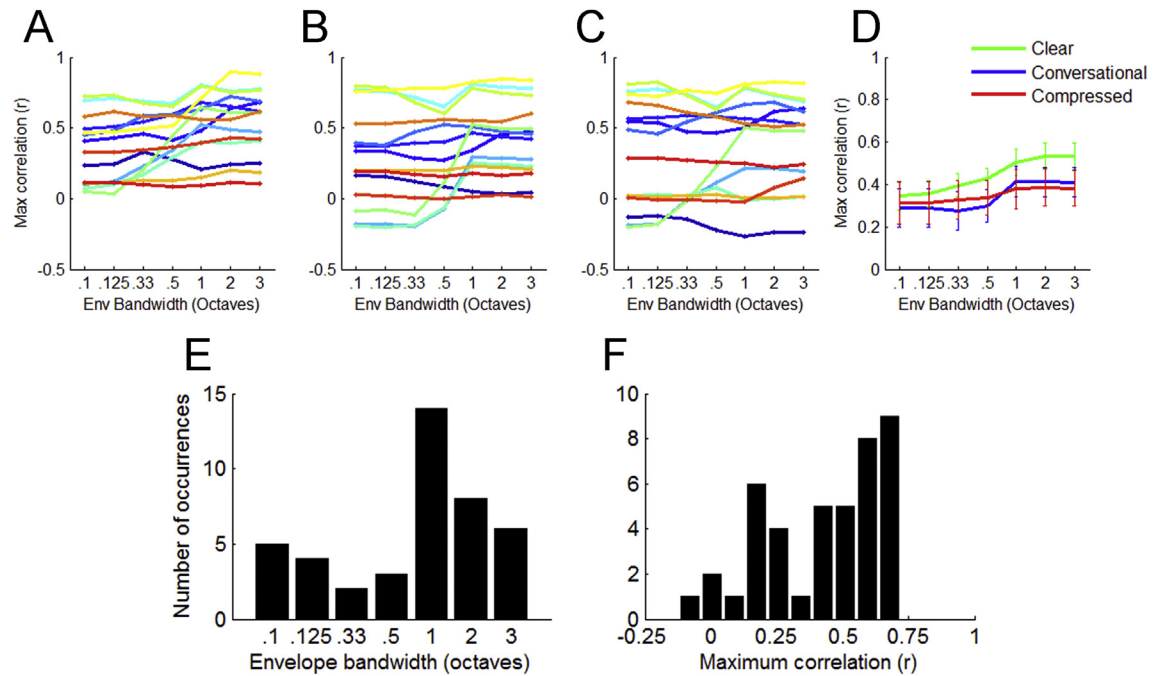


**Fig. 4. Two representative examples of low-pass filtered cortical MUAs and clear stimulus envelopes varying in bandwidth.** Panel A: The envelope of a representative auditory cortical MUA (top, black) and clear stimulus envelopes (bottom, in color). A 1  $\mu$ V scale bar is included in the figure, however absolute amplitudes of MUAs are reduced as a result of LP filtering for this particular analysis (see [Methods, 2.5.1 Speech envelope, bandwidth analysis](#)). Inset: Bandwidth-correlation function for this MUA. The BF for this MUA was 1000 Hz. Panel B: The envelope of a second representative auditory cortical MUA (top, black) and clear stimulus envelopes (bottom, in color). Inset: Bandwidth-correlation function for this MUA. The BF for this MUA was 500 Hz. Low-pass filtered cortical MUAs were computed by band-pass filtering raw MUAs (500–3000 Hz), extracting the envelope of this signal by calculating the absolute value of the Hilbert transformed band-pass signal, then low-pass filtering the resulting signal at 50 Hz. This figure shows that correlation-bandwidth functions vary among auditory cortical recording locations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

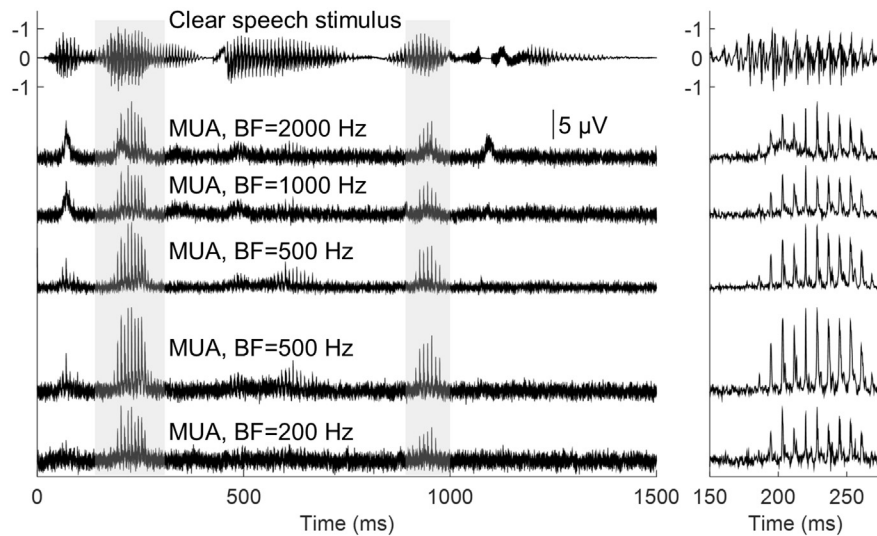
raw MUAs (500–3000 Hz) then extracting the envelope of this signal. A prominent feature visible throughout the stimulus waveform is high amplitude periodicity near 100–120 Hz. This periodicity represents the fundamental frequency ( $f_0$ ) of the stimulus, the acoustic feature that determines the pitch of a speaker's voice. Visual inspection of the MUA reveals periodicity that appears to correspond to the  $f_0$  of the stimulus, most prominent at latencies between 160–260 msec and 900–1000 msec (Fig. 6, highlighted regions). The correspondence between the period of the neural activity and the stimulus  $f_0$  is even more evident in the magnified views of the waveforms (Fig. 6, right). Not all MUA recordings showed pronounced periodicity coding throughout the stimuli. For example, MUAs with BFs of 2000, 1000,

and 200 Hz did not show strong periodic coding between 260 and 900 msec of the clear speech stimulus (Fig. 6). However, periodicity in this particular time range was more strongly represented by the two MUAs with BFs of 500 Hz plotted in the same figure. We interpret these results as showing that neural populations with BF = 500 Hz are more sensitive to the particular spectral and temporal features in this time range of the speech stimuli compared to neural populations with higher (i.e., 1–2 kHz) and lower frequency (i.e., 200 Hz) best frequencies.

The next goal of the analysis was to examine the correspondence between stimulus  $f_0$  periodicity and the frequency of periodic activity in auditory cortical MUAs in response to the speech stimuli. We investigated this correspondance in two manners. First,



**Fig. 5. Correlation-bandwidth functions.** Primary cortical MUAs are presented for clear (panel A), conversational (panel B), and compressed (panel C) speech conditions ( $n = 14$  for each condition; each MUA is shown in a different color, and these colors correspond across panels A–C). Panel D: Mean bandwidth–correlation functions for each speech condition. Note that the y-axis limits are different in panel D compared to panels A–C. Panel E: Histogram showing envelope bandwidth preference over all three stimulus conditions for all primary MUA. Panel F: Histogram showing maximum envelope correlation coefficient between all primary MUA and speech envelope correlations. All three stimulus conditions are included.

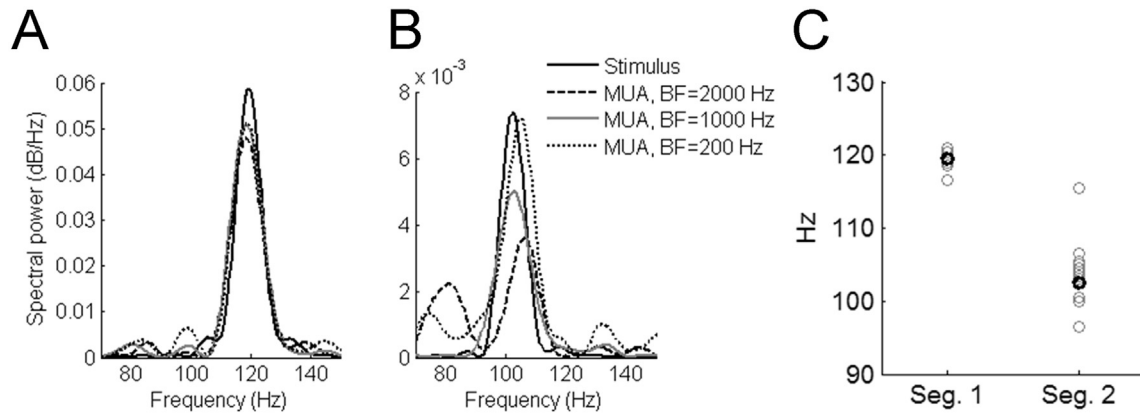


**Fig. 6. Clear stimulus waveform and auditory cortical MUAs.** The clear speech stimulus waveform is plotted on the top row, followed by 5 representative MUAs measured from different recording sites in auditory cortex with different best frequencies in response to the clear speech stimulus. The left-most highlighted region of the waveform in the main panel is enlarged in the right panel. Cortical MUAs were computed by band-pass filtering raw MUAs (500–3000 Hz) and extracting the envelope of this signal by calculating the absolute value of the Hilbert transformed band-passed signal. This figure shows a correspondence between periodicity in the clear speech stimulus  $f_0$  and auditory cortical MUAs.

we performed a frequency analysis of the stimulus and MUAs over the time ranges 160–260 msec and 900–1000 msec (Fig. 7A and 7B, respectively), as MUAs in these particular time ranges show substantial periodic activity. Results of this analysis indicate considerable overlap between the stimulus  $f_0$  and periodic activity in cortical MUAs over these time ranges. We then determined the frequency of maximal amplitude in the speech signal and all MUAs over these time segments, as well as the frequency difference between stimulus  $f_0$  and periodic MUA, and a close correspondence

between these peaks is evident (Fig. 7C). Considering that the stimulus  $f_0$  varied within this speech sample ( $\sim 120$  Hz for 160–260 msec segment;  $\sim 100$  Hz for 900–1000 msec segment) these data show that the dominant frequency of the cortical MUAs closely corresponds with the stimulus  $f_0$  with a high degree of precision.

A second analysis was performed in the time domain as a means of further describing the correspondence between stimulus  $f_0$  periodicity and periodic activity in MUAs. Specifically, we analyzed the latency difference between peaks in the raw stimulus waveform



**Fig. 7. Auditory cortical MUA spectra for the clear speech condition.** Clear stimulus and MUA FFTs for the latency ranges 160–260 (Panel A, Segment 1 in Fig. 6) msec and 900–1000 msec (Panel B, Segment 2 in Fig. 6). Panel C: Peak f0 frequency for stimulus (black circles) and all MUA (grey circles) measured in Segments 1 and 2. This figure shows a correspondence between periodicity in the stimulus f0 and auditory cortical MUAs measured in the frequency domain.

and the MUAs (Fig. 8, top). Peaks in the speech waveform preceded peaks in the MUA by approximately 12 msec. Results indicate that the periodic activity present in MUAs largely represents the stimulus f0, however MUAs fail to represent all of the latency variations expressed in the stimulus waveform. The most obvious discrepancy between peak latency differences in the speech sample and MUAs is between peaks 3–5 (Fig. 8, bottom). During this interval, peaks in the stimulus waveform vary between 8.0 and 9.5 msec while MUA peaks occur in a relatively narrow range between 8.0 and 8.5 msec. This discrepancy suggests that cortical MUAs are not simply tracking the peaks of the stimulus waveform, as this would have resulted in a closer correspondence between stimulus f0 and MUA interval latencies throughout this segment.

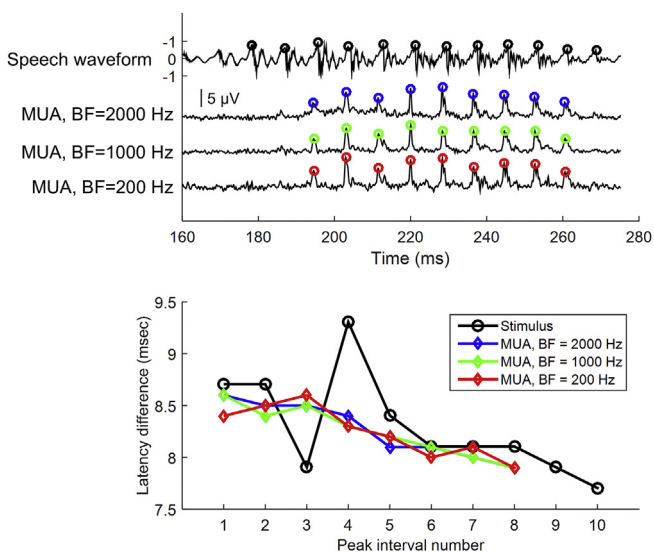
Similar to the clear speech condition, correspondences between periodic features of the raw stimulus waveforms and MUAs were also evident in the conversational and compressed speech

conditions (Fig. 9A and 9B, respectively). Specifically, frequency analyses of the conversational and compressed stimuli and responses indicated a close correspondence between the stimulus f0 and the dominant periodicity within the MUA (Fig. 10A–C for conversational speech; Fig. 10D and 10E for compressed speech). We performed an ANOVA to determine if f0 frequency tracking, defined here as the frequency difference between stimulus f0 and periodic MUA, measured over all 14 recordings, was different between the three stimulus conditions over the segments highlighted in Figs. 6, 9A and 9B. ANOVA results indicate no statistical difference between f0 tracking between these stimulus conditions ( $F_{2,67} = 0.178$ ,  $P = 0.837$ ). Furthermore, time-domain analyses showed a consistent association in peak latency differences between the raw stimulus waveform and the peaks in the MUA (Figs. 8, 11A and 11B). Specifically, peak latency differences between the raw stimulus waveform and the peaks in the MUA did not deviate more than 0.5 msec with the exception of peaks at the beginning and end of the response segments.

In summary, results from the analysis across multiple speech conditions – clear, conversational, and compressed – suggest that the f0 in ongoing speech sentences is represented with periodic neural activity in primary cortex. This representation is resistant to variations in the frequency of the f0, as well as to variations in temporal envelope characteristics across the varying speech conditions. Together, these analyses suggest that periodicity cues in speech are often coded by phase-locked MUAs in primary auditory cortex.

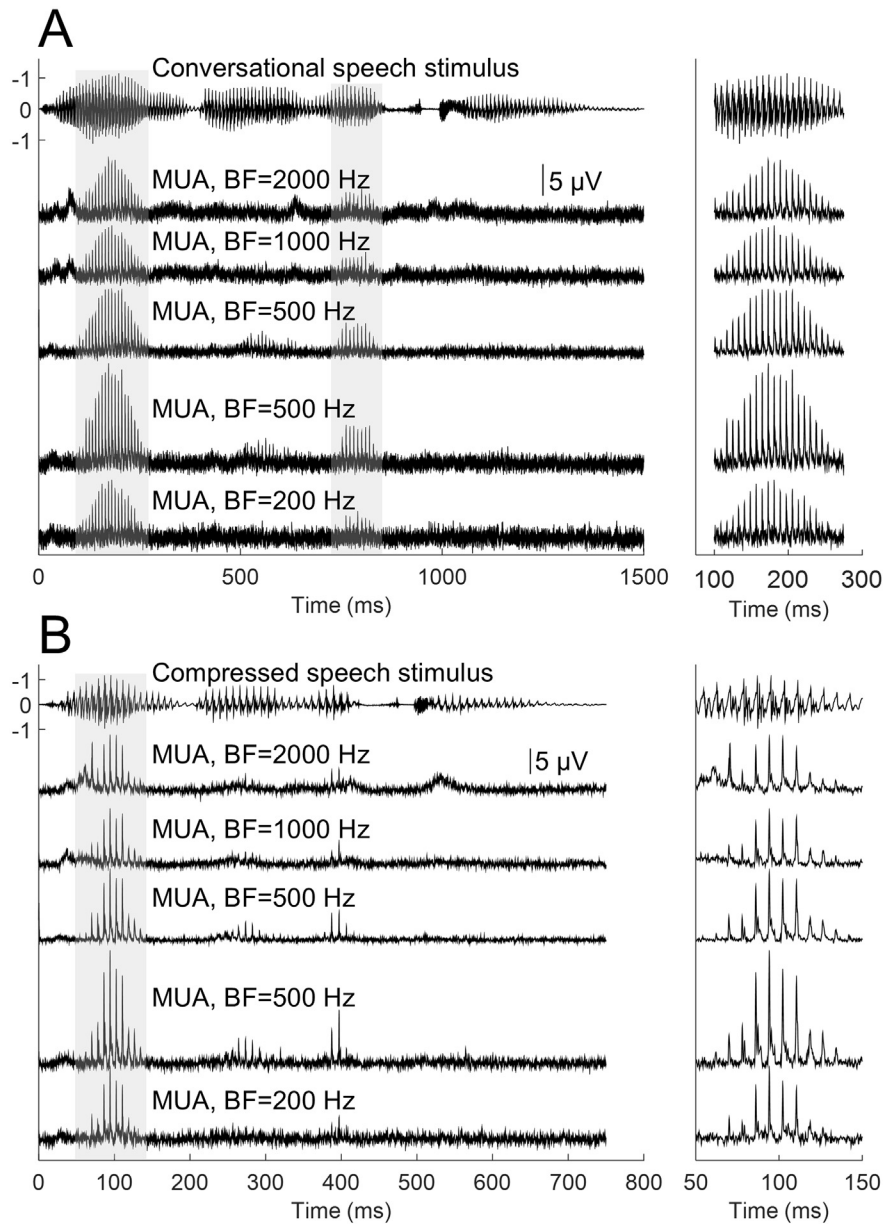
#### 4. Discussion

The goals of this study were to examine cortical processing of two categories of temporal features in speech, the speech envelope and speech periodicity cues (Rosen, 1992), in primary auditory cortex in guinea pig using naturally-produced speech stimuli. The results are summarized as follows: (1) Ensembles in primary auditory cortex track both the speech envelope (Figs. 4 and 5) and the fundamental frequency of speech (Figs. 7 and 10). (2) Ensembles in primary auditory cortex are variable in their representation of the speech envelope: some population responses represented the speech envelope with excellent fidelity, measured by cross-correlation analysis, while others showed poor fidelity in representing this aspect of the signal (Fig. 5A–C). (3) Ensembles in auditory cortex tended to show more precise representation of wider bandwidths ( $\geq 1$  octave) of the speech envelope compared to narrow-bands ( $< 1$  octave; Fig. 5E). (4) MUAs showed variable time-



**Fig. 8. Periodicity analysis for the clear speech condition.** Time-domain phase-locking analysis for the clear speech condition in three representative MUAs. Top panel: peaks used in latency analysis in both the speech stimulus (top) and MUAs are identified with circles. Bottom panel: latency differences between peaks in the speech waveform and MUAs. Cortical MUAs were computed by band-pass filtering raw MUAs (500–3000 Hz) and extracting the envelope of this signal by calculating the absolute value of the Hilbert transformed band-passed signal. This figure shows a correspondence between periodicity in the stimulus f0 and auditory cortical MUAs measured in the time domain.





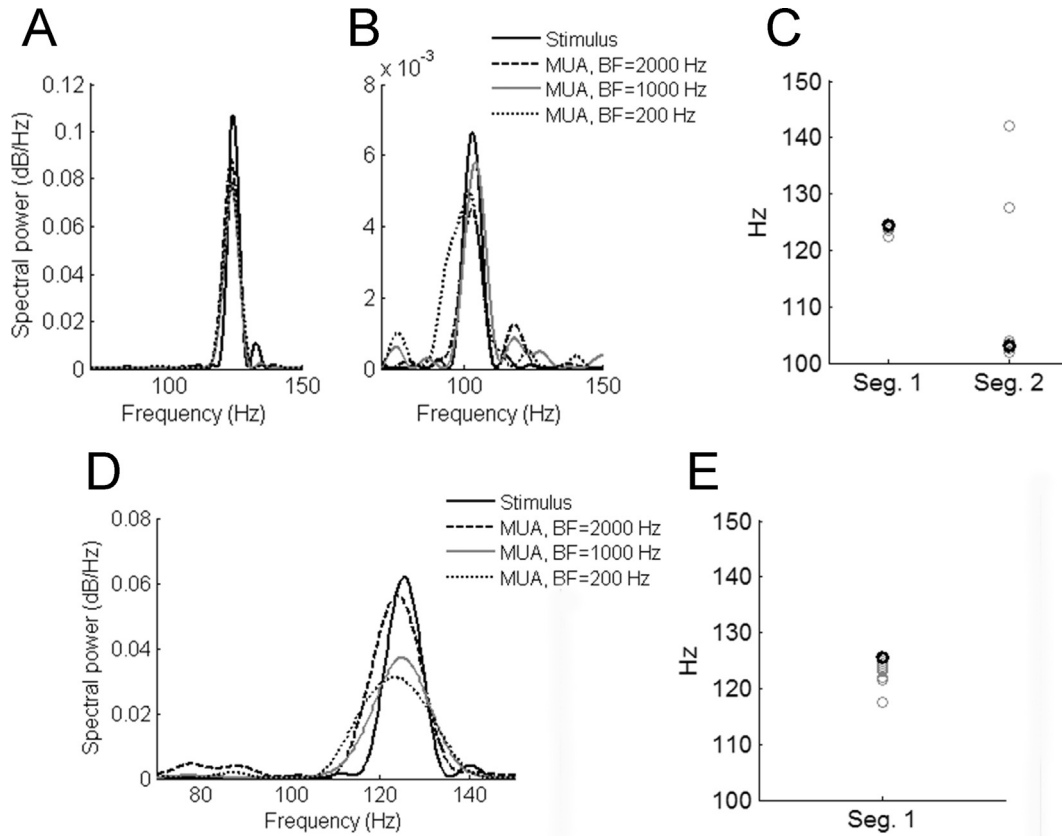
**Fig. 9. Auditory cortical MUAs for conversational and compressed speech conditions.** Panel A: The conversational speech stimulus waveform is plotted on the top row, followed by 5 representative MUAs measured from different recording sites in auditory cortex with different best frequencies in response to the conversational speech stimulus. Panel B: The compressed speech stimulus waveform is plotted on the top row, followed by 5 representative MUAs measured from different regions of auditory cortex with different best frequencies in response to the compressed speech stimulus. For panels A and B, the left-most highlighted region of the waveform in the main panel is enlarged in the right panel. MUA traces shown in panels A and B were measured from the same recording sites as those shown for the clear speech condition in Fig. 6. Cortical MUAs were computed by band-pass filtering raw MUAs (500–3000 Hz) and extracting the envelope of this signal by calculating the absolute value of the Hilbert transformed band-passed signal. This figure shows a correspondence between periodicity in the speech stimulus  $f_0$  and auditory cortical MUAs.

locking to the  $f_0$  of the speech signal (Figs. 8 and 11).

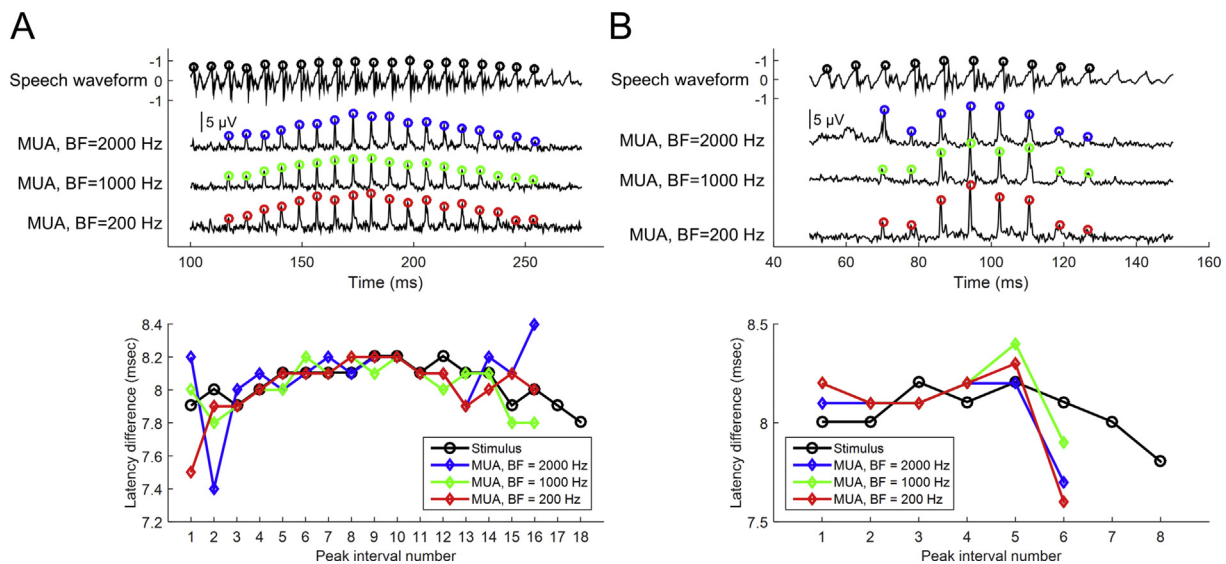
#### 4.1. Cortical representation of multiple temporal features in speech

Results show that neural ensembles in auditory cortex are capable of simultaneously coding multiple categories of temporal features in the speech signal. These features include the speech envelope and periodicity cues, two categories of temporal features that are important for speech perception (Drullman et al., 1994; Shannon et al., 1995; Smith et al., 2002; Zeng et al., 2005) and are highlighted in an influential framework describing the acoustic structure of speech (Rosen, 1992). Specifically, the slowly varying speech envelope is represented in the envelope of cortical

responses, and periodicity in the speech signal associated with the fundamental frequency is represented with phase-locked responses from populations of auditory neurons. From one point of view, this finding might be considered unremarkable: the auditory neurophysiology literature might have predicted this phenomenon based on previous findings investigating both the speech envelope (Kubaneck et al., 2013; Nourski et al., 2009) and periodicity (Steinschneider et al., 1990) cues. From another standpoint, however, this finding represents one of the astounding capabilities of the auditory system: the auditory system is adept at simultaneously resolving multiple temporal features in complex acoustic signals. The contribution of the current study is that we have used a model system to demonstrate how these multiple temporal aspects



**Fig. 10. Primary MUA spectra for conversational and compressed conditions.** Conversational stimulus and MUA FFTs for the latency ranges 100–275 (panel A) and 725–850 msec (panel B) from the highlighted Segments 1 and 2 in Fig. 9A. Fig. 10C: Peak  $f_0$  frequency for stimulus (black circles) and all MUA (grey circles) measured in Segments 1 and 2 in the conversational condition. Fig. 10D: Compressed stimulus and MUA FFTs for the latency range 50–150 msec which is the highlighted segment in Fig. 9B. Fig. 10E: Peak  $f_0$  frequency for stimulus and all MUAa in the compressed condition. MUA spectra shown in panels A–E were measured from the same recording sites as those shown for the clear speech condition in Fig. 7. This figure shows a correspondence between periodicity in the stimulus  $f_0$  and auditory cortical MUAs measured in the frequency domain.



**Fig. 11. Periodicity analysis for conversational and compressed speech conditions.** Time-domain phase-locking analysis for the conversational (panel A) and compressed (panel B) speech conditions in three representative MUAs. Top panels: peaks used in latency analysis in both the speech stimulus (top) and MUAs are identified with circles. Bottom panels: latency differences between peaks in the speech waveform and MUAs. Cortical MUAs were computed by band-pass filtering raw MUAs (500–3000 Hz) and extracting the envelope of this signal by calculating the absolute value of the Hilbert transformed band-passed signal. This figure shows a correspondence between periodicity in the stimulus  $f_0$  and auditory cortical MUAs measured in the time domain.

of speech might be represented in the human auditory cortex during speech perception. It is possible that the cortical mechanisms described here represent a neural correlate to these important perceptual phenomena.

#### 4.2. Comparison to previous studies: speech envelope

A second finding from the current study is that the temporal envelope of population responses measured from auditory cortical neurons closely tracks the speech envelope. Specifically, it was shown that ~50% of auditory cortical sites showed correlation coefficients  $>0.50$  with respect to the speech envelope. Similar correlation values ( $r \sim 0.50$ ) have been shown when human ECoG signals from auditory cortex are measured in response to speech sentences, and resultant ECoG signals are correlated with the broadband speech envelope extracted from those sentence (Kubaneck et al., 2013; Nourski et al., 2009). The current work adds to this literature by showing, for the first time, that many cortical ensembles exhibited a preference for wider bands of the speech envelope (1–3 octaves) compared to narrow bands ( $<1$  octave). Although no previous studies probing speech envelope coding in auditory cortex have examined relative entrainment to envelopes from narrow vs. wide bands, previous studies of conspecific animal calls have addressed this question and have produced somewhat conflicting results. For example, a study of the cortical representation of cat meows showed that there is a strong relationship between cortical responses and the envelope of the vocalization band-pass filtered around the characteristic frequency of the neuron, reflecting specificity of cortical responses to the temporal envelope of relatively narrow bands of the stimulus (Gourevitch and Eggermont, 2007). Moreover, it has also been shown that wide regions of auditory cortex, including tonotopic areas of auditory cortex that would not be predicted to be activated based on the frequency content of the stimulus, show synchronous response patterns in response to temporal features in marmoset twitter calls (Nagarajan et al., 2002). The current results are the first to quantify the bandwidth of the speech envelope that is preferred by auditory neurons. It is hoped that future studies will be able to examine this result in other animal preparations, and possibly ECoG studies in humans, to address whether this phenomenon is specific to population responses or whether this also occurs using other measures of auditory cortical function.

An additional goal of the speech envelope analysis was to examine whether specific response features of MUAs were related to speech envelope bandwidth selectivity, and results showed that neither the BFs of MUAs nor the RMS amplitude of the envelope of MUAs were related to MUAs' bandwidth selectivity. There are additional temporal and spectral response parameters of MUA beyond those examined here that could potentially account for variance in speech envelope bandwidth selectivity and other aspects of speech envelope coding (Depireux et al., 2001; Miller et al., 2002), and it is hoped that future studies will identify response features that contribute to the coding of this perceptually important aspect of the speech signal. It would be particularly compelling to examine this question using ECoG in awake human listeners to explore whether bandwidth selectivity of MUAs, and other envelope-related response features, can be modulated by task parameters, including challenging listening conditions (Abrams et al., 2013; Anderson et al., 2010, 2013; Yi et al., 2014) and vocoded speech (Peelle et al., 2013).

#### 4.3. Comparison with previous studies: periodicity

A major finding of the current work is that cortical ensembles show a close frequency correspondence, and often phase lock to

the fundamental frequency of speech sentence stimuli. This congruence was evident both in the frequency domain, where there was a close correspondence between the  $f_0$  of the stimulus and the peak frequency in the cortical response (Figs. 7 and 10), as well as the time domain, where it was shown that the latency difference between peaks in the cortical responses often mimicked those in the stimulus waveform (Figs. 8 and 11). There was some variability in the strength of time locking between stimulus  $f_0$  and the dominant frequency of the MUA. For example, the timing of one portion of the response shown in Fig. 8 deviates from the stimulus timing while most other portions of MUAs show strong time locking. The source of this variance in time locking is unknown. Ensemble phase locking to the fundamental frequency of speech has been demonstrated in a number of previous studies. For example, Steinschneider and colleagues have shown this phenomenon in studies examining the voice-onset time temporal characteristics of stop consonants in an awake primate preparation (Steinschneider et al., 1995, 2003). These studies have shown that cortical and thalamocortical ensembles represent the fundamental frequency of brief speech segments with phase-locked responses. The current work adds to this literature by also showing this phenomenon using naturally produced speech sentence stimuli. Moreover, results indicate that ensemble phase locking to the  $f_0$  occurs irrespective of variations in the speech envelope characteristics of the stimulus. For example,  $f_0$ -locking was equally prevalent in the clear speech and compressed speech conditions, which have markedly different speech envelope characteristics. Together, results support the hypothesis that phase locking to the  $f_0$  within auditory cortical ensembles is a general mechanism underlying neural coding of speech periodicity (Coffey et al., 2016).

The propensity of ensembles of cortical neurons to follow periodic acoustic signals at these relatively fast rates (~100 Hz) appears to be in contrast with single unit data. Single unit studies in auditory cortex have repeatedly shown that phase-locking is extremely poor for periodicities  $>50$  Hz, even in awake preparations (Lee et al., 2016; Lu et al., 2001). Moreover, it has been shown that single units of auditory cortex represent periodicities  $>50$  Hz with an unsynchronized rate code (Wang et al., 2003). One possible explanation for this discrepancy in the literature is that ensembles are able to phase lock to faster rates than single units because their activity represents a "volley" of activity from a host of single units, similar to what has been proposed for the auditory nerve (Wever and Bray, 1937). Unfortunately, it is not straightforward how the "implicit" mode of rate representation seen in single unit studies for faster periodicities (Wang et al., 2003) might be related to a volley of phase locked activity; perhaps the cortical "volley" and implicit rate representation provide two separate mechanisms for the representation of the  $f_0$ . Unfortunately, the notion that ensembles of cortical neurons are able to phase-lock to faster rates is generally not discussed in studies investigating single-unit activity in auditory cortex. The result is that, despite decades of research on the auditory system, there is no consensus as to whether phase locking provides a viable code for representing the fundamental frequency of speech in the human auditory cortex. It is hoped that future models describing the cortical representation of speech periodicity will incorporate results from single-unit and population responses.

#### 4.4. Caveats and limitations

An important caveat of the current work is that speech is not a behaviorally significant signal to a guinea pig, and consequently these auditory responses would not be expected to produce response enhancements known to result from auditory learning of

complex acoustic signals (Beitel et al., 2003; Kilgard and Merzenich, 1998a, 1998b; Wang and Kadia, 2001). Furthermore, anytime an animal model, particularly in an anesthetized preparation such as that described here, is used to investigate speech processing mechanisms in humans there is no way of knowing whether the same mechanisms are utilized in the human auditory system. It has been shown that temporal aspects of speech are represented similarly in non-human primate auditory cortex (Steinschneider et al., 1999), and an assumption of the current work is that cortical mechanisms are conserved across mammalian auditory systems.

### Conflict of interest

NK is Chief Scientific Officer of Synaural, a company working to develop a user-friendly measure of auditory processing.

### Acknowledgements

We thank Mitra Hartmann, Ph.D. and members of the Auditory Neuroscience Lab and the Sensory and Neural Systems Engineering Lab for thoughtful discussions regarding this work. This work is supported by the National Institutes of Health grant R01 DC01510-10 and National Organization for Hearing Research grant 340-B208.

### References

- Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2008. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J. Neurosci.* 28, 3958–3965.
- Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2009. Abnormal cortical processing of the syllable rate of speech in poor readers. *J. Neurosci.* 29, 7686–7693.
- Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2010. Rapid acoustic processing in the auditory brainstem is not related to cortical asymmetry for the syllable rate of speech. *Clin. Neurophysiol.* 121, 1343–1350.
- Abrams, D.A., Ryali, S., Chen, T., Balaban, E., Levitin, D.J., Menon, V., 2013. Multivariate activation and connectivity patterns discriminate speech intelligibility in Wernicke's, Broca's, and Geschwind's areas. *Cereb. Cortex* 23, 1703–1714.
- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., Merzenich, M.M., 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13367–13372.
- Anderson, S., White-Schwoch, T., Parbery-Clark, A., Kraus, N., 2013. A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hear Res.* 300, 18–32.
- Anderson, S., Skoe, E., Chandrasekaran, B., Zecker, S., Kraus, N., 2010. Brainstem correlates of speech-in-noise perception in children. *Hear Res.* 270, 151–157.
- Beasley, D.S., Bratt, G.W., Rintelmann, W.F., 1980. Intelligibility of time-compressed sentential stimuli. *J. Speech Hear Res.* 23, 722–731.
- Beitel, R.E., Schreiner, C.E., Cheung, S.W., Wang, X., Merzenich, M.M., 2003. Reward-dependent plasticity in the primary auditory cortex of adult monkeys trained to discriminate temporally modulated signals. *Proc. Natl. Acad. Sci. U. S. A.* 100, 11070–11075.
- Bradlow, A.R., Kraus, N., Hayes, E., 2003. Speaking clearly for children with learning disabilities: sentence perception in noise. *J. Speech Lang. Hear Res.* 46, 80–97.
- Coffey, E.B., Herholz, S.C., Chepesiuk, A.M., Baillet, S., Zatorre, R.J., 2016. Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat. Commun.* 7, 11070.
- Cunningham, J., Nicol, T., King, C., Zecker, S.G., Kraus, N., 2002. Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. *Hear Res.* 169, 97–111.
- Depireux, D.A., Simon, J.Z., Klein, D.J., Shamma, S.A., 2001. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J. Neurophysiol.* 85, 1220–1234.
- Drullman, R., Festen, J.M., Plomp, R., 1994. Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* 95, 1053–1064.
- Everest, F.A., 2001. *The Master Handbook of Acoustics*, fourth ed. McGraw-Hill, New York.
- Fishman, Y.I., Reser, D.H., Arezzo, J.C., Steinschneider, M., 2000. Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness. *J. Acoust. Soc. Am.* 108, 235–246.
- Gehr, D.D., Komiya, H., Eggermont, J.J., 2000. Neuronal responses in cat primary auditory cortex to natural and altered species-specific calls. *Hear Res.* 150, 27–42.
- Gourevitch, B., Eggermont, J.J., 2007. Spatial representation of neural responses to natural and altered conspecific vocalizations in cat auditory cortex. *J. Neurophysiol.* 97, 144–158.
- Kilgard, M.P., Merzenich, M.M., 1998a. Plasticity of temporal information processing in the primary auditory cortex. *Nat. Neurosci.* 1, 727–731.
- Kilgard, M.P., Merzenich, M.M., 1998b. Cortical map reorganization enabled by nucleus basalis activity. *Science* 279, 1714–1718.
- King, C., Nicol, T., McGee, T., Kraus, N., 1999. Thalamic asymmetry is related to acoustic signal complexity. *Neurosci. Lett.* 267, 89–92.
- Kraus, N., White-Schwoch, T., 2015. Unraveling the biology of auditory learning: a cognitive-sensorimotor-reward framework. *Trends Cogn. Sci.* 19, 642–654.
- Kubaneck, J., Brunner, P., Gunduz, A., Poeppel, D., Schalk, G., 2013. The tracking of speech envelope in the human cortex. *PLoS One* 8, e53398.
- Lee, C.M., Osman, A.F., Volgushev, M., Escabi, M.A., Read, H.L., 2016. Neural spike-timing patterns vary with sound shape and periodicity in three auditory cortical fields. *J. Neurophysiol.* 115, 1886–1904.
- Lu, T., Liang, L., Wang, X., 2001. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat. Neurosci.* 4, 1131–1138.
- McGee, T., Kraus, N., King, C., Nicol, T., Carrell, T.D., 1996. Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe. *J. Acoust. Soc. Am.* 99, 3606–3614.
- Miller, L.M., Escabi, M.A., Read, H.L., Schreiner, C.E., 2002. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J. Neurophysiol.* 87, 516–527.
- Nagarajan, S.S., Cheung, S.W., Bedenbaugh, P., Beitel, R.E., Schreiner, C.E., Merzenich, M.M., 2002. Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. *J. Neurophysiol.* 87, 1723–1737.
- Nourski, K.V., Reale, R.A., Oya, H., Kawasaki, H., Kovach, C.K., Chen, H., Howard 3rd, M.A., Brugge, J.F., 2009. Temporal envelope of time-compressed speech represented in the human auditory cortex. *J. Neurosci.* 29, 15564–15574.
- Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* 23, 1378–1387.
- Redies, H., Sieben, U., Creutzfeldt, O.D., 1989. Functional subdivisions in the auditory cortex of the guinea pig. *J. Comp. Neurol.* 282, 473–488.
- Rosen, S., 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 336, 367–373.
- Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. *Science* 270, 303–304.
- Smith, Z.M., Delgutte, B., Oxenham, A.J., 2002. Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416, 87–90.
- Steinschneider, M., Arezzo, J., Vaughan Jr., H.G., 1980. Phase-locked cortical responses to a human speech sound and low-frequency tones in the monkey. *Brain Res.* 198, 75–84.
- Steinschneider, M., Arezzo, J.C., Vaughan Jr., H.G., 1990. Tonotopic features of speech-evoked activity in primate auditory cortex. *Brain Res.* 519, 158–168.
- Steinschneider, M., Fishman, Y.I., Arezzo, J.C., 2003. Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. *J. Acoust. Soc. Am.* 114, 307–321.
- Steinschneider, M., Reser, D., Schroeder, C.E., Arezzo, J.C., 1995. Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. *Brain Res.* 674, 147–152.
- Steinschneider, M., Reser, D.H., Fishman, Y.I., Schroeder, C.E., Arezzo, J.C., 1998. Click train encoding in primary auditory cortex of the awake monkey: evidence for two mechanisms subserving pitch perception. *J. Acoust. Soc. Am.* 104, 2935–2955.
- Steinschneider, M., Volkov, I.O., Noh, M.D., Garell, P.C., Howard 3rd, M.A., 1999. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J. Neurophysiol.* 82, 2346–2357.
- Uchanski, R.M., 2005. Clear speech. In: Pisoni, D.B., Remez, R.E. (Eds.), *Handbook of Speech Perception*. Blackwell Publishers, Malden, MA, pp. 207–235.
- Vaissiere, J., 2005. Perception of intonation. In: Pisoni, D.B., Remez, R.E. (Eds.), *The Handbook of Speech Perception*. Blackwell Publishing, Malden, MA.
- Wallace, M.N., Rutkowski, R.G., Palmer, A.R., 2000. Identification and localisation of auditory areas in guinea pig cortex. *Exp. Brain Res.* 132, 445–456.
- Wang, X., Kadia, S.C., 2001. Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *J. Neurophysiol.* 86, 2616–2620.
- Wang, X., Lu, T., Liang, L., 2003. Cortical processing of temporal modulations. *Speech Commun.* 41, 107–121.
- Wang, X., Merzenich, M.M., Beitel, R., Schreiner, C.E., 1995. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J. Neurophysiol.* 74, 2685–2706.
- Warrier, C.M., Abrams, D.A., Nicol, T.G., Kraus, N., 2011. Inferior colliculus contributions to phase encoding of stop consonants in an animal model. *Hear Res.* 282, 108–118.
- Watson, M., Stewart, M., Krause, K., Rastatter, M., 1990. Identification of time-compressed sentential stimuli by good vs poor readers. *Percept. Mot. Ski.* 71, 107–114.
- Wever, E.G., Bray, C.W., 1937. The perception of low tones and the resonance-volley theory. *J. Psychol.* 3, 101–114.
- White-Schwoch, T., Nicol, T., Warrier, C., Abrams, D.A., Kraus, N., 2016. Individual differences in human auditory processing: insights from single-trial auditory midbrain activity in an animal model. *Cereb. Cortex*. <http://dx.doi.org/10.1093/cercor/bhw293>.
- Woodruff Carr, K., White-Schwoch, T., Tierney, A.T., Strait, D.L., Kraus, N., 2014. Beat

- synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14559–14564.
- Yi, H.G., Smiljanic, R., Chandrasekaran, B., 2014. The neural processing of foreign-accented speech and its relationship to listener bias. *Front. Hum. Neurosci.* 8, 768.
- Zeng, F.G., Nie, K., Stickney, G.S., Kong, Y.Y., Vongphoe, M., Bhargave, A., Wei, C., Cao, K., 2005. Speech recognition with amplitude and frequency modulations. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2293–2298.