Neuroscience 258 (2014) 292–306

DETECTION AND IDENTIFICATION OF SPEECH SOUNDS USING CORTICAL ACTIVITY PATTERNS

T. M. CENTANNI,* A. M. SLOAN, A. C. REED, C. T. ENGINEER, R. L. RENNAKER II AND M. P. KILGARD

University of Texas at Dallas, United States

Abstract—We have developed a classifier capable of locating and identifying speech sounds using activity from rat auditory cortex with an accuracy equivalent to behavioral performance and without the need to specify the onset time of the speech sounds. This classifier can identify speech sounds from a large speech set within 40 ms of stimulus presentation. To compare the temporal limits of the classifier to behavior, we developed a novel task that requires rats to identify individual consonant sounds from a stream of distracter consonants. The classifier successfully predicted the ability of rats to accurately identify speech sounds for syllable presentation rates up to 10 syllables per second (up to 17.9 ± 1.5 bits/s), which is comparable to human performance. Our results demonstrate that the spatiotemporal patterns generated in primary auditory cortex can be used to quickly and accurately identify consonant sounds from a continuous speech stream without prior knowledge of the stimulus onset times. Improved understanding of the neural mechanisms that support robust speech processing in difficult listening conditions could improve the identification and treatment of a variety of speech-processing disorders. © 2013 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: classifier, rat, auditory cortex, coding, temporal patterns.

INTRODUCTION

Speech sounds evoke unique spatiotemporal patterns in the auditory cortex of many species (Kuhl and Miller, 1975; Eggermont, 1995; Engineer et al., 2008). Primary auditory cortex (A1) neurons respond to most consonants, which evoke short, transient bursts of neural activity, but respond with different spatiotemporal patterns for different sounds (Engineer et al., 2008). For example, the consonant /d/ evokes activity first in neurons tuned to high frequencies, followed by neurons tuned to lower frequencies. The sound /b/ causes the opposite pattern such that low-frequency neurons fire approximately 20 ms before the high-frequency neurons (Engineer et al., 2008; Shetake et al., 2011; Perez et al., 2012; Ranasinghe et al., 2012b). These patterns of activity can be used to identify the evoking auditory stimulus in both human (Steinschneider et al., 2005; Chang et al., 2010; Pasley et al., 2012) and animal auditory cortex (Engineer et al., 2008; Mesgarani et al., 2008; Huetz et al., 2009; Bizley et al., 2010; Shetake et al., 2011; Perez et al., 2012; Ranasinghe et al., 2012a; Centanni et al., 2013a).

Rats are a good model of human speech sound discrimination as these rodents have neural and behavioral speech discrimination thresholds that are similar to humans. Rats can discriminate isolated human speech sounds with high levels of accuracy (Engineer et al., 2008; Perez et al., 2012; Centanni et al., 2013a). Rats and humans have similar thresholds for discriminating spectrally-degraded speech sounds, down to as few as four bands of spectral information (Ranasinghe et al., 2012b). Rats and humans are both able to discriminate speech sounds when presented at 0-dB signal to noise ratio (Shetake et al., 2011).

In both rats and humans, sounds that evoke different patterns of neural activity are more easily discriminated behaviorally than sounds that evoke similar patterns of activity (Engineer et al., 2008; Shetake et al., 2011; Ranasinghe et al., 2012b). Speech sounds presented in background noise evoke neural response patterns with longer latency and lower firing rate than speech presented in quiet and the extent of these differences is correlated with behavioral performance (Martin and Stapells, 2005; Shetake et al., 2011). Neural activity patterns in anesthetized rats also predict behavioral discrimination ability of temporally degraded speech sounds (Ranasinghe et al., 2012b).

The relationship between neural activity and associated behavior is often analyzed using minimum distance classifiers, but classifiers used in previous studies typically differ from behavioral processes in one key aspect: the classifiers were provided with the stimulus onset time, which greatly simplifies the problem of speech classification (Engineer et al., 2008; Shetake et al., 2011; Perez et al., 2012; Ranasinghe et al., 2012a; Centanni et al., 2013a,b). During natural listening, stimulus onset times occur at irregular intervals. One possible correction allows a classifier to look through an entire recording sweep, rather than only considering activity immediately following stimulus onset. The classifier then guesses the location and

---

*Corresponding author. Address: University of Texas at Dallas, 800W Campbell Road, GR 41, Richardson, TX 75080, United States. Tel: +1-972-883-2376; fax: +1-972-883-2491.

E-mail address: tmcentanni@gmail.com (T. M. Centanni).

Abbreviations: A1, primary auditory cortex; CF, characteristic frequency; CVC, consonant–vowel–consonant; FM, frequency modulated; IR, infrared; ISI, inter-stimulus-interval; NM, normalized metric; SPL, sound pressure level; sps, syllables per second.

0306-4522/13 $36.00
http://dx.doi.org/10.1016/j.neuroscience.2013.11.030

identity of the sound post hoc by picking the location most similar to a template (Shetake et al., 2011). While this method is highly accurate and predicts behavioral ability without the need to provide the onset time, the method could not be implemented in real time and assumes that a stimulus was present. We expected that large numbers of recording sites would be able to accurately identify a sound's onset, since the onset response in A1 to sound is well known (Anderson et al., 2006; Engineer et al., 2008; Dong et al., 2011; Centanni et al., 2013b). We hypothesized that with many recording sites, A1 activity can also be used for identification of the sound with a very brief delay consistent with behavioral performance in humans and animals.

EXPERIMENTAL PROCEDURES

Speech stimuli

For this study, we used the same stimuli as several previous studies in our lab (Engineer et al., 2008; Floody et al., 2010; Porter et al., 2011; Shetake et al., 2011; Ranasinghe et al., 2012b). We used nine English consonant–vowel–consonant (CVC) speech sounds differing only by the initial consonant: (/bad/, /dad/, /gad/, /kad/, /pad/, /sad/, /tad/, /wad/, and /zad/), which were recorded in a double-walled, soundproof booth spoken by a female native-English speaker. The spectral envelope was shifted up in frequency by a factor of two while preserving all spectral information using the STRAIGHT vocoder (Kawahara, 1997) to better accommodate the rat hearing range. The intensity of each sound was calibrated with respect to its length, such that the loudest 100 ms was presented at 60-dB sound pressure level (SPL) and 5 ms on and off ramps were added to prevent any artifacts.

Surgical procedure – Anesthetized recordings

Multi-unit recordings were acquired from the A1 of anesthetized, experimentally-naïve female Sprague–Dawley rats (Charles River, Wilmington, MA, USA). Recording procedures are described in detail elsewhere (Engineer et al., 2008; Shetake et al., 2011; Ranasinghe et al., 2012b). In brief, animals were anesthetized with pentobarbital (50 mg/kg) and were given supplemental dilute pentobarbital (8 mg/ml) as needed to maintain areflexia, along with a 1:1 mixture of dextrose (5%) and standard Ringer's lactate to prevent dehydration. A tracheotomy was performed to ensure ease of breathing throughout the experiment and filtered air was provided through an air tube fixed at the open end of the tracheotomy. Craniotomy and durotomy were performed, exposing right A1. Four Parylene-coated tungsten microelectrodes (1–2 MΩ) were simultaneously lowered to layer (4/5) of right A1 (~600 µm). Electrode penetrations were marked using blood vessels as landmarks.

Brief (25 ms) tones were presented at 90 randomly interleaved frequencies (1–47 kHz) at 16 intensities (0–75 dB SPL) to determine the characteristic frequency (CF) of each site. A set of four stimuli were created using Adobe Audition for comparison to our behavioral task (described below). Each stimulus consisted of a train of six individual speech sounds such that across all four sequences, all 24 possible sound pairs were presented once (/bad bad gad sad tad dad/, /tad tad sad gad bad dad/, /gad gad tad bad sad dad/, /sad sad bad tad gad dad/). The temporal envelope of the stimuli was compressed so that when presented with a 0-s inter-stimulus-interval (ISI), sounds were presented at 2, 4, 5, 6.7, 10 and 20 syllables per second (sps). All speech stimuli were randomly interleaved, and presented at 20 repeats per recording site. All sounds were presented approximately 10 cm from the left ear of the rat. Stimulus generation, data acquisition and spike sorting were performed with Tucker-Davis hardware (RP2.1 and RX5) and software (Brainware).

Surgical procedure – Awake recordings

Rats were anesthetized and implanted with a chronic array of 16 polyimide-insulated 50-µm diameter tungsten microwires. The implantation surgery and microwire arrays have been previously reported in detail (Rennaker et al., 2005). Briefly, subjects were anesthetized with an intramuscular injection of a mixture of ketamine, xylazine and acepromazine (50, 20, 5 mg/kg, respectively). Atropine and dexamethazone were administered subcutaneously prior to and following surgery. A midline incision was made, exposing the top of the skull, and a section of the right temporalis muscle was removed to access A1. Six bone screws were fixed to the dorsal surface of the skull (two in each parietal bone and one in each frontal bone) to provide structural support for the head cap. The two middle screws had attached leads to serve as a reference wire and a grounding wire. Craniotomy and durotomy were performed to expose the cortex in the region of A1. The microwire array was then inserted to a depth of 550–600 µm (layer IV/V) in A1 using a custom-built mechanical inserter (Rennaker et al., 2005). The area was sealed with a silicone elastomer (Kwik-Cast, World Precision Instruments Inc., Sarasota, FL, USA) and the head cap was built with a connector secured with acrylic. Finally, the skin around the implant was sutured in the front and the back of the head cap. Subjects were given prophylactic minocycline in water *ad libitum* for 2 days prior to and 5 days following surgery to lessen immune responses (Rennaker et al., 2005), and were also given Rimadyl tablets for 3 days after surgery to minimize discomfort. Topical antibiotic was applied to the incision to prevent infection. After a minimum of 5 days of recovery, neural activity was collected in a single 2.5-h session and saved using a custom MATLAB program. The session included an abridged tuning curve (to assess each site's best frequency) and the same set of speech sequence stimuli presented to the anesthetized animals. All passive sound sets were created and run through custom MATLAB programming.

Neural analysis and classifier

We designed a classifier that does not require precise information about the sound's onset time by modifying